UTS UNIVERSITY OF TECHNOLOGY SYDNEY

The Identification of Missing Persons in Australia Using Emerging Forensic Genomics Techniques

by Jessica Watson

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Professor Dennis McNevin and Adjunct Associate Professor Jodie Ward

University of Technology Sydney
Faculty of Science

May 2025

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Jessica Lee Watson declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Centre for Forensic Science, Faculty of Science at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note: Signature removed prior to publication.

Date: 06/05/2025

ABSTRACT

The growing number of unidentified and missing persons around the world has left countless individuals without identities and their loved ones without answers. Australian cases can leverage emerging forensic genomic techniques to provide new investigative leads using single nucleotide polymorphisms (SNPs) where routine identification methods have been exhausted. In addition to identity-informative SNPs (iiSNPs) for individualisation, there are several classes of SNPs that can be used to generate genetic intelligence. It is possible to infer an unknown deceased individual's biological sex with X and Y chromosome SNPs, biogeographical ancestry (BGA) with ancestry-informative SNPs, externally visible characteristics with phenotype-informative SNPs and genetic relationships to others with kinship-informative SNPs. These inferences can be used by investigators to narrow a pool of potential candidates. However, forensic laboratories have different technological and analytical capabilities that will determine whether they can conduct boutique analyses in-house or have to outsource these capabilities to an external service provider.

Under the umbrella of the Australian Federal Police (AFP) National DNA Program for Unidentified and Missing Persons, this research evaluated available SNP panels that have been designed for integration into existing forensic laboratories as an end-to-end in-house forensic genomic capability. The ForenSeq® Kintelligence Kit (QIAGEN) was validated for application at the AFP National DNA Program for Unidentified and Missing Persons laboratory and its investigative potential was evaluated for all 10,230 SNPs included in the assay. By testing a range of reference- and casework-type samples of varying quality, the kit and inference pipelines were determined to be robust and the inferred information was consistent with the self-declared information provided by volunteers.

Within a representative group of the Australian population, this research demonstrated that iiSNPs are suitable for application in conjunction with routine short tandem repeat (STR) profiles. Furthermore, optimised pipelines have been developed to estimate eye colour, hair colour and BGA with high success rates. Extended kinship pipelines using either kinship likelihood ratios or haplotype matching are able to accurately infer genetic relationships out to the fifth degree (i.e. second cousin) with complete or partial SNP profiles. This research holds significance for law enforcement, policymakers and family members of missing persons by contributing to the Australian and global development of forensic genomics standard operating procedures and guidelines for implementation.

Australian law enforcement should continue to utilise these forensic genomic techniques to aid unresolved coronial investigations, provide answers to loved ones of missing persons and restore names to unidentified human remains.

PREFACE

The number of unidentified human remains (UHR) and long-term missing persons (LTMP) across the world has come to be known as the silent mass disaster, of which the true extent is largely unknown. Identification efforts are important for providing answers to loved ones of LTMPs as well as contributing to death investigations. In Australia, there are approximately 750 sets of UHR awaiting identification, with cases spanning several decades. The high volume of cases with a variety of complicating factors provides law enforcement with numerous challenges in exploring all possible forensic techniques beyond the traditional dental, fingerprint and DNA analyses.

DNA analysis has been endorsed as one of the primary identification tools for its ability to distinguish between individuals. In its most common application, a short tandem repeat (STR) profile obtained from a UHR is compared to known reference profiles on a law enforcement DNA database to look for a direct or kinship match. In the absence of a DNA match, other forensic techniques have the potential to provide new investigative leads for law enforcement to pursue. Using single nucleotide polymorphisms (SNPs), emerging forensic genomics techniques increase the available intelligence to be gained by interrogating more of the human genome, including the exploration of alternative identity markers and inference of biological sex, externally visible characteristics (EVCs), biogeographical ancestry (BGA) and kinship to genetic relatives.

The first part of this thesis (Chapter One), including the first publication, reviews the current landscape of forensic genomics techniques for their application to UHR and LTMP casework in an Australian context. It compares the informativeness of SNP markers to the traditionally targeted STRs and the possible genotyping technologies that are available. For these technologies, the various operational, expertise and ethical challenges have been highlighted with emphasis on the increased privacy risks associated with DNA intelligence generated through BGA, EVCs and genetic relationships. Each laboratory and jurisdiction will have different operational requirements to consider when implementing SNP genotyping and interpretation pipelines. This review emphasises the importance of centralising the expertise, instrumentation, processes and resources in a dedicated and fit-for-purpose manner to process Australian UHR and LTMP cases. To do this, in-house genomics, bioinformatics and standard operating procedures (SOPs) are required to employ these techniques efficiently and effectively.

It is possible to generate medium- or high-density SNP profiles from challenging forensic samples using either targeted amplicon sequencing (TAS) or whole genome sequencing (WGS) panels, respectively. While WGS can generate the entire human genome of 3 billion bases, TAS applies established forensic massively parallel sequencing (MPS) technology to target a subset of forensically relevant SNP markers. In the second publication (Chapter Three), a range of reference-type and casework-type samples were sequenced using two TAS panels: the ForenSeq® Kintelligence Kit (Verogen, Inc., now a QIAGEN company) and the FORensic Capture Enrichment (FORCE) panel. These pipelines were evaluated for their robustness in UHR casework and the feasibility of adopting each technology in an ISO/IEC 17025 accredited forensic laboratory. The resulting genotypes were compared for concordance and DNA intelligence attained. TAS panels provide an in-house end-to-end solution for medium-density SNP genotyping that maximises the DNA intelligence obtainable from the genome while also enhancing privacy protection through the removal of medically informative markers. Furthermore, the Kintelligence Kit is compatible with law enforcement accessible public genetic genealogy databases, circumventing the need to establish a reference database as required with the FORCE panel.

The third publication (Chapter Four) describes the validation and optimisation of the Kintelligence Kit for UHR casework conducted by the Australian Federal Police National DNA Program for Unidentified and Missing Persons (AFP DNA Program). The panel was internally validated according to the Scientific Working Group for DNA Analysis Methods guidelines to establish sample requirements, analytical thresholds and best practices for application to compromised samples from human remains using a selection of casework-type samples including bones, teeth, blood, hair and nail. Buccal swabs were collected from volunteers with self-declared biological sex, EVCs (hair and eye colour), BGA and genetic relationships to other volunteers to assess the accuracy of the generated DNA intelligence. The panel is robust and able to generate reliable and high-quality profiles from compromised samples of varying DNA quality and quantity. It can accurately predict biological sex and EVCs; however, the in-built BGA pipeline requires refinement to be relevant to an Australian population as the original database was designed for an admixed American population. Utilising the law enforcement accessible public genetic genealogy database GEDmatch PRO™, the medium-density SNP panel is capable of accurately detecting and classifying relatives of fifth degree (e.g. second cousins) or closer. This study was used to establish an internal laboratory and bioinformatics pipeline

for the AFP DNA Program's SOPs and meet the quality assurance and quality control requirements of an ISO/IEC 17025 accredited forensic laboratory.

The identity-informative SNPs (iiSNPs) included in the Kintelligence Kit provide a suitable alternative to STRs for making a direct or kinship match when STR profiling may not be suitable due to low DNA quality or quantity. The fourth publication (Chapter Five) describes a population of Australians with European ancestry and the generation of allele frequencies in order to conduct population genetic analyses. A set of 94 iiSNP markers are sufficient for calculating likelihood ratios (LR), having been found to be in Hardy-Weinberg equilibrium as well as in linkage equilibrium with each other and with the traditional STR markers. These alternative markers improve the discrimination between individuals and produce LRs that exceed those produced with STRs. The combined power of both traditional and alternative identity-informative markers can be useful in identifying partial profiles from compromised or challenging samples and assist in indirect matching to genetic relatives. This data has facilitated the creation of the first iiSNP allele frequency database for Australia.

It is possible to infer the biological sex, EVCs and BGA from DNA and, in the absence of a direct or kinship match on a law enforcement database comprised of relevant STR profiles, these predictions may be used by investigators to narrow a pool of potential candidates for the UHR. The fifth publication (Chapter Six) evaluates the use of the Kintelligence Kit for this purpose through the sex-chromosome markers (X chromosome SNPs and Y chromosome SNPs), phenotype-informative SNPs and ancestry-informative SNPs. Volunteers provided buccal swabs and self-declared their recent ancestry, eye colour and hair colour and the resulting genotypes were analysed using a variety of interpretation pipelines. Biological sex can be accurately inferred from the presence and heterozygosity of the sex chromosome markers. The accuracy of the EVC predictions was improved by using the online HIrisPlex tool as opposed to the in-built Kintelligence algorithm, with 80% of hair colour predictions and 97% of eye colour inferences being consistent with the self-declared phenotypes. The in-built Kintelligence Kit ancestry algorithm was not suitable for a non-American population; the combination of multiple ancestry inference models corrected the previously inconclusive results with the inclusion of additional population datasets and the ability to detect and characterise admixture. This study will inform further refinement of the interpretation guidelines to improve the informativeness of DNA intelligence for investigators.

The final studies of this thesis explore the two primary methods for inferring relatedness using forensic genomics. Traditionally, an LR provides probabilistic support for alternative propositions of relatedness using STRs. This can be achieved by conducting pairwise analysis with first degree relationships (i.e. parent/offspring and full sibling pairs) or conducting pedigree analysis with a combination of relatives ranging from first to third degree, depending on availability. Identical-by-descent (IBD) segment matching is an alternative method that approximates the amount of DNA shared between two genotypes and the relationship can be estimated based on the length and frequency of these segments. This method can detect out to ninth degree relatives (i.e. fourth cousins) depending on the density of the SNP panel. The first kinship study and sixth publication (Chapter Seven) presents the analysis of volunteer family groups using LR calculations and IBD segment matching to detect and classify genetic relationships between the traditional STR, alternative iiSNP and Kintelligence genotypes. Results showed that LR calculations are only informative for determining first and second degree relationships using either STR or iiSNP genotypes. The IBD segment matching algorithms for Kintelligence profiles on GEDmatch PRO™ can efficiently detect out to fifth degree relatives (i.e. second cousins) when searching the database of approximately 2 million profiles. Furthermore, one-to-one comparison of Kintelligence profiles with over ten thousand SNPS produces powerful LRs supporting hypotheses of relatedness for first to fifth degree relationships.

The second kinship study and seventh publication (Chapter 8) tested the robustness of these methods for dealing with incomplete profiles. The Kintelligence profiles of the central individual to each family group underwent simulated locus and allele dropout. The loss of up to 30% of the possible SNPs did not significantly impact the ability to detect and accurately classify first to fifth degree relatives for IBD segment matching approaches and only slightly diminished the statistical power of the LRs. However, allele dropout resulted in the IBD segment matching algorithm detecting less IBD segments and inferring the relationships were more distant than they truly were, while the LRs decreased substantially with increasing allele dropout. These kinship studies will assist in the development of analytical and interpretation guidelines for extended kinship analysis in forensic casework to be used for forensic investigative genetic genealogy.

This thesis evaluates and optimises the potential applications of the Kintelligence Kit to generate DNA intelligence and utilise alternative DNA databases to aid in the identification of Australia's UHR. Chapter 9 explores the sequencing, bioinformatics,

interpretation and reporting guidelines and standards required to operationalise these emerging forensic genomics techniques and provides recommendations for establishing in-house SNP genotyping capabilities. In practice, these studies have aided the AFP DNA Program in an Australian first application of SNPs to provide new investigative leads for unresolved UHR cases and resulted in several identifications before the AFP DNA Program concluded.

ACKNOWLEDGEMENTS

First and foremost, I am profoundly grateful to my supervisors, Professor Dennis McNevin and Adjunct Associate Professor Jodie Ward for believing in me and making this project possible. They provided invaluable advice and unwavering support over the last few years, and I tremendously enjoyed completing this research with them. I was fortunate to receive the unique perspective of the intersection of academia, research and industry through the unique opportunities my supervisors provided during this research.

The Australian Federal Police (AFP) National DNA Program for Unidentified and Missing Persons (DNA Program) team consisting of my supervisors, Dr Kelly Grisedale, Michelle Spiden, Shelley Seddon, Alicia Gagliardi and Louise Coakley were incredibly supportive and encouraging of my research, even after the conclusion of the AFP DNA Program in 2024. In particular, I would like to thank Kelly for the knowledge and training she imparted to me while at the AFP for my research and professional development. I am also grateful for the understanding and support received from members of AFP Forensics during my research including, but not limited to, Dr Nathan Scudder, Dr Paul Roffey and Kaymann Cho.

My gratitude extends to the School of Mathematics and Physical Sciences at the University of Technology Sydney and the Australian Government Research Training Program Scholarship for funding the opportunity to undertake my studies. Additionally, I am thankful for the support provided by the AFP Innovation Fund grant and the International Society for Forensic Genetics for awarding me the short-term fellowship and travel bursary to facilitate discussion with experts and visiting other laboratories. Through this, I was able to engage and learn from the incredible Emerging Technologies team at the Armed Forces DNA Identification Laboratory.

Finally, I would not have completed this research without the endless patience and support of my family and friends. They attended my presentations and entertained my (usually one-sided) talks about my work and kept me grounded for the last few years.

LIST OF PUBLICATIONS

This document has been prepared as a thesis by compilation comprising of multiple published works.

The thesis includes seven papers with are co-authored with supervisors and other researchers. Articles that have been submitted are marked as under review or accepted by the journal. The references for these papers are:

- **Chapter 1: Watson J**, McNevin D and Ward J (2024) 'Genetic kinship testing techniques for human remains identification and missing persons investigations', Forensic Genomics, 4(1), DOI:10.1089/forensic.2023.0018.
- **Chapter 3: Watson JL**, Grisedale K, McNevin D and Ward J (2025) 'Evaluation of the ForenSeq® Kintelligence Kit and the FORensic Capture Enrichment Panel for Unidentified and Missing Persons Casework', International Journal of Legal Medicine, DOI: 10.1007/s00414-025-03492-4.
- **Chapter 4: Watson J**, McNevin D, Grisedale K, Spiden M, Seddon S and Ward, J (2024) 'Operationalisation of the ForenSeq® Kintelligence Kit for Australian unidentified and missing persons casework', Forensic Science International: Genetics, 68, DOI:10.1016/jfsigen.2023.102972.
- **Chapter 5: Watson, JL**, Cho K, Grisedale K, Ward J and McNevin D (2024) 'Characterisation of Identity-Informative Genetic Markers in the Australian Population with European Ancestry.' Forensic Science International: Genetics, 103169, DOI: 10.1016/j.fsigen.2024.103169.
- **Chapter 6: Watson JL**, Grisedale K, Ward J and McNevin D. (2025) 'DNA Intelligence Using Sex-Chromosome, Phenotype-Informative and Ancestry-Informative Markers in an Australian Population'. Australian Journal of Forensic Sciences, 1-18. DOI: 10.1080/00450618.2025.2491376.
- **Chapter 7: Watson JL**, Grisedale K, Coakley L, McNevin D and Ward J. (2025) 'Extended Kinship Inference. Part 1: Evaluation of Short Tandem Repeats and Single Nucleotide Polymorphisms using Likelihood Ratios and Haplotype Matching'. Forensic Genomics, 5(1), DOI: 10.1089/forensic.2025.0001.
- **Chapter 8: Watson JL**, Grisedale K, Coakley L, McNevin D and Ward J (2025) 'Extended Kinship Inference. Part 2: Evaluation of the Impact of Information Loss on Likelihood Ratios and Haplotype Matching', Forensic Genomics, 5(1), DOI: 10.1089/forensics.2025.0002.

The research in this thesis contributed to additional published works that were co-authored with supervisors and other researchers. This included contribution to a comparison between targeted amplicon sequencing panels (Thermo Fisher Scientific). The reference for this paper is:

- McNevin, D., **Watson, J.**, Grisedale, K., Dahal, A., Goodwin, C., & Ward, J. (2025). Comparison of commercial targeted amplicon sequencing assays for human remains identification casework. International Journal of Legal Medicine, 139, 49-60. DOI: 10.1007/s00414-024-03335-8.

The research assisted in the operationalisation of forensic investigative genetic genealogy capabilities within the Australian Federal Police (AFP) as part of the AFP National DNA Program for Unidentified and Missing Persons. Published case studies that have utilised this capability were co-authored with supervisors and other researchers. The reference for this paper is:

- Ward, J., Coakley, L., Grisedale, K., Seddon, S., Spiden, M., **Watson, J. L.**, & McNevin, D. (2024). Operationalization of the National DNA Program for Unidentified and Missing Persons' Forensic Investigative Genetic Genealogy Capability for Human Remains Identification in Australia. *Forensic Genomics*, *4*(1), 32-40. DOI: 10.1089/forensic.2023.0019.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AFP	Australian Federal Police
AFTER	Australian Facility for Taphonomic Experimental Research
aiSNP	Ancestry-informative SNP
ALFRED	Allele FREquency Database
AM	Antemortem
ASO	Allele-specific oligonucleotide
AT	Analytical threshold
atDNA	Autosomal DNA
AUC	Area under the receiver operator curve
BGA	Biogeographical ancestry
CDI	Combinatorial dual index
CE	Capillary electrophoresis
cM	Centimorgan
CODIS	Combined DNA Index System
CPM	Combined probability of matching
DBLR	Database Likelihood Ratio
DI	Degradation index
DNA	Deoxyribonucleic acid
DNA Program	National DNA Program for Unidentified and Missing Persons
DRS	Direct reference sample
DTC	Direct to consumer
EDTA	Ethylenediaminetetraacetic acid
EVC	Externally visible characteristics
FBI	Federal Bureau of Investigation
FGG	Forensic genetic genealogy
FGGS	Forensic genetic genealogical DNA analysis and searching
FIGG	Forensic investigative genetic genealogy
FORCE	FORensic Capture Enrichment
FROG-kb	Forensic Resource Reference on Genetics knowledge base
FRS	Familial reference sample
GSA	Global Screening Array
Hexp	Expected heterozygosity
Hobs	Observed heterozygosity
HREC	Human Research Ethics Committee
HWE	Hardy Weinberg equilibrium
IBD	Identical by descent

IBS	Identical by state
ICMP	International Commission on Missing Persons
ICRC	International Committee of the Red Cross
IGG	Investigative genetic genealogy
iiSNP	Identity-informative SNP
ILB	Intralocus balance
INTERPOL	International Criminal Policing Organization
ISO/IEC	International Organization for Standardization / International Electrotechnical Commission
IT	Interpretation threshold
kiSNP	Kinship-informative SNP
LA	Large autosomal
LE (Chapter Five)	Linkage equilibrium
LE (Chapter One)	Law enforcement
LR	Likelihood ratio
LTMP	Long-term missing person
MAF	Minor allele frequency
MCMC	Markov chain Monte Carlo
MDS	Multidimensional scaling
MLR	Multinomial logistic regression
MPS	Massively parallel sequencing
mtDNA	Mitochondrial DNA
NC	Negative control
NCIC	National Crime Information Center
NCIDD	National Criminal Investigation DNA Database
NPV	Negative predictive value
NRC	National Research Council
NRY	Non-recombinant region
PC	Positive control
PCA	Principle component analysis
PCoA	Principle coordinate analysis
PCR	Polymerase chain reaction
PD	Power of discrimination
PE	Power of exclusion
PG	Possible genotypes
PIA	Privacy impact assessment
PIC	Polymorphism information content
piSNP	Phenotype-informative SNP

PM	Probability of matching
PM	Postmortem
PMI	Postmortem index
POI	Person of interest
PPV	Positive predictive value
Q30	Quality score of 30
qPCR	Quantitative PCR
RFU	Relative fluorescence unit
RMP	Random match probability
RTP	Research training program
SA	Small autosomal
SAV	Sequencing Analysis Viewer
SFC	Standard flow cell
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
STRAF	STR Analysis for Forensics
SWGDAM	Scientific Working Group on DNA Analysis Methods
TAS	Targeted amplicon sequencing
TE	Tris-EDTA
TPI	Typical paternity index
UAS	Universal Analysis Software
UDI	Unique dual index
UHR	Unidentified human remains
UK	United Kingdom
UMI	Unique molecular index
UNTCHI	University of North Texas Center for Human Identification
US	United States
UTS	University of Technology Sydney
WGS	Whole genome sequencing
X SNP	X chromosome SNP
X STR	X chromosome STR
Y SNP	Y chromosome SNP
Y STR	Y chromosome STR

LIST OF FIGURES

CHAPTER ONE:

LIST OF TABLES

CHAPTER ONE:

CHAPTER TWO:

CHAPTER FIVE:

APPENDICES:

1. CHAPTER ONE

INTRODUCTION AND LITERATURE REVIEW

This chapter has been published in part in the following work:

Contribution of authors:

Watson J, McNevin D and Ward J (2024) 'Genetic kinship testing techniques for human remains identification and missing persons investigations', Forensic Genomics, 4(1), DOI:10.1089/forensic.2023.0018.

Jessica Watson (Candidate)

Carried out literature review, prepared manuscript, edited manuscript following review by other co-authors, referenced manuscript, carried out manuscript's submission and incorporated feedback from peer review process.

Signed: Production Note: Signature removed prior to publication.

Dennis McNevin (Co-supervisor)

Conceptualised literature review, contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Jodie Ward (Co-supervisor)

Conceptualised literature review, contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

1.1 Unidentified and Missing Persons

The growing number of unidentified and missing persons cases around the world has become known as the “silent mass disaster” and their resolution is highly important for administrative, legal and humanitarian reasons.[1-3] When an individual is reported missing to the police, an investigation is undertaken to collect information about the individual along with relevant biometric, dental and medical records, as well as reference DNA samples. This information can be compared, directly or through database searching, with equivalent information obtained from unidentified human remains (UHR).[4-8] In Australia, when an individual has been missing for a period greater than three months, they are referred to as a long-term missing person (LTMP).[9]

UHR refers to the remains of a deceased individual with no known identity. The most commonly employed forensic methods for identification are fingerprint, dental and DNA analysis.[2, 10] However, as the postmortem interval (PMI) increases and remains become decomposed and degraded, forensic investigations typically have to rely on skeletal and dental features and DNA to make an identification.[11-13]

Records for unidentified and missing persons cases can be managed at a local, state or national level depending on the jurisdiction requirements and resources. The United States (US) has two reporting bodies that manage unidentified and missing persons cases; the National Crime Information Center (NCIC) which is maintained by the Federal Bureau of Investigation (FBI) has a greater number of LTMP records than the National Missing and Unidentified Persons System (NamUs), but fewer UHR records (Table 1.1).[14, 15] Ritter (2007) estimated that there were approximately 40,000 UHR that were awaiting identification, of which only 6,000 had been entered into the NCIC database by 2007.[1] However, it has been acknowledged by US and United Kingdom (UK) law enforcement agencies that LTMP and UHR records may be underestimated and incomplete due to underreporting and database management.[1, 16-18]

Table 1.1 Numbers of unidentified and missing persons reported in the United States, United Kingdom and Australia. Updated from Watson et al. (2024).[19]

Country	Reporting Body	Year	Missing Persons Reported	Long-Term Missing Persons	Unidentified Human Remains
United States	National Crime Information Center[14]	2023	563,389	96,955	8,521
	National Missing and Unidentified Persons System[15]	2024	~600,000	25,553	15,168
United Kingdom	UK Missing Persons Unit[20]	2023	170,854	4,778	1,033
Australia	National Missing Persons Coordination Centre[21]	2022	~55,000	2,500	750

There are several multidisciplinary international bodies that have been established to assist in the resolution of missing persons investigations and identification of UHR through the application of various forensic techniques. Such organisations include the International Commission on Missing Persons (ICMP), International Committee of the Red Cross (ICRC) and International Criminal Policing Organization (INTERPOL).[3, 5, 6] The ICMP is active in approximately forty countries and manages the Integrated Data Management System (iDMS) which holds the records of over 100,000 missing persons.[22] The ICRC deploys to areas of armed conflict and mass disasters to apply forensic methods as part of their humanitarian action to search for missing persons and identify UHR.[2, 5, 23]

INTERPOL encourages and facilitates communication between jurisdictions of their member countries and manages several international databases such as I-Familia to aid in unidentified and missing persons cases.[24] They encourage member countries to establish national identification programs that are dedicated to unidentified and missing persons investigations.[6] Several countries have followed suit, including the National Centre for Missing Persons and Unidentified Remains in Canada, the University of North Texas Center for Human Identification and NamUs in the US, and the UK Missing Persons Unit.[18, 25, 26]

In 2020, the Australian Federal Police (AFP) established the National DNA Program for Unidentified and Missing Persons (DNA Program) to provide a multidisciplinary forensic approach to unidentified and missing persons cases across all Australian jurisdictions.[9, 27] Their audit of open LTMP and UHR cases in 2022 revealed approximately 2,500 and

750, respectively (Table 1.1).[21, 28] Some of these cases date back over fifty years and some UHR have since been interred and are no longer available for forensic testing. A fit-for-purpose laboratory was established at AFP Forensics and specialists in forensic anthropology, forensic odontology, forensic geochemistry (e.g. isotope testing, radiocarbon dating), craniofacial reconstruction, forensic genetics and forensic investigative genetic genealogy (FIGG) were employed to scientifically link UHR and LTMP cases.[27, 29, 30] While all of these capabilities can be integrated into a multidisciplinary approach for resolving UHR investigations, this research focussed specifically on evaluating emerging forensic genomics techniques that offer new DNA tools and investigative leads to link LTMP and UHR cases.

1.2 Forensic Genetic Markers

1.2.1 Short Tandem Repeats

Short tandem repeat (STRs) markers were introduced in the 1990s and are non-coding segments of repeated DNA motifs, typically two to six base pairs in length.[31, 32] The number of repeats is defined as an allele and these are highly variable within a population, allowing for differentiation between individuals. Their highly polymorphic nature is partially due to their relatively high mutation rates of $1.2x10^{-3}$ per meiosis.[31]

Due to their ability to individualise, STR profiling has become the gold standard method for human identification in both criminal investigations, identifying a living or deceased person of interest who has been associated with a crime, and coronial investigations for UHR cases.[33] In many jurisdictions around the world, STRs are the only genetic markers acknowledged by the legal system to be used for DNA identification.[34] Combining multiple STRs into a panel increases the discrimination power and, therefore, the uniqueness of the DNA profiles. There are multiple panels that are available for forensic applications and their discriminatory power is estimated to be in the order of 10^{27} for a panel targeting 24 STRs.[35, 36]

1.2.2 Single Nucleotide Polymorphisms

Single nucleotide polymorphisms (SNPs) are variations in the genome at a single base that can be insertions, deletions or substitutions.[37] There are approximately 5 million SNPs across the human genome and they are typically biallelic, only having two possible alleles.[38, 39] The mutation rate is lower than those of STRs in the order of 10^{-8} per meiosis.[40, 41] Between the low mutation rate and fewer possible alleles, SNPs have a lower discriminatory power than STRs and therefore require larger panels to produce

profiles with a similar power of discrimination. Kidd et al. (2006) constructed a SNP panel showing that 50 – 100 SNPs are required in a panel to produce an equivalent discrimination power to a standard 10 – 16 STR profile.[40]

SNPs for forensic applications have been divided into four categories based on their characteristics and functions in a panel. Kidd et al. (2006) defined these categories as: identity-informative SNPs (iiSNPs), ancestry-informative SNPs (aiSNPs), phenotype-informative (piSNPs) and kinship-informative SNPs (kiSNPs).[40] Due to their low mutation rates relative to STRs, SNPs have become popular forensic genetic markers as they are less likely to mutate between generations.[42-45].

iiSNPs are most beneficial for individualisation due to their high heterozygosity and low population heterogeneity so that fewer SNPs are required to reach the high discriminatory powers observed with STR profiles. The low population heterogeneity minimises the differences in allele frequencies among different populations, which has been shown to result in match probabilities being nearly constant irrespective of the population in question.[46]

For the generation of DNA intelligence, aiSNPs have low heterozygosity and high population heterogeneity, the opposite of iiSNPs.[47] These SNPs have alleles that are shared by individuals within a population group yet differ to other population groups to allow the inference of biogeographical ancestry (BGA).[47, 48] Externally visible characteristics (EVCs), particularly pigmentation, can be inferred from the genotype of piSNPs.[49-52] In recent years, kiSNPs have been included in medium- to large-density SNP panels to be utilised for extended kinship analysis.[42-45]

1.2.3 Sex Chromosome Markers

The X and Y chromosomes are the two sex chromosomes in the human genome and determine the biological sex of an individual. Biological females will typically have two X chromosomes, whereas biological males will typically have an X and Y chromosome. Markers that have been utilised for forensic applications include Y chromosome STRs (Y STRs), X chromosome STRs (X STRs), Y chromosome SNPs (Y SNPs) and X chromosome SNPs (X SNPs).[53-57]

The Y chromosome is a small acrocentric chromosome consisting of a pseudoautosomal region and a non-recombinant region (NRY), of which the NRY makes up approximately 95%.[58-61] This region is of particular interest in forensics for paternity, kinship analyses involving paternal lineage and deconvolution of male-female mixtures.[54, 62] There are

approximately 200 Y STRs that have been identified in research and several forensic specific kits are available to profile 15 – 30 Y STRs.[63] The mutation rate of Y-STRs is in the order of 10^{-4} to 10^{-2}, and due to their direct paternal inheritance, have a low power of discrimination as male relatives will typically share identical profiles.[60, 61] Y SNPs, although not as extensively applied in forensic investigations, have been documented to have a lower mutation rate in the order of 10^{-8} per base pair per generation.[54, 63]

The X chromosome is considered one of the most stable nuclear chromosomes, with a single copy present in biological males and a pair in biological females. As a result, biological females inherit one X chromosome almost entirely unchanged from their father.[55, 64, 65] Since the mother carries two X chromosomes, these can undergo recombination during meiosis before being transmitted to either the biological male or female offspring.[65, 66] While X STRs and X SNPs are not commonly applied in forensic investigations, X STRs have been used for deficient paternity cases where a sibling or half-sibling is available and where autosomal and Y STRs provide insufficient statistical support for an identification to be made.[66-69]

1.3 DNA Identification

1.3.1 Likelihood Ratio

For an identification to be made through forensic DNA analysis, a quantitative or qualitative statement is required in the report.[70] In order to calculate statistical support, the frequencies of the observed alleles for a locus are required.[71, 72] Forensic databases containing allele frequency information have been developed for various population and subpopulation groups, as allele frequencies can vary between population groups.[73, 74]

The likelihood ratio (LR) is the most commonly applied statistical method in forensic DNA analysis.[75, 76] It is calculated as the ratio of conditional probabilities of observing DNA evidence given two alternative propositions, providing statistical support for the first hypothesis if the LR is greater than 1 and statistical support for the second hypothesis if the LR is less than 1.[70] A qualitative statement can then convey the statistical support using the verbal scale published by the Scientific Working Group on DNA Analysis Methods (SWGDAM).[70]

The LR for an individual locus can be estimated using population allele frequencies for the alleles that appear in the DNA profile. The LR for a multilocus DNA profile is then the product of the LRs for each individual locus. For use within a population group, forensic genetic population studies are necessary to confirm that the markers are in Hardy-

Weinberg equilibrium (within locus independence) and linkage equilibrium (between locus independence), and to establish the population-specific allele frequencies.[77, 78] Using the observed allele frequencies of the population group, the frequency of the DNA profile in question can be calculated, referred to as the random match probability.[77] The inverse of this random match probability is used to calculate the LR for direct comparison between a single source reference and an unknown profile, based on the hypothesis that the DNA originated from the same person.

Numerous population allele frequency databases have been curated for STRs and are routinely used for forensic casework.[71, 79-81] Equivalent databases for SNPs have been catalogued and made available through international projects including the 1000 Genomes Project and the International HapMap Project.[82, 83] However, SNPs have not been extensively applied to forensic casework outside of research. Recently, studies of iiSNPs have led to the publication of a number of allele frequency databases for populations and subpopulations in the US, Europe, Asia and South America.[43, 80, 81, 84-89]

1.3.2 Direct Reference Samples

During a missing persons investigation, antemortem (AM) samples may be collected by the police so that a DNA profile of the individual can be uploaded to relevant state or national law enforcement databases.[8] These are referred to as direct reference samples (DRS) and can be compared to postmortem (PM) DNA profiles obtained from UHR to search for an exact match.[90-92] DRS may be collected in the form of a retained medical specimen, biological keepsakes or personal items.

In some jurisdictions, medical specimens can be retained for diagnostic testing or research in medical science.[93, 94] In Canada, the presumptive identification of a deceased female was confirmed after investigators compared the PM profile to the STR profile obtained from cytological (PAP) smears retained by a provincial cancer agency.[95] Newborn screening cards are blood samples that are obtained from a heel prick of a newborn and stored on a special filter paper for genetic screening, retesting (if required) and research (if consent is provided).[94, 96] In Australia and Sweden, these specimens are stored in biobanks with records dating back up to 40 years (depending on the state), however, only Australia allows a Coroner to request the cards for DNA identification purposes.[93] Medical specimens are preferred as DRS and considered the most reliable source of biological material as they have a documented chain of custody due to the nature of sample collection, are less likely to be contaminated and are appropriately

stored to prevent degradation.[96] Newborn screening cards were successfully used to identify victims of the Victorian Bushfire disaster in 2009.[96, 97]

Family members may retain biological keepsakes such as exfoliated primary teeth or baby hair.[98] Teeth serve as a stable biological material for preserving DNA when stored at room temperature. A pilot study of 18 exfoliated primary teeth found sufficient DNA quantities (1.33 – 154 ng/μL) were recovered after being stored for up to 18 years, so could therefore be used for direct comparison to DNA profiles obtained from UHR samples.[98] Retained hair samples are more difficult to obtain a usable STR profile from depending on the presence of the root, where living cellular material contains nuclear DNA.[99] Rootless hair samples are usually only suitable for mitochondrial DNA (mtDNA) testing; however, optimised methods have been developed to improve the results obtained from downstream nuclear DNA testing.[100, 101]

In the absence of these types of samples, investigators can collect personal items of the missing person including their toothbrush, razors, hairbrush, shavers or combs, if available.[91, 102] These types of items have been extensively used for DRS as they contain biological material shed during the item's use. Personal items like these were widely used during disaster victim identification (DVI) events, including the World Trade Center attack in 2001 and the South East Asia tsunami in 2004.[102-105] However, the utility of these items often depends on time since collection, sample storage conditions and household living arrangements. Lau et al. (2005) highlighted the importance of verifying ownership of a personal item, as any household member may have used the item, potentially resulting in either a single source DNA profile that is not from the person of interest or a mixed DNA profile which could complicate its use.[103]

1.3.3 Familial Reference Samples

It is also recommended during a missing persons investigation to collect family reference samples (FRS) from close genetic relatives.[91, 92, 106] STR testing is typically limited to short-range kinship analysis due to the small panel size.[106-108] As a result, genetic relatives typically targeted are first degree relatives – parents, offspring and full siblings.[109] These profiles can be uploaded to relevant state and national databases for comparison to PM samples, potentially resulting in a partial or indirect match. Ge et al. (2011) recommends collecting multiple FRS as this can increase the LR and provide strong support for an identification.[108] If necessary, additional lineage testing using the Y chromosome or mtDNA can be conducted to confirm or refute the relationship and identification.[108, 110]

SNPs are increasingly being used for extended kinship analysis with FRS to supplement routinely used STRs or because the FRS are too distant for STRs to be informative. When attempting to identify 402 skeletal remains recovered from a 1948 mass disaster in the Jeju Province of Korea, several DNA markers were utilised (autosomal STRs, Y STRs and mtDNA) and resulted in 74 identifications.[111] Cho et al. (2020) obtained SNP profiles for comparison to a reference database of FRS SNP profiles, resulting in an additional 51 UHR being identified.[111]

1.4 Forensic Genomics

As sequencing technologies and bioinformatics advance, forensic DNA analysis has continuously improved, producing more sensitive, accurate, discriminatory and informative DNA profiles.[112-114] This progress has led to the increased application of SNPs as an alternative to STRs, extending forensic genomics capabilities beyond identification. There are four genotyping technologies that are currently applied to forensic casework: fragment length analysis, microarray genotyping, whole genome sequencing (WGS) and targeted amplicon sequencing (TAS).

Where forensic genomics technologies are not available in-house, agencies often outsource these capabilities to private forensic genomics service providers under contracts that specify testing deliverables, timeframes and costs. Due to the commercial nature of these arrangements, service providers may not be able to disclose full details of their laboratory, bioinformatics or kinship analysis procedures. When assessing potential service providers, agencies should consider a range of factors based on their specific needs, such as evidence type, specialised methodologies, database access, privacy and data security safeguards, validation and accreditation requirements and forensic casework and court reporting experience.[115] Additional logistical considerations include chain of custody, sample transport arrangements, sequencing turnaround times and costs of testing.

1.4.1 Fragment Length Analysis

Fragment length analysis is a genotyping method commonly used in forensic laboratories for STR analysis.[116] In this method, a DNA sample is amplified through polymerase chain reaction (PCR) with primers containing fluorescent dye labels and then the DNA fragments are separated by size using a capillary electrophoresis (CE) platform.[117-119] The DNA fragments are moved through capillaries via an electrokinetic injection, with a high-voltage charge applied to the buffered sequencing reaction. As the

DNA fragments pass through the detection window, a laser beam causes the dye labels to fluoresce. The fluorescent signals are detected by a charge-coupled device camera to generate an electropherogram, consisting of fluorescent signals that represent the alleles for each locus based on the length of the DNA fragments.[112, 119]

1.4.2 Microarray Genotyping

Microarray genotyping technology is primarily applied in medical genetics research and by genetic genealogy companies.[120, 121] This method involves the DNA sample being amplified, fragmented and immobilised on the surface of a silicon chip via oligonucleotide probes.[113, 122, 123] Fluorescently-labelled dideoxynucleic triphosphates are then used to determine the genetic sequence through single base extension. Reference genome sequences can be used to fill in any missing genotypes through imputation.[122, 124] This method can target anywhere from half a million to 2.5 million SNPs in a single assay, but manufacturer recommendations require high DNA input amounts (200 ng) that are not usually obtained from forensic samples. Several studies have tested microarray methods with low DNA input amounts from forensic samples, with as little as 0.2 ng of DNA producing high quality SNP genotypes.[125-127] However, microarray genotyping is severely impaired by low DNA quality, which is commonly encountered in UHR investigations where the DNA has degraded and fragmented.[126]

1.4.3 Whole Genome Sequencing

Whole genome sequencing was originally designed for clinical applications and sequences all three billion nucleotides in the human genome.[128, 129] This can be achieved using either short-read sequencing technologies, which involve fragmenting DNA through sonication or restriction enzymes, or long-read sequencing technologies, which sequence individual DNA molecules spanning several thousand base pairs.[122, 128, 130] One of the most commonly employed instruments for WGS is the NovaSeq™ 6000 System (Illumina), which is capable of sequencing up to 48 human genomes in 44 hours.[131] The resulting data requires extensive bioinformatics expertise to generate genotypes and impute incomplete sequences using haplotype patterns from a reference panel.[122] When higher amounts of template DNA are used, higher coverage is achieved, which reduces the need for imputation. While WGS provides extensive genetic information that forensic investigations can leverage, it has not yet been operationalised within a law enforcement forensic laboratory and the capability is typically outsourced to private forensic laboratories that have validated WGS.[44]

1.4.4 Targeted Amplicon Sequencing

Massively parallel sequencing (MPS) is a genotyping technology that can sequence millions of DNA fragments from multiple samples in a single sequencing run.[132] An increasing number of forensic laboratories are adopting benchtop MPS instruments to implement TAS applications.[133, 134] Forensic TAS panels have been designed to target several thousand forensically-relevant SNPs for a range of applications including identification, inferring BGA and EVC and extended kinship analysis.[135-137] Privacy risks associated with MPS can be mitigated by excluding SNPs that may reveal medical information about the individual.[138, 139] Additionally, TAS can be used to enhance STR profiling applications by determining both the length and DNA sequence variations of the target STRs, as opposed to only determining the STR alleles based on their length with fragment length analysis methods.[140]

1.5 DNA Intelligence

Forensic genomics enables investigators to infer valuable information about an unknown individual based on their genotype at specific loci. This can be beneficial to investigations where comparison to known references, either directly or through database searching, do not yield possible identities.[141, 142] The inferred characteristics, such as biological sex, EVCs and BGA, can generate new investigative leads that may help to narrow the pool of potential candidates. In UHR casework, this information can assist law enforcement in filtering through the large number of LTMP reports or identifying communities near where UHRs were recovered.[143-145] The DNA intelligence information can also complement findings from other forensic disciplines, such as forensic anthropology, craniofacial reconstruction, forensic odontology and forensic geochemistry, where biological sex, visible features and ancestral origins may also be inferred.[28, 146, 147]

1.5.1 Biological Sex

Biological sex estimation has long been used in forensic DNA analysis for criminal and coronial cases through DNA quantification and STR profiling. Both methods detect the presence of the male-specific Y chromosome to indicate the DNA originated from a biologically male individual.[35, 148] Forensic STR panels typically target the Amelogenin gene, identifying two possible alleles for the insertion/deletion (indel): X, representing the X chromosome, and Y, representing the Y chromosome.[35, 149] Several panels also include additional Y STRs, providing further support for the presence of DNA from a biological male individual. MPS panels that contain multiple classes of SNPs to generate

different forms of genetic intelligence, may also incorporate Y SNPs and possibly X SNPs to provide biological sex information and, if male, paternal lineage.[135-137, 150]

While the disclosure of an individual's biological sex in criminal investigations may raise privacy concerns regarding their gender identity[151], in UHR investigations, this information can significantly reduce the number of possible missing persons that might be linked to the UHR. It can also help to support or refute the biological sex inferred from circumstantial information or forensic anthropological examination of the remains. For example, in 2018, disarticulated human remains recovered on a beach in Cape Town were initially inferred to be female during the autopsy and anthropological examination and the police focussed on missing women in the area.[152] However, when forensic DNA analysis, including DNA quantification and STR profiling using fragment length analysis and MPS, was employed to determine if the disarticulated remains were from the same individual, the results indicated the remains were biologically male. This information provided the police with new investigative leads to pursue and they were able to link the UHR to a male reported missing at sea.

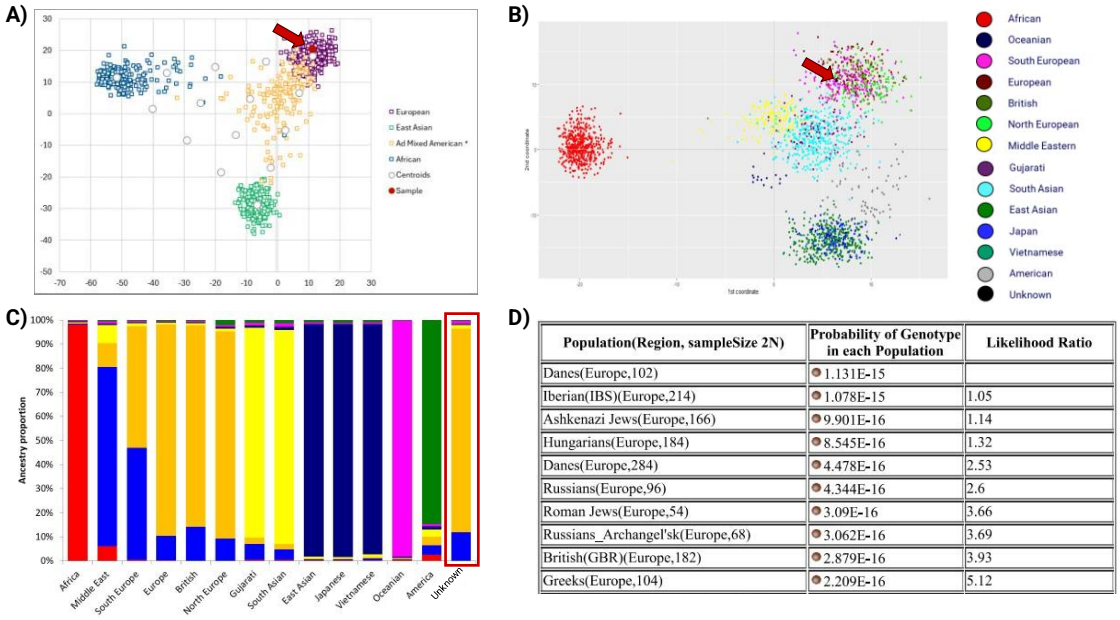
1.5.2 Biogeographical Ancestry

BGA refers to the ability to infer the likely recent ancestry of an individual from their aiSNP genotype.[153] This class of SNP is characterised by their low heterozygosity and high population heterogeneity, increasing the probability of the SNPs having alleles shared by individuals in the same population group but not shared among individuals of different population groups.[47] As forensic applications are limited to small- or medium-density SNP panels combining multiple SNP classes, limited subsets of aiSNPs are used to infer an individual's BGA in relation to superpopulation groups that encompass a continent (e.g. African, European, East Asian).[153, 154] Inferring BGA requires a reference database of genotypes with known ancestry against which a prediction algorithm can compare the queried genotype. However, due to colonialism, migration and modern mating practices, many individuals have recent ancestry from more than one population group (i.e. admixed ancestry), making BGA inference more complex.[154]

There are several prediction algorithms available for the inference of BGA depending on the number of aiSNPs to be evaluated.[155] The most widely used aiSNP panel is the Kidd Lab Panel which targets 55 aiSNPs and is compatible for analysis with most BGA inference pipelines.[47, 156] Multidimensional scaling (MDS) methods such as principle component analysis (PCA) and principle coordinate analysis (PCoA) reduces genotypes to two or three coordinates using eigenvalue decomposition.[141, 155, 157, 158] The resulting

scatter plot is then used to visualise the variance amongst samples, with individuals who share genetic similarities clustering together, reflecting the different population groups in the reference database (Figure 1.1).

Model-based likelihood estimators, such as Structure, estimate the ancestral population contributions to a questioned genotype by analysing a reference database of known ancestral genotypes.[159, 160] These population contributions can be graphed and compared to the contributions assigned to individuals in the reference database (Figure 1.1). The Forensic Research/Reference on Genetics-knowledge base (FROG-kb) calculates the probability of observing the questioned genotype within population groups from the Allele FREquency Database (ALFRED) and the output ranks the population groups by random match probability (RMP; Figure 1.1).[161, 162]



Population(Region, sampleSize 2N)	Probability of Genotype in each Population	Likelihood Ratio
Danes(Europe,102)	1.131E-15	
Iberian(IBS)(Europe,214)	1.078E-15	1.05
Ashkenazi Jews(Europe,166)	9.901E-16	1.14
Hungarians(Europe,184)	8.545E-16	1.32
Danes(Europe,284)	4.478E-16	2.53
Russians(Europe,96)	4.344E-16	2.6
Roman Jews(Europe,54)	3.09E-16	3.66
Russians_Archangel'sk(Europe,68)	3.062E-16	3.69
British(GBR)(Europe,182)	2.879E-16	3.93
Greeks(Europe,104)	2.209E-16	5.12

Figure 1.1. Biogeographical ancestry (BGA) inference of an individual with known European ancestry using different pipelines with 55 aiSNPs: A) principle component analysis (PCA) on the Universal Analysis Software (UAS), the red arrow indicates the questioned genotype; B) principle coordinate analysis (PCoA) in RStudio, the red arrow indicates the questioned genotype; C) Structure, the red box indicates the questioned genotype; D) Forensic Research/Reference on Genetics-knowledge base (FROG-kb) with the ten highest ranking subpopulation groups by random match probability and likelihood ratio (LR).

BGA inference has been used in UHR cases of coronial significance and historical investigations to assist in determining the appropriate management and repatriation process to follow.[163-165] For example, partial skeletal remains were recovered in the Himalayas in 2016, where approximately 30 foreign tourists had gone missing over the

past 20 years.[165] The BGA of the UHR was inferred to be of Asian origin from analysis of the Y haplogroup and aiSNPs using FROG-kb. This information allowed the investigating authority to focus on linking the UHR to a local missing persons case and circumvent contacting family members of the missing tourists unnecessarily.

1.5.3 Externally Visible Characteristics

In forensic genomics, it is becoming increasingly common to infer the phenotype or EVCs of an individual based on their genotype at piSNPs to provide new investigative leads.[166] The most successful applications have involved inference of human pigmentation traits such as hair colour, eye colour and skin colour.[49-52, 167] However, the expression of these pigmentation traits can be influenced by environmental factors and ageing.[166] The combination of craniofacial reconstruction for a UHR and inferences for pigmentation can provide law enforcement with more accurate facial depictions to be circulated within communities to assist in identifying the individual. Beyond pigmentation, other EVCs are currently being investigated to determine whether they can generate reliable inferences for traits such as body height, facial morphology and male pattern baldness.[168-172] Recent studies have also demonstrated that it is possible to estimate the age of an unknown individual by analysing epigenetic markers including methylated DNA.[173, 174]

The HIrisPlex panel is comprised of 24 piSNPs associated with hair colour (blond, brown, red and black), hair shade (light and dark) and eye colour (blue, intermediate and brown).[50, 52] It employs an online tool that uses a multinomial logistic regression (MLR) model trained on genotype data and known hair and eye colours of individuals within a reference database.[50, 52, 167] The accuracy of each test is expressed as an area under the receiver operating characteristic curve (AUC) and probability values (p-values) are reported for each possible pigmentation category based on analysis of the questioned genotype. These 24 piSNPs have been incorporated into several forensic kits for MPS applications in criminal and coronial casework.[135-137, 150, 175] The panel has since been expanded to target 41 piSNPs for the additional inference of skin colour (very pale, pale, intermediate, dark and dark-to-black) in the HIrisPlex-S panel.[45, 46, 48]

Similar to BGA inference, EVC inference pipelines have been applied to UHR of coronial significance and historical cases to provide additional information to aid identification. Chaitanya et al. (2017) used the HIrisPlex system on DNA samples extracted from bones and teeth obtained from World War II excavations in Slovenia, in conjunction with STR analysis, to identify the UHR.[146] Living relatives of the identified individuals confirmed that the HIrisPlex inferences were consistent with their EVCs, demonstrating applicability

of this capability for UHR, missing persons and DVI investigations. More recently, a craniofacial reconstruction was generated by the AFP DNA Program of a UHR recovered in 1986 at Kangaroo Island, South Australia.[176] The pigmentation applied to the reconstruction was informed by the piSNP and aiSNP profiles to generate inferences of the individual's hair colour, eye colour and BGA.

1.6 Genetic Kinship Testing Techniques for Human Remains Identification and Missing Persons Investigations

Forensic Genomics
Volume 4, Number 1, 2024
© Mary Ann Liebert, Inc.
DOI: 10.1089/forensic.2023.0018

Forensic Genomics

Open camera or QR reader and scan code to access this article and other resources online.

REVIEW ARTICLE

Genetic Kinship Testing Techniques for Human Remains Identification and Missing Persons Investigations

Jessica Lee Watson, BForSc (Hons),[1,2,*] Dennis McNevin, PhD,[1,2] and Jodie Ward, PhD[1,2]

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

Abstract

Medium- and long-range familial/kinship testing is demonstrating its potential to assist identification of unknown deceased persons when standard DNA identification techniques fail. Up until now, unidentified and missing persons casework has benefited from the use of well-established short-range familial/kinship testing techniques; however, law enforcement (LE) agencies and forensic DNA laboratories have begun to leverage new forensic genomics tools to extend kinship testing. Forensic investigative genetic genealogy (FIGG) is a form of long-range familial/kinship testing that has become increasingly popular in recent years. There are, however, considerations that need to be addressed before this technique can be routinely applied to the large numbers of unresolved identification cases. The majority of forensic DNA laboratories are unable to meet the specific technological and data analytical requirements for high-density single nucleotide polymorphism (SNP) genotyping. As a result, several private laboratories now provide FIGG services to LE agencies, including SNP genotyping and genealogical research. The development of bespoke targeted amplicon sequencing panels for conducting medium-range familial/kinship testing now presents forensic DNA laboratories with the capacity and capability to perform this testing in-house. Outside of the laboratory, governance frameworks will need to reflect the goals of human remains identification and missing persons investigations, while addressing privacy impacts associated with the handling of personal and genetic information by LE agencies.

Keywords: kinship testing, forensic genomics, forensic investigative genetic genealogy, unidentified human remains, missing persons, DNA identification

Introduction

The identification of human remains is highly important for administrative, legal, and humanitarian reasons, contributing evidence for death investigations and providing answers to family and friends of long-term missing persons (LTMPs).[1] A number of international agencies[1–3] endorse dental, fingerprint, and DNA analysis as the primary scientific methods for human remains identification. DNA analysis, or genotyping, is a powerful identification tool because of its ability to differentiate among individuals.

In its simplest form, this involves comparing a short tandem repeat (STR) profile recovered from unidentified human remains (UHRs) to an STR profile recovered from direct reference sample/s (DRS) of a nominated LTMP. The STR profile could also be uploaded to a relevant state, national, and/or international DNA database to be

[1]*Centre for Forensic Science, School of Mathematical and Physical Sciences, Faculty of Science, University of Technology Sydney, Ultimo, Australia.*
[2]*National DNA Program for Unidentified and Missing Persons, National Missing Persons Coordination Centre, Australian Federal Police, Majura, Australia.*

*Address correspondence to: *Jessica Lee Watson, BForSc (Hons), Centre for Forensic Science, School of Mathematical and Physical Sciences, Faculty of Science, University of Technology Sydney, Broadway, Ultimo, NSW 2007, Australia*, Email: jessica.watson@student.uts.edu.au

Table 1. Terminology used in this review

Term	*Definition*
UHR	The remains of a deceased individual with no known identity.
Missing Person	An individual reported missing to the police, whose whereabouts are unknown, and where there are concerns for the safety and welfare of that person.
LTMP	A missing person who has been missing for more than 3 months.
Coronial Investigation	An investigation to determine the identity of a deceased person (and when, where, and how they died).
Criminal Investigation	An investigation to determine the identity of a living or deceased person who has been associated with a crime.
MPS	The DNA sequencing technology where millions of fragments of DNA from multiple samples are sequenced simultaneously in one device. Also known as NGS. MPS will be the term used in this review.
Direct Match	A DNA profile exactly matches another DNA profile at every genetic marker (barring mutation), inferring the DNA profiles originated from the same individual with a high level of confidence.
Indirect Match	A DNA profile partially matches another DNA profile at a proportion of genetic markers, inferring the DNA profiles could have originated from genetically related individuals.
Kinship Testing	The use of indirect matching in a coronial investigation.
Familial Testing	The use of indirect matching in a criminal investigation.
Short-Range Familial/ Kinship Testing[a]	The process of searching a database using a low-density DNA profile with a limited number of markers to identify indirect matches with individuals who could be first-order genetic relatives (parent/offspring or full sibling).
Medium-Range Familial/ Kinship Testing[a]	The process of searching a database using a medium-density DNA profile with an intermediate number of markers to identify indirect matches with individuals who could be first- to fifth-order genetic relatives (out to second cousins).
Long-Range Familial/ Kinship Testing[a]	The process of searching a database using a high-density DNA profile with an extensive number of markers to identify indirect matches with individuals who could be first- to ninth-order genetic relatives (out to fourth cousins). Also known as FGG, IGG, or FIGG. FIGG will be the term used in this review.

[a]While a distinction is drawn here between short-, medium-, and long-range familial/kinship testing to reflect terms emerging in the literature, the differences are arbitrary as there exists a continuum from short- to long-range testing.

FGG, forensic genetic genealogy; FIGG, forensic investigative genetic genealogy; IGG, investigative genetic genealogy; LTMP, long-term missing person; MPS, massively parallel sequencing; NGS, next generation sequencing; UHR, unidentified human remain.

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

compared against a collection of DRS to obtain a direct match.[4]

In the absence of a direct match to a DRS, genetic kinship testing can be applied to compare the STR profile recovered from UHR with the family reference sample/s (FRS) collected for a nominated LTMP, or the UHR DNA profile can be searched against relevant FRS database/s to obtain an indirect DNA match.[5,6] In such cases, an unknown person can be identified through their genetic relatives by using the principles of genetic inheritance.[7] This can be applied to the identification of a person of interest in criminal investigations (referred to as familial testing hereafter) or the identification of UHR in coronial investigations (referred to as kinship testing hereafter). Table 1 introduces terminology used in this review.

In recent years, the incorporation of genetic genealogy into investigations of unresolved cases has allowed familial/kinship testing to extend beyond close genetic relatives. As such, an increasing number of law enforcement (LE) agencies are establishing policies and procedures for using forensic investigative genetic genealogy (FIGG), also known as long-range familial/kinship analysis, to assist unidentified and missing persons investigations when all other conventional identification methods have been exhausted.

In 2020, the Australian Federal Police (AFP) launched the National DNA Program for Unidentified and Missing Persons (hereafter called the AFP DNA Program).[8] The AFP DNA Program uses a multidisciplinary forensic approach to scientifically link UHR and LTMP cases in Australia, including emerging forensic genomics tools such as FIGG.[9,10]

Unidentified and Missing Persons Investigations

The large number of UHR and LTMP across the world has come to be known as a ''silent mass disaster.''[11] The true global extent is difficult to verify given discrepancies in reporting, database management, and standards for identification.[12,13] The volume of cases presents a workload challenge for LE agencies at the local, state, and national levels. In the United States, it is estimated that there are more than 40,000 UHR awaiting identifications,[11] yet fewer than 9000 have been entered into the National Crime Information Centre, a centralized national database managed by the Federal Bureau of Investigation (FBI) (Table 2).

In Australia, a recent case audit facilitated by the AFP DNA Program has recorded ~750 sets of UHR

Table 2. Numbers of reported unidentified and missing persons in the United States, the United Kingdom, and Australia

Country	*Reporting body*	*Year*	*Total population (million)*	*Missing persons reported*	*LTMPs*	*UHR*
United States	National Crime Information Centre[209]	2022	331.9	550,000	97,127	8450
	National Missing and Unidentified Persons System[210]	2022	331.9	600,000	22,593	14,332
United Kingdom	UK Missing Persons Unit[14]	2022	67.6	152,000	5295	1096
Australia	National Missing Persons Coordination Centre[211]	2022	25.7	55,000	2500	750

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

nationally,[9] whereas in the United Kingdom the latest reported number of UHRs is just over 1000[14] (Table 2). Some LE agencies acknowledge that their case numbers are likely to be significantly underestimated, either due to LTMP not being reported to the police, UHR not being catalogued accurately, or incomplete and inadequate reporting mechanisms.[12,14]

At the time a person is reported missing, a police investigation should be undertaken to collect detailed information about the individual, including the collection of vital antemortem (AM) records and samples. Ward[15] recommends that DRS should include a retained medical specimen (e.g., newborn screening card or biopsy sample) if available or at least two samples from personal items (e.g., hairbrush and toothbrush). In addition, FRS should be sought from multiple close genetic relatives if available.[16] The resulting DNA profile/s can be then stored in a relevant LE DNA database for comparison to postmortem (PM) DNA collected from UHR.

However, there are numerous challenges that are often faced with the collection and profiling of AM and PM DNA samples that complicate identification efforts, affecting both contemporary and historical cases. The most common is a lack of DRS or close genetic relatives to provide a suitable FRS. In historical cases, AM DNA samples were rarely collected or retained. In addition, PM DNA samples are often of insufficient quality and/or quantity.

The application of forensic genomics techniques, such as FIGG, addresses many of these issues. This includes decreasing required template amounts through DNA enrichment steps, increasing sensitivity of massively parallel sequencing (MPS) approaches, managing data loss through imputation, and circumnavigating a lack of reference DNA samples by increasing the potential FRS pool available for searching in international genetic genealogy databases.

The remit of several multidisciplinary international bodies involves assisting countries to resolve the identity of large numbers of unknown deceased persons, including through the application of forensic human identification techniques such as DNA testing. They include the International Commission on Missing Persons (ICMP)[17] and the International Committee of the Red Cross.[18]

To facilitate communication between jurisdictions, the International Criminal Policing Organization (INTERPOL) recommends their member countries implement a nationally coordinated DNA identification program[19] and utilize INTERPOL databases such as I-Familia[20] to aid missing person investigations. Similarly, after reviewing literature concerning the humanitarian crisis of worldwide UHR numbers, Reid et al.[13] supported the use of standardized identification procedures and the establishment of accessible, centralized national and international databases to cross-match UHR and LTMP data.

Examples of countries who have formed dedicated national programs to coordinate identification efforts across multiple jurisdictions and agencies include the National Centre for Missing Persons and Unidentified Remains in Canada,[21] the University of North Texas Center for Human Identification (UNTCHI) in the United States,[22] and, most recently, the AFP DNA Program in Australia.[10]

STRs and Short-Range Familial and Kinship Testing

Short-range familial testing was first developed to identify indirect matches in a criminal DNA database for the purpose of identifying close genetic relatives of the source of a crime scene sample.[23] The targets typically consist of first-order relatives—parents, offspring, and full siblings—whose DNA profiles already exist in the database.[7,24] Due to general and genetic privacy concerns about the use of genetic data, familial searching has utilized non-coding STRs*—segments of short, repeated DNA motifs—and supplementary lineage markers.[6,27]

Short-range familial testing was first implemented using the UK National DNA Database in 2002, with the first suspect identification and conviction made in 2004.[28–30] In Australia, it first resulted in a conviction in 2016.[31] Many countries do not allow this search method, and those that do ensure it is highly regulated.[28,32,33] However, if familial searching were used more frequently and the relevant references were present on the DNA database (e.g., convicted offenders), it is purported that many cases would have been resolved earlier, and without needing to employ more specialized methods such as FIGG.[30]

The use of short-range kinship testing to identify UHR through close genetic relatives for coronial investigations has been employed in numerous scenarios from domestic LTMP investigations (e.g., Kingscliff Beach jawbone[34]) to large-scale disaster victim identification efforts (e.g., 2001 World Trade Centre attack[35]). The process of short-range kinship testing is outlined in Figure 1. Identification can be made by applying two primary algorithms: parent/offspring and sibling.[28]

The former creates a subset of candidates by selecting only DNA profiles that share at least a single allele at each locus, allowing definitive exclusion if there are no alleles in common at a marker.[36] The latter creates a list of potential siblings ranked by the number of shared alleles or, preferably, likelihood ratio (LR).[37] Parents are the optimum FRS recommended, followed by offspring, then full siblings.[38] In simulations with STR data, Ge et al.[6] observed that when a second biological parent is available in combination with an offspring, the LR increases by more than 100 times.

This type of pedigree was used to identify the remains of Josef Mengele, following DNA testing of his wife

*The premise that STRs are non-coding has recently been challenged.[25,26]

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

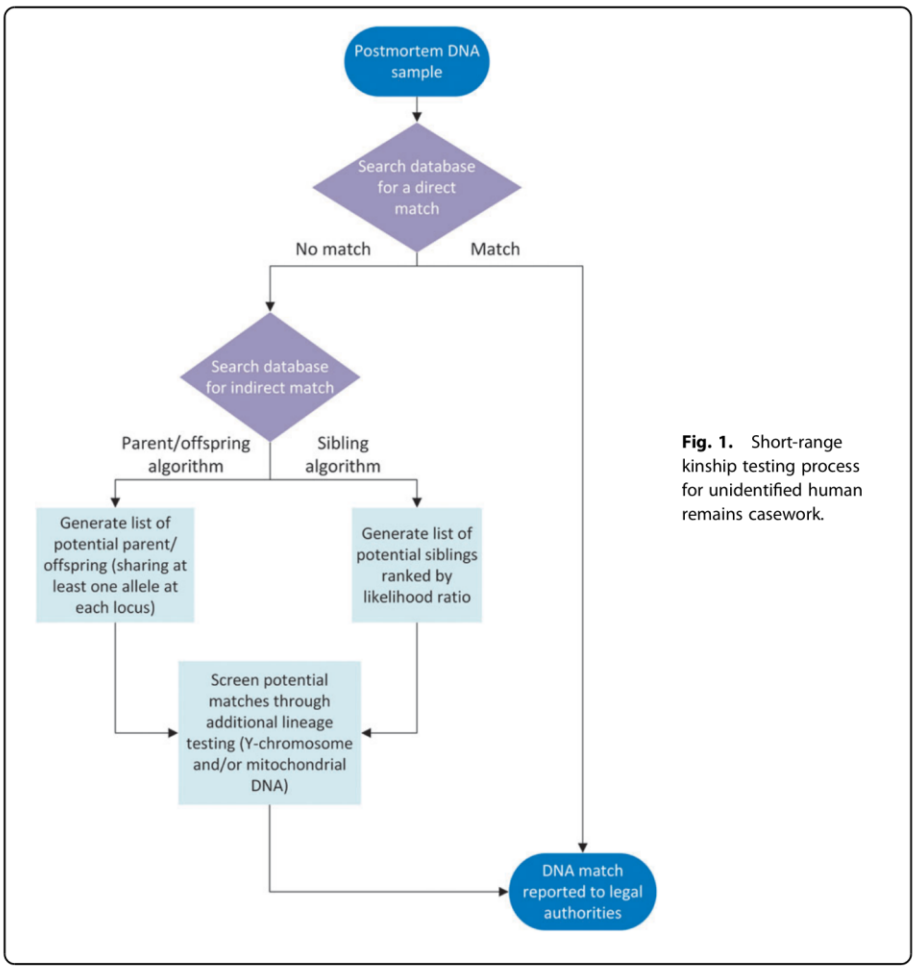


Fig. 1. Short-range kinship testing process for unidentified human remains casework.

and son.[39] For putative relatives, additional testing of supplementary genetic markers, such as lineage-based Y-chromosome (Y-DNA) and mitochondrial DNA (mtDNA) markers, can be conducted to confirm or refute relatedness.[23,40]

The success of short-range kinship testing is dependent on the type and number of FRS available.[6,41] The disadvantage of this method is that it is usually restricted to first-order relatives as they have a high enough probability of sharing more alleles than unrelated individuals. The limited number of STR markers in standard DNA profiles, together with their higher mutation rates (relative to single nucleotide polymorphisms [SNPs]), typically means more distant genetic relatives cannot be distinguished from unrelated individuals.[6]

However, in many cases, FRS from first-order relatives may not be available for comparison because the LTMP has not been reported (and therefore a FRS was never collected), the relatives are now deceased, or the relatives decline to provide an FRS.

There have been cases where indirect matches have occurred beyond first order using a combination of DNA markers. In Australia, a jawbone was found on a Sydney beach in 2011 and an STR profile was recovered.

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

At this time, there were no direct matches in the National Criminal Investigation DNA Database (NCIDD) or indirect matches to FRS housed on the New South Wales (state-based) DNA database.[34]

A convicted offender's STR profile was uploaded to the database in 2020 and was subsequently identified as a potential male relative following a short-range familial search. The individual had a paternal uncle who had been lost at sea in 1979 and his STR profile partially matched the UHR STR profile. Y-chromosome STR (Y-STR) testing was conducted to support the identification by confirming the paternal lineage.

SNPs as an Alternative to STRs

Expanding the reach of kinship testing to include a wider range of useful genetic relatives can overcome the restrictions experienced when searching only first-order relatives. One way of achieving this is through the adoption of alternative marker sets such as SNPs.[42,43] These are single points of variation in the genome and are more abundant than STR markers.[44] Using medium- to high-density SNP profiles, kinship testing capabilities can be extended beyond what is achievable with STR profiling. There are three technologies that can produce these SNP profiles: microarray genotyping, whole genome sequencing (WGS), and targeted amplicon sequencing (TAS) (Table 3).

Microarray genotyping

Microarray technology was developed nearly 20 years ago for application in medical genetics research.[45,46] The whole genome is amplified and fragmented, and the oligonucleotides are hybridized to complementary oligonucleotide probes that are immobilized on the surface of a silicon chip. Allele-specific oligonucleotide (ASO) microarrays (originally marketed as Affymetrix) employ ASO probes that bind fluorescently labeled DNA fragments.

Silicon bead chip arrays employ primers that bind to fragmented DNA, which then undergo single base extension with fluorescently labeled dideoxynucleotide triphosphates at the SNP position.[47] Genotyping may also involve imputation, a method using reference genome sequences with complete genetic information to fill in missing genotypes according to linkage patterns.[48,49]

Due to their fast, reliable, and cost-effective coverage of up to 2.5 million SNPs, microarrays have been assessed for forensic implementation for the past 15 years.[50–52] However, most forensic samples are unable to meet the recommended input amount of 200 ng of DNA, which is significantly greater than the standard 1 ng DNA input for forensic assays used routinely.[53] This is further complicated by a limited tolerance of degradation and inhibition.[54]

In a recent study, however, Russell et al.[55] demonstrated that accurate genotypes with SNP call rates >95% can be obtained with Illumina Global Screening Array (GSA) microarrays from DNA input as low as 0.2 ng, significantly less than the manufacturer's recommended input of 200 ng. Similarly, de Vries et al.[56] explored the robustness of microarray technology with blood and bone samples of varying quality and quantity. A kinship classification success rate of 98.5% was achieved with as little as 0.25 ng DNA input for first- to third-order relatives (i.e., first cousins), and 1 ng for fifth-order relatives (i.e., second cousins). Decreasing DNA quality resulted in increased fragmentation and lower accuracy, leading to relationships being undetected. Another study using microarrays to identify UHR concluded that high-quality samples are a necessity for successful analysis.[57]

Whole genome sequencing

WGS is designed to yield all 3 billion nucleotides in the human genome from which high-density SNP profiles can be extracted.[58,59] WGS was designed for clinical applications and operates by fragmenting DNA through sonication or restriction enzymes.[60] Adapter sequences and oligonucleotide barcodes are ligated to the fragments before they are sequenced. As opposed to microarrays, WGS requires a high level of bioinformatics expertise and computational workload for genotype calling. Low coverage WGS often requires a higher degree of imputation than what is used for microarray genotyping.[48]

The primary advantage of WGS over microarray techniques is the lower DNA input requirement, with most workflows requiring an input of 0.05–50 ng.[61] Lower

Table 3. Comparison of the genotyping technologies for generating medium- to high-density single nucleotide polymorphism profiles

Metric	*Microarray genotyping*[212]	*Whole genome sequencing*[58,61]	*Targeted amplicon sequencing*[140]
SNP Typing Method	Hybridization and single base extension	Single base sequencing	Single base sequencing
Manufacturer Recommended DNA Input	≥200 ng	0.05–50 ng	≥1 ng
SNP Markers	650,000–2,500,000	≤3,000,000,000	100–10,000
Suitable for Degraded Samples	No	Sometimes	Yes
Bioinformatics Expertise	Optional	Required	Kit dependent
Data Manipulation Options	Analytical and interpretation thresholds, imputation	Analytical and interpretation thresholds, imputation	Analytical and interpretation thresholds

SNP, single nucleotide polymorphism.

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

input amounts, however, usually result in lower coverage and an increased requirement for imputation. With the higher density of SNP data produced, Li et al.[59] concluded that WGS has the potential to increase detection power for distant genetic relatives by 5–15% over microarrays. This small gain is tempered by the fact that current WGS technologies are not commonplace in LE and government forensic laboratories.[62]

Targeted amplicon sequencing

TAS is beginning to be utilized by an increasing number of forensic laboratories that employ MPS, making it the most easily integrated genotyping technology for forensic SNP analysis.[62] The design of condensed subsets of SNPs (''panels'') ensures the targeting of more informative SNPs and the removal of superfluous medical information not relevant to forensic applications.

Moreover, the smaller number of SNPs, combined with limiting sequencing to the short fragments of DNA surrounding the target SNPs, improves the recovery of DNA from forensic samples of varying quantity and quality (see Development of TAS Kits section). While medium-density TAS genotypes decrease the range of genetic relationships able to be detected compared with microarrays and WGS, they are able to be generated in accredited forensic laboratories.[63]

Kinship Analysis: From Identity by State to Identity by Descent

There are two major analytical methods that can be utilized to infer relatedness between two individuals based on inherited DNA. The first is the use of an LR to provide probabilistic support for two alternative propositions about kinship. For example:

H_1: The donor of profile A and the donor of profile B are first cousins.

H_2: The donor of profile A and the donor of profile B are unrelated members of a specified population.

Some of the genetic markers in profiles A and B will be identical by state (IBS), and some will be different.[41,64,65] The more closely related the donors, the more markers are likely to be IBS. This is the method used for short-range kinship testing with STRs, but it can also be used for short- and medium-range kinship testing with SNPs.[66]

However, just because an SNP is IBS in two different profiles does not mean that the alleles are inherited from a common ancestor and are therefore identical by descent (IBD). The mutations responsible for them being IBS may have occurred independently in two different lineages. It is possible to estimate the proportion of SNPs that will be IBD for different orders of relationship (Table 4).

SNPs in close proximity on the genome that are IBS form a haplotype, and the larger or longer the haplotype, the greater the probability that they are IBD.[67] This is because the genetic process of recombination, which shuffles autosomal DNA (atDNA) on chromosomes during meiosis, is less likely to have disrupted the haplotypes for individuals that are separated by less meioses (i.e., are more closely related).

While the LR approach is suitable for closely related individuals that share a high enough number of DNA markers by descent (and, hence, are more likely to have enough markers IBS), second-order relatives and beyond are more difficult to differentiate from unrelated individuals. The preferred inference method for detecting relatives with greater genetic distance is the number and length of shared uninterrupted haplotypes.[68]

This is the approach used for FIGG. Two genetic relatives will share a distribution of uninterrupted haplotypes (IBD segments) of various lengths and frequencies, depending on their order of relationship. When IBD segments shared by two individuals are longer and more frequent, there is a shorter lineage path between the individuals as the haplotypes are less likely to have been disrupted by recombination events over time.[69,70] Once the kinship order is determined between two individuals, the challenge is specifying the type of biological

Table 4. Theoretical proportion of DNA shared by descent as a function of kinship order and the corresponding possible genetic relationships

Kinship order	*Relationship between individuals*	*Theoretical proportion of DNA shared by descent (%)*
0	Self, identical twins	100
1	Parents, offspring, siblings	50
2	Grandparents, grandchildren, aunts/uncles, nieces/nephews, half-siblings	25
3	First cousins, great grandparents, great grandchildren, great aunts/uncles, great nieces/nephews	12.5
4	First cousins once removed	6.25
5	First cousins twice removed, second cousins	3.13
6	First cousin thrice removed, second cousins once removed	1.56
7	Second cousins twice removed, third cousins	0.78
8	Second cousins thrice removed, third cousins once removed	0.40
9	Third cousins twice removed, fourth cousins	0.20

Source: The H600 Project.[213]

relationship.[67] For example, grandparents and half-siblings of the same individual are both second-order relatives of that individual.

Genetic Genealogy

Traditional genealogy is the study of family pedigrees through the use of historical records to track ancestors and their descendants.[71] This research is highly dependent on the availability of written records, with the search typically stalling when records are lost, destroyed, or not kept.[72] The combination of traditional genealogical research with DNA testing, referred to as genetic genealogy, is a more robust application of standard ancestry research.[73]

Lineage markers recovered from Y-DNA and mtDNA were first employed for genetic genealogy testing as they are uniparentally inherited from the biological father and mother, respectively, and do not undergo recombination, unlike atDNA. Their non-recombined transfer from generation to generation makes them a useful tool for inferring direct paternal or maternal relatedness.[74–78] Y-DNA originally inspired the genetic genealogy revolution, with many individuals conducting their own surname studies.[79]

atDNA testing employing IBD segment matching was introduced to genetic genealogy in 2009, expanding the possibilities beyond patrilineal or matrilineal searching.[80,81] With the ability to detect more complex and distant relationships with greater accuracy, atDNA profiles can generate a list of matches with the predicted kinship order consistent with the matched IBD segments. As the cost of high-density SNP genotyping has decreased and genetic genealogy databases have grown, genetic genealogy has become an increasingly useful tool for extended kinship testing.[80]

Since it emerged in the early 2000s, direct-to-consumer (DTC) genetic testing has been an ever-growing phenomenon, enticing the public to access information derived from their genetics without involving a physician.[82] Online tools allow consumers to compare their genetic information with others in the companies' databases to find genetic relatives.[83]

The rising popularity and use of DTC genetic testing has been accompanied by an increase of users in these genetic genealogy databases.[30] As a result, genealogical records are becoming progressively more available online to supplement kinship testing. Erlich et al.[84] demonstrated theoretically that only 2% of the population is required to be included in such databases before the probability of finding a third or fourth cousin approaches 100% for any sufficiently high-density SNP profile.

The four leading companies offering these services, in order of database holdings, are AncestryDNA®, 23andMe®, MyHeritage™, and FamilyTreeDNA.[85–88] Over 40 million people have taken a DTC saliva or buccal swab test in their homes to be sent off to private laboratories and analyzed using bioinformatics.[89] The returned results include estimates of biogeographical ancestry (BGA), kinship, and genetic disposition for diseases.[90]

Nearly all major DTCs use high-density SNP microarrays to generate DNA profiles comprising 650,000 to 1 million SNPs for each consumer.[91] Genetic relatives are located by measuring shared IBD segment lengths but each company uses different detection and match thresholds,[84,92] meaning that relationship predictions can differ between companies.[30]

As well as the four major companies previously mentioned, there are a number of genetic genealogy databases that can be utilized by consumers and/or LE agencies to identify genetic relatives.[30,93] Once genetic analysis has been completed for any consumer, they are able to download their genetic data from their DTC company of choice and then upload it to specific genetic genealogy databases for cross-comparison with DNA profiles generated by other DTC companies. They can also upload personal information and constructed family trees and genealogical records.

While AncestryDNA and 23andMe do not facilitate upload of DNA profiles from other DTC companies, FamilyTreeDNA, MyHeritage, and GEDmatch allow uploads of DNA profiles from other sources.[91,94] For example, GEDmatch uses a One-to-Many Kinship Tool to compare a user's DNA profile (produced by any DTC company) to all public DNA profiles (produced by any DTC company) in their database, returning a list of users that are ranked by the total length of shared IBD segments. Concerns about the privacy of personal and genetic data on these databases have led DTC companies to clearly articulate their privacy policies and terms of service.[49,95–98]

Forensic Investigative Genetic Genealogy

Long-range familial/kinship testing, also referred to as FIGG, utilizes specific private and/or public genetic genealogy databases for comparing the DNA profile recovered from UHR (or crime scene samples) to the database population of DNA profiles from consenting individuals to find genetic relatives of the unknown sample donor.[99–102] This technique has proven useful in cases where the individual or their close genetic relatives are not present in an LE DNA database, instead providing investigative leads to infer their identity through genetic genealogy.[103,104]

The process of conducting FIGG testing is outlined in Figure 2 and is typically only used after all other avenues of inquiry have been exhausted.[100,105] A medium- to high-density SNP profile is uploaded to one or more genetic genealogy databases that allow LE matching. If suitable genetic relatives appear in the match list, then the identity of the unknown individual may be able to be inferred by building out their family tree.

By combining the database search results with genealogy research using public, historical, and government

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

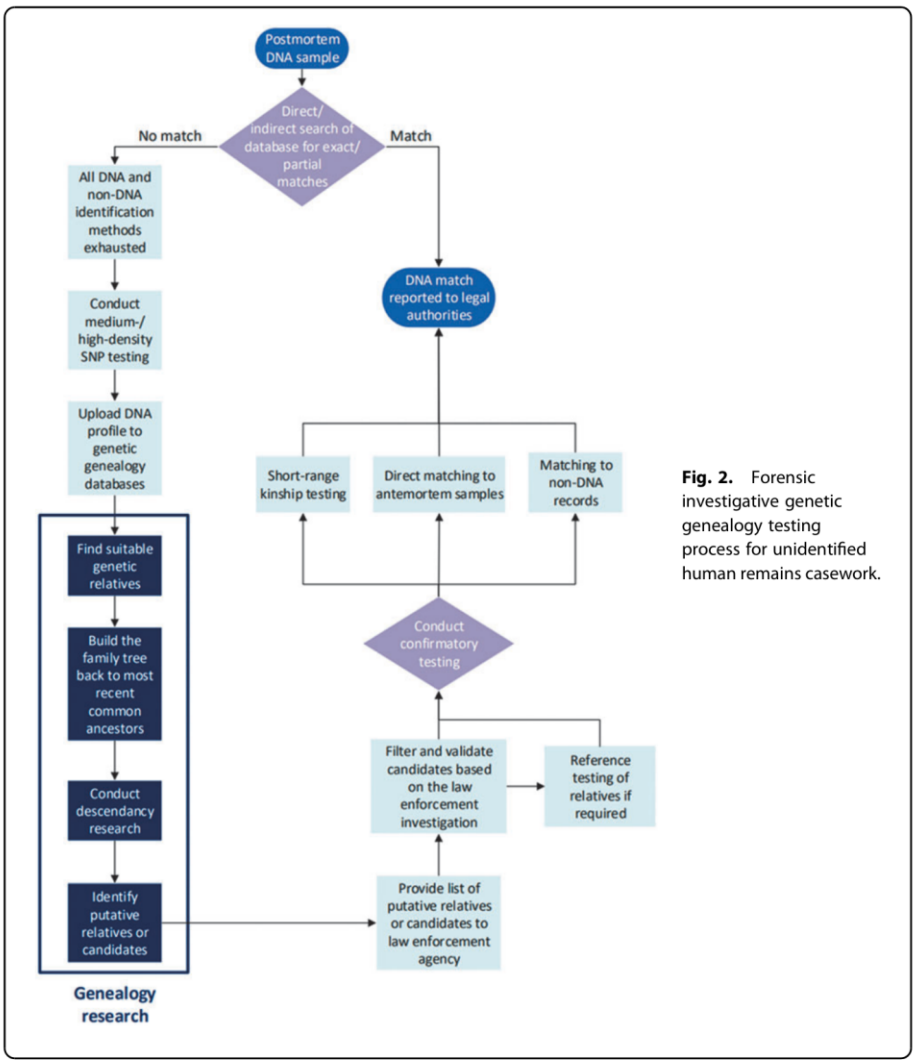


Fig. 2. Forensic investigative genetic genealogy testing process for unidentified human remains casework.

records, a genetic genealogist, or FIGG practitioner, can link the person of interest with firstly their distant and then close genetic relatives. A list of relevant putative genetic relatives or candidates can then be provided to the LE agency to investigate. LE can use investigative techniques, non-public records, and reference testing of putative relatives to populate and refine the tree.

The aim is to discover the most recent common ancestor/s shared by the UHR and the putative genetic relatives and then research the descendants until a present-day family with a missing or unaccounted-for relative is found. The identity of the UHR will then be confirmed using an accepted scientific method such as kinship analysis of STR profiles produced from the UHR and first-order relative/s.[106]

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

LE agency use of FIGG

In the United States, LE agencies considering using FIGG in a criminal investigation, including for the purpose of identifying the remains of a suspected homicide victim, should be adhering to guidance published by the Department of Justice in their 2019 interim policy and endorsed by LE accessible genetic genealogy databases.[100,107,108] Subsequent to this, the American Society of Crime Laboratory Directors released a position statement outlining eight criteria that need to be met before FIGG should be considered for use in a criminal case.[109] Since then, numerous agencies around the globe have released guidelines for the use of FIGG.[105,110,111]

The two public genetic genealogy databases currently accessible to LE are GEDmatch (via GEDmatch PRO™) and FamilyTreeDNA. LE agency access is authorized for investigating UHR and violent crimes such as homicide, sexual assault, manslaughter, robbery, and common assault.[107,112] Users of these two databases can consent to have their DNA profiles used by LE for the purpose of identifying UHR; this is the default for users of GEDmatch but requires users of FamilyTreeDNA to opt in. Users of GEDmatch may further opt in for their profiles to be used to identify the perpetrators of violent crimes, whereas there is no distinction between these two LE uses for FamilyTreeDNA.

Therefore, LE can access all GEDmatch profiles and those of individuals who have opted in to LE matching on FamilyTreeDNA for UHR cases, whereas LE is restricted to only searching against DNA profiles from users who have specifically opted in for LE matching on both of these databases for violent criminal cases.

At the time of writing, the GEDmatch database contains over 1.8 million profiles uploaded by the public;[102] all of which are available for identifying UHR. After GEDmatch was acquired by Verogen, Inc. (now a QIAGEN company) in December 2019, GEDmatch PRO was created as a dedicated LE portal to improve regulation and oversight of LE use.[113] GEDmatch PRO allows LE to upload an unknown DNA profile and search against the GEDmatch database for potential genetic relatives.

The FamilyTreeDNA database also allows LE matching against DNA profiles of consenting users.[114] At the time of writing, the FamilyTreeDNA database has ~1.7 million profiles uploaded by the public; however, only ~95% of these are available for identifying UHR.[102] FamilyTreeDNA is operated by Gene by Gene, Ltd. (a subsidiary of myDNA, Inc.); in January 2024, they announced they have partnered with Othram, Inc., to enhance the database's search capabilities for FIGG.[115]

Use of public genetic genealogy databases by LE has been, and continues to be, highly contentious. In the wake of the Golden State Killer case solve announcement in 2018, and the subsequent proliferation of FIGG use, GEDmatch and FamilyTreeDNA changed their site terms and policies in 2019 to require users to opt in or out of LE matching.[93]

GEDmatch opted all existing users out, requiring them to manually change their settings to be included; FamilyTreeDNA, however, opted all users in by default. These actions, although very different, were designed to enhance the informed consent process. However, Ram and Roberts[112] noted that people are unlikely to change the default setting, and this corrupts the notion of ''informed consent.'' In 2021, Verogen, Inc. modified the terms of service at GEDmatch PRO to allow all UHR profiles uploaded for identification purposes to be searched against the entire database, instead of just those individuals who have opted in specifically to LE matching.[107]

Currently, this still stands for GEDmatch users, whereas FamilyTreeDNA users still have the option to opt out of LE use for both UHR and criminal cases. In light of very recent reports of the improper use of public databases by some FIGG practitioners,[116] companies such as Verogen, Inc. have initiated new measures to protect the privacy and trust of their database users.[117] For example, cancelling all FIGG practitioners access to GEDmatch PRO and compelling them to re-consent to the terms of service by signing a new binding contract before database access is reinstated.

There are two genetic genealogy databases that are not available for public use and established specifically for LE matching. DNASolves® is a private genetic genealogy database created by the private company Othram, Inc., in 2019.[118] Individuals can populate the database and assist investigations by either submitting a saliva sample for sequencing at Othram, Inc., or providing their existing DTC genetic data.[48]

It is not known how many profiles are currently housed in DNASolves. DNA Justice™ is a private genetic genealogy database created by the non-profit organization DNA Justice Foundation in 2023.[119] At the time of writing, over 500 profiles have been uploaded by the public since the database was launched in early 2023. These LE exclusive genetic genealogy databases may serve to minimize some of the genetic privacy concerns raised to date surrounding LE agencies' use of public genetic genealogy databases if and when they increase to a useable size.

FIGG service providers

The majority of forensic DNA laboratories do not have the capability to generate the medium- to high-density SNP profiles required for FIGG. At the time of writing, the main service providers available for outsourcing one or more components of FIGG in the United States are primarily private laboratories; however, there are a growing number of not-for-profit, LE, and public laboratories beginning to offer FIGG services globally.

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

Bode Technology™ (Bode Cellmark Forensics, Inc.), DNA Doe Project, Inc., DNA Labs International, Inc., Identifinders International, LLC., Othram, Inc., Parabon NanoLabs, Inc., Signature Science, LLC., Gene by Gene, Ltd., Intermountain Forensics, and Astrea Forensics have accepted FIGG cases from LE for SNP genotyping for a number of years;[120–129] with the AFP DNA Program[8] and UNTCHI[130] more recently able to provide in-house FIGG services for LE and medicolegal agencies in Australia and Texas, respectively. These laboratories offer either microarray genotyping, WGS or TAS, with various sample acceptance criteria and marketed areas of specialization.

Some of these service providers conduct in-house genealogical research and others outsource this capability to individual FIGG practitioners or a genetic genealogy company to return investigative leads to the LE agency. The number and diversity of laboratories across the world validating and offering FIGG-related services could be expected to increase in the future.

First use of FIGG for UHRs

At the time of writing, over 260 predominantly US-based UHR cases have been reported to be solved by different investigating entities using FIGG (including government, private, and not-for-profit), resulting in the resolution of many LTMP cases.[131] In 2018, the first LE case was reported to be resolved using FIGG and highlighted its utility where all other forensic identification techniques had failed. The ''Buckskin Girl'' was found on the side of a road in Ohio in April of 1981, having suffered blunt force trauma and strangulation.[30]

After the police failed to identify her using a suite of forensic tools, she was buried as a Jane Doe later that year with no leads on her identity or the individual/s responsible for her death.[103] Forensic tests applied to this case included dental and fingerprint analysis, craniofacial reconstruction, pollen analysis, stable isotope analysis, and STR and mtDNA profiling.[132,133]

The DNA Doe Project adopted her case and used publicly raised funds to conduct high-density SNP genotyping on a retained blood sample. When uploaded to the GEDmatch database in 2018, a first cousin once removed (fourth-order relative) was found.[133] Within hours, FIGG practitioners had generated solid leads for the investigating LE agency, resulting in Marcia Lenore King being identified after 37 years. Her identity was confirmed through kinship analysis using an FRS from a first-order relative.[134]

Development of TAS Kits

A recent advancement to make FIGG more accessible to LE agencies is the use of TAS to generate a medium-density SNP profile, making it possible to embed all analytical and genealogical processes within an LE or forensic facility. With the reduction in SNPs, the range of detectable relationships also decreases, defining medium-range familial/kinship testing.

While not as powerful as longer range genetic genealogy methods, it can still provide important leads to LE who are investigating the identity of an unknown person. Further, medium-range familial/kinship testing can be used for within case matches where the available relative is more distant than is suitable for short-range testing with STRs. For the purposes of this review, medium-range familial/kinship testing is defined as inferring out to fifth-order relatives (Table 1).

The primary advantage of these methods over microarrays and WGS is that only regions determined as relevant for forensic applications are sequenced, such as identity-, phenotype-, ancestry-, and kinship-informative SNPs.[48] This means that in addition to generating an SNP profile suitable for extended kinship testing, other genetic intelligence can be generated simultaneously; phenotype-informative SNPs can be used to predict externally visible characteristics (EVCs), and ancestry-informative SNPs can be used to predict BGA.

Further, medically informative SNPs are excluded as a privacy enhancing feature[95,135] and all SNPs are suitable for calculation of LRs, assuming that population SNP allele frequencies are available, as well as IBD segment matching. Further, keeping all processes associated with FIGG in-house ensures physical control over the sample and all information generated, providing data quality assurance and chain of custody compliance.[105]

This approach addresses privacy concerns regarding all aspects of data handling by third parties on or offshore. It also makes in-house data storage and back up achievable for most laboratories that might otherwise struggle with the huge datasets associated with WGS.

There are three TAS kits that have been developed for use in unidentified and missing persons investigations thus far; however, only two are now currently commercially available (Table 5). The first panel, the ICMP Missing Persons SNP Panel (QIAGEN), comprised 1270 tri-allelic SNPs but it is no longer available.[136–138] The second is the ForenSeq® Kintelligence Kit (Verogen, Inc.) that analyzes over ten thousand SNPs in a single assay.[139] The third kit is the FORensic Capture Enrichment (FORCE) panel, which interrogates about half the number of SNPs in the Kintelligence Kit but places more emphasis on non-kinship SNPs of forensic importance.[63]

ForenSeq Kintelligence Kit

The Kintelligence Kit was launched by Verogen, Inc. in February 2021 as a fully integrated forensic-specific SNP kit for extended kinship testing.[140] The Kintelligence workflow is compatible with the MiSeq FGx®

Table 5. Currently available forensic single nucleotide polymorphism panels designed for medium-range familial/kinship testing and genetic intelligence applications

Kit	*FORCE panel*[63,155]	*ForenSeq® Kintelligence Kit*[139]
Total Number of SNPs	5422	10,230
Kinship SNPs	3931[a]	9867
Identity SNPs	138[a]	94
Phenotype SNPs	41[b]	22[c]
Ancestry SNPs	237[b]	56[c]
X Chromosome SNPs	246	106
Y Chromosome SNPs	829	85
Sequencing Methods Available	Hybridization capture; primer extension capture; multiplexed PCR	Multiplexed PCR
Recommended DNA Input	1 ng	1 ng
Kinship Order	≤Fifth-order relatives	≤Fifth-order relatives
Kinship SNPs Required	≥25% for beyond first-order relatives	≥70% for upload to GEDmatch PRO™

[a]Two SNPs overlap the kinship and identity categories in the FORCE panel and are only counted in the kinship category.
[b]Four SNPs overlap the phenotype and ancestry categories in the FORCE panel and are only counted in the phenotype category.
[c]Two SNPs overlap the phenotype and ancestry categories in the Kintelligence Kit and are only counted in the ancestry category.
FORCE, FORensic Capture Enrichment; PCR, polymerase chain reaction.

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

Sequencing System, using established ForenSeq library preparation methods.[141] The manufacturer recommends at least 1 ng of input DNA to produce reliable results; however, published internal validations have tested as little as 50 pg of DNA with high call rates and genotype concordance.[142–145]

The kit targets 10,230 SNPs, including 9867 kinship SNPs that have been curated from those present on GEDmatch. Forensic-specific SNPs to infer identity, BGA and EVCs, as well as sex-chromosome SNPs, have also been included to maximize the genetic information produced in a single sequencing run. Antunes[143] discussed the advantage of the smaller amplicon lengths utilized in this kit (the average below 150 base pairs), allowing high call rates for degraded samples.

Degraded blood and contemporary bone samples, with degradation indices as high as 158 and 14 respectively, have been profiled by the Kintelligence Kit.[140] Staadig et al.[144] demonstrated that bone samples ($n=4$) experienced higher rates of allele dropout and highlighted the importance of setting appropriate thresholds for analysis. When testing 11 bone samples ranging from 0.1 to 1.0 ng input, Peck et al.[145] reported high call rates exceeding 96.9% for seven of the samples, much higher than those recorded by Staadig et al.[144] The Kintelligence Kit has been tested on a wide range of compromised forensic samples, including bone, teeth, blood, and nail, all of which had call rates greater than 90%, demonstrating its utility for UHR casework.[146]

Verogen, Inc. have ensured that the corresponding Kintelligence Analysis Module produces forensic-appropriate reports on BGA and EVCs, as well as kinship reports formatted for direct upload to GEDmatch PRO.[147] However, a minimum of 70% of the relevant SNPs need to be genotyped in order for the report to be generated. In addition to IBD segment matching, GEDmatch PRO employs a windowed kinship algorithm specifically for Kintelligence genotypes and a simulation study reported high performance for kinship determination for first- to fourth-order relatives.[148] When the algorithm was tested with Kintelligence genotypes generated for a known family group, the GEDmatch PRO One-to-Many tool was able to detect and accurately predict relationships for first- to fifth-order relatives.[146]

A number of forensic DNA laboratories have now validated and accredited the Kintelligence Kit as an SNP genotyping method for FIGG in accordance with ISO/IEC 17025:2017 requirements. The private laboratory DNA Labs International, Inc. was the first reported accredited laboratory to validate and successfully employ the Kintelligence Kit for FIGG casework, announcing the resolution of a 1998 UHR case in collaboration with Oregon State Police in February 2022.[149]

Toward the end of that year, the private laboratory Signature Science, LLC published their internal validation of the kit[127,145] and the not-for-profit laboratory Intermountain Forensics[150] reported receiving accreditation for MPS-based SNP genotyping on their website; both laboratories are now offering Kintelligence Kit sequencing as one of their FIGG services.

In 2023, the AFP DNA Program became the first accredited LE laboratory (and the first laboratory outside of the United States) to operationalize the kit[146] and has since used it to successfully resolve multiple cold and current UHR cases in Australia.[151,152] Also in 2023, UNTCHI became the first public laboratory in the United States to gain accreditation for SNP genotyping (including using the Kintelligence Kit) and will likely begin using this kit for select FIGG cases in Texas.[153]

FORCE panel

The FORCE panel, developed in 2021 by Tillmar et al.,[63] is a custom hybridization capture assay with extended kinship capabilities. Similar to the Kintelligence Kit, this panel includes an assortment of forensically relevant

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

SNPs for inference of BGA, EVCs, and identity, making a total of 5422 SNPs. The incorporation of hybridization capture has been adopted from ancient DNA analysis techniques, providing the ability to analyze highly degraded DNA fragments that are not suitable for traditional polymerase chain reaction (PCR) enrichment.[154,155]

For historical bone samples, Tillmar et al.[63] reported there was a 380-fold increase in reads using hybridization capture over shotgun sequencing. In 2023, a QIAseq assay of the FORCE panel using unique molecular indices was evaluated and was successful in generating accurate genotypes for compromised research samples.[156]

The FORCE panel was found to be suitable for inferring relationships between an unknown individual and first- to fifth-order relatives.[63,156] To test this, skeletal DNA extracts were analyzed from previously identified World War II service members against two or three FRS.[63] While the SNP genotype call rate was ~44.4% for the bone samples, kinship predictions were consistent with the expected relationships when at least 25% of the kinship SNPs were typed; below this threshold, only first-order relatives were detected.

Operational Challenges

Accredited forensic laboratories are required to thoroughly validate new methods for casework. The requirement for gaining or maintaining appropriate accreditation status has prompted an increase in testing of medium- and high-density SNP genotyping technologies applied to forensic samples.[56,57,63,142,144] Quality assurance and control processes become complicated when third-party contractors are engaged to conduct SNP genotyping, leaving LE agencies with little visibility of their validation, sequencing, and bioinformatic analysis processes.[62]

Tillmar et al.[157] discussed the importance of forensic facilities maintaining responsibility for processing cases rather than sending samples to third-party companies for SNP analysis, where possible. To this end, the Scientific Working Group on DNA Analysis Methods (SWGDAM) in the United States recently published the first guidelines for the interpretation and reporting of SNP profiles developed using MPS by forensic DNA testing laboratories.[158]

Operationalizing this technique in accordance with the validation, quality, and evidential requirements by which accredited forensic laboratories are typically governed will facilitate the court admissibility of SNP genotyping results generated using MPS.[134,159] This was demonstrated in December 2023, when SNP profiles produced by the Kern Regional Crime Laboratory using the ForenSeq DNA Signature Prep Kit (Verogen, Inc.) were admitted as a reliable form of evidence for the first time in the U.S. court system.[160]

However, the implementation of end-to-end FIGG workflows in forensic laboratories will continue to be challenged by the requirements for new genotyping technologies, bioinformatics capabilities, processing of compromised evidence samples, and genealogical expertise.

Technology requirements

With the increased forensic interest in MPS technology, Alonso et al.[161] surveyed laboratories from 25 European countries in 2016 and found that 52% had purchased an instrument for SNP analysis. With more laboratories expressing interest in and pursuing MPS, medium-range familial/kinship testing will become more accessible as a result of the availability of TAS panels and benchtop instruments.[139]

In the United States, 163 laboratories already have a MiSeq FGx Sequencing System.[162] The specialized instrumentation required for microarray and WGS methods is not readily available in accredited forensic laboratories; however, the instrumentation exists in non-forensic genomics facilities that may also offer their services to LE.

Alonso et al.[161] also noted that implementation of MPS was hindered primarily due to concerns around reagent/consumable costs and underdevelopment of the technology.[161] A cost-benefit analysis study in the United States concluded that while the cost of implementing TAS methods may be high, the operational and societal benefits justify the investment.[162] The generation of DNA profiles using dense SNP multiplexes, coupled with the higher sensitivity of MPS technology, will result in improved DNA identification outcomes for both criminal and coronial casework.

For human remains identification, the application of SNP genotyping will reduce the rate of false positive and negative identifications made when close genetic relatives are unavailable for STR profiling.[162]

An Australian pilot study assessed the ability of microarray technology to analyze low quantity and quality blood and bone samples, concluding that other analysis methods (i.e., WGS or TAS) should be pursued for UHR casework.[57] This study demonstrated that 0.1 ng of high-quality DNA could produce a high call rate; however, only 25% of samples (those with over 12 ng of input DNA) returned the expected matches when analyzed using GEDmatch PRO kinship algorithms. Bioinformatics were applied to several samples to enrich the SNP data by combining datasets for analysis without any improvement.

WGS was assessed in a forensic setting when applied to a set of UHR recovered in 2003 in Sweden, known as the "Ekeby man."[157] Approximately 3 billion reads were obtained from 3 ng of skeletal DNA, and 1 million SNPs were selected for upload to GEDmatch using extensive bioinformatics analysis. The Ekeby man was identified through genealogical research of 36 candidate relatives identified on GEDmatch with Croatian ancestry.

In showing that ancestry and kinship investigative leads could be successfully generated using this in-house method, the authors emphasized the importance of maintaining control of FIGG cases.

However, there are several cost and time considerations that forensic laboratories will need to evaluate before deciding to move forward with establishing such advanced sequencing capabilities, including those associated with procuring specialist equipment, software and reagents, completing validation studies, data storage and retention, and seeking ISO/IEC 17025:2017 accreditation. The NovaSeq™ 6000 System and its successor, the NovaSeq X Series, have list prices in the order of US$1 million.[163]

While the cost per genome is low (about US$200), the upfront cost is high and represents a substantial capital investment for a laboratory that is not regularly performing WGS. A single run can generate six terabytes of data that must be stored securely and backed up regularly.[163] Further, validating and maintaining such instrumentation may not be cost-effective. Due to these factors and the expected low throughput, WGS is unlikely to be introduced in the majority of forensic laboratories.

Compromised samples

LE agencies seeking to utilize FIGG need to consider sample quality and quantity on a case-by-case basis. When processing compromised samples from UHR, forensic analysis can be complicated by the age and degradation of the remains and the presence of PCR inhibitors. The ancient DNA community has developed workflows for such samples, successfully recovering DNA from skeletal remains over 100,000 years old.[164] The insights gained from their protocols have been adopted for the identification of UHR in historical and contemporary casework.[165]

Both extraction and library preparation protocols are critical to the success of downstream analyses, and the Dabney extraction protocol[166] coupled with hybridization capture[154] has the potential to increase sequencing success rates. The Dabney extraction protocol was developed to enable recovery of ultrashort (≥25 base pairs) DNA fragments from ancient specimens.[166] When compared with the total demineralization protocol[167] employed by many forensic laboratories for skeletal samples, multiple studies reported greater success using the Dabney protocol when analyzing highly degraded DNA with MPS approaches.[164,165,168,169]

However, total demineralization yields more DNA overall because it tolerates higher DNA input amounts. Further, a recent study on DNA quantification methods demonstrated that commercially available quantitative PCR (qPCR) assays are a successful screening method to predict the success of downstream analysis methods, including SNP genotyping.[170]

Both PCR and hybridization capture target specific DNA sequences for MPS library preparation. Hybridization capture differs from PCR, however, by using oligonucleotide probes (''baits'') immobilized on beads to bind target sequences of interest without amplification.[171] This technique is favored for its success with compromised specimens, as degraded DNA fragments may not always be amenable to PCR due to their reduced size.[154] Hybridization capture approaches show great promise, whereas PCR-based methods have received greater attention for forensic applications to date, once again because of compatibility with existing laboratory instrumentation and procedures.

Genealogical expertise

A critical component of FIGG is conducting the genealogical research for the observed kinship matches.[80,172] The skills required for genetic genealogy include the ability to interpret not only historical records and tracing lineage, but also genetic data. The increased availability of online records has accelerated genealogical research as several companies have digitized and indexed millions of records for consumers pursuing recreational genetic genealogy.

The DNA Doe Project reports that case outcomes depend on the kinship distance of genetic matches and the completeness of records, with some cases requiring hundreds of hours of researching to resolve.[121] The complexity of accurately positioning an unknown individual within a family tree makes this process a limiting factor of success and thus requires suitable expertise to collect and correlate documentary evidence.[102,172]

Before the rise in popularity of genetic genealogy, genealogists taught themselves by gaining experience with their own families.[173] Now there are several education and training opportunities, all of which emphasize practical experience as the most effective way to learn.[174] Current LE FIGG programs have chosen to either perform the genealogy research in-house by training existing employees in genetic genealogy (e.g., FBI[102]) or recruiting genetic genealogists (e.g., Toronto Police Service[175]), or by contracting the services of external, vetted FIGG practitioners (e.g., AFP DNA Program[152]).

However, it is important that engaged genetic genealogists are suitably skilled and experienced, adhere to agency-prescribed operational, legal, ethical, privacy, and confidentiality parameters, abide by genetic genealogy database providers' terms of service and privacy policies, document their research in accordance with relevant genealogy and forensic standards, and are capable of testifying in court as required.

A recent article brought attention to the necessity for standards and certification of FIGG practitioners to ensure public trust is maintained.[176] The authors adopted the Board of Certification for Investigative Genetic

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

Genealogy to ensure proficient and ethical practice by outlining general frameworks for standards, certification, and education.[177]

Policy Considerations for Familial and Kinship Testing

There are a number of ethical issues that have been the subject of debate in published literature that need to be considered when implementing and regulating familial and kinship testing. It is widely recognized that transparency, accountability, and oversight need to be clearly established in legislation or policy for the proper use of such techniques.[27,178,179] Issues that have been raised previously include the privacy and security of genetic data, the proper acquisition of consent to use such data, and the criteria for when a case sample qualifies for familial/kinship testing.

The majority of literature and legislative discourse to date has focused on the use of this technique for criminal investigations, rather than for coronial investigations. A UK study observed the majority of surveyed stakeholders, both public and professional, support the use of FIGG in criminal investigations; however, the public were largely ignorant of the privacy risks involved.[180]

There are two published surveys that assessed the level of public support for the use of FIGG in various forensic applications, including UHR casework. A 2018 survey in the United States found 77% of participants supported the use of FIGG for identifying human remains.[181] More recently, an international survey recorded higher acceptability and support, with 85.2% endorsing LE use of public and private genetic genealogy databases for coronial investigations.[182]

Legislation and regulation

Despite the use of genetic kinship testing for criminal and coronial investigations over the past two decades, it is still largely unregulated and lacking clear guidelines in most jurisdictions. There is limited legislation addressing the use of short-range familial/kinship testing in the United States, Poland, and the Netherlands;[28,183,184] yet other jurisdictions such as the United Kingdom and Australia do not explicitly permit or prohibit its use.[28,183]

The lack of authority and, subsequently, the lack of accountability for employing this technique has further complicated the implementation of extended kinship testing procedures. When determining how FIGG can be best utilised for casework, it will be important to establish clear guidelines that are adaptable to accommodate the rapid development of this field.

A number of jurisdictions have approved FIGG use for UHR cases, but not criminal cases, due to the more complex privacy and ethical considerations with application to criminal investigations. This is because privacy protections are usually only triggered for identified, living persons. In 2020, the National Police Chiefs' Council Homicide Working Group in the United Kingdom reported FIGG would not be used in cold cases due to the ethical and operational risks involved.[185,186]

However, a later report by the British Biometric and Forensic Ethics Group[105] took a different view. This report evaluated the use of FIGG for both criminal and coronial investigations, recommending only the latter to test the validity of the under-regulated procedure in the United Kingdom. Further, some jurisdictions believe that FIGG should not be applied to every unresolved case, and that the specific circumstances need to be considered when determining the probative value of conducting such searches.[111]

Maryland was the first state in the United States to introduce an FIGG-related law. In May 2021, the ''Criminal Procedure—Forensic Genetic Genealogical DNA Analysis, Searching, Regulation and Oversight'' bill was passed into law (effective date October 1, 2021) to regulate the use of FIGG in criminal and coronial procedures.[187] The bill defines forensic genetic genealogical DNA analysis and searching (FGGS) and establishes the requirements and procedures LE agencies must undertake.

It permits the use of FGGS for criminal investigations, restricted to serious crimes, and human remains identification for coronial investigations. A second bill was passed in May 2022 (effective date October 1, 2022), titled ''Genetic Information Privacy—Consumer Protection and Forensic Genealogy,'' to address genetic privacy concerns for consumers of DTC genetic testing companies, including informed consent for inclusion in LE matching.[188] However, since the law has become active, several parts have yet to roll out, including the report for best practices and minimum qualifications required.[189,190]

The second state, Montana, has now introduced two relevant laws. In May 2021, the ''Consumer DNA Database Searches—Familial DNA Searches—Warrant Required'' bill was introduced (effective date October 1, 2021) to legislate the requirement for a search warrant to be obtained for an LE search of DTC DNA databases, unless the privacy rights of the consumer have been previously waived.[191,192]

Then in June 2023, Montana enacted the ''Genetic Information Privacy Act'' (effective date October 1, 2023) as a mechanism to provide its citizens with greater protection around the collection, use, or disclosure of their genetic data by DTC genetic testing companies.[193] In March 2023, Utah became the third state to pass a related law, the ''Investigative Genetic Genealogy Modifications'' bill (effective date May 3, 2023), to address some of the concerns for LE agencies' use of FIGG by defining terms and establishing the case requirements that they need to meet.[194]

At the time of writing, the Florida Senate is progressing the ''Forensic Investigative Genetic Genealogy Grant Program'' bill to establish a specialized financial stream to assist LE agencies in pursuing this technology for eligible cases in accordance with Florida Department of LE rules.[195]

Genetic privacy

The analysis of an individual's whole genome yields much more information about them than required for kinship, and this extends to the individual's relatives due to the large proportion of DNA they share.[196] When an individual gives consent for their DNA to be used for FIGG by LE, they also make available some of the DNA of their genetic relatives.

With information about certain genetic traits such as BGA and EVCs able to be analyzed by those with the necessary skills, there is the potential for this data to be misused for genetic discrimination in employment or insurance.[95,197] Theoretical modeling demonstrates that de-identified genetic information can be re-identified with only a few items of personal information.[198,199] For example, 99.98% of Americans could be correctly re-identified in any dataset using 15 demographic attributes.[200]

Therefore, the sensitive personal information that is generated and/or accessed during extended kinship testing, for both the genetic analysis and genealogical research, warrants all agencies undertaking FIGG to conduct adequate privacy, risk, and cybersecurity assessments to ensure compliance with appropriate genetic data handling practices.

The function of coronial investigations is to identify UHR and while there are no privacy protections triggered for deceased individuals, the genetic privacy of their living relatives must be protected. A privacy impact assessment (PIA) is one mechanism for reviewing an agency's proposed FIGG process, to both assess its impact on the privacy of individuals and recommend actions for managing, minimizing, or eliminating any identified privacy impacts.[201,202]

PIAs have been conducted previously to inform the implementation of other new forensic DNA tools such as forensic DNA phenotyping in Australia[203] and the use of rapid DNA systems in the United States.[204] In 2021, the AFP DNA Program undertook a PIA to ensure both its proposed insourced and outsourced FIGG workflows were compliant with the Australian Privacy Principles (consistent with the Commonwealth Privacy Act 1988) before commencing FIGG casework. The modeling of this approach in other countries will allow privacy risks to be managed alongside associated legal, contractual, and confidentiality issues.

FIGG programs for LE

In 2019, the SWGDAM released an overview of FIGG, outlining the procedures involved and recommendations for database use, informed consent, and data privacy for LE agencies seeking to employ the technique.[101] Similarly, Scudder et al.[104] proposed a number of operational challenges that would need to be overcome before implementing FIGG in Australia. The adoption of FIGG by U.S. agencies has increased rapidly since then (e.g., Refs.[174,205–207]), with several jurisdictions outside the United States now currently using or exploring FIGG.

In 2022, following completion of a PIA and implementation of recommended privacy and legal safeguards, the AFP DNA Program began outsourcing FIGG services for select UHR cases to an external service provider bound by contractual and confidentiality requirements, while simultaneously assessing genotyping technologies, bioinformatics pipelines, and local genetic genealogy expertise for an insourced FIGG capability.

As of 2023, the AFP DNA Program now offers its state and territory LE partners both pipelines. In 2021, the Swedish National Forensic Centre published a report discussing the potential pathways to utilize FIGG in criminal investigations.[110] Following the success of a pilot case study, the report recommended that UHR casework be included in plans to implement the technology. Lastly, the Netherlands are conducting a pilot study on the identification of UHR using FIGG, the results of which are yet to be published.[208]

Conclusions

For LE agencies considering implementation of medium- or long-range kinship testing, the technological and analytical capabilities of the laboratory need to be evaluated. Of the SNP genotyping technologies available, both WGS and TAS have proven suitable for processing samples commonly encountered in UHR casework. When determining whether to establish an insourced FIGG pipeline or outsource the sequencing and/or genealogical research components to an external service provider, LE agencies need to consider the: condition and availability of typical forensic samples; quantity and quality of genetic information required; accessibility of genetic genealogy databases; availability of a suitable level of internal or external expertise; and data handling requirements.

The lack of legislation guiding application of extended kinship testing in some jurisdictions leaves individual agencies to grapple with issues of genetic privacy and establish best practice guidelines. As a starting point, it is good practice to meet or exceed the recommendations of a PIA commissioned to evaluate specific familial/kinship analysis pipelines.

With the growing international backlog of unidentified and missing persons cases, many of which are decades old, it is more important than ever to consider using techniques such as extended kinship testing. Challenges faced by the passage of time, such as the absence of AM samples or unavailability of first-order relatives, can be

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

overcome by broadening the pool of suitable relatives for comparison. Members of the public continue to populate genetic genealogy databases through DTC companies, increasing the likelihood of finding genetic relatives of UHR over time.

Despite ethical and legal concerns still being raised regarding genetic privacy, data security, and governance frameworks for these techniques being used in a forensic context, it appears that the public supports its use for human remains identification and missing persons investigations. To routinely apply genetic kinship testing techniques to coronial casework and maintain public trust, the understandable desire of relatives to know the fate of their missing loved one must be balanced against the rights of other relatives to have their genetic privacy respected.

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

Authors' Contributions

J.L.W.: writing—original draft (lead); writing—review and editing (equal). D.M.: conceptualization; writing—review and editing (equal). J.W.: conceptualization (lead); funding acquisition (lead); and writing—review and editing (equal).

Author Disclosure Statement

There are no interests, funding, or employment that influence or affect the integrity of this submission.

Funding Information

J.L.W. is supported by an Australian Government Research Training Program (RTP) Scholarship and the AFP National DNA Program for Unidentified and Missing Persons.

References

1. International Committee of the Red Cross. The Forensic Human Identification Process: An Integrated Approach. Geneva, Switzerland; 2022. Available from: https://www.icrc.org/en/publication/4590-forensic-human-identification-process-integrated-approach [Last accessed: May 29, 2022].
2. International Criminal Police Organisation. INTERPOL Disaster Victim Identification Guide. 2018. Available from: https://www.interpol.int/en/How-we-work/Forensics/Disaster-Victim-Identification-DVI [Last accessed: September 29, 2023].
3. Robins S. Analysis of Best Practices on the Identification of Missing Migrants: Implications for the Central Mediterranean. Central Mediterranean Route Thematic Report Series. International Organization for Migration: Geneva, Switzerland; 2018.
4. Wickenheiser RA. Expanding DNA database effectiveness. Forensic Sci Int Synerg 2022;4:100226; doi: 10.1016/j.fsisyn.2022.100226
5. Bieber FR, Brenner CH, Lazer D. Finding criminals through DNA of their relatives. Science 2006;312(5778):1315–1316; doi: 10.1126/science.1122655
6. Ge J, Budowle B, Chakraborty R. Choosing relatives for DNA identification of missing persons. J Forensic Sci 2011;56:S23–S28; doi: 10.1111/j.1556-4029.2010.01631.x
7. Kayser M, De Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 2011;12(3):179–192; doi: 10.1038/nrg2952
8. Australian Federal Police. National DNA Program for Unidentified and Missing Persons. Commonwealth of Australia: Canberra; 2022. Available from: https://www.missingpersons.gov.au/support/national-dna-program-unidentified-and-missing-persons [Last accessed: December 11, 2023].
9. Dahal A, McNevin D, Chikhani M, et al. An interdisciplinary forensic approach for human remains identification and missing persons investigations. Wiley Interdiscip Rev Forensic Sci 2023;5:e1484; doi: 10.1002/wfs2.1484
10. Ward J. A multifaceted, multijurisdictional, multiagency, and multidisciplinary approach to investigating unidentified and missing persons cases in Australia. Forensic Sci Int Genet Suppl Ser 2022;8:53–55; doi: 10.1016/j.fsigss.2022.09.020
11. Ritter N. Missing persons and unidentified remains: The nation's silent mass disaster. NIJ J 2007;256(7):2–7.
12. Rodriguez AL, Smiley-McDonald HM, Cummings MS, et al. Understanding unidentified human remains investigations through the United States census data. Forensic Sci Int Synerg 2022;4:100225; doi: 10.1016/j.fsisyn.2022.100225
13. Reid KM, Martin LJ, Heathfield LJ. Understanding the burden of unidentified bodies: A systematic review. Int J Legal Med 2023;137(4):1193–1202; doi: 10.1007/s00414-023-02968-5
14. UK Missing Persons Unit. UKMPU Missing Persons Data Report 2021/22: National Crime Agency; 2022. Available from: https://missingpersons.police.uk/en-gb/resources/downloads/missing-persons-statistical-bulletins [Last accessed: August 17, 2023].
15. Ward J. Best practice recommendations for the establishment of a national DNA identification program for missing persons: A global perspective. Forensic Sci Int Genet Suppl Ser 2017;6:e43–e45; doi: 10.1016/j.fsigss.2017.09.009
16. Leclair B. Large-scale comparative genotyping and kinship analysis: Evolution in its use for human identification in mass fatality incidents and missing persons databasing. Int Congress Ser 2004;1261:42–44; doi: 10.1016/s0531-5131(03)01661-3
17. International Commission on Missing Persons. DNA-Led Human Identification 2016. Available from: https://www.icmp.int/news/dna-led-human-identification [Last accessed: June 5, 2022].
18. International Committee of the Red Cross. Missing Persons | Internationa Committee of the Red Cross; 2022. Available from: https://www.icrc.org/en/war-and-law/protected-persons/missing-persons [Last accessed: June 5, 2022].
19. International Criminal Police Organisation. Best Practice Principles: Recommendation on the Use of DNA for the Identification of Missing Persons and Unidentified Human Remains. INTERPOL DNA Monitoring Expert Group: Lyon, France; 2015.
20. Laurent F-X, Fischer A, Oldt R, et al. Streamlining the decision-making process for international DNA kinship matching using worldwide allele frequencies and tailored cutoff log10LR thresholds. Forensic Sci Int Genet 2022;57:102634; doi: 10.1016/j.fsigen.2021.102634
21. Royal Canadian Mounted Police. National Home Page | Royal Canadian Mounted Police; 2022. Available from: https://www.rcmp-grc.gc.ca [Last accessed: June 5, 2022].
22. University of North Texas Center for Human Identification. Center for Human Identification—A Globally-Recognised Leader in Forensic Identification: The University of North Texas Health Science Center at Fort Worth; 2022. Available from: https://www.unthsc.edu/center-for-human-identification [Last accessed: June 5, 2022].
23. National Institute of Justice. Familial DNA searching: Current approaches. Natl Inst Justice 2015:1–82.
24. Scientific Working Group on DNA Analysis Methods. Recommendations from the SWGDAM Ad Hoc Working Group on Familial Searching. SWGDAM Communications: Unites States; 2013.
25. Wyner N, Barash M, McNevin D. Forensic autosomal short tandem repeats and their potential association with phenotype. Front Genet 2020;11:884; doi: 10.3389/fgene.2020.00884
26. Bañuelos MM, Zavaleta YJA, Roldan A, et al. Associations between forensic loci and expression levels of neighboring genes may compromise medical privacy. Proc Natl Acad Sci U S A 2022;119(40):e2121024119; doi: 10.1073/pnas.2121024119
27. Murphy E. Law and policy oversight of familial searches in recreational genealogy databases. Forensic Sci Int 2018;292:e5–e9; doi: 10.1016/j.forsciint.2018.08.027
28. Maguire CN, McCallum LA, Storey C, et al. Familial searching: A specialist forensic DNA profiling service utilising the National DNA Database® to identify unknown offenders via their relatives—The UK experience. Forensic Sci Int Genet 2014;8(1):1–9; doi: 10.1016/j.fsigen.2013.07.004
29. Prainsack B. Key Issues in DNA Profiling and Databasing: Implications for Governance. In: Genetic Suspects: Global Governance of Forensic DNA Profiling and Databasing. (Hindmarsh R, Prainsack B. eds.) Cambridge

University Press: Cambridge, United Kingdom; 2010; pp. 153–174; doi: 10.1017/CBO9780511778193.003
30. Kennett D. Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes. Forensic Sci Int 2019;301:107–117; doi: 10.1016/j.forsciint.2019.05.016
31. Abarno D, Sobieraj T, Summers C, et al. The first Australian conviction resulting from a familial search. Aust J Forensic Sci 2019;51(Suppl 1):S56–S59; doi: 10.1080/00450618.2019.1568553
32. Smith M, Urbas GF. Regulating new forms of forensic DNA profiling under Australian legislation: Familial matching and DNA phenotyping. Aust J Forensic Sci 2012;44(1):63–81; doi: 10.1080/00450618.2011.581250
33. Thomas L. Nothing to hide, something to fear?: The use of partial DNA matching in criminal investigations. J Law Inf Sci 2006;17:72–93.
34. Shand A. Case: Lost at Sea: Ep 1—Kingscliff. NSW Police State Crime Command—Investgiations: LiSTNR; 2021.
35. Brenner CH, Weir BS. Issues and strategies in the DNA identification of World Trade Center victims. Theor Popul Biol 2003;63(3):173–178; doi: 10.1016/S0040-5809(03)00008-X
36. Scientific Working Group on DNA Analysis Methods. Guidelines for Missing Persons Casework. SWGDAM Communications: United States; 2014.
37. Curran JM, Buckleton JS. Effectiveness of familial searches. Sci Justice 2008;48(4):164–167; doi: 10.1016/j.scijus.2008.04.004
38. Alonso A, Martín P, Albarrán C, et al. Challenges of DNA profiling in mass disaster investigations. Croat Med J 2005;46(4):540–548.
39. Jeffreys AJ, Allen MJ, Hagelberg E, et al. Identification of the skeletal remains of Josef Mengele by DNA analysis. Forensic Sci Int 1992;56(1):65–76; doi: 10.1016/0379-0738(92)90148-p
40. Ge J, Chakraborty R, Eisenberg A, et al. Comparisons of familial DNA database searching strategies. J Forensic Sci 2011;56(6):1448–1456; doi: 10.1111/j.1556-4029.2011.01867.x
41. Leclair B, Shaler R, Carmody GR, et al. Bioinformatics and human identification in mass fatality incidents: The World Trade Center disaster. J Forensic Sci 2007;52(4):806–819; doi: 10.1111/j.1556-4029.2007.00456.x
42. Oefelein RH. Seven years of SNPs: An assessment of methods utilized for generating profiles for forensic genetic genealogy. Forensic Sci Int Genet Suppl Ser 2022;8:347–348; doi: 10.1016/j.fsigss.2022.11.002
43. Amorim A, Pereira L. Pros and cons in the use of SNPs in forensic kinship investigation: A comparative analysis with STRs. Forensic Sci Int 2005;150(1):17–21; doi: 10.1016/j.forsciint.2004.06.018
44. Butler JM, Coble MD, Vallone PM. STRs vs. SNPs: Thoughts on the future of forensic DNA testing. Forensic Sci Med Pathol 2007;3(3):200–205; doi: 10.1007/s12024-007-0018-1
45. Rajagopalan RM, Fujimura JH. Variations on a chip: Technologies of difference in human genetics research. J Hist Biol 2018;51(4):841–873; doi: 10.1007/s10739-018-9543-x
46. Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. Brief Funct Genomic Proteomic 2009;8(3):174–183; doi: 10.1093/bfgp/elp013
47. ATDBio. Sequencing, Forensic Analysis and Genetic Analysis. 2021. Available from: https://atdbio.com/nucleic-acids-book/Sequencing-forensic-analysis-and-genetic-analysis#DNA-Fingerprinting [Last accessed: January 7, 2022].
48. Kling D, Phillips C, Kennett D, et al. Investigative genetic genealogy: Current methods, knowledge and practice. Forensic Sci Int Genet 2021;52:102474; doi: 10.1016/j.fsigen.2021.102474
49. Erlich Y, Narayanan A. Routes for breaching and protecting genetic privacy. Nat Rev Genet 2014;15(6):409–421; doi: 10.1038/nrg3723
50. Balogh MK, Bender K, Schneider PM. Application of Nanogen microarray technology for forensic SNP analysis. Int Congress Ser 2006;1288:43–45; doi: 10.1016/j.ics.2005.10.047
51. Divne A-M, Allen M. A DNA microarray system for forensic SNP analysis. Forensic Sci Int 2005;154(2–3):111–121; doi: 10.1016/j.forsciint.2004.09.134
52. Keating B, Bansal AT, Walsh S, et al. First all-in-one diagnostic tool for DNA intelligence: Genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip. Int J Legal Med 2013;127(3):559–572; doi: 10.1007/s00414-012-0788-1
53. Urtiaga GO, Domingues WB, Komninou ER, et al. DNA microarray for forensic intelligence purposes: High-density SNP profiles obtained directly from casework-like samples with and without a DNA purification step. Forensic Sci Int 2022;332:111181; doi: 10.1016/j.forsciint.2022.111181
54. Bever R, Byrne M. Forensic Genealogy: Unlocking the Science of Genealogy. Forensic Genealogy—Webinar Series: Exploring the Impact, Processes, Successes and Challenges of Forensic Genealogy. Bode Technology: Virginia, United States; 2020.
55. Russell DA, Gorden EM, Peck MA, et al. Developmental validation of the Illumina Infinium Assay using the global screening array on the iScan System for use in forensic laboratories. Forensic Genom 2023;3(1):15–24; doi: 10.1089/forensic.2022.0013
56. de Vries JH, Kling D, Vidaki A, et al. Impact of SNP microarray analysis of compromised DNA on kinship classification success in the context of investigative genetic genealogy. Forensic Sci Int Genet 2022;56:102625; doi: 10.1016/j.fsigen.2021.102625
57. Davawala A, Stock A, Spiden M, et al. Forensic Genetic Genealogy using microarrays for the identification of human remains: The need for good quality samples—A pilot study. Forensic Sci Int 2022;334:111242; doi: 10.1016/j.forsciint.2022.111242
58. Ilumina. Whole-Genome Sequencing (WGS) 2021. Available from: https://sapac.illumina.com/techniques/sequencing/dna-sequencing/whole-genome-sequencing.html [Last accessed: January 7, 2022].
59. Li H, Glusman G, Hu H, et al. Relationship estimation from whole-genome sequence data. PLoS Genet 2014;10(1):e1004144; doi: 10.1371/journal.pgen.1004144
60. Centre for Disease Control and Prevention. Whole Genome Sequencing (WGS). 2021. Available from: https://www.cdc.gov/pulsenet/pathogens/wgs.html [Last accessed: January 7, 2022].
61. Takara Bio. ThruPLEX DNA-Seq Kit: All-Purpose DNA Library Preparation for Illumina NGS Platforms. 2022. Available from: https://www.takarabio.com/products/next-generation-sequencing/dna-seq/low-input-library-prep-for-illumina [Last accessed: February 13, 2022].
62. Kling D. On the use of dense sets of SNP markers and their potential in relationship inference. Forensic Sci Int Genet 2019;39:19–31; doi: 10.1016/j.fsigen.2018.11.022
63. Tillmar A, Sturk-Andreaggi K, Daniels-Higginbotham J, et al. The FORCE panel: An all-in-one SNP marker set for confirming investigating genetic genealogy leads and for general forensic applications. bioRxiv 2021; doi: 10.1101/2021.11.30.470354
64. Lange K. The affected sib-pair method using identity by state relations. Am J Hum Genet 1986;39(1):148.
65. Niedzwiecki E, Debus-Sherrill S, Field MB. Understanding Familial DNA Searching: Coming to a Consensus on Terminology. US Department of Justice: Virginia, United States; 2017.
66. Stevens EL, Heckenberg G, Roberson EDO, et al. Inference of relationships in population data using identity-by-descent and identity-by-state. PLoS Genet 2011;7(9):e1002287; doi: 10.1371/journal.pgen.1002287
67. Morimoto C, Manabe S, Fujimoto S, et al. Discrimination of relationships with the same degree of kinship using chromosomal sharing patterns estimated from high-density SNPs. Forensic Sci Int Genet 2018;33:10–16; doi: 10.1016/j.fsigen.2017.11.010
68. Ge J, Budowle B. Forensic investigation approaches of searching relatives in DNA databases. J Forensic Sci 2021;66(2):430–443; doi: 10.1111/1556-4029.14615
69. Sticca EL, Belbin GM, Gignoux CR. Current developments in detection of identity-by-descent methods and applications. Front Genet 2021;12(1725):722602; doi: 10.3389/fgene.2021.722602
70. Speed D, Balding DJ. Relatedness in the post-genomic era: Is it still useful? Nat Rev Genet 2015;16(1):33–44; doi: 10.1038/nrg3821
71. Bouquet M. Genealogy in Anthroplogy. In: International Encyclopedia of the Social & Behavioural Sciences. 2nd ed. (Wright JD. ed.) 2001; pp. 6029–6032; doi: 10.1016/b0-08-043076-7/00876
72. Byrnes J, Myers N, Underhill P. Genetic Genealogy in the Genomic Era. In: Forensic DNA Applications: An Interdisciplinary Perspective. 2nd ed. (Primorac D, Schanfield M. eds.) CRC Press: Florida, United States; 2014; p. 483.
73. Kennett D. Sense About Genealogical DNA Testing: Sense About Science; 2013. Available from: https://archive.senseaboutscience.org/blog.php/41/sense-about-genealogical-dna-testing.html [Last accessed: March 23, 2022].
74. Gomes C, Palomo-Díez S, López-Parra AM, et al. Genealogy: The tree where history meets genetics. Genealogy 2021;5(4):98; doi: 10.3390/genealogy5040098
75. Parker K, Erzurumluoglu AM, Rodriguez S. The Y chromosome: A complex locus for genetic analyses of complex human traits. Genes 2020;11(11):1273; doi: 10.3390/genes11111273

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

76. Claerhout S, Defraye C, Decorte R. Validation of Y-ancestor time calculators for forensic familial searching. Forensic Sci Int Genet Suppl Ser 2019;7(1):411–413; doi: 10.1016/j.fsigss.2019.10.032
77. Calafell F, Larmuseau MHD. The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. Hum Genet 2017;136(5):559–573; doi: 10.1007/s00439-016-1740-0
78. King TE, Jobling MA. Founders, drift, and infidelity: The relationship between Y chromosome diversity and patrilineal surnames. Mol Biol Evol 2009;26(5):1093–1102; doi: 10.1093/molbev/msp022
79. King TE, Jobling MA. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. Trends Genet 2009;25(8):351–360; doi: 10.1016/j.tig.2009.06.003
80. Larmuseau MHD, Van Geystelen A, Van Oven M, et al. Genetic genealogy comes of age: Perspectives on the use of deep-rooted pedigrees in human population genetics. Am J Phys Anthropol 2013;150(4):505–511; doi: 10.1002/ajpa.22233
81. Moore C. The history of genetic genealogy and unknown parentage research: An insider's view. J Genet Geneal 2016;8(1):35–37.
82. Hogarth S, Javitt G, Melzer D. The current landscape for direct-to-consumer genetic testing: Legal, ethical, and policy issues. Annu Rev Genomics Hum Genet 2008;9(1):161–182; doi: 10.1146/annurev.genom.9.081307.164319
83. Allyse MA, Robinson DH, Ferber MJ, et al. Direct-to-consumer testing 2.0: Emerging models of direct-to-consumer genetic testing. Mayo Clin Proc 2018;93(1):113–120; doi: 10.1016/j.mayocp.2017.11.001
84. Erlich Y, Shor T, Pe'Er I, et al. Identity inference of genomic data using long-range familial searches. Science 2018;362(6415):690–694; doi: 10.1126/science.aau4832
85. Ancestry.com. AncestryDNA® | DNA Tests for Ethnicity & Genealogy DNA. 2022. Available from: https://www.ancestry.com.au/dna [Last accessed: March 14, 2022].
86. 23andMe. DNA Genetic Testing & Analysis—23andMe; 2022. Available from: https://www.23andme.com/en-int [Last accessed: February 20, 2022].
87. MyHeritage. Free Family Tree, Genealogy, Family History, and DNA Testing. 2022. Available from: https://www.myheritage.com [Last accessed: February 22, 2022].
88. FamilyTreeDNA. DNA Testing for Ancestry & Genealogy | FamilyTreeDNA; 2022. Available from: https://www.familytreedna.com [Last accessed: February 24, 2022].
89. International Society of Genetic Genealogy Wiki. Autosomal DNA Testing Comparison Chart. 2022. Available from: https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart [Last accessed: June 30, 2022].
90. Majumder MA, Guerrini CJ, McGuire AL. Direct-to-consumer genetic testing: Value and risk. Annu Rev Med. 2021;72(1):151–166; doi: 10.1146/annurev-med-070119-114727
91. Greytak EM, Moore C, Armentrout SL. Genetic genealogy for cold case and active investigations. Forensic Sci Int 2019;299:103–113; doi: 10.1016/j.forsciint.2019.03.039
92. Ball CA, Barber MJ, Byrnes J, et al. Ancestry DNA Matching White Paper. Ancestrycom: Utah, Untied States; 2016; pp. 1–46.
93. Samuel G, Kennett D. Problematizing consent: Searching genetic genealogy databases for law enforcement purposes. N Genet Soc 2021;40:1–21. doi: 10.1080/14636778.2020.1843149
94. GEDmatch. Find DNA Matches for Free | Analyze Your DNA | GEDmatch; 2022. Available from: https://www.gedmatch.com [Last accessed: July 17, 2021].
95. Berkman BE, Miller WK, Grady C. Is it ethical to use genealogy data to solve crimes? Ann Intern Med 2018;169(5):333–334; doi: 10.7326/m18-1348
96. Callaghan TF. Responsible genetic genealogy. Science 2019;366(6462):155; doi: 10.1126/science.aaz6578
97. Greytak EM, Kaye DH, Budowle B, et al. Privacy and genetic genealogy data. Science 2018;361(6405):857; doi: 10.1126/science.aav0330
98. Denise Syndercombe Court. Forensic genealogy: Some serious concerns. Forensic Sci Int Genet 2018;36:203–204; doi: 10.1016/j.fsigen.2018.07.011
99. Terry SF, Caffet M. Free at last? Discussing the future of forensic genealogy. Genet Test Mol Biomarkers 2020;24(4):171–172; doi: 10.1089/gtmb.2020.29051.sjt
100. Department of Justice. Forensic Genetic Genealogical DNA Analysis and Searching. Interim Policy: United States; 2019.
101. Scientific Working Group on DNA Analysis Methods. Overview of Investigative Genetic Genealogy. SWGDAM Communications: United States; 2020.
102. Glynn CL. Bridging disciplines to form a new one: The emergence of forensic genetic genealogy. Genes 2022;13(8):1381; doi: 10.3390/genes13081381
103. Scudder N, McNevin D, Kelty SF, et al. Policy and regulatory implications of the new frontier of forensic genomics: Direct-to-consumer genetic data and genealogy records. Curr Issues Crim Justice 2019;31(2):194–216; doi: 10.1080/10345329.2018.1560588
104. Scudder N, Robertson J, Kelty SF, et al. An international consideration of a standards-based approach to forensic genetic genealogy. Forensic Sci Int Genet Suppl Ser 2019;7(1):512–514; doi: 10.1016/j.fsigss.2019.10.071
105. Jobling M. Denise Syndercombe Court. Should We Be Making Use of Genetic Genealogy to Assist in Solving Crime? A Report on the Feasibility of Such Methods in the UK. Biometrics and Forensics Ethics Group: United Kingdom; 2020.
106. Greytak E, Moore C. Closing Cases with a Single SNP Array: Integrated Genetic Genealogy, DNA Phenotyping, and Kinship Analyses [Poster]. International Symposium on Human Identification: Arizona, United States; 2018.
107. GEDmatch. GEDmatch.com Terms of Service and Privacy Policy: GEDmatch; 2021. Available from: https://www.gedmatch.com/terms-of-service-privacy-policy [Last accessed: July 17, 2021].
108. FamilyTreeDNA. Terms of Service. 2022. Available from: https://www.familytreedna.com/legal/terms-of-service [Last accessed: February 22, 2022].
109. American Society of Crime Laboratory Directors. ASCLD Investigative Genetic Genealogy. ASCLD Member Resoure Committee: North Carolina, United States; 2019.
110. Fagerholm SA, Tillmar A, Staaf J, et al. Forensic DNA Traces and Genealogy: Use of Investigative Genetic Genealogy in Criminal Investigations. The Swedish Police Authority, Swedish National Forensic Centre: Linköping, Sweden; 2021.
111. Scudder N, Daniel R, Raymond J, et al. Operationalising forensic genetic genealogy in an Australian context. Forensic Sci Int 2020;316:110543; doi: 10.1016/j.forsciint.2020.110543
112. Ram N, Roberts JL. Forensic genealogy and the power of defaults. Nat Biotechnol 2019;37(7):707–708; doi: 10.1038/s41587-019-0172-5
113. Kumar SA. GEDmatch: A Platform for Data Driven Forensic Intelligence. 2020. Available from: https://promega.foleon.com/theishireport/november-2020/gedmatch-a-platform-for-data-driven-forensic-intelligence [Last accessed: November 1, 2021].
114. FamilyTreeDNA. FamilyTreeDNA Law Enforcement Guide. 2023. Available from: https://www.familytreedna.com/legal/law-enforcement-guide [Last accessed: July 24, 2023].
115. Othram. Gene by Gene Selects Othram as Exclusive Partner for Forensic Genetic Genealogy Search and Analysis. 2024. Available from: https://othram.com/press_release/2024-01-23-gene-by-gene-selects-othram.html [Last accessed: February 5, 2024].
116. Smith J. Police are getting DNA data from people who think they opted out: The Intercept; 2023 updated 18/08/2023. Available from: https://theintercept.com/2023/08/18/gedmatch-dna-police-forensic-genetic-genealogy [Last accessed: September 21, 2023].
117. Kumar SA. Notice Regarding Investigations into FIGG Practitioners Circumventing GEDmatch Settings and Violating Terms of Service, and Actions Taken. Verogen, Inc.; 2023. Available from: https://verogen.com/notice-regarding-investigations-into-figg-practitioners-circumventing-gedmatch-settings-and-violating-terms-of-service-and-actions-taken [Last accessed: August 21, 2023].
118. DNASolves. Help solve crimes with your DNA, DNASolves, Othram Inc.; 2022. Available from: https://dnasolves.com [Last accessed: February 22, 2022].
119. DNA Justice. DNA Justice | Justice via IGG for All 2023. Available from: https://www.dnajustice.org [Last accessed: July 24, 2023].
120. Bode Technology. Forensic Genealogy Service. 2022. Available from: https://www.bodetech.com/FGS [Last accessed: March 20, 2022].
121. DNA Doe Project. DNA Doe Project. 2021. Available from: https://dnadoeproject.org [Last accessed: February 5, 2022].
122. DNA Labs International. Genealogy. 2021. Available from: https://dnalabsinternational.com/genealogy [Last accessed: March 31, 2022].
123. Identifinders International. Forensic Genealogy Services Fountain Valley, CA. 2022. Available from: https://identifinders.com/services [Last accessed: January 22, 2022].
124. Othram. Digitize DNA Evidence with Forensc-Grade Genome Sequencing®. 2022. Available from: https://othram.com/fggs.html [Last accessed: February 2, 2022].

125. Parabon Nanolabs. Parabon™ Snapshot® DNA Analysis Service. 2022. Available from: https://snapshot.parabon-nanolabs.com/#genealogy-how [Last accessed: March 9, 2022].
126. Parabon Nanolabs. Snapshot Genetic Genealogy: Parabon Computation; 2021. Available from: https://snapshot.parabon-nanolabs.com/genealogy [Last accessed: March 9, 2022].
127. Signature Science. What We Do: Forensic DNA at the Center for Advanced Genomics® Austin, TX. 2022. Available from: https://www.signaturescience.com/what-we-do/forensic-dna [Last accessed: August 17, 2023].
128. Intermountain Forensics. Forensic DNA Testing | Intermountain Forensics; 2022. Available from: https://www.intermountainforensics.com [Last accessed: June 15, 2023].
129. Astrea Forensics. What We Do. 2021. Available from: https://www.astreaforensics.com/what-we-do [Last accessed: September 29, 2022].
130. University of North Texas Center for Human Identification. Forensics—Center for Human Identification. 2024. Available from: https://www.unthsc.edu/center-for-human-identification/forensics [Last accessed: February 5, 2024].
131. Mendeley Data. Forensic Genetic Genealogy Project v. 2022 [Internet]. Mendeley Data; 2023 (cited August 17, 2023). Available from: data.mendeley.com/datasets/jcycgvhm96/1 [Last accessed: November 1, 2023].
132. Augenstein S. 'Buck Skin Girl' Case Break Is Success of New DNA Doe Project. Forensic Mag; 2018.
133. DNA Doe Project. Buckskin Girl 2018. Available from: https://dnadoeproject.org/case/buckskin-girl [Last accessed: November 5, 2021].
134. Scudder N, McNevin D, Kelty SF, et al. Massively parallel sequencing and the emergence of forensic genomics: Defining the policy and legal issues for law enforcement. Sci Justice 2018;58(2):153–158; doi: 10.1016/j.scijus.2017.10.001
135. Hazel JW, Clayton EW, Malin BA, et al. Is it time for a universal genetic forensic database? Science 2018;362(6417):898–900; doi: 10.1126/science.aav5475
136. QIAGEN. Massively Parallel Sequencing Solutions for Human Identity. QIAGEN® Human ID and Forensics: Hilden, Germany; 2020.
137. Cho W-C, Yoo S, Youn BJ, et al. Forensic application of SNP-targeted next-generation sequencing on Korean kinship analysis. 2022; doi: 10.2139/ssrn.4279141
138. Phillips C, Amigo J, Tillmar AO, et al. A compilation of tri-allelic SNPs from 1000 Genomes and use of the most polymorphic loci for a large-scale human identification panel. Forensic Sci Int Genet 2020;46:102232; doi: 10.1016/j.fsigen.2020/102232
139. Verogen. Data Sheet: ForenSeq Kintelligence Kit. ForenSeq Kintelligence Kit: California, United States; 2021.
140. Verogen. High-Quality Outcomes from Low-Quality Samples with ForenSeq Kintelligence [Application Note]. ForenSeq Kintelligence Kit: California, United States; 2021.
141. Antunes J, Radecke S, Snedecor J, et al. Application of a dense SNP PCR multiplex for high throughput kinship determination using Next Generation Sequencing (NGS). Forensic Sci Int Genet Suppl Ser 2022;8:324–325; doi: 10.1016/j.fsigss.2022.10.007
142. Cuenca D. An Improved Method for Forensic Genetic Genealogy in an Operational Crime Lab. Xtalks: Verogen: California, United States; 2021.
143. Antunes J. ISHI News [Internet]: International Symposium on Human Identification. 2021. Available from: https://www.ishinews.com/using-the-forenseq-kintelligence-workflow-with-difficult-samples-for-forensic-genetic-genealogy [Last accessed: September 1, 2023].
144. Staadig A, Sidstedt M, Tillmar A. ForenSeq® Kintelligence A Targeted DNA Sequencing Assay for Forensic Genetic Genealogy. Rättsmedicinalverket National Board of Forensic Medicine: Linköping University, Polisen Swedish Police: Linköping, Sweden; 2021.
145. Peck MA, Koeppel AF, Gorden EM, et al. Internal validation of the ForenSeq Kintelligence Kit for application to forensic genetic genealogy. Forensic Genom 2022;2(4):103–114; doi: 10.1089/forensic.2022.0014
146. Watson JL, McNevin D, Grisedale K, et al. Operationalisation of the ForenSeq® Kintelligence Kit for Australian unidentified and missing persons casework. Forensic Sci Int Genet 2024;68:102972; doi: 10.1016/j.fsigen.2023.102972
147. Verogen. SNP Typing in Universal Analysis Software and Kinship Estimation with GEDmatch PRO [Technical Note]. ForenSeq Kintelligence Kit: California, United States; 2021.
148. Snedecor J, Fennell T, Stadick S, et al. Fast and accurate kinship estimation using sparse SNPs in relatively large database searches. Forensic Sci Int Genet 2022;61:102769; doi: 10.1016/j.fsigen.2022.102769
149. Arenson J. DNA Labs International Aids in the Identification of Remains That Had Been Unidentified for 22 Years [Internet]. DNA Labs International, ed. PRWeb: Deerfield Beach, FL; 2022. Available from: https://www.prweb.com/releases/2022/2/prweb18514476.htm [Last accessed: February 23, 2022].
150. Intermountain Forensics. Pricing. 2023. Available from: https://www.intermountainforensics.com/pricing [Last accessed: August 17, 2023].
151. Ward J, McNevin D. There Are 750 Unidentified Human Remains in Australia. Could Your DNA Help Solve One of These Cold Cases?: The Conversation; Updated October 9, 2023. Available from: https://theconversation.com/there-are-750-unidentified-human-remains-in-australia-could-your-dna-help-solve-one-of-these-cold-cases-210626 [Last accessed: November 1, 2023].
152. Ward J, Coakley L, Grisedale K, et al. Operationalisation of the National DNA Program for Unidentified and Missing Persons' forensic investigative genetic genealogy capability for human remains identification in Australia. Forensic Genom 2024;4(1):32–40.
153. Jordan S. CHI Becomes Nation's First Public Lab to Earn Accreditation to Perform Forensic Genetic Genealogy. University of North Texas Centre for Human Identification; 2023. Available from: https://www.unthsc.edu/newsroom/story/chi-becomes-nations-first-public-lab-to-earn-accreditation-to-perform-forensic-genetic-genealogy [Last accessed: February 5, 2024].
154. Gorden EM, Sturk-Andreaggi K, Marshall C. Capture enrichment and massively parallel sequencing for human identification. Forensic Sci Int Genet 2021;53:102496; doi: 10.1016/j.fsigen.2021.102496
155. Tillmar A, Andreaggi K, Amory C, et al. Advancing Forensic SNP Typing: Insights from an Interlaboratory Study of the FORCE Panel [Conference Session]. 12th Haploid Markers Conference, Budapest, Hungary, May 19, 2023.
156. Staadig A, Hedman J, Tillmar A. Applying unique molecular indices with an extensive all-in-one forensic SNP panel for improved genotype accuracy and sensitivity. Genes 2023;14(4):818; doi: 10.3390/genes14040818
157. Tillmar A, Sjölund P, Lundqvist B, et al. Whole-genome sequencing of human remains to enable genealogy DNA database searches—A case report. Forensic Sci Int Genet 2020;46:102233; doi: 10.1016/j.fsigen.2020.102233
158. Scientific Working Group on DNA Analysis Methods. SWGDAM Interpretation Guidelines for Single Nucleotide Polymorphism (SNP) Analysis by Forensic DNA Testing Laboratories. SWGDAM Communications: United States; 2024.
159. Oosthuizen T, Howes LM. The development of forensic DNA analysis: New debates on the issue of fundamental human rights. Forensic Sci Int Genet 2022;56:102606; doi: 10.1016/j.fsigen.2021.102606
160. Taylor M. Next Generation Sequencing Accepted in Court for First Time: Forensic on the Scene and in the Lab. 2024. Available from: https://www.forensicmag.com/610513-Next-Generation-Sequencing-Deemed-Admissible-in-Court-for-First-Time [Last accessed: February 5, 2024].
161. Alonso A, Müller P, Roewer L, et al. European survey on forensic applications of massively parallel sequencing. Forensic Sci Int Genet 2017;29:e23–e25; doi: 10.1016/j.fsigen.2017.04.017
162. Budowle B, Arnette A, Sajantila A. A cost–benefit analysis for use of large SNP panels and high throughput typing for forensic investigative genetic genealogy. Int J Legal Med 2023;137(5):1595–1614; doi: 10.1007/s00414-023-03029-7
163. Genetic Engineering & Biotechnology News. Illumina Reveals New High-Throughput Instrument, NovaSeq X. 2022. Available from: https://www.genengnews.com/news/illumina-reveals-new-high-throughput-instrument-novaseq-x [Last accessed: February 5, 2024].
164. Zavala EI, Thomas JT, Sturk-Andreaggi K, et al. Ancient DNA methods improve forensic DNA profiling of Korean War and World War II unknowns. Genes 2022;13(1):129; doi: 10.3390/genes13010129
165. Xavier C, Eduardoff M, Bertoglio B, et al. Evaluation of DNA extraction methods developed for forensic and ancient DNA applications using bone samples of different age. Genes 2021;12(2):146; doi: 10.3390/genes12020146
166. Rohland N, Glocke I, Aximu-Petri A, et al. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nat Protoc 2018;13(11):2447–2461; doi: 10.1038/nprot.2007.247
167. Loreille OM, Diegoli TM, Irwin JA, et al. High efficiency DNA extraction from bone by total demineralization. Forensic Sci Int Genet 2007;1(2):191–195; doi: 10.1016/j.fsigen.2007.02.006
168. Xavier C, Parson W. Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit—MPS forensic application for the MiSeq FGx™ benchtop

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

sequencer. Forensic Sci Int Genet 2017;28:188–194; doi: 10.1016/j.fsigen.2017.02.018
169. Emery MV, Bolhofner K, Winingear S, et al. Reconstructing full and partial STR profiles from severely burned human remains using comparative ancient and forensic DNA extraction techniques. Forensic Sci Int Genet 2020;46:102272; doi: 10.1016/j.fsigen.2020.102272
170. Thomas JT, Cavagnino C, Kjelland K, et al. Evaluating the usefulness of human DNA quantification to predict DNA profiling success of historical bone samples. Genes 2023;14(5):994; doi: 10.3390/genes14050994
171. Hofreiter M, Sneberger J, Pospisek M, et al. Progress in forensic bone DNA analysis: Lessons learned from ancient DNA. Forensic Sci Int Genet 2021;54:102538; doi: 10.1016/j.fsigen.2021.102538
172. Ertürk MS, Fitzpatrick C, Press M, et al. Analysis of the genealogy process in forensic genetic genealogy. J Forensic Sci 2022;67(6):2218–2229; doi: 10.1111/1556-4029.15127
173. Board for Certification of Genealogists. Genetic Genealogy Education. 2015. Available from: https://bcgcertification.org/genetic-genealogy-education-for-us [Last accessed: March 20, 2022].
174. Novroski NMM. The Idaho student homicides and the future of forensic genetic genealogy. Forensic Genom 2023;3(1):3–8; doi: 10.1089/forensic.2023.0004
175. Hendry L. Genetic Genealogy Is Cracking Cases Once Thought Unsolvable. Not All Police Forces Can Afford to Use It. CBC News; 2024. Available from: https://www.cbc.ca/news/canada/montreal/genetic-genealogy-montreal-police-svpm-longueuil-police-1.7107656 [Last accessed: February 9, 2024].
176. Gurney D, Press M, Moore C, et al. The need for standards and certification for investigative genetic genealogy, and a notice of action. Forensic Sci Int 2022;341:111495; doi: 10.1016/j.forsciint.2022.111495
177. Board of Certification for Investigative Genetic Genelaogy. Board of Certification for Investigative Genetic Genelaogy. 2022. Available from: https://bcigg.org [Last accessed: December 12, 2022].
178. Temkin J, Prainsack B, Moore C. Ethical Principles. Biometrics & Forensics Ethics Group: United Kingdom; 2018.
179. Roffey P, Scudder N. Privacy implications of the new "omic" technologies in law enforcement. WIREs Forensic Sci 2021;4(3):e1445; doi: 10.1002/wfs2.1445
180. Samuel G, Kennett D. The impact of investigative genetic genealogy: Perceptions of UK professional and public stakeholders. Forensic Sci Int Genet 2020;48:102366; doi: 10.1016/j.fsigen.2020.102366
181. Guerrini CJ, Robinson JO, Petersen D, et al. Should police have access to genetic genealogy databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique. PLoS Biol 2018;16(10):e2006906; doi: 10.1371/journal.pbio.2006906
182. Quinton AR, Kelty SF, Scudder N. Attitudes towards police use of consumer/private DNA databases in investigations. Sci Justice 2022;62(3):263–271; doi: 10.1016/j.scijus.2022.02.009
183. Machado H, Granja R. Emerging DNA Technologies and Stigmatisation. In: Forensic Genetics in the Governance of Crime. (Machado H, Granja R. eds.) Springer: Singapore, 2020; pp. 85–104.
184. Granja R, Machado H. Ethical controversies of familial searching: The views of stakeholders in the United Kingdom and in Poland. Sci Technol Hum Values 2019;44(6):1068–1092; doi: 10.1177/0162243919828219
185. Beckford M, Bridge M. Police Won't Use Genealogy Sites for Cold Cases. *The Times* [Internet], 2020. Available from: https://www.thetimes.co.uk/article/police-wont-use-genealogy-sites-for-cold-cases-vvk0rbhqg [Last accessed: March 31, 2022].
186. FIND Strategy Board, Open Minutes of Meeting [Press Release]. March 24 2020.
187. Maryland General Assembly. Criminal Procedure—Forensic Genetic Genealogical DNA Analysis, Searching, Regulation, and Oversight, Maryland, HB0240, 2021. Available from: https://mgaleg.maryland.gov/mgawebsite/Legislation/Details/hb0240?ys=2021RS [Last accessed: November 28, 2023].
188. Maryland General Assembly. Genetic Information Privacy—Consumer Protection and Forensic Genealogy, Maryland, HB0866, 2022. Available from: https://mgaleg.maryland.gov/mgawebsite/Legislation/Details/hb0866?ys=2022RS [Last accessed: November 28, 2023].
189. Terpstra P. Maryland Quietly Shelves Part of Genealogy Privacy Law. 2022. Available from: https://www.wmar2news.com/infocus/maryland-quietly-shelves-parts-of-genealogy-privacy-law [Last accessed: December 12, 2022].
190. Newsy. Pioneering Privacy Law Runs into More Trouble in Maryland. 2022. Available from: https://www.ktvh.com/news/national/pioneering-privacy-law-runs-into-more-trouble-in-maryland [Last accessed: December 12, 2022].
191. Taylor M. Maryland, Montana Become First States to Restrict Genetic Genealogy Searches: Forensic on the Scene and in the Lab; 2021. Available from: https://www.forensicmag.com/577298-Maryland-Montana-Become-First-States-to-Restrict-Genetic-Genealogy-Searches [Last accessed: December 12, 2022].
192. Consumer DNA Database Searches—Familial DNA Searches—Warrant Required, Montana, Section 44-6-104, 2023. Available from: https://www.leg.mt.gov/bills/mca/title_0440/chapter_0060/part_0010/section_0040/0440-0060-0010-0040.html [Last accessed: February 5, 2024].
193. Genetic Information Privacy Act, Montana, SB0351, 2023. Available from: https://leg.mt.gov/bills/2023/billhtml/SB0351.htm [Last accessed: February 5, 2024].
194. Investigative Genetic Genealogy Modifications, Utah, SB156, 2023. Available from: https://le.utah.gov/~2023/bills/static/SB0156.html [Last accessed: February 5, 2024].
195. Forensic Investigative Genetic Genealogy Grant Program, Florida, SB678, 2024. Available from: https://www.flsenate.gov/Session/Bill/2024/678 [Last accessed: February 5, 2024].
196. Stajano F, Bianchi L, Liò P, et al. Forensic Genomics: Kin Privacy, Driftnets and Other Open Questions. In: Proceedings of the 7th ACM Workshop on Privacy in the Electronic Society, ACM Press: New York, United States; 2008; doi: 10.1145/1456403.1456407
197. Moran KS. Damned by DNA—Balancing personal privacy with public safety. Forensic Sci Int 2018;292:e3–e4; doi: 10.1016/j.forsciint.2018.09.011
198. Erlich Y, Shor T, Carmi S, et al. Re-identification of genomic data using long range familial searches. Science 2018;362(6415):690–694; doi: 10.1126/science.aau4832
199. Schwab AP, Luu HS, Wang J, et al. Genomic privacy. Clin Chem 2018;64(12):1696–1703; doi: 10.1373/clinchem.2018.289512
200. Rocher L, Hendrickx JM, De Montjoye Y-A. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 2019;10(1):3069; doi: 10.1038/s41467-019-10933-3
201. Wright D, De Hert P. Introduction to Privacy Impact Assessment. In: Privacy Impact Assessment, Vol. 6. (Wright D, De Hert P. eds.) Springer: Dordrecht, 2012; pp. 3–32.
202. Machado H, Silva S. Investigative genetic genealogy: An ethical and privacy assessment framework tool is needed. Forensic Sci Rev 2022;34(1):17–19.
203. Scudder N, McNevin D, Kelty SF, et al. Forensic DNA phenotyping: Developing a model privacy impact assessment. Forensic Sci Int Genet 2018;34:222–230; doi: 10.1016/j.fsigen.2018.03.005
204. Miles C. Privacy Impact Assessment for the Rapid DNA System. US Department of Homeland Security: United States; 2013.
205. Dowdeswell TL. Forensic genetic genealogy: A profile of cases solved. Forensic Sci Int Genet 2022;58:102679; doi: 10.1016/j.fsigen.2022.102679
206. Ford-Monroe J. Why familial searches of civilian DNA databases can and should survive carpenter. Hastings Law J 2021;72(6):1717.
207. Butler JM. Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019–2022. Forensic Sci Int Synerg 2023;6:100311; doi: 10.1016/j.fsisyn.2022.100311
208. de Groot NF, Van Beers BC, Meynen G. Commercial DNA tests and police investigations: A broad bioethical perspective. J Med Ethics 2021;47(12):788–795; doi: 10.1136/medethics-2021-107568
209. National Crime Information Center. 2022 NCIC Missing Person and Unidentified Person Statistics. 2022. Available from: https://www.fbi.gov/file-repository/2022-ncic-missing-person-and-unidentified-person-statistics.pdf/view [Last accessed: March 20, 2022].
210. National Missing and Unidentified Persons System. NamUs Case Statistics—All Cases [December 2022]: National Institute of Justice; 2022. Available from: https://namus.nig.ojp.gov/library/reports-and-statistics [Last accessed: May 14, 2023].
211. National Missing Persons Coordination Centre. About | National Missing Persons Coordination Centre; 2023. Available from: https://www.missingpersons.gov.au/about [Last accessed: August 17, 2023].
212. Illumina. Infinium™ Global Screening Array-24 v3.0 BeadChip, Data Sheet, Pub. No. 370-2016-016-G. 2020.
213. The H600 Project. Consanguinity. 2018. Available from: https://h600.org/wiki/Consanguinity [Last accessed: November 2, 2021].

Downloaded by Jodie Ward from www.liebertpub.com at 03/18/24. For personal use only.

1.7 Research Aims

The research presented in this thesis aimed to evaluate and operationalise a SNP genotyping technology for the AFP DNA Program that could be applied to Australia's unidentified and missing persons casework and generate new investigative leads. By using volunteer samples with self-declared metadata, research PM samples and UHR casework samples approved for these studies by Australian law enforcement jurisdictions, the laboratory, analysis and inference pipelines were optimised and implemented.

The project aimed to achieve this by:

- Evaluating and comparing the available SNP genotyping panels employing TAS technology to provide recommendations for forensic laboratory implementation.
- Validating and optimising the ForenSeq® Kintelligence Kit for unidentified and missing persons casework.
- Evaluating identity-informative markers for applications in Australian populations and establishing SNP population allele frequencies.
- Optimising DNA intelligence pipelines for the inference of biological sex, hair colour, eye colour, BGA and kinship.
- Evaluating STRs and SNPs for kinship analysis using LR calculations and haplotype matching.
- Assessing the impact of information loss on extended kinship analysis capabilities with the ForenSeq® Kintelligence Kit.

1.8 References

1. Ritter N. Missing persons and unidentified remains: The nation's silent mass disaster. NIJ Journal. 2007;256(7):2-7.
2. International Committee of the Red Cross. The Forensic Human Identification Process: An Integrated Approach Geneva, Switzerland2022. [Last accessed: 29/05/2022]. Available from: https://www.icrc.org/en/publication/4590-forensic-human-identification-process-integrated-approach.
3. International Commission on Missing Persons. DNA-Led Human Identification 2016. [Last accessed: 05/06/2022]. Available from: https://www.icmp.int/news/dna-led-human-identification/.
4. Bricknell S. Missing persons: who is at risk? Australian Institute of Criminology Research Report. 2017:1-40.

5. International Committee of the Red Cross. Missing persons | Internationa Committee of the Red Cross 2022. [Last accessed: 05/06/2022]. Available from: https://www.icrc.org/en/war-and-law/protected-persons/missing-persons.
6. INTERPOL (DNA Monitoring Expert Group). Best Practice Principles: Recommendations on the Use of DNA for the Identification of Missing Persons and Unidentified Human Remains 2015. Available from: https://www.interpol.int/content/download/4877/file/ENGLISH%20MPUHR%20DNA%20RECOMMENDATION%20INTERPOL.pdf.
7. Ward J. The past, present and future state of missing persons investigations in Australia. Aust J Forensic Sci. 2018:1-15. doi: 10.1080/00450618.2018.1466535.
8. Ward J. Best practice recommendations for the establishment of a national DNA identification program for missing persons: A global perspective. Forensic Sci Int Genet Suppl Ser. 2017;6:e43-e5. doi: 10.1016/j.fsigss.2017.09.009.
9. Australian Federal Police. National DNA Program for Unidentified and Missing Persons Canberra: Commonwealth of Australia; 2022. [Last accessed: 11/12/2023]. Available from: https://www.missingpersons.gov.au/support/national-dna-program-unidentified-and-missing-persons.
10. International Criminal Police Organisation. INTERPOL Disaster Victim Identification Guide. 2018.
11. International Committee of the Red Cross. Missing people, DNA analysis and identification of remains. A guide for best practice in armed conflicts and other situations of armed violence. 2009.
12. International Commission on Missing Persons. Standard Operating Procedure for Sampling Bone ad Tooth Specimens from Human Remains for DNA Testing at the ICMP. 2015.
13. Alvarez-Cubero MJ, Saiz M, Martinez-Gonzalez LJ, Alvarez JC, Eisenberg AJ, Budowle B, et al. Genetic Identification of Missing Persons: DNA Analysis of Human Remains and Compromised Samples. Pathobiology. 2012;79(5):228-38. doi: 10.1159/000334982.
14. National Crime Information Center. 2023 NCIC Missing Person and Unidentified Person Statistics: Federal Bureau of Investigation; 2024. [Last accessed: 17/01/2025]. Available from: https://le.fbi.gov/file-repository/2023-ncic-missing-person-and-unidentified-person-statistics.pdf.

15. National Missing and Unidentified Persons System. NamUs Monthly Case Report - December 2024: National Institute of Justice; 2024. [Last accessed: 17/01/2025]. Available from: https://namus.nig.ojp.gov/library/reports-and-statistics.
16. Rodriguez AL, Smiley-McDonald HM, Cummings MS, Wire S, Slack D, Williams CL, et al. Understanding unidentified human remains investigations through the United States census data. Forensic Sci Int Synerg. 2022;4:100225. doi: 10.1016/j.fsisyn.2022.100225.
17. Reid KM, Martin LJ, Heathfield LJ. Understanding the burden of unidentified bodies: a systematic review. Int J Legal Med. 2023:1-10. doi: 10.1007/s00414-023-02968-5.
18. UK Missing Persons Unit. Home | UK Missing Persons Unit 2022. Available from: https://missingpersons.police.uk/en-gb/home#.
19. Watson JL, McNevin D, Ward J. Genetic Kinship Testing Techniques for Human Remains Identification and Missing Persons Investigations. Forensic Genom. 2024;4(1):4-23. doi: 10.1089/forensic.2023.0018.
20. UK Missing Persons Unit. UKMPU Missing Persons Data Report 2022/23: National Crime Agency; 2023. [Last accessed: 17/01/2025]. Available from: https://missingpersons.police.uk/en-gb/resources/downloads/missing-persons-statistical-bulletins.
21. National Missing Persons Coordination Centre. About | National Missing Persons Coordination Centre 2023. [Last accessed: 17/08/2023]. Available from: https://www.missingpersons.gov.au/about.
22. International Commission on Missing Persons. Data Systems and Data Coordination 2024. [Last accessed: 22/01/2025]. Available from: https://icmp.int/what-we-do/data-systems-and-data-coordination/.
23. International Committee of the Red Cross. The Forensic Identification of Human Remains 2013. Available from: https://www.icrc.org/en/publication/4590-forensic-human-identification-process-integrated-approach.
24. International Criminal Police Organization. INTERPOL unveils new global database to identify missing persons through family DNA 2021. Available from: https://www.interpol.int/en/News-and-Events/News/2021/INTERPOL-unveils-new-global-database-to-identify-missing-persons-through-family-DNA.
25. Royal Canadian Mounted Police. National Centre for Missing Persons and Unidentified Remains (NCMPUR) 2018. [Last accessed: 17/01/2025]. Available

from: https://www.rcmp-grc.gc.ca/en/national-centre-missing-persons-and-unidentified-remains-ncmpur.

26. University of North Texas Center for Human Identification. Center for Human Identification - A globally-recognised leader in forensic identification: The University of North Texas Health Science Center at Fort Worth,; 2022. [Last accessed: 05/06/2022]. Available from: https://www.unthsc.edu/center-for-human-identification/.
27. Ward J, Coakley L, Grisedale K, Seddon S, Spiden M, Watson JL, et al. Operationalisation of the National DNA Program for Unidentified and Missing Persons' forensic investigative genetic genealogy capability for human remains identification in Australia [in press, accepted 27 January 2024]. Forensic Genom. 2024.
28. Dahal A, McNevin D, Chikhani M, Ward J. An interdisciplinary forensic approach for human remains identification and missing persons investigations. Wiley Interdisciplinary Reviews: Forensic Science. 2023:e1484. doi: 10.1002/wfs2.1484.
29. Australian Federal Police. National DNA Program for Unidentified and Missing Persons Forensic Techniques 2020. [Last accessed: 19/09/2021]. Available from: https://www.missingpersons.gov.au/sites/default/files/National%20DNA%20Program%20for%20Unidentified%20and%20Missing%20Persons%20%20Forensic%20Techniques.pdf.
30. Australian Federal Police. National DNA Program for Unidentified and Missing Persons DNA Testing 2020. [Last accessed: 19/09/2021]. Available from: https://www.missingpersons.gov.au/sites/default/files/FINAL%20FACT%20SHEET_%20National%20DNA%20Program%20for%20Unidentified%20and%20Missing%20Persons%20%20DNA%20Testing.pdf.
31. Weber JL, Wong C. Mutation of human short tandem repeats. Hum Mol Genet. 1993;2(8):1123-8. doi: 10.1093/hmg/2.8.1123.
32. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, et al. Identification of the remains of the Romanov family by DNA analysis. Nat Genet. 1994;6(2):130-5.
33. Puch-Solis R, Pope S. Interpretation of DNA data within the context of UK forensic science—evaluation. Emerg Top Life Sci. 2021. doi: 10.1042/ETLS20200340.

34. Australian Law Reform Commission. Report 96: Essentially yours - the protection of human genetic information in Australia. 2003.
35. Thermo Fisher Scientific. GlobalFiler PCR Amplification Kits. 2019.
36. Promega. PowerPlex(R) 21 System for use on the Applied Biosystems(R) Genetic Analyzers [technical manual]. 2017.
37. Budowle B, Van Daal A. Forensically relevant SNP classes. Biotechniques. 2008;44(5):603-10. doi: 10.2144/000112806.
38. Angers A, Kagkli D, Oliva L, Petrillo M, Raffael B. Study on DNA Profilng Technology for Its Implementation in Central Schengen Information System: EUR29766 Publications Office of the EU, Luxembourg; 2019. Available from: https://op.europa.eu/en/publication-detail/-/publication/7ae614b0-b274-11e9-9d01-01aa75ed71a1/language-en.
39. Decorte R. Genetic identification in the 21st century—current status and future developments. Forensic Sci Int. 2010;201(1-3):160-4.
40. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SLB, Karoma NJ, et al. Developing a SNP panel for forensic identification of individuals. Forensic Sci Int. 2006;164(1):20-32. doi: 10.1016/j.forsciint.2005.11.017.
41. Butler JM, Coble MD, Vallone PM. STRs vs. SNPs: thoughts on the future of forensic DNA testing. Forensic Science, Medicine, and Pathology. 2007;3(3):200-5. doi: 10.1007/s12024-007-0018-1.
42. Antunes J, Radecke S, Snedecor J, Padmabandu G, Stephens KM. Application of a dense SNP PCR multiplex for high throughput kinship determination using Next Generation Sequencing (NGS). Forensic Sci Int Genet Suppl Ser. 2022;8:324-5. doi: 10.1016/j.fsigss.2022.10.007.
43. Grandell I, Samara R, Tillmar A. A SNP panel for identity and kinship testing using massive parallel sequencing. Int J Legal Med. 2016;130:905-14. doi: 10.1007/s00414-016-1341-4.
44. Kling D. On the use of dense sets of SNP markers and their potential in relationship inference. Forensic Sci Int Genet. 2019;39:19-31. doi: 10.1016/j.fsigen.2018.11.022.
45. Pontes ML, Fondevila M, Laréu MV, Medeiros R. SNP Markers as Additional Information to Resolve Complex Kinship Cases. Transfus Med Hemother. 2015;42(6):385-8. doi: 10.1159/000440832.

46. Wei Y-L, Li C-X, Jia J, Hu L, Liu Y. Forensic Identification Using a Multiplex Assay of 47 SNPs*. J Forensic Sci. 2012;57(6):1448-56. doi: 10.1111/j.1556-4029.2012.02154.x.
47. Kidd JR, Friedlaender FR, Speed WC, Pakstis AJ, De La Vega FM, Kidd KK. Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. Investig Genet. 2011;2(1):1. doi: 10.1186/2041-2223-2-1.
48. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis. 2006;27(9):1713-24. doi: 10.1002/elps.200500671.
49. Walsh S, Chaitanya L, Breslin K, Muralidharan C, Bronikowska A, Pospiech E, et al. Global skin colour prediction from DNA. Hum Genet. 2017;136:847-63. doi: 10.1007/s00439-017-1808-5.
50. Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L, et al. Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet. 2014;9:150-61. doi: 10.1016/j.fsigen.2013.12.006.
51. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M. IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet. 2011;5(3):170-80. doi: 10.1016/j.fsigen.2010.02.004.
52. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet. 2013;7(1):98-115. doi: 10.1016/j.fsigen.2012.07.005.
53. Andersen MM, Balding DJ. How convincing is a matching Y-chromosome profile? PLoS Genet. 2017;13(11):e1007028. doi: 10.1371/journal.pgen.1007028.
54. Kayser M. Forensic use of Y-chromosome DNA: a general overview. Hum Genet. 2017;136(5):621-35. doi: 10.1007/s00439-017-1776-9.
55. Pinto N, Gusmão L, Amorim A. X-chromosome markers in kinship testing: A generalisation of the IBD approach identifying situations where their contribution is crucial. Forensic Sci Int Genet. 2011;5(1):27-32. doi: 10.1016/j.fsigen.2010.01.011.

56. Tomas C, Sanchez JJ, Castro JA, Børsting C, Morling N. Forensic usefulness of a 25 X-chromosome single-nucleotide polymorphism marker set. Transfusion. 2010;50(10):2258-65. doi: 10.1111/j.1537-2995.2010.02696.x.
57. Wise AS, Shi M, Weinberg CR. Family-based multi-SNP X chromosome analysis using parent information. Fron Genet. 2016;7:20. doi: 10.3389/fgene.2016.00020.
58. Gomes C, Palomo-Díez S, López-Parra AM, Arroyo-Pardo E. Genealogy: The Tree Where History Meets Genetics. Genealogy. 2021;5(4):98. doi: 10.3390/genealogy5040098.
59. Claerhout S, Defraye C, Decorte R. Validation of Y-ancestor time calculators for forensic familial searching. Forensic Sci Int Genet Suppl Ser. 2019;7(1):411-3. doi: 10.1016/j.fsigss.2019.10.032.
60. Claerhout S, Vanpaemel S, Gill MS, Antiga LG, Baele G, Decorte R. YMrCA: Improving Y-chromosomal ancestor time estimation for DNA kinship research. Hum Mutat. 2021. doi: 10.1002/humu.24259.
61. Calafell F, Larmuseau MHD. The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. Hum Genet. 2017;136(5):559-73. doi: 10.1007/s00439-016-1740-0.
62. Gettings KB. Forensic ancestry and phenotype SNP analysis and integration with established forensic markers: The George Washington University; 2013.
63. King TE, Jobling MA. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. Trends Genet. 2009;25(8):351-60. doi: 10.1016/j.tig.2009.06.003.
64. Schaffner SF. The X chromosome in population genetics. Nat Rev Genet. 2004;5(1):43-51. doi: 10.1038/nrg1247.
65. Coakley L. X-DNA's Helpful Inheritance Patterns: Genie1 & Genetic Genealogy; 2019. [Last accessed: 06/01/2022]. Available from: https://www.genie1.com.au/x-dnas-helpful-inheritance-patterns/.
66. Barbaro A, Cormaci P. X-STR typing for an identification casework. International Congress Series. 2006;1288:513-5.
67. Alzate L, Agudelo N, Builes J. X-STRs as a tool for missing persons identification using only siblings as reference. Forensic Sci Int Genet Suppl Ser. 2015;5:e636-e7. doi: 10.1016/j.fsigss.2015.10.008.

68. Iwamura ESM, Guimarães MA, Evison MP. DNA Methods to Identify Missing Persons. In: Morewitz SJ, Sturdy Colls C, editors. Handbook of Missing Persons. Cham: Springer International Publishing; 2016. p. 337-52.
69. Auler-Bittencourt E, Iwamura ESM, Lima MJM, Da Silva IDCG, Dos Santos SEB. Exploring the applicability of analysing X chromosome STRs in Brazilian admixed population. Sci Justice. 2015;55(5):323-8. doi: 10.1016/j.scijus.2015.03.004.
70. Scientific Working Group on DNA Analysis Methods. Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Resulted Reported as Likelihood Ratios. 2016.
71. Taylor D, Bright J-A, McGovern C, Neville S, Grover D. Allele frequency database for GlobalFiler™ STR loci in Australian and New Zealand populations. Forensic Sci Int Genet. 2017;28:e38-e40. doi: 10.1016/j.fsigen.2017.02.012.
72. Butler JM. Forensic DNA typing: biology, technology, and genetics of STR markers: Elsevier; 2005.
73. Machado H, Granja R. DNA Databases and Big Data. Springer Singapore; 2020. p. 57-70.
74. Gill P, Werrett DJ, Budowle B, Guerrieri R. An assessment of whether SNPs will replace STRs in national DNA databases - joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). Sci Justice. 2004;44(1):51-3. doi: 10.1016/s1355-0306(04)71685-8.
75. Gill P, Hicks T, Butler JM, Connolly E, Gusmão L, Kokshoorn B, et al. DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence-Guidelines highlighting the importance of propositions: Part I: evaluation of DNA profiling comparisons given (sub-) source propositions. Forensic Sci Int Genet. 2018;36:189-202. doi: https://doi.org/10.1016/j.fsigen.2018.07.003.
76. Biology Specialist Advisory Group. Biology Specialist Advisory Group (BSAG) Statistics Scientific Working Group (StatSWG) recommendations for the interpretation of DNA, Version 3. BSAG StatSWG. 2016.
77. National Research Council Committee on DNA Forensic Science. The Evaluation of Forensic DNA Evidence. Washington (DC): National Academies Press (US). 1996. doi: 10.17226/5141.

78. Crow JF, Kimura M. An introduction to population genetics theory. New York: Harper and Row; 1970.
79. Buckleton J, Curran J, Goudet J, Taylor D, Thiery A, Weir BS. Population-specific FST values for the forensic STR markers: A worldwide survey. Forensic Sci Int Genet. 2016;23:91-100. doi: 10.1016/j.fsigen.2016.03.004.
80. Delest A, Godfrin D, Chantrel Y, Ulus A, Vannier J, Faivre M, et al. Sequenced-based French population data from 169 unrelated individuals with Verogen's ForenSeq DNA signature prep kit. Forensic Sci Int Genet. 2020;47:102304. doi: 10.1016/j.fsigen.2020.102304.
81. Guevara EK, Palo JU, King JL, Buś MM, Guillen S, Budowle B, et al. Autosomal STR and SNP characterization of populations from the Northeastern Peruvian Andes with the ForenSeq™ DNA Signature Prep Kit. Forensic Sci Int Genet. 2021;52:102487. doi: 10.1016/j.fsigen.2021.102487.
82. The International Genome Sample Resource. 1000 Genomes Project 2021. [Last accessed: 17 October]. Available from: https://www.internationalgenome.org/.
83. The International HapMap Consortium. The International HapMap Project. Nature. 2003;426:789-96. doi: 10.1038/nature02168.
84. Davenport L, Devesse L, Court DS, Ballard D. Forensic identity SNPs: Characterisation of flanking region variation using massively parallel sequencing. Forensic Sci Int Genet. 2023;64:102847. doi: 10.1016/j.fsigen.2023.102847.
85. Tao R, Dong X, Zhen X, Xia R, Qu Y, Liu S, et al. Population genetic analyses of Eastern Chinese Han nationality using ForenSeq™ DNA Signature Prep Kit. Molecular Genetics and Genomics. 2024;299(1):9. doi: 10.1007/s00438-024-02121-w.
86. Aguilar-Velázquez JA, Duran-Salazar MÁ, Córdoba-Mercado MF, Coronado-Avila CE, Salas-Salas O, Martinez-Cortés G, et al. Characterization of 58 STRs and 94 SNPs with the ForenSeq™ DNA signature prep kit in Mexican-Mestizos from the Monterrey city (Northeast, Mexico). Molecular Biology Reports. 2022;49(8):7601-9. doi: 10.1007/s11033-022-07575-y.
87. Kiesler KM, Borsuk LA, Steffen CR, Vallone PM, Gettings KB. US Population Data for 94 Identity-Informative SNP Loci. Genes. 2023;14:1071. doi: 10.3390/genes14051071.

88. Augustinus D, Gahan ME, McNevin D. Development of a forensic identity SNP panel for Indonesia. Int J Legal Med. 2015;129:681-91. doi: 10.1007/s00414-014-1055-4.
89. Guo F, Liu Z, Long G, Zhang B, Liu D, Yu S. Performance and characterization of 94 identity-informative SNPs in Northern Han Chinese using ForenSeq™ DNA signature prep kit. J Forensic Legal Med. 2024;103:102678. doi: 10.1016/j.jflm.2024.102678.
90. Leclair B. Large-scale comparative genotyping and kinship analysis: evolution in its use for human identification in mass fatality incidents and missing persons databasing. International Congress Series. 2004;1261:42-4. doi: 10.1016/s0531-5131(03)01661-3.
91. Alonso A, Martín P, Albarrán C, Garcí P, De Simón LF, Iturralde MJ, et al. Challenges of DNA profiling in mass disaster investigations. Croat Med J. 2005;46(4):540-8.
92. Scientific Working Group on DNA Analysis Methods. Guidelines for Missing Persons Casework. SWGDAM Communications. 2014.
93. Montelius K, Lindblom B. DNA analysis in disaster victim identification. Forensic Science, Medicine, and Pathology. 2012;8(2):140-7. doi: 10.1007/s12024-011-9276-z.
94. Knoppers B, Avard D, Sénécal K. Newborn screening programmes: Emerging biobanks? Nor Epidemiol. 2012;21(2).
95. Sweet D, Hildebrand D, Phillips D. Identification of a skeleton using DNA from teeth and a PAP smear. Journal of Forensic Science. 1999;44(3):630-3. doi: 10.1520/JFS14522J.
96. Hartman D, Benton L, Morenos L, Beyer J, Spiden M, Stock A. The importance of Guthrie cards and other medical samples for the direct matching of disaster victims using DNA profiling. Forensic Sci Int. 2011;205(1-3):59-63. doi: 10.1016/j.forsciint.2010.05.023.
97. Hartman D, Drummer O, Eckhoff C, Scheffer JW, Stringer P. The contribution of DNA to the disaster victim identification (DVI) effort. Forensic Sci Int. 2011;205(1-3):52-8. doi: 10.1016/j.forsciint.2010.09.024.
98. Xavier M, Bento A, Costa A, Corte-Real A, Veloso C, Sampaio L, et al. Primary teeth as DNA reference sample in disaster victim identification (DVI). Forensic Sci Int Genet Suppl Ser. 2011;3(1):e381-e2. doi: 10.1016/j.fsigss.2011.09.052.

99. Grisedale KS, Murphy GM, Brown H, Wilson MR, Sinha SK. Successful nuclear DNA profiling of rootless hair shafts: a novel approach. Int J Legal Med. 2018;132:107-15. doi: 10.1007/s00414-017-1698-z.
100. Kincaid KH. Solve Cold Cases with DNA from Rootless Hair using Genetic Genealogy The ISHI Report: The ISHI Report; 2020. [Last accessed: 2 February]. Available from: https://promega.foleon.com/theishireport/november-2020/solve-cold-cases-with-dna-from-rootless-hair-using-genetic-genealogy/.
101. Kapp JD, Green RE, Shapiro B. A fast and efficient single-stranded genomic library preparation method optimized for ancient DNA. Journal of Heredity. 2021;112(3):241-9. doi: 10.1093/jhered/esab012.
102. De Valck E. Major incident response: collecting ante-mortem data. Forensic Sci Int. 2006;159:S15-S9. doi: 10.1016/j.forsciint.2006.02.004.
103. Lau G, Tan WF, Tan PH. After the Indian Ocean tsunami: Singapore's contribution to the international disaster victim identification effort in Thailand. Ann Acad Med Singap. 2005;34(5):341. doi: https://pubmed.ncbi.nlm.nih.gov/16021223/.
104. Morgan OW, Sribanditmongkol P, Perera C, Sulasmi Y, Van Alphen D, Sondorp E. Mass Fatality Management following the South Asian Tsunami Disaster: Case Studies in Thailand, Indonesia, and Sri Lanka. PLoS Med. 2006;3(6):e195. doi: 10.1371/journal.pmed.0030195.
105. Brenner CH, Weir BS. Issues and strategies in the DNA identification of World Trade Center victims. Theor Popul Biol. 2003;63(3):173-8. doi: 10.1016/S0040-5809(03)00008-X.
106. Ge J, Budowle B, Chakraborty R. Choosing Relatives for DNA Identification of Missing Persons. J Forensic Sci. 2011;56:S23-S8. doi: 10.1111/j.1556-4029.2010.01631.x.
107. Amorim A, Pereira L. Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. Forensic Sci Int. 2005;150(1):17-21. doi: 10.1016/j.forsciint.2004.06.018.
108. Ge J, Chakraborty R, Eisenberg A, Budowle B. Comparisons of Familial DNA Database Searching Strategies. J Forensic Sci. 2011;56(6):1448-56. doi: 10.1111/j.1556-4029.2011.01867.x.
109. Ge J, Budowle B. How many familial relationship testing results could be wrong? PLoS Genet. 2020;16(8):e1008929. doi: 10.1371/journal.pgen.1008929.

110. National Institute of Justice. Familial DNA Searching: Current Approaches. National Institute of Justice. 2015:1-82.
111. Cho S, Kim M-Y, Lee JH, Lee HY, Lee SD. Large-scale identification of human bone remains via SNP microarray analysis with reference SNP database. Forensic Sci Int Genet. 2020;47:102293. doi: 10.1016/j.fsigen.2020.102293.
112. Jordan D, Mills D. Past, Present, and Future of DNA Typing for Analyzing Human and Non-Human Forensic Samples. Front Ecol Evol. 2021;9(172). doi: 10.3389/fevo.2021.646130.
113. ATDBio. Sequencing, forensic analysis and genetic analysis 2021. [Last accessed: 07/01/2022]. Available from: https://atdbio.com/nucleic-acids-book/Sequencing-forensic-analysis-and-genetic-analysis#DNA-Fingerprinting.
114. Jeffreys AJ. The man behind the DNA fingerprints: an interview with Professor Sir Alec Jeffreys. Investig Genet. 2013;4(1):21. doi: 10.1186/2041-2223-4-21.
115. Tillmar A, Sjölund P, Lundqvist B, Klippmark T, Älgenäs C, Green H. Whole-genome sequencing of human remains to enable genealogy DNA database searches – A case report. Forensic Sci Int Genet. 2020;46:102233. doi: 10.1016/j.fsigen.2020.102233.
116. McCord BR, Gauthier Q, Cho S, Roig MN, Gibson-Daw GC, Young B, et al. Forensic DNA Analysis. Anal Chem. 2019;91(1):673-88. doi: 10.1021/acs.analchem.8b05318.
117. Durney BC, Crihfield CL, Holland LA. Capillary electrophoresis applied to DNA: determining and harnessing sequence and structure to advance bioanalyses (2009–2014). Anal Bioanal Chem. 2015;407(23):6923-38. doi: 10.1007/s00216-015-8703-5.
118. Shewale JG, Qi L, Calandro LM. Principles, practice, and evolution of capillary electrophoresis as a tool for forensic DNA analysis. In: Shewale JG, Liu Q, editors. Forensic DNA Analysis: Current Practices and Emerging Technologis: CRC Press; 2012. p. 131.
119. Thermo Fisher Scientific. What is Fragment Analysis? 2021. Available from: https://www.thermofisher.com/au/en/home/life-science/sequencing/sequencing-learning-center/capillary-electrophoresis-information/what-is-fragment-analysis.html.
120. Rajagopalan RM, Fujimura JH. Variations on a Chip: Technologies of Difference in Human Genetics Research. J Hist Biol. 2018;51(4):841-73. doi: 10.1007/s10739-018-9543-x.

121. Hurd PJ, Nelson CJ. Advantages of next-generation sequencing versus the microarray in epigenetic research. Briefings in Functional Genomics and Proteomics. 2009;8(3):174-83. doi: 10.1093/bfgp/elp013.
122. Kling D, Phillips C, Kennett D, Tillmar A. Investigative genetic genealogy: Current methods, knowledge and practice. Forensic Sci Int Genet. 2021;52:102474. doi: 10.1016/j.fsigen.2021.102474.
123. Illumina. Infinium™ Global Screening Array-24 v3.0 BeadChip, Data Sheet, pub. no. 370-2016-016-G. 2020.
124. Erlich Y, Narayanan A. Routes for breaching and protecting genetic privacy. Nat Rev Genet. 2014;15(6):409-21. doi: 10.1038/nrg3723.
125. de Vries JH, Kling D, Vidaki A, Arp P, Kalamara V, Verbiest MMPJ, et al. Impact of SNP microarray analysis of compromised DNA on kinship classification success in the context of investigative genetic genealogy. Forensic Sci Int Genet. 2022;56:102625. doi: 10.1016/j.fsigen.2021.102625.
126. Davawala A, Stock A, Spiden M, Daniel R, McBain J, Hartman D. Forensic Genetic Genealogy using microarrays for the identification of human remains: the need for good quality samples – a pilot study. Forensic Sci Int. 2022:111242. doi: 10.1016/j.forsciint.2022.111242.
127. Keating B, Bansal AT, Walsh S, Millman J, Newman J, Kidd K, et al. First all-in-one diagnostic tool for DNA intelligence: genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip. Int J Legal Med. 2013;127(3):559-72. doi: 10.1007/s00414-012-0788-1.
128. Centre for Disease Control and Prevention. Whole Genome Sequencing (WGS) 2021. [Last accessed: 07/01/2022]. Available from: https://www.cdc.gov/pulsenet/pathogens/wgs.html.
129. Ilumina. Whole-Genome Sequencing (WGS) 2021. [Last accessed: 07/01/2022]. Available from: https://sapac.illumina.com/techniques/sequencing/dna-sequencing/whole-genome-sequencing.html.
130. Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu Q, et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nature methods. 2021;18(2):165-9. doi: 10.1038/s41592-020-01041-y.

131. Illumina. NovaSeq System 2024. [Last accessed: 8 Aug]. Available from: https://www.illumina.com/systems/sequencing-platforms/novaseq.html.
132. Illumina. An Introduction to Next-Generation Sequencing Technology 2017. [Last accessed: 27 May]. Available from: https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.
133. Verogen. MiSeq FGx Sequencing System Reference Guide. MiSeq FGx Sequencing System. 2021.
134. Alonso A, Müller P, Roewer L, Willuweit S, Budowle B, Parson W. European survey on forensic applications of massively parallel sequencing. Forensic Sci Int Genet. 2017;29:e23-e5. doi: 10.1016/j.fsigen.2017.04.017.
135. Tillmar A, Sturk-Andreaggi K, Daniels-Higginbotham J, Thomas JT, Marshall C. The FORCE panel: An all-in-one SNP marker set for confirming investigating genetic genealogy leads and for general forensic applications. bioRxiv. 2021. doi: 10.1101/2021.11.30.470354.
136. Antunes J, Walichiewicz P, Forouzmand E, Barta R, Didier M, Han Y, et al. Developmental Validation of the ForenSeq® Kintelligence Kit, MiSeq Fgx® Sequencing System and ForenSeq Universal Analysis Software. Forensic Sci Int Genet. 2024;71:103055. doi: 10.1016/j.fsigen.2024.103055.
137. Thermo Fisher Scientific. Ion AmpliSeq NGS Panels for Targeted Sequencing 2024. [Last accessed: 14 March]. Available from: https://www.thermofisher.com/au/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-select-targets/ampliseq-target-selection.html.
138. Greytak EM, Kaye DH, Budowle B, Moore C, Armentrout SL. Privacy and genetic genealogy data. Science. 2018;361(6405):857.1-. doi: 10.1126/science.aav0330.
139. Hazel JW, Clayton EW, Malin BA, Slobogin C. Is it time for a universal genetic forensic database? Science. 2018;362(6417):898-900. doi: 10.1126/science.aav5475.
140. Xavier C, Parson W. Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit – MPS forensic application for the MiSeq FGx™ benchtop sequencer. Forensic Sci Int Genet. 2017;28:188-94. doi: 10.1016/j.fsigen.2017.02.018.

141. Kayser M, Branicki W, Parson W, Phillips C. Recent advances in Forensic DNA Phenotyping of appearance, ancestry and age. Forensic Sci Int Genet. 2023;65:102870. doi: 10.1016/j.fsigen.2023.102870.
142. Kayser M, De Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet. 2011;12(3):179-92. doi: 10.1038/nrg2952.
143. Novroski NMM. The use of massively parallel sequencing in forensic DNA typing: what will the future hold? 2nd Forensic Sciences Virtual Event: Labroots 2020.
144. Scudder N, McNevin D, Kelty SF, Walsh SJ, Robertson J. Massively parallel sequencing and the emergence of forensic genomics: Defining the policy and legal issues for law enforcement. Sci Justice. 2018;58(2):153-8. doi: 10.1016/j.scijus.2017.10.001.
145. Ward J. To investigate specialist DNA techniques for the identification of compromised human remains. The Winston Churchill Memorial Trust of Australia. 2016.
146. Chaitanya L, Pajnič IZ, Walsh S, Balažic J, Zupanc T, Kayser M. Bringing colour back after 70 years: Predicting eye and hair colour from skeletal remains of World War II victims using the HIrisPlex system. Forensic Sci Int Genet. 2017;26:48-57. doi: 10.1016/j.fsigen.2016.10.004.
147. Ward J. A multifaceted, multijurisdictional, multiagency, and multidisciplinary approach to investigating unidentified and missing persons cases in Australia. Forensic Sci Int Genet Suppl Ser. 2022;8:53-5. doi: 10.1016/j.fsigss.2022.09.020.
148. Thermo Fisher Scientific. Quantifiler™ HP and Trio DNA Quantification Kits User Guide [User Guide]. 2017.
149. Hennessy LK, Mehendale N, Chear K, Jovanovich S, Williams S, Park C, et al. Developmental validation of the GlobalFiler® express kit, a 24-marker STR assay, on the RapidHIT® System. Forensic Sci Int Genet. 2014;13:247-58.
150. Churchill JD, Schmedes SE, King JL, Budowle B. Evaluation of the Illumina ® Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling. Forensic Sci Int Genet. 2016;20:20-9. doi: 10.1016/j.fsigen.2015.09.009.
151. Lynch V, Heathfield LJ, Budowle B. Disclosure of biological sex may impact individual privacy. Forensic Sci Int Genet. 2025;76:103213. doi: 10.1016/j.fsigen.2024.103213.

152. Finaughty C, Reid KM, Alli IH, Heathfield LJ. A first for forensic genetics in Africa: successful identification of skeletal remains from the marine environment using massively parallel sequencing. Forensic Sci Int Genet. 2020;49:102370. doi: 10.1016/j.fsigen.2020.102370.
153. Cheung EY, Gahan ME, McNevin D. Predictive DNA analysis for biogeographical ancestry. Aust J Forensic Sci. 2018;50(6):651-8.
154. Schneider PM, Prainsack B, Kayser M. The use of forensic DNA phenotyping in predicting appearance and biogeographic ancestry. Deutsches Ärzteblatt International. 2019;116(51-52):873. doi: 10.3238/arztebl.2019.0873.
155. McNevin D. Forensic inference of biogeographical ancestry from genotype: The Genetic Ancestry Lab. Wiley Interdisciplinary Reviews: Forensic Science. 2020;2(2):e1356. doi: 10.1002/wfs2.1356.
156. Kidd KK, Speed WC, Pakstis AJ, Furtado MR, Fang R, Madbouly A, et al. Progress toward an efficient panel of SNPs for ancestry inference. Forensic Sci Int Genet. 2014;10:23-32. doi: 10.1016/j.fsigen.2014.01.002.
157. Gaspar HA, Breen G. Probabilistic ancestry maps: a method to assess and visualize population substructures in genetics. BMC Bioinformatics. 2019;20(1):1-11.
158. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004;20(2):289-90. doi: 10.1093/bioinformatics/btg412.
159. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. Genetics. 2000;155(2):945-59. doi: 10.1093/genetics/155.2.945.
160. Porras-Hurtado L, Ruiz, Y. Santos, C., Phillips, C., Carracedo, A., Lareu, M. An overview of STRUCTURE: Applications, parameter settings, and supporting software. Fron Genet. 2013;4(98):1-13. doi: 10.3389/fgene.2013.00098.
161. Kidd KK, Soundararajan U, Rajeevan H, Pakstis AJ, Moore KN, Ropero-Miller JD. The redesigned Forensic Research/Reference on Genetics-knowledge base, FROG-kb. Forensic Sci Int Genet. 2018;33:33-7. doi: 10.1016/j.fsigen.2017.11.009.
162. Rajeevan H, Soundararajan U, Pakstis AJ, Kidd KK. Introducing the Forensic Research/Reference on Genetics knowledge base, FROG-kb. Investig Genet. 2012;3(1):18. doi: 10.1186/2041-2223-3-18.

163. Ambers A. Modern DNA Technology for Historical Cold Cases: Forensic Genetic Investigation of Two Adult Male Skeletons Recovered from the 17th Century La Belle Shipwreck Using Massively Parallel Sequencing (MPS) 2020. Available from: https://promega.foleon.com/theishireport/november-2020/modern-dna-technology-for-historical-cold-cases-forensic-genetic-investigation-of-two-adult-male-skeletons-recovered-from-the-17th-century-la-belle-shipwreck-using-massively-parallel-sequencing-mps/.
164. Ambers A, Bus MM, King JL, Jones B, Durst J, Bruseth JE, et al. Forensic genetic investigation of human skeletal remains recovered from the La Belle shipwreck. Forensic Sci Int. 2020;306:110050. doi: 10.1016/j.forsciint.2019.110050.
165. Ambers A. The ISHI Report [Internet]. New Haven: Promega. 2019. Available from: https://promega.foleon.com/theishireport/april-2019-final/biogeographic-ancestry-in-absence-of-dna-database/.
166. Haddrill PR. Developments in forensic DNA analysis. Emerg Top Life Sci. 2021;5(3):381-93. doi: 10.1042/etls20200304.
167. Chaitanya L, Breslin K, Zuñiga S, Wirken L, Pośpiech E, Kukla-Bartoszek M, et al. The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. Forensic Sci Int Genet. 2018;35:123-35. doi: 10.1016/j.fsigen.2018.04.004.
168. Marano LA, Fridman C. DNA phenotyping: current application in forensic science. Res Rep Forensic Med Sci. 2019;9:1-8. doi: 10.2147/rrfms.s164090.
169. Atwood L, Raymond J, Sears A, Bell M, Daniel R. From identification to intelligence: An assessment of the suitability of forensic DNA phenotyping service providers for use in Australian law enforcement casework. Fron Genet. 2021:1596. doi: 10.3389/fgene.2020.568701.
170. Greytak EM, Armentrout S, cartographers. DNA Phenotyping: Predicting Ancestry and Physical Appearance from Forensic DNA2015.
171. Kayser M. Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet. 2015;18:33-48. doi: 10.1016/j.fsigen.2015.02.003.
172. MacPhetridge A. DNA as an Eyewitness: How Phenotyping is Being Used to Predict a Suspect's Appearance 2020. Available from: https://promega.foleon.com/theishireport/august-2020/dna-as-an-eyewitness-how-phenotyping-is-being-used-to-predict-a-suspects-appearance/.

173. Correia Dias H, Corte-Real F, Cunha E, Manco L. DNA methylation age estimation from human bone and teeth. Aust J Forensic Sci. 2022;54(2):163-76. doi: 10.1080/00450618.2020.1805011.
174. Maulani C, Auerkari EI. Age estimation using DNA methylation technique in forensics: A systematic review. Egyptian Journal of Forensic Sciences. 2020;10(1):38. doi: 10.1186/s41935-020-00214-2.
175. Verogen. ForenSeq Imagen Kit. 2022.
176. Wakelin J, Mendes J. Cold cases solved: The rollout of Australia's first forensic genealogy capabilities: ABC News; 2023. [Last accessed: 12 November]. Available from: https://www.abc.net.au/news/2023-10-08/how-genetic-genealogy-is-solving-australias-coldest-cases/102870058.

2. CHAPTER TWO

METHODS

2.1 Introduction

This chapter details the methods that are in common for all results chapters of this thesis (ethics approval, sample collection, sample preparation, DNA extraction and DNA quantitation). All individual results chapters have method sections specific to the studies involved.

2.2 Ethics Approval

Ethics approval for this research was granted by the University of Technology (UTS) Human Research Ethics Committee (HREC) with the approval number UTS HREC NO. ETH21-5821 (Appendix 1). This was later updated to UTS HREC NO. ETH21-6606 (Appendix 2). Samples collected from the decomposed and skeletonised human remains at the Australian Facility for Taphonomic Experimental Research (AFTER) were used in accordance with an existing ethics approval (UTS HREC NO. ETH21-2999). Approval to include additional research and casework samples submitted to the AFP DNA Program was given by the Program Lead of the AFP DNA Program and affiliated Australian law enforcement agencies.

2.3 Sample Collection

2.3.1 Reference-Type Samples

Volunteers were recruited through emails distributed to the Centre for Forensic Science at UTS and the Forensics Command at the AFP. The recruitment email included a participation information sheet explaining what their genetic information would be used for, the risks involved, and how to withdraw from the study (see Appendix 3). All volunteers signed a consent form and indicated their approvals for certain genetic tests (see Appendix 4). A questionnaire was completed to provide self-declared BGA for each participant as well as their parents and grandparents, EVCs, and other genetic relatives involved in the study (see Appendix 5).

Buccal swabs (n = 73) were collected in person at UTS or the AFP Forensics Facility. Due to COVID-19 travel restrictions, sample collection packages were distributed to volunteers who were unable to travel. This included gloves, sterile buccal swabs and collection instructions. Each sample was de-identified and received a unique sample code.

AFP Biology research samples (n = 57) that were previously extracted, quantified and sequenced by staff members were provided for use in the identity-informative markers study (Chapter Five).

2.3.2 Casework-Type Samples

A range of PM samples of varying conditions and ages were collected from AFTER, previously analysed AFP research samples and approved casework samples submitted to the AFP DNA Program (Table 2.1). Additional AM samples were collected from volunteers (Table 2.1).

Table 2.1. Casework-type human remains samples from volunteers, the Australian Facility for Taphonomic Experimental Research (AFTER) and Australian Federal Police National DNA Program for Unidentified and Missing Persons (AFP DNA Program) research and casework samples.

Sample ID	Source	Condition
Hair 1	Volunteer (Family 1)	Baby rootless hair cutting, stored for 38 years at room temperature (25ºC)
Blood 1	Volunteer (Family 1)	Blood collected on FTA card, stored for 18 months at room temperature (25ºC)
Blood 2	AFP DNA Program research sample	Whole blood, stored for 6 months at -20ºC
Nail 1	AFP DNA Program research sample	Nail clipping, stored for 2 years at room temperature (25ºC)
Nail 2	AFTER	Nail clipping, PMI 2 months
Tooth 1	Volunteer (Family 2)	Primary tooth, stored for 25 years at room temperature (25ºC)
Tooth 2	AFP DNA Program research sample	Permanent tooth, stored for 25 years at room temperature (25ºC)
Bone 1	AFP DNA Program casework sample	Femur sample, PMI unknown, saltwater exposure
Bone 2	AFP DNA Program casework sample	Femur sample, PMI approximately 230 years, buried
Bone 3	AFP DNA Program casework sample	Mandible sample, PMI unknown, saltwater exposure
Bone 4	AFP DNA Program casework sample	Petrous sample, PMI unknown, surface decomposition
Bone 5	AFP DNA Program casework sample	Petrous sample, PMI approximately 70 years, buried
Bone 6	AFTER	Petrous sample, PMI 2 years, surface decomposition
Bone 7	AFTER	Petrous sample, PMI 2 years, surface decomposition
Bone 8	AFTER	Petrous sample, PMI 2 years, surface decomposition
Bone 9	AFTER	Petrous sample, PMI 2 years, shallow burial
Bone 10	AFTER	Petrous sample, PMI 2 years, shallow burial
Bone 11	AFTER	Petrous sample, PMI 2 years, shallow burial

2.4 Sample Preparation

2.4.1 Hair and Nail Samples

Nail clippings were cut with sterile scissors and a 0.5 x 0.5 cm section was collected in a sample tube. A cutting of hair shaft approximately 2 cm in length was cut into two pieces and collected in a sample tube. Hair and nail samples underwent cleaning by agitating the sample tube with a vortex for 5 minutes for the following washes: 5% Terg-a-zyme® enzyme detergent, 10% bleach, Milli-Q water (three washes) and 70% ethanol. The samples were transferred to a fresh tube between washes using tweezers. The samples were exposed to ultraviolet (UV) light for 15 minutes on each side and were then transferred to a microcentrifuge tube.

2.4.2 Bone and Tooth Samples

Bone samples were collected by cutting a section of the skeletal element weighing 0.5 to 1.0 g using a Dremel® 4000 Rotary Tool (Dremel®) and Dremel® Diamond Point Cutting Wheel (Dremel®). The surface was sanded using the rotatory tool and Dremel® Sanding Bands (Dremel®) to remove any contaminants in the outer cortex. For the permanent tooth sample (Tooth 2), the crown was removed by cutting a groove at the cementoenamel junction with the rotary tool and diamond wheel, and a chisel and hammer were used to separate the crown and root. The primary tooth (Tooth 1) was sampled whole.

Bone and teeth samples underwent cleaning by agitating the sample tube with a vortex for 5 minutes for the following washes: 10% bleach, Milli-Q water (three washes) and 70% ethanol. The solution was decanted into a waste container between washes. The samples were exposed to ultraviolet (UV) light for 15 minutes on each side. Samples were then crushed using a pre-chilled BioPulverizer 59014N (Biospec Products). This involved firstly chilling the BioPulverizer in liquid nitrogen for 10 minutes before the sample was added, then chilling the sample inside the mortar for an additional 1 minute and then using the mallet to strike the pestle, rotating the pestle between blows.

2.5 DNA Extraction

2.5.1 Buccal Swabs

Buccal swabs either underwent manual DNA extraction or automated DNA extraction. For the manual extraction, the QIAamp DNA Investigator® Kit (QIAGEN) was used following the manufacturer's protocols for Omni Swabs and DNA was eluted with 100 µL

of Buffer ATE.[1] The automated extraction followed the manufacturer's recommended protocols with the EZ1® DNA Investigator Kit (QIAGEN) on the EZ1® Advanced XL (QIAGEN) with DNA eluted with 100 µL of TE buffer.[2,3]

2.5.2 Blood Samples

The blood samples underwent DNA extraction by using the EZ1® DNA Investigator Kit following the manufacturer's protocol for casework and reference samples [2]. Automated purification was performed on the EZ1® Advanced XL (QIAGEN) using the trace protocol.[3] The DNA was eluted with 100 µL of TE buffer.

2.5.3 Hair and Nail Samples

DNA from the hair and nail samples was extracted using the QIAamp DNA Investigator® Kit (QIAGEN) following the manufacturer's protocols.[1] The DNA was eluted with 50 µL of Buffer ATE.

2.5.4 Bone and Tooth Samples

Skeletal DNA was extracted using a modified demineralisation method validated by the AFP DNA Program using the MinElute® PCR Purification Kit (QIAGEN).[4] Samples were incubated overnight at 56°C and 750 rpm in a thermomixer with proteinase K and demineralisation buffer (0.5M EDTA, 1% n-lauroylsarcosine).[5] The lysate was decanted into an Amicon® Ultra-15 concentrator (Millipore®) and underwent purification according to the MinElute® PCR Purification Kit handbook.[4,6] The DNA was eluted with 50 µL of EB buffer.

2.6 DNA Quantification

Quantitation for the 214 bp large autosomal (LA), 80 bp small autosomal (SA) and 75 bp Y chromosome target fragments was conducted using the Quantifiler™ Trio DNA Quantitation Kit (Applied Biosystems™) on a QuantStudio™ 5-Real-Time PCR System (Applied Biosystems™) according to manufacturer instructions.[7,8] The results were analysed on the HID Real-Time PCR Analysis Software (Applied Biosystems™) to construct the standard curve and determine quantity of DNA (ng/µL) for each sample and target. Samples were deemed to be inhibited if the internal positive control (IPC) for that sample had a cycle threshold greater than 31. The degradation index (DI) was calculated using the following equation and sample quality inferred according to Table 2.2.

$$\mathrm{DI} = \frac{\text{Concentration of SA target (ng/µL)}}{\text{Concentration of LA target (ng/µL)}}$$

Table 2.2. Degradation index (DI) scale and sample quality.

DI Range	Sample Quality
< 1	Not degraded
1 – 5	Slightly degraded
5 – 10	Moderately degraded
> 10	Significantly degraded

The DNA quantities and DI for all samples utilised in the following studies (n = 161) are listed in Appendix 6 (Supplementary Table 1) and which studies they were used for.

2.7 References

1. QIAGEN. QIAamp® DNA Investigator Handbook. 2020. [Last accessed: 14/12/2024]. Available from: https://www.qiagen.com/us/products/human-id-and-forensics/investigator-solutions/qiaamp-dna-investigator-kit
2. QIAGEN. EZ1&2 DNA Investigator Kit. 2022. [Last accessed: 14/12/2024]. Available from: https://www.qiagen.com/us/products/human-id-and-forensics/investigator-solutions/ez1-2-dna-investigator-kit?catno=952034
3. QIAGEN, EZ1 Advanced XL. 2024. [Last accessed: 14/12/2024]. Available from: https://www.qiagen.com/us/products/diagnostics-and-clinical-research/solutions-for-laboratory-developed-tests/ez1-advanced-xl-instrument
4. QIAGEN. MinElute® Handbook. 2020. [Last accessed: 14/12/2024]. Available from: https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/dna-clean-up/minelute-pcr-purification-kit
5. Loreille OM, Diegoli TM, Irwin JA, Coble MD, Parsons TJ. High efficiency DNA extraction from bone by total demineralization. Forensic Sci Int Genet. 2007;1(2):191-5. doi: 10.1016/j.fsigen.2007.02.006.
6. Sigma-Aldrich. Amicon® Ultra-15 Centrifugal Filters 30 kDa MWCO Millipore 2023. [Last accessed: 02/08/2024]. Available from: https://www.sigmaaldrich.com/AU/en/product/mm/ufc9030
7. Applied Biosystems. Quantifiler™ HP and Trio DNA Quantification Kits User Guide [User Guide]. Thermo Fisher Scientific. 2017. [Last accessed: 14/12/2024]. Available from: https://www.thermofisher.com/order/catalog/product/4482910

8. Applied Biosystems. QuantStudio™ 5 Real-Time PCR Instrument (for Human Identification) [User Guide]. Life Technologies. 2017. [Last accessed: 14/12/2024]. Available from: https://www.thermofisher.com/order/catalog/product/A28139

3. CHAPTER THREE

EVALUATION OF THE FORENSEQ® KINTELLIGENCE KIT AND THE FORENSIC CAPTURE ENRICHMENT PANEL FOR UNIDENTIFIED AND MISSING PERSONS CASEWORK

Contributions of authors:

Watson JL, Grisedale K, McNevin D and Ward J (2025) 'Evaluation of the ForenSeq® Kintelligence Kit and the FORensic Capture Enrichment Panel for Unidentified and Missing Persons Casework', International Journal of Legal Medicine, DOI: 10.1007/s00414-025-03492-4.

Jessica Watson (Candidate)

Conceptualised and assisted in the experimental design for the study. Carried out sample collection, sample preparation, DNA testing and data analysis. Prepared manuscript, edited manuscript following review by other co-authors, referenced manuscript, carried out manuscript's submission and incorporated feedback from peer review process.

Signed: Production Note: Signature removed prior to publication.

Kelly Grisedale (AFP DNA Program staff)

Assisted in the experimental design for the study. Contributed to DNA testing and data analysis. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Dennis McNevin (Co-supervisor)

Conceptualised and assisted in the experimental design for the study. Contributed to data analysis. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Jodie Ward (Co-supervisor)

Conceptualised and assisted in the experimental design for the study. Acquired funding for the study. Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

International Journal of Legal Medicine
https://doi.org/10.1007/s00414-025-03492-4

ORIGINAL ARTICLE

Evaluation of the ForenSeq® Kintelligence Kit and the FORensic Capture Enrichment Panel for Unidentified and Missing Persons Casework

Jessica L. Watson[1,2,3] · Kelly Grisedale[1,2] · Dennis McNevin[2,3] · Jodie Ward[2,3]

Received: 16 January 2025 / Accepted: 1 April 2025
© Crown 2025

Abstract
Targeted amplicon sequencing (TAS) employs massively parallel sequencing technology to generate profiles comprised of several thousand single nucleotide polymorphisms (SNPs) to assist in identifying an individual and generating investigative leads. By targeting a range of SNP classes, profiles are able to be analysed to infer biological sex, externally visible characteristics, biogeographical ancestry, paternal lineage and relationships to other individuals. Such leads can be beneficial for human remains identification where antemortem data is not available for comparison. This study evaluated the performance and requirements of two in-house TAS pipelines: the ForenSeq® Kintelligence Kit and the FORensic Capture Enrichment (FORCE) panel. Both TAS pipelines demonstrated suitability for a range of samples typically encountered in missing persons cases, including buccal, bone, tooth and nail samples. There was a high degree of concordance between the TAS genotypes and the majority of the genetic intelligence produced was consistent with the self-declared information provided by DNA donors. This study highlights the requirements for each pipeline to be considered by forensic laboratories seeking to establish a forensic genomics capability for unidentified and missing persons casework.

Keywords Targeted amplicon sequencing · Single nucleotide polymorphism · Unidentified human remains · Missing persons · Whole genome sequencing · Kinship

Introduction

DNA analysis is one of the primary identification methods employed in unidentified human remains (UHR) investigations due to its ability to differentiate individuals [1–3]. Routine short tandem repeat (STR) profiles are uploaded to a law enforcement DNA database to be compared against known profiles in an attempt to obtain a direct or kinship match [4]. UHR DNA analysis can be complicated by the postmortem interval of the remains, degradation of the DNA and the unavailability of suitable direct or familial reference samples. In addition, STR profiles are only suitable for 1 st or 2nd degree kinship inferences. When routine STR testing and law enforcement database searches do not result in an identification, forensic genomics techniques can be employed to provide new investigative leads. [3]

Single nucleotide polymorphisms (SNPs) are single points of variation in the genome and can be categorised into SNP classes based on the information that can be yielded, including: identity-informative SNPs (iiSNPs) for individualisation; phenotype-informative SNPs (piSNPs) for estimating externally visible characteristics (EVCs); ancestry-informative SNPs (aiSNPs) for estimating biogeographical ancestry (BGA); kinship-informative SNPs (kiSNPs) for detecting close and distant genetic relatives; and sex-chromosome SNPs from the Y chromosome (Y SNPs) and X chromosome (X SNPs) for inferring biological sex and paternal lineage. [5–7]

DNA sequencing techniques have been continuously developed since their first application for forensic identification purposes in order to produce more sensitive, accurate,

✉ Jessica L. Watson
Jessica.Watson@student.uts.edu.au

[1] Forensics Command, Australian Federal Police, Canberra, Australia

[2] National DNA Program for Unidentified and Missing Persons, Australian Federal Police, Canberra, Australia

[3] Centre for Forensic Science, School of Mathematical & Physical Science, Faculty of Science, University of Technology Sydney, Sydney, Australia

Published online: 07 April 2025

discriminatory and informative DNA profiles [8–10]. Targeted amplicon sequencing (TAS) employs massively parallel sequencing (MPS) technology, widely used for SNP typing due to its multiplexing power and ability to sequence millions of reads from multiple samples simultaneously [11]. Due to the use of benchtop MPS instruments, TAS is the most easily integrated genotyping technology for an ISO/IEC 17025 accredited forensic laboratory. Despite having lower density SNP coverage than other genomics technologies such as microarray and whole genome sequencing (WGS), forensic TAS panels can target several thousands of SNPs that have been specifically curated for forensic applications. They produce medium density SNP genotypes, as opposed to the high density genotypes produced by microarrays and WGS. Consequently, some privacy risks can be mitigated if medically informative SNPs, sequenced using other approaches, are deliberately excluded. [3, 12]

The ForenSeq® Kintelligence Kit (QIAGEN, Hilden, Germany) and the FORensic Capture Enrichment (FORCE) panel are two TAS kits available for forensic genomics [13–15]. Verogen, Inc. (now a QIAGEN company) released the Kintelligence Kit in 2021 for use on the MiSeq FGx® Sequencing System [16]. This panel was designed for processing a variety of forensic casework samples and targets over ten thousand SNPs. The FORCE panel was originally developed by Tillmar et al. (2021) as a hybridisation capture assay for highly degraded contemporary and historical UHR and targets over five thousand SNPs [15]. The FORCE workflow has been subsequently redesigned to be agnostic for multiple library preparation and sequencing chemistries, including Illumina (QIAseq) and Ion Torrent [17, 18] Table 1 compares the number of SNPs by class for the Kintelligence Kit and FORCE panel.

The SNP genotypes produced by the TAS pipelines can be used to assist in identifying an individual and inferring biological sex, EVCs, BGA, paternal lineage and extended kinship. Extended kinship can be conducted in two ways:

1. by examining the number and length of DNA segments shared by two individuals where the segments consist of multi-SNP haplotypes that are identical by descent (IBD), or
2. by calculating a likelihood ratio (LR), which compares likelihoods for two alternative kinship scenario propositions for two individuals based on SNPs that are identical by state (IBS) and their allele frequencies within the population in question. [19–23]

Since 2018, extended kinship analysis has been applied to forensic investigative genetic genealogy (FIGG) to detect close and distant genetic relatives on law enforcement accessible public genetic genealogy databases [24, 25]. A SNP profile of an unknown person can be compared to profiles uploaded by consenting members of the public. FIGG can be a powerful intelligence tool where a person has not been reported as missing, or in cases where antemortem data or family reference samples are not available.

A laboratory should endeavour to validate and accredit a forensic genomics workflow that is best suited for their typical sample types, laboratory capacity, available expertise and intended genetic intelligence applications to advance unresolved casework. This study evaluated in-house TAS pipelines to inform an optimal forensic genomics strategy for the Australian Federal Police (AFP) National DNA Program for Unidentified and Missing Persons. Several reference- and casework-type samples were sequenced using the Kintelligence Kit and the FORCE panel with the QIAseq workflow. These findings may assist other forensic laboratories seeking to establish a SNP typing capability.

Table 1 Number of SNPs by class in the ForenSeq® Kintelligence Kit and the FORCE panel with the QIAseq workflow as well as overlapping SNPs between the panels. [14, 17]

SNP Class	ForenSeq® Kintelligence Kit	FORCE Panel (QIAseq Workflow)	Overlapping SNPs
X SNPs	106	246	2
Y SNPs	85	883	11
piSNPs	24[a]	41[b]	22[d]
aiSNPs	54[a]	254[b]	54[d]
iiSNPs	94	137[c]	93
kiSNPs	9,867	3,936[c]	810
Total	**10,230**	**5,497**	**992**

[a] Two SNPs overlap the piSNP and aiSNP classes in the Kintelligence Kit and are only counted in the piSNPs

[b] Three SNPs overlap the piSNP and aiSNP classes in the FORCE panel and are only counted in the piSNPs

[c] Two SNPs overlap in the iiSNP and kiSNP classes in the FORCE panel and are only counted in the kiSNPs

[d] The two SNPs that overlap the piSNP and aiSNP classes are only counted in the piSNPs

Methods

Ethics Approval and Sample Preparation

Ethics approval for this forensic genomics research and collection of samples from the Australian Facility for Taphonomic Experimental Research (AFTER) was granted by the University of Technology Sydney (UTS) Human Research Ethics Committee (HREC); UTS HREC NO. ETH21 - 5821 and UTS HREC NO. ETH18 - 2999, respectively.

Reference-type samples included buccal swabs collected from genetically related volunteers spanning 1 st to 5 th degree relationships (n = 5, Fig. 1). Casework-type samples

included bone (n = 2), tooth (n = 2) and nail (n = 1) samples sourced from AFTER, volunteers and approved research and casework samples submitted to the AFP National DNA Program for Unidentified and Missing Persons. The positive control (PC) DNA sample was NA24385 lymphoblastoid cell line (provided with the Kintelligence Kit) [16] and the negative control (NC) was nuclease-free water.

DNA was extracted from buccal swabs and nail samples using the QIAamp DNA Investigator® Kit (QIAGEN, Hilden, Germany) [26]. For bone and tooth samples, 500 mg of pulverised bone or tooth powder underwent total demineralisation lysis, concentration using an Amicon® 30 K Ultra Centrifugal Filter (Sigma-Aldrich, St. Louis, MO, US) and purification with the MinElute® PCR Purification Kit (QIAGEN) [27–29]. All samples were quantified using the Quantifiler™ Trio DNA Quantification Kit (Thermo Fisher Scientific, Waltham, MA, US) on a QuantStudio™ 5 Real-Time PCR System (Thermo Fisher Scientific) according to the manufacturer's protocol [30, 31]. The PC (NA24385) was quantified with the QuantiFluor® ONE dsDNA System (Promega, Madison, WI, US) on the Quantus™ Fluorometer (Promega). [32, 33]

Library Preparation, Sequencing and Bioinformatics

For all pipelines, the DNA input amount was calculated from the large autosomal (LA) target concentration, to avoid over-diluting degraded samples. Extracted DNA was then diluted accordingly with nuclease-free water. Table 2 outlines the quantification results, degradation index (DI) and the DNA input calculated for the Kintelligence Kit (maximum of 1 ng in 25 µL) and the FORCE panel (maximum of 10 ng in 18.43 µL) workflows. The pipelines for library preparation are outlined in Fig. 2. The Veriti™ 96-Well Fast Thermal Cycler (Thermo Fisher Scientific) was used for the Kintelligence and FORCE pipelines.

ForenSeq® Kintelligence Kit

The libraries were prepared manually in batches of 10 samples together with a PC (NA24385) and NC [36]. The amplified products were purified and barcoded using unique dual indices (UDIs), before being individually normalised to 0.75 ng/µL using the QuantiFluor® ONE dsDNA System (Promega, Madison, WI, US) on the Quantus™ Fluorometer (Promega) [32, 33]. The normalised libraries were then pooled in batches of three samples. The PC and NC were included in the first pool only. Sequencing of each pool was performed on the MiSeq FGx® Sequencing System with the standard flow cell (SFC) [37]. Different UDI combinations were used on subsequent runs to limit the effect of sample carryover.

The sequencing run metrics were assessed using the Universal Analysis Software (UAS) v2.5 (Verogen, Inc.) and the Sequencing Analysis Viewer (SAV; Illumina, San Diego, CA, US) for cluster density, clusters passing filter, reads passing the quality score of 30 (Q30) threshold and approximation of adapter dimers present on the SFC [38, 39]. The genotypes were exported from the UAS and analysed according to the optimised thresholds and Microsoft Excel macro workbook previously published by Watson et al. (2023) to generate the final genotype [36]. This

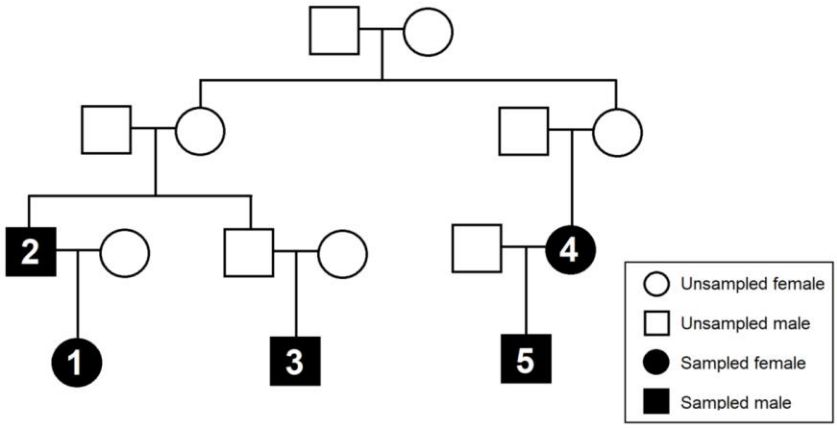


Fig. 1 Pedigree chart of the genetically related volunteers; individuals sampled are filled in (black) and marked 1 through 5, representing samples Family 1 through Family 5

Table 2 Sample DNA concentrations, degradation indices and calculated input amount for the ForenSeq® Kintelligence Kit and FORensic Capture Enrichment (FORCE) panel

Sample ID	Sample Type	Large Autosomal (LA) Target (ng/µL)	Degradation Index (DI)	DNA Input Amount (ng)	
				ForenSeq® Kintelligence Kit	FORCE Panel
NA24385	Control DNA	10.000 [a]	N/A	1.00	10.00
Family 1	Buccal	0.006	3.33	0.15	0.11
Family 2	Buccal	0.037	1.15	0.94	0.69
Family 3	Buccal	0.203	1.46	1.00	3.74
Family 4	Buccal	0.095	2.53	1.00	1.76
Family 5	Buccal	0.058	1.14	1.00	1.06
Tooth 1	Tooth	0.007	5.49	0.17	0.12
Tooth 2	Tooth	0.379	1.08	1.00	6.99
Bone 1	Bone	0.322	1.01	1.00	5.94
Bone 2	Bone	0.009	82.94	0.74	0.16
Nail 1	Nail	0.747	1.26	1.00	10.00

[a] Concentration of double stranded DNA in solution

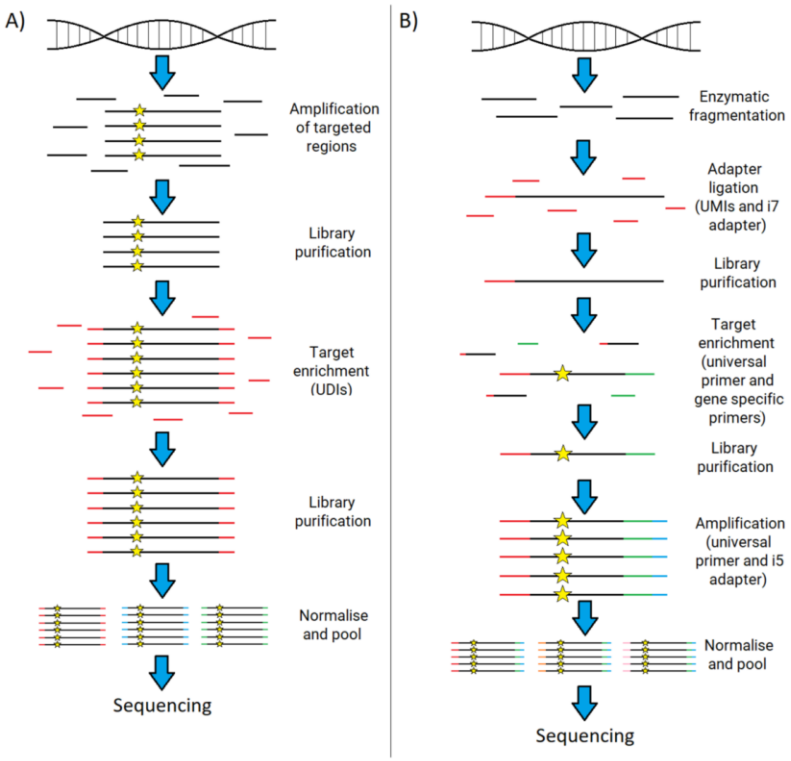


Fig. 2 Library preparation workflows for the A) ForenSeq® Kintelligence Kit and B) FORensic Capture Enrichment (FORCE) panel with the QIAseq workflow. The star represents the SNP being targeted. [16, 17, 34, 35]

included a total read threshold of 20 reads and relative allele frequency thresholds for homozygotes (0.95–1.00), heterozygotes (0.10–0.90), sequencing error (< 0.10) and ambiguous variants (with relative frequencies in the ranges 0.05–0.10 and 0.90–0.95).

FORensic Capture Enrichment Panel

The FORCE panel library preparation followed the single primer extension workflow specified in the May 2017 version of the QIAseq Targeted DNA Panel handbook unless otherwise stated [34]. All reagent volumes were based on the standard DNA option and increased by 10% for all parts of the workflow to allow for transferring the libraries between 96-well semi-skirted and midi plates for the purifications. The libraries were prepared manually in a batch of 24 samples together with a PC (NA24385) and NC.

Fragmented DNA was tagged with unique molecular indices (UMIs) and i7 adapters from the QIAseq 12-index I set of combinatorial dual index (CDI) adapters. The targets were purified, enriched using the universal primer and gene specific primers and amplified after being tagged with the universal primer and i5 adapter. Libraries were normalised to 1.00 ng/µL using the QuantiFluor® ONE dsDNA System on the Quantus™ Fluorometer. The normalised libraries were then pooled in batches of six samples. Sequencing of each pool was performed on the MiSeq FGx® Sequencing System with the SFC in "Research Use Only" mode. [37]

The sequencing run metrics were assessed using the SAV for cluster density, clusters passing filter, reads passing the Q30 threshold and approximation of adapter dimers present on the SFC. FASTQ files were exported from the MiSeq FGx® Sequencing System, imported into the CLC Genomics Workbench v22.0.2 (QIAGEN) and analysed using a custom workflow published by Tillmar et al. (2021). [15, 40] The FORCE genotypes were exported in CSV files and additional relative allele frequency thresholds were applied for homozygotes (0.90–1.00), heterozygotes (0.20–0.80), sequencing error (< 0.10) and ambiguous variants (with relative frequencies in the ranges 0.10–0.20 and 0.80–0.90).

Genotype Analysis

For the SNPs that were shared between the Kintelligence Kit and FORCE panel (Table 1), the genotypes were analysed for coverage, call rates, autosomal heterozygosity and genotype concordance. Differences in the genotypes that could not be explained by sequencing on the opposite strand were classified as non-concordant SNPs.

Phenotype and Ancestry Estimation

For the Kintelligence genotypes, the reported phenotypes and ancestries were exported from the UAS. This included the probabilities for hair colours (blond, brown, red and black) and eye colours (brown, blue and intermediate) using an embedded multinomial logistic regression (MLR) algorithm. An in-built principal component analysis (PCA) is used to infer BGA from four population clusters (European, East Asian, African and Admixed American).

Text files of the FORCE genotypes were imported to FamLink2 for analysis with the Phenotype/Ancestry/Haplogroup tool [21]. The FORCE panel genetic map and allele frequencies for several populations (African, American, East Asian, European, Middle Eastern, Oceanian and South Asian) were sourced from the FamLink2 database and imported into FamLink2 for BGA inference [41]. A CSV file was exported from FamLink2 for upload to HIrisPlex-S to infer hair colour (blond, brown, red or black), hair shade (light or dark), eye colour (brown, blue or intermediate) and skin colour (very pale, pale, intermediate, dark or dark-to-black). [42–44]

Kinship Analysis

To calculate the LR for kinship analysis, the conditional probabilities of observing the genotypes given two alternative propositions were compared for each pairwise combination of the five genetically related individuals. The first proposition corresponded with the true relationship between the individuals and the second proposed that they were unrelated. For example:

H_1: The donor of profile 1 and the donor of profile 2 are full siblings.

H_2: The donor of profile 1 and the donor of profile 2 are unrelated members of the European population.

LRs were calculated using DBLR™ v1.3 (STRmix™) [45]. The genetic linkage map and European population allele frequency data were downloaded from the FamLink2 database [41]. The Kintelligence Kit map was constructed using the Map Interpolator of the Rutger's Map v.3 with the sex-average centimorgan (cM) positions [46]. The allele frequencies for the autosomal SNPs targeted in the Kintelligence Kit were sourced from the 1000 Genomes Project for a European population. [47]

Results and Discussion

Sequencing Performance

For the Kintelligence Kit pipeline, the 12 samples were distributed over five sequencing runs in pools of three

libraries. The majority of the sequencing metrics for the runs were within the recommended ranges for sequencing on the MiSeq FGx® Sequencing System; cluster density was slightly higher than the recommended range for one sequencing run (Table 3) [37, 38]. The higher densities of the Kintelligence runs correlated with lower percentages of adapter dimers present on the SFC.

For the FORCE pipeline, the samples were sequenced in batches of six libraries over four sequencing runs. The first sequencing run failed, which included the Family 1 reference-type sample, and was excluded from analysis. The FORCE runs had lower cluster densities and a greater presence of adapter dimers compared to the Kintelligence runs (Table 3). However, the clusters passing filter and reads passing the Q30 were, on average, higher than the Kintelligence runs.

An evaluation of 15 sequencing runs with the FORCE QIAseq workflow by Staadig et al. (2023) revealed a significantly higher average cluster density (1032 K/mm^2 ± 318 K/mm^2), with 89% of reads passing the Q30 threshold [17]. The number of libraries pooled by Staadig et al. (2023) varied from 3 to 8 libraries per pool with a DNA input of 1 to 10 ng, as opposed to 6 libraries with DNA input ranging from 0.11 to 10 ng in this study. The low DNA inputs of the samples tested in this study likely resulted in the low cluster density of the sequencing runs despite the similar sequencing plexity. The large proportion of adapter dimers in the FORCE runs may also result from the reduced amount of template DNA for the samples that were below the recommended DNA input of 10ng, resulting in higher concentrations of adapters and primers, facilitating formation of dimers.

Quality and Quantity of the Genetic Information

Of the ten Kintelligence genotypes, 70% of samples met the recommended DNA input amount of 1.0 ng (Table 2) and at least 92% of loci were called (Fig. 3). There was no discernible impact of low DNA inputs on the average call rate (98.0% ± 2.2%) or autosomal heterozygosity (47.4% ± 0.7%) for the samples. Previous evaluations of the Kintelligence Kit have demonstrated that it is able to produce high quality SNP genotypes from low template amounts, particularly for sample types commonly encountered in UHR casework [13, 36, 48]. The NC had two SNPs called and this is in line with the rate of sequencing error observed in NCs from other studies and unlikely to be the result of contamination. [13, 36, 48]

Only 20% of the samples (NA24385 and Nail 1) met the 10.0 ng DNA input amount recommended for the FORCE panel (Table 2). Only one sample produced a call rate below 94%; Tooth 1 had a low DNA input amount of 0.12 ng and call rate of 77.5% (Fig. 3). Similar to the Kintelligence genotypes, there was no discernible impact of low DNA inputs on the average call rate (94.7% ± 6.4%) or autosomal heterozygosity (33.6% ± 7.2%). The FORCE NC had no SNPs called.

Both panels produced high quality genotypes for Bone 2 despite the severe degradation (DI = 83, Kintelligence call rate = 96.9%, FORCE call rate = 95.0%). This is consistent with the previous observations from validations of both pipelines [13, 36, 48]. In the developmental validation of the Kintelligence Kit, 69% (11) of bone samples – including burned, cremated, embalmed, buried and ancient bones – produced profiles with call rates over 90% [13]. Dilutions of the buried bone samples improved the sequencing performance, indicating that inhibitors were present. The internal validation of the Kintelligence Kit by Peck et al. (2022) reported 64% (7) of bone samples with DNA inputs ranging from 0.1 to 1.0 ng produced call rates of at least 96.9% [48]. Furthermore, call rates exceeding 90% were obtained for a range of UHR casework-type samples tested by the AFP National DNA Program for Unidentified and Missing Persons as part of their Kintelligence Kit validation. [36]

Similarly, the FORCE panel was designed for sequencing poor quality samples with severely degraded DNA; initially using hybridisation capture to target DNA fragments less than 75 bp long [15]. When testing 12 bone samples using different DNA extraction methods, an average of 44.4% of SNPs were called during the developmental validation [15]. An evaluation of the FORCE panel's QIAseq workflow using 1.0 ng of input DNA for testing bone and tissue samples revealed an average call rate of 97.6% and 90.7% of SNPs, respectively. [17]

The 992 SNPs in common between the Kintelligence Kit and FORCE panel were assessed for call rate, autosomal heterozygosity and genotype concordance (Table 4). The Kintelligence genotypes produced higher call rates than the FORCE panel for 7 of the 10 samples; however, there were no substantial deviations with the exception of Tooth 1. The Kintelligence Kit was less susceptible to locus dropout of aiSNPs, iiSNPs and kiSNPs, returning higher call rates than the FORCE panel for these SNP

Table 3 Performance metrics for the ForenSeq® Kintelligence Kit (n = 5) and FORensic Capture Enrichment (FORCE) panel (n = 3) sequencing runs

Metric	ForenSeq® Kintelligence Kit	FORCE Panel
Cluster Density (K/mm^2)	1037–1893	199–451
Clusters Passing Filter (%)	75.0–89.9	91.0–93.5
Reads Passing Q30[a] (%)	63.4–85.4	81.7–84.5
Adapter Dimers (%)	11–27	65–70
Libraries per SFC	3	6

[a] Quality score of 30 where there is an error rate of 1 in 1000, with a corresponding call accuracy of 99.9%

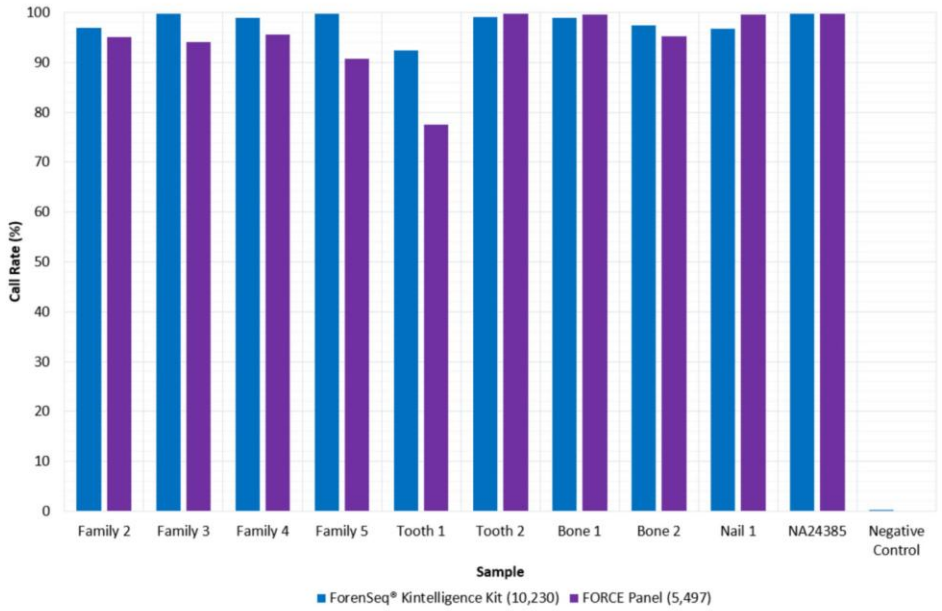


Fig. 3 Sample call rates for sequencing with the ForenSeq® Kintelligence Kit and FORensic Capture Enrichment (FORCE) panel

Table 4 Call rate, autosomal heterozygosity and concordance for the 992 SNPs in common between the ForenSeq® Kintelligence Kit and FORensic Capture Enrichment (FORCE) panel

Sample ID	Biological Sex	Call Rate (%)		Autosomal Heterozygosity (%)		Concordance (%)
		ForenSeq® Kintelligence Kit	FORCE Panel	ForenSeq® Kintelligence Kit	FORCE Panel	
Family 2	Male	97.5	94.4	43.5	41.1	89.4
Family 3	Male	99.9	93.3	47.0	38.7	86.6
Family 4	Female	98.7	93.6	45.0	38.2	86.9
Family 5	Male	99.8	92.8	49.5	34.3	78.6
Tooth 1	Female	91.6	77.7	42.9	19.5	51.5
Tooth 2	Female	98.8	98.3	45.4	45.8	97.5
Bone 1	Male	99.2	99.5	47.7	48.3	97.7
Bone 2	Male	96.9	95.0	43.4	42.6	88.8
Nail 1	Male	96.4	99.6	42.2	44.8	94.3
NA24385	Male	99.6	99.6	44.9	44.6	98.0

classes. The FORCE panel recovered the piSNPs with greater consistency. When assessing the autosomal SNPs (979), the average heterozygosity was 45.2% ± 2.2% for Kintelligence and 39.8% ± 7.8% for FORCE genotypes. It is likely there was substantial allele dropout in the FORCE genotypes for Tooth 1, as only 19.5% of autosomal SNPs were heterozygous.

The average concordance between the Kintelligence and FORCE genotypes was 86.9% ± 13.2%, with the greatest concordance observed with the casework-type samples (Fig. 4). Tooth 1 had the lowest concordance rate of 51.5%, whereas the control DNA NA24385 had the highest at 98.0%. When compared to the known Kintelligence genotype for NA24385, both Kintelligence and FORCE

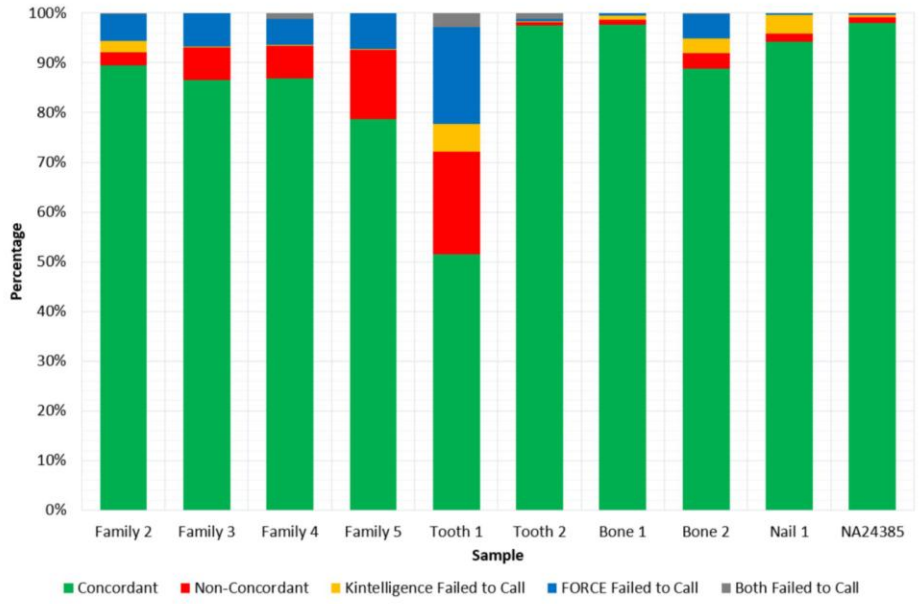


Fig. 4 Distribution of concordance, non-concordance and no-calls for the 992 SNPs in common between the ForenSeq® Kintelligence Kit and FORensic Capture Enrichment (FORCE) panel

had a genotype concordance of 99.6% [38]. Across the ten samples, the average non-concordance rate was 5.8% ± 6.3%, ranging from 0.7% (Tooth 2) to 20.7% (Tooth 1).

Of the 573 non-concordant SNPs observed over the 10 samples, 81% of SNPs were called as heterozygous by the Kintelligence Kit and homozygous for only one of those alleles by the FORCE panel, suggesting a high incidence of allele dropout for the FORCE panel (Fig. 5). Conversely, only 5% of non-concordant SNPs were called as heterozygous in the FORCE genotypes. The remaining non-concordance could not be explained by sequencing on the opposite strand; 14% were called homozygous for different alleles by both pipelines and only two kiSNPs were called as heterozygous by both panels. These two kiSNPs exhibited very low coverage when sequenced with the FORCE panel. They were rs2706586 for Family 4 (Kintelligence typed GT from 2312 reads; FORCE typed CG from 3 reads) and rs4660390 for Family 5 (Kintelligence typed CT from 160 reads; FORCE typed CG from 3 reads). There were no discernible trends in the non-concordance or locus dropout by SNP class except for rs1805005, a piSNP on the MC1R gene. In the eight samples where both pipelines called this SNP, the FORCE genotype was TT for all samples and the Kintelligence genotype was either GG or TT.

When comparing the read data produced for the 992 overlapping SNPs, the libraries prepared with the Kintelligence Kit returned significantly higher total sequencing reads per sample than those prepared with the FORCE panel ($p < 0.01$; Table 5). Furthermore, the median reads per SNP was significantly higher for the Kintelligence genotypes ($p < 0.001$). It is important to note, however, that the FORCE reads are unique reads, each one corresponding with a unique UMI. The higher median reads per SNP for the Kintelligence Kit include PCR duplicates which are combined into unique reads for the FORCE panel.

Phenotype, Ancestry and Kinship Inferences

Table 6 outlines the EVC and BGA inferences generated using the Kintelligence Kit and FORCE panel analytical pipelines against the self-declared data provided by the volunteers. The Kintelligence genotype for Tooth 1 did not yield all 24 piSNPs and, subsequently, the embedded MLR algorithm was unable to generate p-values for hair and eye colour inference.

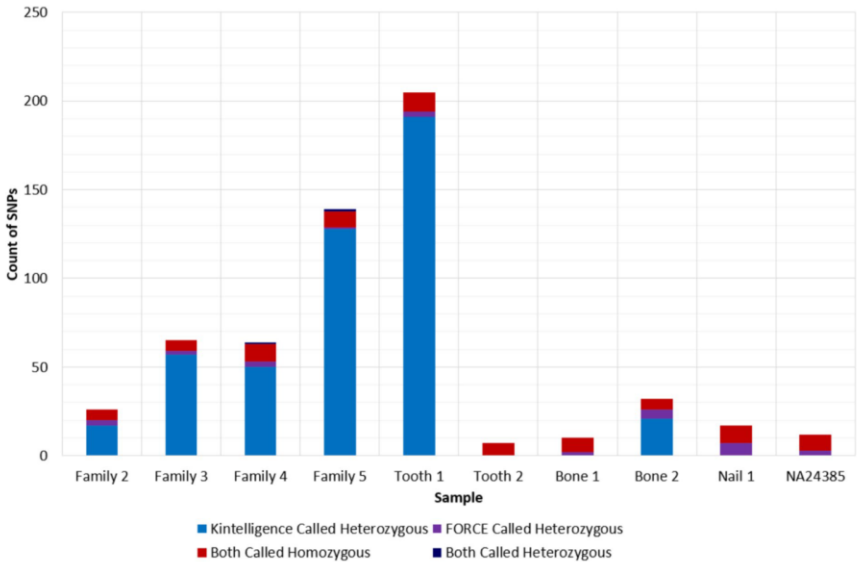


Fig. 5 Types of non-concordant SNPs observed in each sample between the ForenSeq® Kintelligence Kit and FORensic Capture Enrichment (FORCE) panel

Table 5 Total sequencing reads and median reads per SNP for just the 992 overlapping SNPs for each sample sequenced with the ForenSeq® Kintelligence Kit and FORensic Capture Enrichment (FORCE) panel

Sample ID	ForenSeq® Kintelligence Kit		FORCE Panel	
	Total Sequencing Reads	Median Reads per SNP	Total Sequencing Reads	Median Reads per SNP
Family 2	638,043	432	8,623	8
Family 3	1,583,992	1,047	6,916	6
Family 4	1,126,906	783	8,797	8
Family 5	836,849	569	4,282	4
Tooth 1	323,234	214	1,937	2
Tooth 2	1,531,494	1,047	43,823	44
Bone 1	372,427	268	46,174	45
Bone 2	2,369,358	1,425	14,383	13
Nail 1	225,068	151	52,839	51
NA24385	1,691,205	1,163	72,406	71

All BGA inferences for both workflows were consistent with the self-declared ancestry (European in all cases). For the UAS, all Kintelligence aiSNP genotypes were exclusively within the European cluster on the PCA plot, indicating the genotypes were more consistent with European ancestry than East Asian or African. It has been observed in previous studies that the PCA method on the UAS is limited by the small number of available populations [36]. As the volunteers in this study all self-declared European ancestry, this was not problematic for our study. However, this limitation could impact the inference of ancestries for individuals outside of the European, Asian and African populations, as well as for the interpretation of admixture. [36, 49, 50]

FamLink2 uses a naïve Bayesian approach to estimate the likelihood probability for each reference population [21]. All FORCE genotypes were more likely to be of European ancestry than African, American, East Asian, Middle Eastern, Oceanic or South Asian ancestry. It has previously been demonstrated that BGA inferences derived from the FORCE panel using this pipeline were consistent with self-declared ancestry for nine genotypes of European, Hispanic and African American ancestry. [15, 17]

The eye, hair and skin colour inferences presented in Table 6 are based on the highest p-value obtained from the Kintelligence and FORCE analytical pipelines. All eye colour inferences were consistent across both pipelines, indicating these individuals were more likely to have brown or blue eyes. There were two inferences that were inconsistent

Table 6 Phenotype and ancestry inferences based on the highest p-values for the genotypes produced with the ForenSeq® Kintelligence Kit and FORensic Capture Enrichment (FORCE) panel compared to the self-declared phenotype and ancestry

Sample	ForenSeq® Kintelligence Kit	FORCE Panel	Self-Declared
Family 2	Blue eye colour (0.91) Brown hair colour (0.83) European ancestry	Blue eye colour (0.90) Brown hair colour (0.81) Light hair shade (0.82) Pale skin colour (0.75) European ancestry	Grey eye colour Brown hair colour Pale skin colour European ancestry
Family 3	Brown eye colour (0.83) Brown hair colour (0.64) European ancestry	Brown eye colour (0.84) Brown hair colour (0.73) Light hair shade (0.75) Intermediate skin colour (0.79) European ancestry	Brown eye colour Brown hair colour Intermediate skin colour European ancestry
Family 4	Blue eye colour (0.89) Blond hair colour (0.73) European ancestry	Blue eye colour (0.88) Red hair colour (0.49) Light hair shade (0.99) Pale skin colour (0.70) European ancestry	Green eye colour Brown hair colour Intermediate skin colour European ancestry
Family 5	Blue eye colour (0.94) Blond hair colour (0.78) European ancestry	Blue eye colour (0.93) Red hair colour (0.47) Light hair shade (1.00) Pale skin colour (0.75) European ancestry	Blue eye colour Red hair colour Very pale skin colour European ancestry
Tooth 1	European ancestry	Brown eye colour (0.40) Brown hair colour (0.84) Dark hair shade (0.88) Pale skin colour (0.42) European ancestry	Hazel eye colour Brown hair colour Pale skin colour European ancestry
Tooth 2	Blue eye colour (0.92) Blond hair colour (0.88) European ancestry	Blue eye colour (0.91) Blond hair colour (0.71) Light hair colour (1.00) Pale skin colour (0.64) European ancestry	Blue eye colour Blond hair colour Intermediate skin colour European ancestry

with the self-declared data. The first individual, Family 4, self-declared having intermediate (green) eye colour and the p-values for both panels were highest for blue eye colour. The second individual, Tooth 1, self-declared intermediate (hazel) eye colour and only the FORCE panel was able to generate p-values, which were highest for brown eye colour. Previous EVC studies have shown that the HIrisPlex algorithm has lower success rates for inferring intermediate eye colours [36, 51, 52]. For the Kintelligence Kit, all inconsistent eye colour inferences corresponded with individuals who had self-declared intermediate eye colours [36]. In the development of the FORCE panel, the sole volunteer who self-declared an intermediate eye colour was inferred to have blue eyes. [15]

Hair colour inferences were consistent between analytical pipelines and the self-declared data for three of the genotypes (Table 6). The inference generated with the FORCE genotype for Tooth 1 was also consistent with the self-declared data, whereas analysis of the Kintelligence genotype was unable to produce p-values without a full piSNP profile. The panels were inconsistent with each other for inferring the hair colour for Family 4 and Family 5. Family 5 has self-declared red hair and was inferred as red with the FORCE pipeline but inferred as blond with the Kintelligence pipeline. Further interrogation of the Kintelligence SNP data showed that there was alignment ambiguity at N29 insA (also denoted rs312262906) on the MC1R gene from position 16:89,919,340 to 16:89,919,344 leading to a heterozygous SNP (A/C) being called at 16:189,919,342 instead of a heterozygous A insertion at 16:89,919,344. This is one of the SNPs used to infer red hair colour [53]. The piSNP genotype was analysed using the online HIrisPlex tool to determine the impact of this alignment ambiguity [42, 44, 54]. The Kintelligence genotype called NA29 insA as homozygous C and produced a p-value of 0.111 for red hair colour; when the genotype was adjusted to a heterozygous A insertion, the p-value for red hair colour increased to 0.997.

Skin colour inferences were only possible with the FORCE genotypes as the panel included the additional 17 piSNPs in the HIrisPlex-S panel [42–44]. The largest p-values were for pale and intermediate skin colours, and these were largely consistent with the self-declared skin colours. Skin colour can be complex to infer depending on population-specific influences and environmental factors. [42, 43]

LRs were calculated for six pair-wise combinations of the four related volunteers using the full autosomal SNP potential of each panel (Family 1 was excluded). As a result of the Kintelligence Kit targeting nearly twice as many SNPs

than the FORCE panel, it was expected that the LRs would be significantly higher (Fig. 6). The LR for the true parent/offspring relationship (1 st degree) was unable to be calculated using DBLR™ as there was not an allele in common at every locus. Due to the high density of these panels, it is likely that at least one instance of allele dropout will occur in samples such as the buccal swabs collected from volunteers in this study.

The LRs were generated for each of the relationship pairs and for each of the Kintelligence and FORCE genotypes to test the 2nd through 5 th degree relationships. According to the forensic verbal equivalency scale, the LRs for all Kintelligence relationships and FORCE 2nd and 3rd degree relationships provided "very strong support" (LR > 1,000,000) [55]. For the remaining FORCE genotype pairings, 4 th degree relationship LRs provided "moderately strong support" (LR > 1,000) and the 5 th degree relationship LR provided "moderate support" (LR > 100).

Gettings et al. (2024) compared the Kintelligence Kit and FORCE panel for disaster victim identification (DVI) applications and evaluated kinship predictions using data simulations for 1 st to 5 th degree relationships using the iiSNPs and kiSNPs available (9,959 SNPs for the Kintelligence Kit and 4,073 SNPs for the FORCE panel) [56]. They observed the median $\log_{10}$ LR decreased by a factor of three for each increase in relationship degree. The $\log_{10}$ LRs calculated for the samples in this study for all autosomal SNPs (10,039 SNPs for the Kintelligence Kit and 4,368 SNPs for the FORCE panel) were consistent with those published by Gettings et al. for 2nd to 5 th degree relationships [56]. However, LRs were unable to be generated in this study for 1 st degree relationships (parent/offspring) due to allele dropout, whereas Gettings et al. generated median $\log_{10}$ LRs of 1300 for these relationships using simulated samples with no allele dropout. [56]

Considerations for Implementation

There are several advantages to the implementation of medium-density TAS SNP panels. As these pipelines can be operationalised with existing MPS workflows and instrumentation, the entire laboratory and analytical process can be performed in-house, maintaining control over the chain of custody and quality processes [3, 57, 58]. Table 7 summarises the requirements for implementing both TAS pipelines.

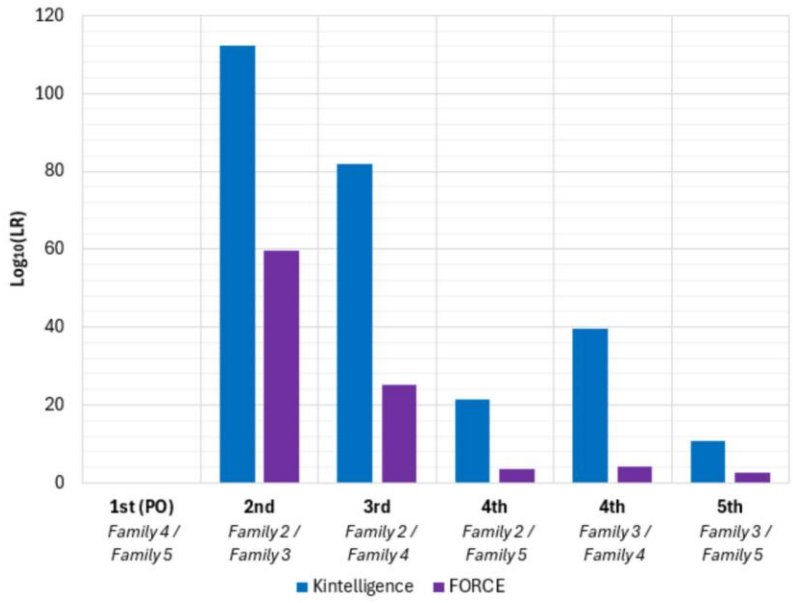


Fig. 6 Logarithms of the likelihood ratios (LR) generated for six pairwise combinations of pedigree members using the ForenSeq.® Kintelligence Kit and FORCE panel. LRs were unable to be calculated for the 1 st degree relationship between Family 4 and Family 5 (parent/offspring (PO))

Table 7 Facility, sample and bioinformatics requirements and generatable genetic intelligence for the ForenSeq® Kintelligence Kit and FORensic Capture Enrichment (FORCE) panel with the QIAseq workflow

Component	ForenSeq® Kintelligence Kit	FORCE Panel
TAS Workflow	PCR amplicon-based	Single primer extension
Library Preparation Cost (USD)[a]	$15,720 [56]	$1,745 (QIAseq Kit) [38] $1,580 (Index Kit) [69]
Sequencing Cost (USD)[b]	$2,085 [56]	$2,085 [56]
Sample Requirements		
Input Volume	25 µL	16.75 µL
Recommended DNA Input	≥ 1 ng	≥ 10 ng
Sample Quality	Can be severely degraded	Can be severely degraded
Recommended Library Preparation Plexity	12	48 (antemortem) 16 (postmortem) [56]
Recommended Sequencing Plexity	3	48 (antemortem) 16 (postmortem) [56]
Genetic Intelligence		
Bioinformatics Required	No	Yes
Individual Identification	Yes	Yes
Y Haplogroup Prediction	Yes	Yes
X Chromosome Analysis	Yes	Yes
Externally Visible Characteristics Inference	Yes; hair and eye colour	Yes; hair, eye and skin colour
Biogeographical Ancestry Inference	Yes	Yes
Extended Kinship Analysis	Yes; 1 st to 5 th degree relatives	Yes; 1 st to 5 th degree relatives
Genetic Genealogy Database Compatibility	GEDmatch PRO™ and FamilyTreeDNA	No

[a] Kit component costs calculated for 12 libraries included in the Kintelligence and FORCE QIAseq Kits

[b] Cost of MiSeq FGx® Reagent Kit for operation on the MiSeq FGx® Sequencing System

Facility Requirements

Both TAS pipelines require similar consumables and storage requirements for the components for library preparation and sequencing. Furthermore, similar laboratory skills are required such as the handling of magnetic beads for library purification. The differences between the TAS workflows are most discernible in the setup of pre-amplification and post-amplification laboratory spaces, as the FORCE panel requires use of a thermocycler for adapter ligation and target enrichment before the DNA is amplified, requiring equipment typically housed in the post-amplification space to be accessible in the pre-amplification space. [34]

For incorporation into a laboratory environment, both TAS workflows are compatible with sequencing on the MiSeq FGx® Sequencing System which has a user-friendly interface and generates exportable data in a FASTQ format. The advantage of using a benchtop sequencer and MPS technology is that more forensic laboratories are expressing interest in pursuing MPS capabilities and acquiring this instrumentation [59, 60]. A 2023 study of US forensic laboratories revealed 163 facilities already have a MiSeq FGx® Sequencing System for forensic genomics applications. [60]

Routine STR typing employed by forensic biology laboratories for human identification is relatively inexpensive compared to emerging technologies. For example, the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific) costs US$4,350 for autosomal STR typing of 200 samples [61]. At the time of publication, the cost of the Kintelligence Kit (US$15,720) was greater than the components required for FORCE (US$3,325) and both kits have enough reagents to prepare 12 libraries [62, 63]. The MiSeq FGx® Reagent Kit (Verogen, Inc.) required for sequencing costs US$2,085, but the cost per library is dependent on the sequencing plexity validated by the laboratory. [64]

The recently released ForenSeq® Kintelligence HT Kit offers a higher throughput workflow, reducing the cost per sample by increasing the sequencing plexity to up to 12 postmortem and 36 antemortem samples per SFC [65]. At the time of writing, the kit contains reagents for 96 samples at a cost of approximately US$88,800 [66]. However, as sequencing plexity increases, the number of SNPs typed deceases, which limits the downstream applications. For example, to perform FIGG, at least 70% of Kintelligence SNPs are required for upload to the law enforcement genetic genealogy databases. [67]

Sample Requirements

Forensic laboratories seeking to operationalise an in-house TAS pipeline will need to consider the typical sample types

and conditions encountered in casework. The amount of DNA recovered from UHR can vary due to differences in DNA preservation rates, degradation from environmental exposure and the postmortem interval [1]. The TAS panels tested in this study have been specifically designed for forensic applications and have demonstrated broad suitability for samples commonly encountered in UHR and missing persons cases. Both pipelines demonstrated suitability for reference- and casework-type samples, including bone, tooth, nail and buccal samples. The success with an array of compromised samples aligns with findings from previous studies. [13, 15, 17, 36, 48]

Furthermore, while manufacturer protocols recommended DNA inputs of 1.0 ng and 10.0 ng for Kintelligence and FORCE library preparations, respectively, previous sensitivity studies have observed high quality genotypes for lower input amounts [13, 17, 36, 48]. In this study, samples with inputs as little as 0.12 ng still produced nearly complete Kintelligence genotypes suitable for further analysis. However, the performance of FORCE was negatively impacted by the lower than recommended DNA inputs in this study, resulting in comparatively high numbers of adapter dimers and low cluster density for the sequencing runs and allele dropout in the FORCE genotypes. This is important, as DNA input might be limited by the volume of DNA extract remaining for a sample, especially since SNP genotyping is typically conducted after STR typing. [68]

Bioinformatics Requirements and Genetic Intelligence Applications

The Kintelligence and FORCE analytical pipelines differ substantially in their methods for producing genotypes and associated genetic intelligence. The UAS circumvents the need for an external bioinformatics system when sequencing with the Kintelligence Kit by integrating multiple bioinformatic processes to generate a final genotype from raw sequencing data, infer hair and eye colour using an MLR algorithm and infer BGA using PCA [13, 38]. This streamlined approach is presented through a user-friendly interface that requires no prior bioinformatics knowledge or experience to operate and any additional analyses performed in this study for genotype generation utilised Microsoft Excel. With the UAS v2.7 update, the piSNP genotype can be exported and uploaded to the HIrisPlex website for hair and eye colour inference [69]. The HIrisPlex tool calculates area under the curve values to estimate the impact of locus dropout on the generated p-values. [42–44]

In contrast, the FORCE workflow requires additional bioinformatics systems for analysing the FASTQ files generated by the MiSeq FGx® Sequencing System in "Research Use Only" mode. Such systems, including the CLC Genomics Workbench used in this study, incur additional costs to the laboratory [40]. Although the FORCE panel targets approximately half the number of SNPs compared to the Kintelligence Kit, it includes a greater number of non-kinship SNPs, enabling expanded genetic intelligence, including skin colour inference, higher resolution BGA inference and Y haplogroup prediction [15]. The FamLink2 software, a freely available application, allows for further analysis of the finalised genotype [21]. This includes the comparison of the questioned genotype to population data using a naive Bayesian approach for BGA inference, the production of probabilities for hair and eye colours and the capability to generate an input file for upload to the HIrisPlex-S website. [42–44]

Both TAS pipelines were able to produce genetic intelligence for hair colour, eye colour and BGA, as well as skin colour for the FORCE panel, that were largely consistent with the self-declared data provided by the volunteers. While both panels generated X and Y SNP genotypes, these were not evaluated in this study. When testing the kinship capabilities of these pipelines, relationships beyond the 5 th degree could not be assessed with the sampled family group. Both panels produced LRs exceeding 10^6 for 2nd and 3rd degree relatives; however, only the Kintelligence Kit demonstrated medium-range kinship analysis capability for 4 th and 5 th degree relationships.

Kintelligence genotypes are compatible for upload to two law enforcement accessible genetic genealogy databases, GEDmatch PRO™ and FamilyTreeDNA, provided that more than 70% of SNPs are typed [70, 71]. These databases contain the SNP profiles of consenting individuals produced by direct-to-consumer genetic testing companies. In GEDmatch PRO™, UHR SNP profiles can be searched against the entire database, while SNP profiles derived from evidence in specific criminal cases and all cases uploaded to FamilyTreeDNA can only be searched against profiles that the consumer has opted in for law enforcement searching. [71, 72]

The GEDmatch PRO™ windowed kinship algorithm utilised for Kintelligence uploads has previously been shown to be an efficient and effective method for detecting and classifying relationships using simulated data and volunteer DNA samples [36, 73, 74]. This method produces a cM value, which estimates the total length of the genome shared between matched profiles, and a whole kinship coefficient to approximate relatedness. Furthermore, GEDmatch PRO™ facilitates database searching and direct comparisons to known profiles. If a known profile is generated from a reference sample using the Kintelligence Kit, it can be directly compared to unknown profiles using an LR generating tool, such as DBLR™ (used in this study) [45]. With the UAS v2.6 update, a kinship database can be constructed in-house using either the Kintelligence or Kintelligence HT kits. [75]

The FORCE panel has also been found to be suitable for extended kinship analysis in previous studies; however, the resulting profiles are not suitable for upload to law enforcement accessible genetic genealogy databases [15, 17, 56]. This workflow requires sequencing and analysing reference profiles for either direct comparisons or the building of an internal database to conduct direct or kinship searches. This pipeline could be particularly beneficial for localised identification efforts such as DVI, where references from family members of missing persons or antemortem samples can be collected and stored in a closed database [56]. Similar to Kintelligence, an LR generating tool can be used to compare profiles for extended kinship analysis.

Conclusions

TAS technologies have been effectively implemented for unidentified and missing persons casework, presenting new investigative avenues in the absence of suitable antemortem or close familial reference samples. The SNP genotypes produced can be utilised for identification, inferring EVCs (i.e. hair, eye and skin colour), estimating BGA and extended kinship analysis. The Kintelligence Kit performed better for low DNA input samples than the FORCE panel with the QIAseq workflow in this study, with FORCE sequencing runs exhibiting a higher proportion of adapter dimers that were likely due to low amounts of template DNA for the majority of samples. There was substantial non-concordance between the Kintelligence and FORCE genotypes for the 992 shared SNPs, likely attributable to allele dropout in the FORCE panel. Additionally, the low DNA input for library preparation resulted in low coverage for FORCE sequencing runs and substantial locus dropout in the FORCE genotypes. Further optimisation of the FORCE panel with the QIAseq workflow may be required to ensure optimal performance for compromised samples where the DNA input is substantially lower than recommended amount.

Before validating an in-house forensic genomics pipeline, forensic laboratories must consider several key factors, including the facility and financial requirements, quality and quantity of DNA from typical forensic samples, availability of bioinformatics capabilities and genomics expertise and access to required databases for extended kinship analysis or the infrastructure to create their own. Additionally, laboratories should evaluate the types of genetic intelligence that can be derived from the genotype and whether it meets their operational needs. These two TAS pipelines offer in-house, end-to-end solutions for forensic genomics, ensuring transparency and accountability throughout the entire process.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data Availability Data are stored at the Australian Federal Police and may be made available to approved entities upon written request and subject to consent provisions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Watherston J, McNevin D, Gahan ME, Bruce D, Ward J (2018) Current and emerging tools for the recovery of genetic information from post mortem samples: New directions for disaster victim identification. Forensic Sci Int Genet 37:270–282. https://doi.org/10.1016/j.fsigen.2018.08.016
2. Dahal A, McNevin D, Chikhani M, Ward J (2023) An interdisciplinary forensic approach for human remains identification and missing persons investigations. Wiley Interdiscipl Rev Forensic Sci 5:e1484. https://doi.org/10.1002/wfs2.1484
3. Watson JL, McNevin D, Ward J (2024) Genetic Kinship Testing Techniques for Human Remains Identification and Missing Persons Investigations. Forensic Genom 4(1):4–23. https://doi.org/10.1089/forensic.2023.0018
4. Wickenheiser RA (2022) Expanding DNA database effectiveness. Forensic Sci Int Synerg 4:100226. https://doi.org/10.1016/j.fsisyn.2022.100226
5. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SLB, Karoma NJ et al (2006) Developing a SNP panel for forensic identification of individuals. Forensic Sci Int 164(1):20–32. https://doi.org/10.1016/j.forsciint.2005.11.017
6. Seo SB, King JL, Warshauer DH, Davis CP, Ge J, Budowle B (2013) Single nucleotide polymorphism typing with massively parallel sequencing for human identification. Int J Legal Med 127(6):1079–1086. https://doi.org/10.1007/s00414-013-0879-7
7. Butler JM, Coble MD, Vallone PM (2007) STRs vs. SNPs: thoughts on the future of forensic DNA testing. Forensic Sci Med Pathol 3(3):200–5. https://doi.org/10.1007/s12024-007-0018-1
8. Jordan D, Mills D (2021) Past, Present, and Future of DNA Typing for Analyzing Human and Non-Human Forensic Samples. Front Ecol Evol 9:172. https://doi.org/10.3389/fevo.2021.646130
9. ATDBio. Sequencing, forensic analysis and genetic analysis 2021. [Last accessed: 07/01/2022]. Available from: https://atdbio.com/nucleic-acids-book/Sequencing-forensic-analysis-and-genetic-analysis#DNA-Fingerprinting.
10. Jeffreys AJ (2013) The man behind the DNA fingerprints: an interview with Professor Sir Alec Jeffreys. Investig Genet 4(1):21. https://doi.org/10.1186/2041-2223-4-21
11. Illumina. An Introduction to Next-Generation Sequencing Technology 2017. [Last accessed: 27 May]. Available from: https://

www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.
12. Greytak EM, Kaye DH, Budowle B, Moore C, Armentrout SL (2018) Privacy and genetic genealogy data. Science 361(6405):857.1
13. Antunes J, Walichiewicz P, Forouzmand E, Barta R, Didier M, Han Y et al (2024) Developmental Validation of the ForenSeq® Kintelligence Kit, MiSeq Fgx® Sequencing System and ForenSeq Universal Analysis Software. Forensic Sci Int Genet 71:103055. https://doi.org/10.1016/j.fsigen.2024.103055
14. QIAGEN. Verogen ForenSeq Kintelligence Kit 2024. [Last accessed: 24 May]. Available from: https://www.qiagen.com/us/products/human-id-and-forensics/nextgeneration-sequencing/verogen-forenseq-kintelligence-kit.
15. Tillmar A, Sturk-Andreaggi K, Daniels-Higginbotham J, Thomas JT, Marshall C. The FORCE panel: An all-in-one SNP marker set for confirming investigating genetic genealogy leads and for general forensic applications. bioRxiv. 2021. https://doi.org/10.1101/2021.11.30.470354.
16. Verogen. ForenSeq Kintelligence Kit Reference Guide. ForenSeq Kintelligence Kit. 2021.
17. Staadig A, Hedman J, Tillmar A (2023) Applying Unique Molecular Indices with an Extensive All-in-One Forensic SNP Panel for Improved Genotype Accuracy and Sensitivity. Genes 14(4):818. https://doi.org/10.3390/genes14040818
18. Tillmar A, Andreaggi K, Amory C, Ballard D, Bergseth EF, Børsting C, et al. Advancing forensic SNP typing: Insights from an interlaboratory study of the FORCE panel [Conference session]. 12th Haploid Markers Conference, Budapest, Hungary. 2023, May 19.
19. Ge J, Budowle B, Chakraborty R (2011) Choosing Relatives for DNA Identification of Missing Persons. J Forensic Sci 56:S23–S28. https://doi.org/10.1111/j.1556-4029.2010.01631.x
20. Kling D (2019) On the use of dense sets of SNP markers and their potential in relationship inference. Forensic Sci Int Genet 39:19–31. https://doi.org/10.1016/j.fsigen.2018.11.022
21. Kling D, Egeland T, Tillmar AO (2012) FamLink–a user friendly software for linkage calculations in family genetics. Forensic Sci Int Genet 6(5):616–620. https://doi.org/10.1016/j.fsigen.2012.01.012
22. Lewis KE. Statistical Methods for Kinship Analysis. National Institute of Standards and Technology Biochemical Science Division, Applied Genetics Group - DNA Biometrics Project: Towson University; 2010.
23. Kling D, Tillmar A, Egeland T, Mostad P (2015) A general model for likelihood computations of genetic marker data accounting for linkage, linkage disequilibrium, and mutations. Int J Legal Med 129:943–954. https://doi.org/10.1007/s00414-014-1117-7
24. Kling D, Phillips C, Kennett D, Tillmar A (2021) Investigative genetic genealogy: Current methods, knowledge and practice. Forensic Sci Int Genet 52:102474. https://doi.org/10.1016/j.fsigen.2021.102474
25. Glynn CL (2022) Bridging Disciplines to Form a New One: The Emergence of Forensic Genetic Genealogy. Genes 13(8):1381. https://doi.org/10.3390/genes13081381
26. QIAGEN. QIAamp® DNA Investigator Handbook. 2020.
27. Loreille OM, Diegoli TM, Irwin JA, Coble MD, Parsons TJ (2007) High efficiency DNA extraction from bone by total demineralization. Forensic Sci Int Genet 1(2):191–195. https://doi.org/10.1016/j.fsigen.2007.02.006
28. Sigma-Aldrich. Amicon® Ultra-15 Centrifugal Filters 30 kDa MWCO Millipore 2023. [Last accessed: 2 August]. Available from: https://www.sigmaaldrich.com/AU/en/product/mm/ufc9030.
29. QIAGEN. MinElute® Handbook. 2020.
30. Thermo Fisher Scientific. QuantStudio™ 5 Real-Time PCR Instrument(for Human Identification) [User Guide]. Life Technologies. 2017.
31. Thermo Fisher Scientific. Quantifiler™ HP and Trio DNA Quantification Kits User Guide [User Guide]. 2017.
32. Promega. QuantiFluor® ONE dsDNA System. 2022.
33. Promega. Quantus™ Fluorometer. 2022.
34. QIAGEN. QIAseq™ Targeted DNA Panel Handbook. 2017.
35. Ilumina. Whole-Genome Sequencing (WGS) 2021. [Last accessed: 07/01/2022]. Available from: https://sapac.illumina.com/techniques/sequencing/dna-sequencing/whole-genome-sequencing.html.
36. Watson JL, McNevin D, Grisedale K, Spiden M, Seddon S, Ward J (2023) Operationalisation of the ForenSeq® Kintelligence Kit for Australian Unidentified and Missing Persons Casework. Forensic Sci Int Genet 68:102972. https://doi.org/10.1016/j.fsigen.2023.102972
37. Verogen. MiSeq FGx Sequencing System Reference Guide. MiSeq FGx Sequencing System. 2021.
38. Verogen. Universal Analysis Software v2.0 and above, Reference Guide. Universal Analysis Software. 2021.
39. Illumina. Sequencing Analysis Viewer 2023. [Last accessed: 2 December]. Available from: https://support.illumina.com/sequencing/sequencing_software/sequencing_analysis_viewer_sav.html.
40. QIAGEN. QIAGEN CLC Genomics Workbench 2024. [Last accessed: 28 May]. Available from: https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-clc-genomics-workbench/.
41. Kling D, Egeland T, Tillmar A. FamLink Databases 2024. [Last accessed: 13 February]. Available from: https://famlink.se/f_databases.html.
42. Chaitanya L, Breslin K, Zuñiga S, Wirken L, Pośpiech E, Kukla-Bartoszek M et al (2018) The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. Forensic Sci Int Genet 35:123–135. https://doi.org/10.1016/j.fsigen.2018.04.004
43. Walsh S, Chaitanya L, Breslin K, Muralidharan C, Bronikowska A, Pospiech E et al (2017) Global skin colour prediction from DNA. Hum Genet 136:847–863. https://doi.org/10.1007/s00439-017-1808-5
44. Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L et al (2014) Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet 9:150–161. https://doi.org/10.1016/j.fsigen.2013.12.006
45. STRmix™. DBLR™ 2023. [Last accessed: 29 January]. Available from: strmix.com/dblr/.
46. University R. Map Interpolator of the Rutgers Map: Matise Laboratory of Computational Genetics; 2014. [Last accessed: 20 October]. Available from: http://compgen.rutgers.edu/map_interpolator.shtml.
47. The International Genome Sample Resource. 1000 Genomes Project 2021. [Last accessed: 17 October]. Available from: https://www.internationalgenome.org/.
48. Peck MA, Koeppel AF, Gorden EM, Bouchet JL, Heaton MC, Russell DA et al (2022) Internal Validation of the ForenSeq Kintelligence Kit for Application to Forensic Genetic Genealogy. Forensic Genom. https://doi.org/10.1089/forensic.2022.0014
49. Cheung EY, Gahan ME, McNevin D (2018) Prediction of biogeographical ancestry in admixed individuals. Forensic Sci Int Genet 36:104–111
50. Cheung EY, Gahan ME, McNevin D (2018) Predictive DNA analysis for biogeographical ancestry. Aust J Forensic Sci 50(6):651–658

51. Salvo NM, Janssen K, Kirsebom MK, Meyer OS, Berg T, Olsen G-H (2021) Predicting eye and hair colour in a Norwegian population using Verogen's ForenSeqTM DNA Signature Prep Kit. Forensic Sci Int Genet 56:102620. https://doi.org/10.1016/j.fsigen.2021.102620
52. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M (2011) IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5(3):170–180. https://doi.org/10.1016/j.fsigen.2010.02.004
53. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A et al (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet 7(1):98–115. https://doi.org/10.1016/j.fsigen.2012.07.005
54. Department of Genetic Identification of Erasmus MC. HIrisPlex-S Eye, Hair and Skin Colour DNA Phenotyping Webtool 2022. [Last accessed: 19th April]. Available from: https://hirisplex.erasmusmc.nl.
55. Scientific Working Group on DNA Analysis Methods. Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Resulted Reported as Likelihood Ratios. 2016.
56. Gettings KB, Tillmar A, Sturk-Andreaggi K, Marshall C (2024) Review of SNP assays for disaster victim identification: Cost, time, and performance information for decision-makers. J Forensic Sci 69(5):1546–1557. https://doi.org/10.1111/1556-4029.15585
57. Kling D, Tillmar A (2019) Forensic genealogy—A comparison of methods to infer distant relationships based on dense SNP data. Forensic Sci Int Genet 42:113–124. https://doi.org/10.1016/j.fsigen.2019.06.019
58. Tillmar A, Sjölund P, Lundqvist B, Klippmark T, Älgenäs C, Green H (2020) Whole-genome sequencing of human remains to enable genealogy DNA database searches – A case report. Forensic Sci Int Genet 46:102233. https://doi.org/10.1016/j.fsigen.2020.102233
59. Alonso A, Müller P, Roewer L, Willuweit S, Budowle B, Parson W (2017) European survey on forensic applications of massively parallel sequencing. Forensic Sci Int Genet 29:e23–e25. https://doi.org/10.1016/j.fsigen.2017.04.017
60. Budowle B, Arnette A, Sajantila A (2023) A cost–benefit analysis for use of large SNP panels and high throughput typing for forensic investigative genetic genealogy. Int J Legal Med 137(5):1595–614. https://doi.org/10.1007/s00414-023-03029-7
61. Thermo Fisher Scientific. GlobalFiler PCR Amplification Kits. 2019.
62. QIAGEN. QIAseq Index Kits, QIAseq 12-Index I (48) 2023. [Last accessed: 23 March]. Available from: https://www.qiagen.com/us/products/discovery-and-translational-research/next-generation-sequencing/dna-sequencing/germline-panels/qiaseq-indexes?catno=333714.
63. QIAGEN. QIAseq Targeted DNA Pro Custom Panels 2024. [Last accessed: 24 May]. Available from: https://www.qiagen.com/us/products/discovery-and-translational-research/next-generation-sequencing/dna-sequencing/targeted-dna-panels/qiaseq-targeted-dna-pro-custom-panels?catno=333491.
64. QIAGEN. Verogen MiSeq FGx Sequencing Reagents 2024. [Last accessed: 15 August]. Available from: https://www.qiagen.com/us/products/human-id-and-forensics/nextgeneration-sequencing/verogen-miseq-fgx-sequencing-reagents.
65. QIAGEN. ForenSeq® Kintelligence HT Kit Technical Information 2024. Available from: https://www.qiagen.com/br/resources/download.aspx?id=63849dbe-ab1f-4dbe-968f-e468055637a6&lang=en.
66. QIAGEN. Verogen ForenSeq Kintelligence HT Kit 2024. [Last accessed: 15 August]. Available from: https://www.qiagen.com/us/products/human-id-and-forensics/nextgeneration-sequencing/verogen-forenseq-kintelligence-ht-kit.
67. Verogen. SNP Typing in Universal Analysis Software and Kinship Estimation with GEDmatch PRO [Technical Note]. ForenSeq Kintelligence Kit. 2021.
68. Ward J (2022) A multifaceted, multijurisdictional, multiagency, and multidisciplinary approach to investigating unidentified and missing persons cases in Australia. Forensic Sci Int Genet Suppl Ser 8:53–55. https://doi.org/10.1016/j.fsigss.2022.09.020
69. QIAGEN. ForenSeq® Signature Plus Kit and Universal Analysis Software version 2.7. 2024.
70. GEDmatch PRO. GEDmatch PRO 2024. Available from: https://pro.gedmatch.com/.
71. FamilyTreeDNA. FamilyTreeDNA Law Enforcement Guide 2024. [Last accessed: 25/11/2024]. Available from: https://www.familytreedna.com/legal/law-enforcement-guide.
72. GEDmatch. GEDmatch.com Terms of Service and Privacy Policy: GEDmatch; 2024. [Last accessed: 25/11/2024]. Available from: https://www.gedmatch.com/terms-of-service-privacy-policy.
73. Antunes J, Radecke S, Snedecor J, Padmabandu G, Stephens KM (2022) Application of a dense SNP PCR multiplex for high throughput kinship determination using Next Generation Sequencing (NGS). Forensic Sci Int Genet Suppl Ser 8:324–325. https://doi.org/10.1016/j.fsigss.2022.10.007
74. Snedecor J, Fennell T, Stadick S, Homer N, Antunes J, Stephens K et al (2022) Fast and accurate kinship estimation using sparse SNPs in relatively large database searches. Forensic Sci Int Genet 61:102769. https://doi.org/10.1016/j.fsigen.2022.102769
75. QIAGEN. Release Note: ForenSeq® Universal Analysis Software (UAS) v2. 2024.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

4. CHAPTER FOUR

OPERATIONALISATION OF THE FORENSEQ® KINTELLIGENCE KIT FOR AUSTRALIAN UNIDENTIFIED AND MISSING PERSONS CASEWORK

Contributions of authors:

Watson J, McNevin D, Grisedale K, Spiden M, Seddon S and Ward, J (2024) 'Operationalisation of the ForenSeq® Kintelligence Kit for Australian unidentified and missing persons casework', Forensic Science International: Genetics, 68, DOI:10.1016/jfsigen.2023.102972.

Jessica Watson (Candidate)

Conceptualised and assisted in the experimental design for the study. Carried out sample collection, sample preparation, DNA testing and data analysis. Prepared manuscript, edited manuscript following review by other co-authors, referenced manuscript, carried out manuscript's submission and incorporated feedback from peer review process.

Signed: Production Note: Signature removed prior to publication.

Dennis McNevin (Co-supervisor)

Conceptualised and assisted in the experimental design for the study. Contributed to data analysis. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Kelly Grisedale (AFP DNA Program staff)

Assisted in the experimental design for the study. Contributed to sample preparation and DNA testing. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Michelle Spiden (AFP DNA Program staff)

Contributed to sample collection, sample preparation and DNA testing. Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

Shelley Seddon (AFP DNA Program staff)

Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

Jodie Ward (Co-supervisor)

Conceptualised and assisted in the experimental design for the study. Acquired funding for the study. Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

Forensic Science International: Genetics 68 (2024) 102972

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Operationalisation of the ForenSeq® Kintelligence Kit for Australian unidentified and missing persons casework

Jessica Watson [a,b,*], Dennis McNevin [a,b], Kelly Grisedale [a], Michelle Spiden [a], Shelley Seddon [a], Jodie Ward [a,b]

[a] Australian Federal Police National DNA Program for Unidentified and Missing Persons, Australia
[b] Centre for Forensic Science, School of Mathematical & Physical Sciences, Faculty of Science, University of Technology Sydney, Australia

ARTICLE INFO

Keywords:
Genetic kinship
Biogeographical ancestry
Externally visible characteristics
Forensic investigative genetic genealogy
Unidentified human remains
Missing persons
Massively parallel sequencing
Genetic intelligence

ABSTRACT

Single nucleotide polymorphism (SNP) genotyping technologies can generate investigative leads for human remains identification, including estimation of biological sex, biogeographical ancestry (BGA), externally visible characteristics (EVCs), identity, uniparental lineage and extended kinship. The ForenSeq® Kintelligence Kit provides forensic laboratories with the ability to apply this suite of genetic tools to forensic samples using one panel targeting 10,230 SNPs (including 56 ancestry-informative, 24 phenotype-informative, 94 identity-informative, 106 X chromosome, 85 Y chromosome and 9867 kinship-informative SNPs) sequenced on the MiSeq FGx® Sequencing System. The ForenSeq® Kintelligence Kit has been internally validated, optimised and operationalised by the Australian Federal Police National DNA Program for Unidentified and Missing Persons (AFP Program) for coronial casework. The internal validation was conducted according to the Scientific Working Group on DNA Analysis Methods guidelines (excluding mixture analysis), focussing on sample types typically encountered in human remains identification casework, such as bones, teeth, nail, blood and hair. The workflow was optimised for a high throughput library preparation and sequencing workflow, and additional analytical thresholds were developed to improve genotyping accuracy for low DNA input samples. Additionally, the genetic intelligence generated from the kit was compared to the self-declared biological sex, EVCs and BGA of the DNA donors to assess concordance. The kit was able to produce high quality SNP profiles from 1.0 ng down to 0.1 ng of DNA, with high repeatability and reproducibility, and minimal background noise. The prediction accuracy for biological sex (95%), hair colour (58%), eye colour (74%) and BGA inferences (consistent: 74%; partially consistent: 10%; inconclusive: 16%) was determined based on self-declared data. Additionally, SNP profiles from a volunteer family group of ten related individuals were uploaded to GEDmatch PRO™ to assess kinship accuracy. The kit was capable of detecting (97%) and accurately classifying (90%) genetic relationships spanning from first to fifth degree. The Kintelligence Kit provides the AFP Program with a robust and reliable genetic intelligence tool for unidentified and missing persons investigations, which has been designed to sequence multiple challenging samples in a single multiplexed assay using existing laboratory instrumentation.

1. Introduction

Australia currently has approximately 750 unidentified human remains (UHR) and 2500 long-term missing persons (LTMP) [1]. The Australian Federal Police (AFP) National DNA Program for Unidentified and Missing Persons (AFP Program) was established in 2020 to assist Australian police, forensic and coronial agencies identify these UHR, resolve LTMP cases and provide families with answers [2]. The AFP Program has implemented a suite of forensic technologies to achieve this, including new forensic genomics tools to generate investigative leads when identification is not possible using fingerprint, dental, medical or DNA data comparisons and/or searches of law enforcement databases [3].

For DNA identification, short tandem repeat (STR) profiles and law enforcement DNA databases are primarily used to find direct matches between an unidentified and missing person, or a kinship match between an unidentified person and a first degree relative (i.e. parent, offspring and full siblings) [4–7]. In the absence of a DNA match,

* Corresponding author at: Australian Federal Police National DNA Program for Unidentified and Missing Person, Australia.
E-mail address: Jessica.Watson@student.uts.edu.au (J. Watson).

https://doi.org/10.1016/j.fsigen.2023.102972
Received 14 August 2023; Received in revised form 3 October 2023; Accepted 24 October 2023
Available online 30 October 2023
1872-4973/© 2023 Elsevier B.V. All rights reserved.

emerging forensic genomics techniques have the potential to provide other investigative leads for law enforcement to pursue. Such leads include the estimation of biological sex, biogeographical ancestry (BGA), externally visible characteristics (EVCs), identity, uniparental lineage and kinship relative to putative family members. Through the use of single nucleotide polymorphisms (SNPs), these predictions can be used to narrow the potential identity of the unknown individual from a pool of possibilities.

When analysing old and degraded biological samples, typical in unidentified and missing persons casework, the shorter amplicon lengths in SNP panels improve the chances of obtaining full DNA profiles [8–10]. SNPs can be classified according to the intended application or predictions gained from the genotype [11]. This includes X chromosome SNPs (X SNPs) and Y chromosome SNPs (Y SNPs) from the sex chromosomes, ancestry-informative SNPs (AI SNPs), phenotype-informative SNPs (PI SNPs), kinship-informative SNPs (KI SNPs) and identity-informative SNPs (II SNPs) [12]. Similar to an STR panel targeting Amelogenin, biological sex can be predicted from a SNP panel using X and Y SNPs. While the X chromosome is present in both male and female DNA, it is haploid in males and the Y chromosome is male specific. Both can be used as an indicator of the donor's biological sex.

BGA and EVC inference currently provides an estimate of an unknown individual's genetic ancestry and pigmentation (e.g. hair and eye colour) [13–16]. AI SNP alleles are ideally shared by individuals within population groups and differ from other groups and this allows estimation of BGA or ancestral origins [17,18]. The majority of panels incorporating BGA prediction require at least several dozen AI SNPs to predict whether an individual can be included or excluded from a population. The most common super populations readily distinguishable from each other are European, Asian and African. One pipeline commonly used for inferring BGA uses principal component analysis (PCA) which clusters individuals by ancestry within two or three dimensions [19]. Another common pipeline involves the use of the Structure model-based likelihood estimator [20,21]. The HIrisPlex System was developed by Walsh et al. [15,16] for predicting hair and eye colour from 24 PI SNPs using multinominal logistic regression (MLR). This system predicts three eye colours (blue, brown and intermediate) and four hair colours (blond, brown, red and black).

The use of KI SNPs in a medium- or high-density panel allows medium- and long-range kinship analysis to be conducted, respectively. Long-range kinship analysis is also referred to as forensic investigative genetic genealogy (FIGG). The use of high-density SNP genotypes in UHR cases increases the number of potential relatives that could be detected and identified. The more closely related two individuals are, the more genetic markers will be shared that have been inherited from a most recent common ancestor [22]. To conduct FIGG, a dense SNP profile is uploaded to a law enforcement accessible public or private genetic genealogy database to detect potential relatives, followed by genealogy research to construct a family tree to ultimately narrow the search to a present-day family with a missing or unaccounted-for relative [23–25].

GEDmatch PRO™ is a law enforcement portal that permits the upload of SNP profiles from forensic samples for FIGG [26,27]. The One-to-Many Kinship Tool enables an unknown SNP profile to be searched against the GEDmatch database to link with potential genetic relatives based on the amount of shared DNA. For unidentified and missing persons casework, the entire GEDmatch database is included in the search. For defined criminal investigations (e.g. homicides and aggravated sexual assaults), only the SNP profiles of consenting individuals can be compared to the uploaded law enforcement SNP profiles. The GEDmatch database is populated by DNA data generated by direct-to-consumer (DTC) consumer genomics companies including AncestryDNA®, FamilyTreeDNA, MyHeritage™, 23andMe® and Living DNA [28]. These companies utilise microarray sequencing, targeting upwards of 500,000 SNPs across the human genome. These high density SNP profiles make up the majority of profiles populating the GEDmatch database [29]; however, similar density SNP profiles generated using whole genome sequencing approaches also exist on the database [28].

One of the techniques capable of producing SNP profiles of sufficient density to allow all of these inferences to be made is targeted amplicon sequencing (TAS). As it is a massively parallel sequencing (MPS) approach, TAS can be utilised in an increasing number of forensic laboratories that already have bench-top sequencers for in-house implementation [30]. One such instrument is the MiSeq FGx® Sequencing System (Verogen Inc., now a QIAGEN company), which was designed specifically for application of DNA sequencing in forensic casework [28,31]. It employs the ForenSeq® suite of forensic MPS assays for inference of identity (using STRs and SNPs), BGA and EVCs [32,33].

The ForenSeq® Kintelligence Kit (Verogen Inc.) was released in February 2021 as an end-to-end solution for extended kinship analysis within a forensic laboratory [10,32,34]. It was designed to be implemented on the MiSeq FGx® Sequencing System using established ForenSeq® library preparation chemistry and methods. The Kintelligence Kit targets 10,230 SNPs across the whole human genome to provide a range of genotypes useful for forensic applications. Using the Universal Analysis System (UAS) (Verogen Inc.), it is possible to infer biological sex, BGA and EVCs. The UAS has embedded PCA (for BGA) and MLR (for EVCs) algorithms. For kinship inference, the UAS generates a report compatible for direct upload to GEDmatch PRO™ [26,27] or submission to FamilyTreeDNA [35].

This study aimed to validate the Kintelligence Kit and optimise the laboratory and bioinformatics analysis methods for operationalisation by the AFP Program for coronial casework. This kit was assessed by the AFP Program as it had the potential to provide an end-to-end, in-house SNP-based solution for extended kinship analysis that is not currently available in an accredited Australian forensic laboratory. Furthermore, this kit would enable the AFP Program to offer its stakeholders an accessible, effective and efficient FIGG capability, as well as other genetic intelligence to generate investigative leads for human remains identification.

2. Materials and methods

2.1. Sample preparation, extraction and quantification

Volunteers were recruited to provide biological samples and complete a questionnaire (refer to supplementary material) recording their biological sex, BGA, EVCs and relationship to genetic relatives also included in the study. There were ten genetically related individuals (Fig. 1) and six random unrelated individuals selected for inclusion in the study. Additional case-type human remains samples were sourced from donors at the Australian Facility for Taphonomic Experimental Research (AFTER), as well as from approved research and casework samples submitted to the AFP Program. All samples are detailed in Table 1.

Buccal swab, nail (1x clipping) and hair shaft (2 cm) samples were extracted using the QIAamp DNA Investigator® Kit (QIAGEN) [36]. Blood (FTA: 0.5 cm^2; whole blood: 50 µL) samples were extracted using the EZ1® DNA Investigator Kit (QIAGEN) [37]. For bone and tooth samples, 0.5 g of pulverised material was extracted using the MinElute® PCR Purification Kit [38] following total demineralisation lysis [39] and Amicon® 30 K concentration steps [40].

Quantification was performed using the Quantifiler™ Trio DNA Quantitation Kit (Applied Biosystems™) on a QuantStudio™ 5 Real-Time PCR System (Applied Biosystems™) according to the manufacturer's protocol [41]. Refer to Table S1 for the quantification results.

2.2. Library preparation and sequencing

Libraries were prepared according to the manufacturer's protocol for the Kintelligence Kit with the exception of sample throughput (see Section 2.6.2 Sample Throughput). If required, the samples were diluted

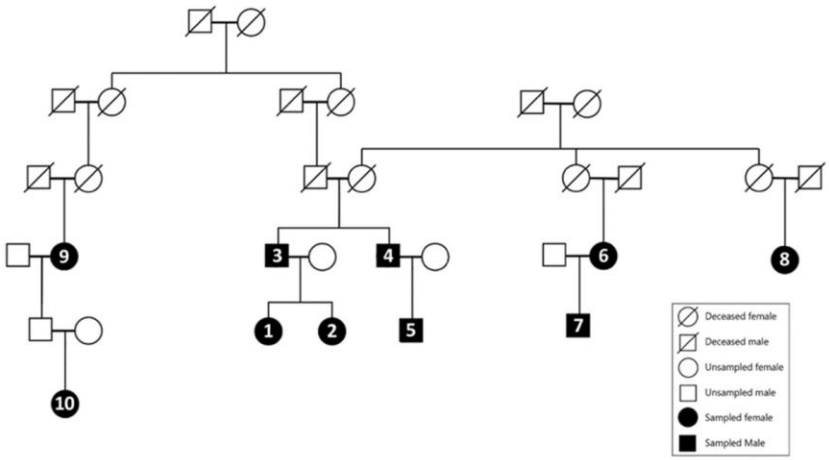


Fig. 1. Pedigree chart of the selected volunteers comprising the family group. Individuals included in this study are filled in (black), with living relatives linking the family members not filled (white). Deceased relatives are crossed out. The individuals numbered 1 through 10 correspond with the samples Family 1 through 10 in Table 1.

to the recommended 1 ng input using the large autosomal (LA) target concentration. The QuantiFluor® ONE dsDNA System (Promega®) with the Quantus™ Fluorometer (Promega®) were used to normalise the libraries to 0.75 ng/μL for pooling [42]. Following manufacturer recommendations, three individual libraries were selected for pooling and the pooled libraries were stored at − 20°C prior to sequencing [43].

Each pooled library was sequenced on the MiSeq® FGx System using the MiSeq® FGx Reagent Kit (Verogen Inc.) and standard flow cell (SFC) according to the manufacturer's protocol. Different index combinations were used on successive runs to limit the effect of sample carryover. Sequencing run quality metrics were assessed on the Kintelligence Module of the UAS (Table 2).

2.3. Bioinformatic analysis

For each sample, the Sample, Phenotype & Ancestry and GEDmatch PRO™ Reports were generated on the UAS using the default analysis method (UAS 20X), a lowered analysis method (UAS 10X) and a minimum UAS thresholds method (Table 3). Alleles that did not meet the minimum read counts were not typed by the software. The processed read information for each SNP was exported from the Sample Report using the minimum UAS thresholds and analysed in a bespoke Microsoft Excel macro workbook to generate a genotype with the additional thresholds optimised in this validation (Table 4).

The biological sex and contributor status were determined by the UAS using the manufacturer thresholds for X and Y SNP calls. The UAS has an inbuilt PCA method for BGA inference (i.e. European, East Asian, African) and an MLR algorithm modelled from the HirisPlex webtool [44] to infer hair and eye colour. These inferences are summarised an exportable Phenotype & Ancestry Report.

2.4. GEDmatch PRO™

For kinship analysis, the UAS generated GEDmatch PRO™ Reports for the family group samples were uploaded to GEDmatch PRO™. Each sample was given a unique kit number and analysed with the One-to-Many Kinship Tool. The estimated shared centimorgan (cM) values, longest segment length, mean segment length, number of segments, number of SNPs overlapping between the kits and the predicted relationship degree were analysed.

2.5. Validation studies

The validation was conducted in accordance with the Scientific Working Group on DNA Analysis Methods (SWGDAM) Validation Guidelines for DNA Analysis Methods [45,46]. With the purpose of applying the Kintelligence Kit to AFP Program casework, samples typically encountered in UHR (e.g. bones, teeth, nail and hair) and missing persons (e.g. buccal swabs, stored biological samples) cases were tested. Mixture analysis was excluded from this validation as mixtures are not commonly recovered from human remains when applying AFP Program procedures. Additionally, any mixtures are usually detected by STR profiling prior to use of the Kintelligence Kit and not submitted for further downstream analysis.

2.5.1. Sensitivity study

The control DNA NA24385 was diluted in duplicate to provide DNA inputs of 1.0 ng (manufacturer's recommendation), 0.5 ng and 0.1 ng. For consistency, a high quality sample was prepared and sequenced as the third library on each SFC alongside the two replicates of each control DNA dilution. The samples were analysed for call rate, accuracy, precision, intralocus balance (ILB) and heterozygosity. Allele frequency plots were generated using dplyr [47] and ggplot2 [48] packages in R [49], so the range of allele frequencies and coverage could be visualised simultaneously.

2.5.2. Repeatability and reproducibility study

To assess repeatability between sequencing runs, a triplicate and two duplicates of the control DNA NA24385 at 1.0 ng DNA input underwent library preparation in the same batch but were sequenced on different SFCs. Reproducibility was assessed by two operators (A and B) sequencing five replicates of the control DNA NA24385 on two different SFCs at 1.0 ng DNA input. The accuracy (concordance rate) and precision (degree of similarity between replicates) was assessed within each operator's samples and reproducibility was assessed between operators.

Table 1
Volunteer samples, the Australian Facility for Taphonomic Experimental Research (AFTER) samples and Australian Federal Police (AFP) Program research and casework samples.

Sample Name	Sample Information[1]	Source
NA24385	Control DNA	ForenSeq® Kintelligence Kit
Family 1	Buccal swab	Volunteer (family group)
Family 2	Buccal swab	Volunteer (family group)
Family 3	Buccal swab	Volunteer (family group)
Family 4	Buccal swab	Volunteer (family group)
Family 5	Buccal swab	Volunteer (family group)
Family 6	Buccal swab	Volunteer (family group)
Family 7	Buccal swab	Volunteer (family group)
Family 8	Buccal swab	Volunteer (family group)
Family 9	Buccal swab	Volunteer (family group)
Family 10	Buccal swab	Volunteer (family group)
Unrelated 1	Buccal swab	Volunteer
Unrelated 2	Buccal swab	Volunteer
Unrelated 3	Buccal swab	Volunteer
Unrelated 4	Buccal swab	Volunteer
Unrelated 5	Buccal swab	Volunteer
Unrelated 6	Buccal swab	Volunteer
Hair 1	Baby rootless hair cuttings from Family 1, stored for 38 years (25 °C)	Volunteer (family group)
Blood 1	Blood collected on FTA card from Family 1, stored for 18 months (25 °C)	Volunteer (family group)
Blood 2	Whole blood, stored for 6 months (−20°C)	AFP Program research sample
Tooth 1	Primary tooth from Family 2, stored for approximately 25 years (25 °C)	Volunteer (family group)
Tooth 2	Permanent tooth, stored for 25 years (25 °C)	AFP Program research sample
Bone 1	Bone (femur), postmortem interval (PMI) unknown, saltwater exposure	AFP Program casework sample
Bone 2	Bone (femur), PMI approximately 230 years, buried	AFP Program casework sample
Bone 3	Bone (mandible), PMI unknown, saltwater exposure	AFP Program casework sample
Bone 4	Bone (petrous), PMI unknown, surface deposition	AFP Program casework sample
Bone 5	Bone (petrous), PMI approximately 70 years, buried	AFP Program casework sample
Bone 6	Bone (petrous), PMI 2 years, surface deposition	AFTER
Bone 7	Bone (petrous), PMI 2 years, surface deposition	AFTER
Bone 8	Bone (petrous), PMI 2 years, surface deposition	AFTER
Bone 9	Bone (petrous), PMI 2 years, shallow burial	AFTER
Bone 10	Bone (petrous), PMI 2 years, shallow burial	AFTER
Bone 11	Bone (petrous), PMI 2 years, shallow burial	AFTER
Nail 1	Nail clipping, stored for 2 years (25 °C)	AFP Program research sample
Nail 2	Nail clipping, PMI 2 months	AFTER

[1]All buccal swabs were stored at room temperature until dried, and then stored at 4 °C prior to extraction within two weeks of collection.

2.5.3. Positive control performance

The average performance of the positive control (PC) samples (1.0 ng of control DNA NA24385) was assessed for six replicates from different library preparation batches. The samples were analysed for total coverage, call rate, concordance and heterozygosity.

Table 2
Target values for the run quality metrics.

Metric	Target Value
Cluster Density	400 – 1600 K/mm^2
Clusters Passing Filter	≥ 80%
Phasing	≤ 0.25%
Prephasing	≤ 0.15%
Read 1	Pass
Index 1	Pass
Index 2	Pass
Read 2	Pass
Human Sequencing Control (HSC) Overall Intensity	Pass
HSC Zero Discordant Loci	Pass

Table 3
Analytical methods within the Universal Analysis System (UAS).

Thresholds	Default UAS Method (UAS 20X)	Lowered UAS Method (UAS 10X)	Minimum UAS Method
Minimum Read Count	20	10	10
Analytical Threshold (AT)	3%	1.5%	1%
Interpretation Threshold (IT)	3%	1.5%	1%
Intralocus Balance	50%	50%	50%

Table 4
Additional thresholds applied in the Microsoft Excel macro workbook.

Threshold	Value	Interpretation
Minimum Allele Frequency (MAF) Threshold	Varied from 0.00 to 0.30 in increments of 0.05	Minor alleles with a frequency less than this threshold were not reported. Major alleles with a frequency greater than 1.00 minus this threshold were typed as homozygous. Loci with both allele frequencies above this threshold were typed as heterozygous.
Total Coverage Threshold	Varied from 10 to 50 in increments of 10	Loci with the total number of reads below this threshold were not reported.

2.5.4. Contamination study

The NC and PC samples, and any dilutions of the PC, were analysed for allele drop-in and background noise. Any drop-in observed was compared to other samples in the library preparation batch and sequencing run to attempt to identify the source of the potential contamination.

2.5.5. Casework-type and reference-type samples

A range of sample types commonly submitted to the AFP Program were tested. This included several buccal swabs as reference-type samples, and a selection of casework-type samples including bone, tooth, nail, hair and blood samples which ranged in DNA quantity, degradation index (DI), postmortem interval (PMI) and environmental exposure (Table 1). Samples sourced from the same DNA donor were compared for precision.

The Quantifiler™ Trio DNA Quantitation kit quantifies DNA using a small autosomal (SA) target of 80 bp and a large autosomal (LA) target of 214 bp. A highly degraded bone sample (Bone 5, DI = 82.94) was sequenced using DNA input amounts derived from both the SA and LA target quantitation values to assess which quantitation target is more appropriate to use when determining the DNA input for the sequencing of degraded samples. The SA target concentration for Bone 5 was 0.705 ng/μL, therefore the sample was diluted to 0.04 ng/μL and the recommended 1.0 ng of DNA input was used. The LA target

concentration for Bone 5 was 0.0085 ng/μL. Dilution was not required based on this and the maximum volume (25 μL) was used for a DNA input of 0.213 ng. The equivalent quantity according to the SA target for this sample would have been 17.625 ng. The sequencing results were compared for coverage, call rate, heterozygosity and concordance between the genotypes.

2.6. Optimisation studies

2.6.1. Optimisation of analytical thresholds

Replicates of the control DNA NA24385 diluted to 100 pg input were analysed at the UAS 20X and 10X thresholds for concordance to the known reference genotype. Using the Sample Reports generated with the minimum UAS thresholds, the samples were then analysed using different coverage and minor allele frequency (MAF) thresholds in the macro workbook. The proportion of concordant, non-concordant (allele dropout or drop-in) and untyped SNPs were recorded at each threshold. The pROC package [50] in R [49] was used to determine the area under the receiver operating characteristic curve (AUC) for a range of MAF thresholds, with the concordance of autosomal SNPs varying as a function of the coverage threshold.

2.6.2. Sample throughput

The Kintelligence Kit was designed to sequence 12 samples in four batches, with each batch consisting of a PC, an NC and an unknown sample [43]. Using this approach, four unknown samples are able to be sequenced per kit. In order to evaluate variations to multiplexing [46], a high throughput method was investigated, where 12 individual libraries were prepared in a single batch (a negative and positive control and ten samples; Fig. 2). These 12 individual libraries were divided into four sequencing batches of three pooled libraries, with the PC and NC in the first sequencing run.

To assess the impact of the high throughput sequencing method on sequencing efficiency, the sequencing run metrics were compared between the Sequencing Run 1 (following the manufacturer's SFC recommendations) and Sequencing Runs 2–4. Additionally, the control DNA NA24385 was included in Sequencing Run 4 and compared to the PC in Sequencing Run 1.

2.6.3. Genotyping errors

The UAS determines the genotype at a locus according to which strand is being sequenced against a set of expected alleles. Unexpected alleles are flagged as genotyping errors in the system and automatically removed from the profile. These errors were analysed across 54 samples of varying sample type, input amount and sample quality for frequency.

2.6.4. Locus and allele dropout

Locus dropout was analysed across 49 samples for trends in chromosome and SNP category. The dropout percentage was calculated as the proportion of observed dropouts across all samples, with the exception of Y SNPs which were calculated from male samples only (n = 30). Allele dropout was analysed across the 15 control DNA

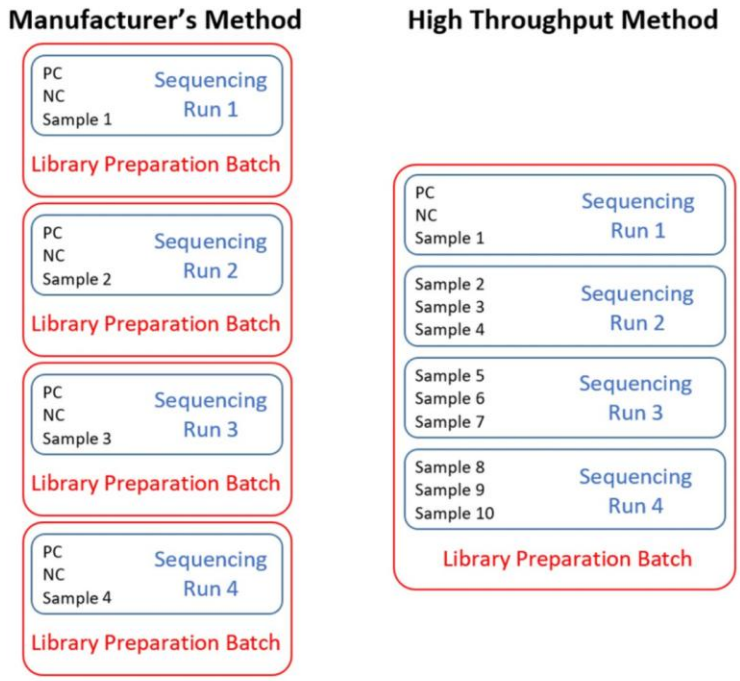


Fig. 2. The difference between the manufacturer's recommended method for preparing and sequencing three libraries in a single sequencing batch (left hand side) with the high throughput method for preparation of 12 libraries across four sequencing batches (right hand side). Each library preparation batch includes a positive control (PC) and negative control (NC).

NA24385 samples ranging from 0.1 to 1.0 ng of DNA input.

2.7. Assessment of sex, externally visible characteristics and biogeographical ancestry predictions

The UAS provides predictions of biological sex, BGA and EVCs as detailed in the Phenotype & Ancestry Report. To assess consistency of these predictions, they were compared to the self-declared data provided by the DNA donors in the questionnaires. The DNA donors provided their biological sex, eye colour, hair colour and the ancestries of themselves, their parents and grandparents. Given significant hair colour changes can occur in childhood and senior years for some individuals, donors were asked to provide their hair colour at 20 years of age.

2.7.1. Prediction of biological sex

The UAS determines biological sex based on SNP count where male genotypes have $\geq$ 10 Y SNPs and female genotypes have no Y SNPs and an overall call rate of $\geq$ 50% [51]. Genotypes that do not satisfy these criteria return an inconclusive biological sex prediction. This prediction algorithm was compared to a validated pipeline embedded in the macro workbook (Table 5). The reported biological sex for samples with no self-declared data was determined by the quantitation data, using the male DNA target, or the typing of Amelogenin and Y STRs using the GlobalFiler™ PCR Amplification Kit [52].

2.7.2. Prediction of externally visible characteristics

The UAS has an embedded MLR algorithm to predict eye and hair colour using the 24 PI SNPs. The DNA donors provided their self-declared eye colour (blue, grey, green, hazel or brown) and hair colour (black, dark brown, brown, dark blond, blond, red or grey).

2.7.3. Prediction of biogeographical ancestry

The UAS uses an in-built PCA to infer BGA from four population clusters (European, East Asian, African and Admixed American). The Admixed American cluster is super-imposed over the other three population clusters, which is a limitation of the two dimensional representation. Admixed American was determined not to be relevant for the inference of BGA within an Australian population and was not used. To be included in a population, the sample coordinates had to be closer to the population cluster centroid than at least one other member of that population. If this did not apply to any of the population clusters, all populations were excluded and BGA was deemed inconclusive.

2.8. Assessment of kinship determination on GEDmatch PRO™

The GEDmatch PRO™ Reports generated for the buccal swabs collected from the genetically related volunteer group were uploaded to GEDmatch PRO™ and analysed in the One-to-Many Kinship Tool [26]. Approval was obtained from the GEDmatch Support Team (Verogen, Inc.) to upload the reports as UHR samples prior to the database update allowing upload of reports pertaining to validation samples. The predicted relationship degree was compared to the self-declared pedigree chart (Fig. 1) to determine the detection capability and accuracy of GEDmatch PRO™ using the Kintelligence Kit. The blood sample from Family 1 (Blood 1) and primary tooth from Family 2 (Tooth 1) were uploaded to simulate casework samples and direct reference samples from missing persons. Another DNA donor (Family 8) had previously uploaded an AncestryDNA® profile to GEDmatch and gave permission to report any matches to that kit.

Table 5
Biological sex prediction thresholds applied by the Universal Analysis System (UAS) and the validated pipeline.

Biological Sex	UAS Thresholds	Validated Pipeline Thresholds
Male	$\geq$ 10 Y SNPs typed.	$\geq$ 10 Y SNPs typed.
Female	0 Y SNPs typed and $\geq$ 50% overall call rate.	$\leq$ 5 Y SNPs and $\geq$ 50% X SNPs typed.
Inconclusive	Sample is a negative control, a mixture, or male and female rules are not met.	$<$ 10 Y SNPs and $<$ 50% X SNPs typed.

3. Results

3.1. Validation and optimisation of the Kintelligence Kit

3.1.1. Run quality metrics

Sequencing performance was assessed across 21 runs using the UAS default run quality metrics. For 19 runs (90.48%), all run quality metrics passed. The average cluster density across all runs was 1221 $\pm$ 50 K/mm^2 and cluster passing filter was 86.85 $\pm$ 0.92%. Of the other two runs, the first was over-clustered (cluster density: 1893 K/mm^2; clusters passing filter: 74.97%) but all other metrics passed. The second run had only the Read 2 metric not passing, likely caused by the run also being over-clustered. Despite these metrics not passing, there was no noticeable impact on the sample genotypes.

3.1.2. Optimisation of analytical thresholds

For each sample, the concordant, non-concordant and untyped SNPs were recorded with each combination of MAF and total coverage threshold, as well as according to the UAS 20X and 10X analysis methods (Fig. 3). The proportion of non-concordant SNPs decreased with an increasing total coverage threshold, while the proportion of untyped SNPs increased. Both of the UAS analysis methods were outperformed by some of the threshold combinations tested. At a 20X total coverage threshold and MAF $<$ 0.15, approximately 87.5% of SNPs were concordant and 10% were untyped, with only 2.5% non-concordant with the reference genotype. For all MAF thresholds, 95% of non-concordance was allele dropout and the remaining 5% was allele drop-in.

The total coverage thresholds were applied to the NCs to determine the impact on potential contamination and background noise (Fig. 4). Both the 10X total coverage threshold and UAS 10X had an average number of SNPs typed per NC exceeding the recommended interpretation threshold defined in the reference guide [43].

Fig. 3 demonstrates that there is little difference in genotyping accuracy for MAF $<$ 0.15. Similarly, the AUC values were greater than 0.9 for the UAS 10X analysis method and MAF thresholds $\leq$ 0.15 (Fig. 5). With little difference between the 0.05 and 0.10 MAF thresholds, the MAF window 0.05 – 0.10 was deemed an "ambiguous" region where genotypes were not reported, meaning the major allele had an ambiguous window of 0.90 – 0.95 that was also not reported. Applying these window (referred to as the optimised analytical method, OM) produced the highest AUC of 0.940.

The OM thresholds demonstrated higher concordance than the UAS analytical methods for a range of DNA inputs (Fig. 6). At the 1.0 ng DNA input, the concordance rates with the OM was significantly greater than those produced with the default UAS 20X analysis method (Kruskal-Wallace, $p = 5.4 \times 10^{-4}$). These thresholds were used for the remainder of the validation.

3.1.3. Sample throughput

The run metrics were compared between Sequencing Run 1 (following the manufacturer's SFC recommendations) and Sequencing Runs 2–4 (Fig. 2, Table 6). The runs were not compromised by excluding an NC, with only one run being over-clustered. In fact, the high throughput method resulted in a significantly lower proportion of dimers compared to Sequencing Run 1 ($p = 0.045$). There were no other significant differences between the two formats.

As expected, the coverage was reduced for the NA24385 samples in Sequencing Run 4 (n = 4) when compared to the PC in Sequencing Run 1 (n = 6, Table 7). However, there were no significant differences

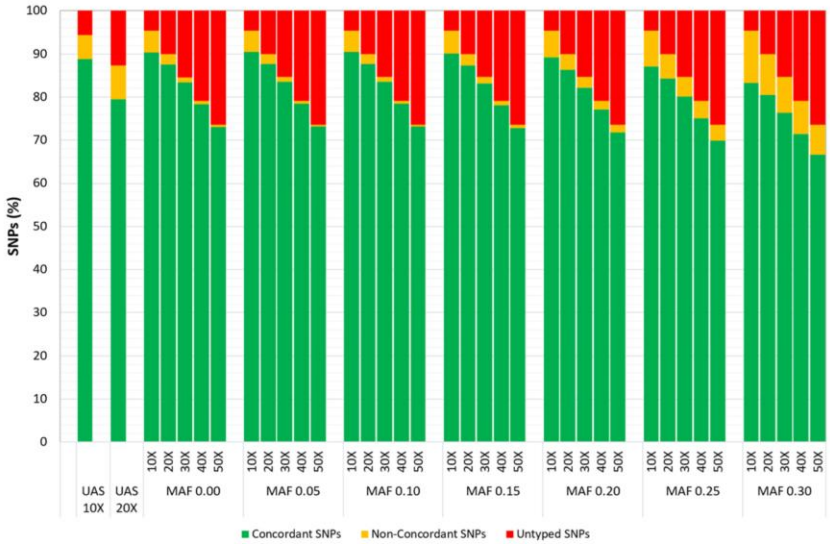


Fig. 3. Concordant, non-concordant and untyped SNPs as functions of the default UAS 20X, UAS 10X and combinations of the minimum allele frequency (MAF) and total coverage thresholds for NA4385 samples at 0.1 ng DNA input (n = 3).

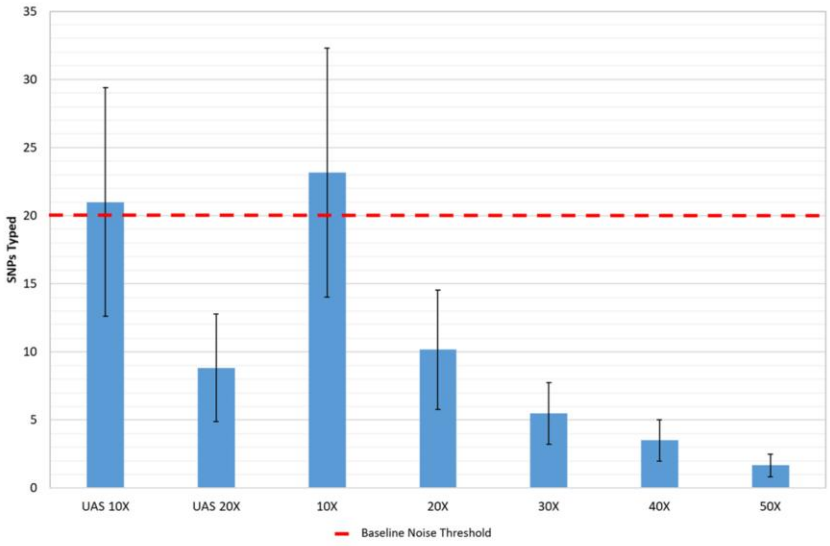


Fig. 4. Average number of SNPs typed in the negative control (NC) samples for the default UAS 20X, UAS 10X and various total coverage thresholds, together with standard errors. The baseline noise threshold is 20 typed SNPs.

observed in call rate, concordance rate and heterozygosity between the manufacturer's method and the high throughput method. For Sequencing Run 1, approximately one half of the total reads on the flow cell were attributed to the PC, which decreased to approximately one third of the total reads for Sequencing Run 4. No allele drop-in was observed in the samples for either method.

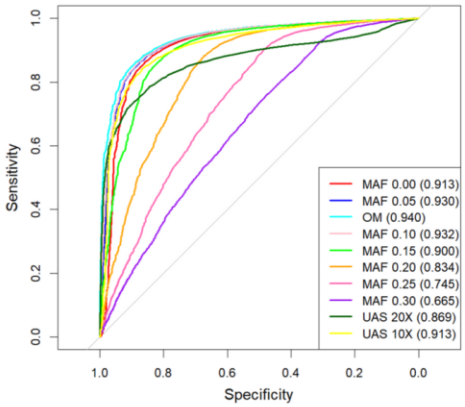


Fig. 5. Area under the receiver operating characteristic curve (AUC) analysis comparing the default UAS 20X, UAS 10X and minimum allele frequency (MAF) thresholds, including the optimised analytical threshold (OM; MAF: 0.05; ambiguous windows 0.05 – 0.10 and 0.90 – 0.95), for genotyping accuracy.

3.1.4. *Sensitivity study*

For the manufacturer's recommended 1.0 ng and the 0.5 ng DNA inputs, both call and concordance rates exceeded 99% (Fig. 7). At the lowest DNA input of 0.1 ng, the call and concordance rates dropped to approximately 90%. Allele dropout, resulting in false homozygous genotypes, was observed at all DNA input amounts (0.08 – 1.96%). Allele drop-in, resulting in false heterozygous genotypes, was only observed in the 0.1 ng samples at a low rate. For both the 1.0 ng and 0.5 ng DNA input samples, the heterozygosity was close to the true value of the NA24385 reference profile (46.97%), but the accuracy decreased for the 0.1 ng samples (44.23%).

When categorised by SNP class, PI SNPs were most consistently recovered across the different DNA input amounts (Fig. 8). Similar degrees of locus dropout were observed across X SNPs, AI SNPs, II SNPs and KI SNPs, with the high call rates observed in the 1.0 and 0.5 ng DNA input samples dropping substantially in the 0.1 ng DNA input samples. Y SNPs suffered the steepest loss of data as call rates fell below 80% at the lowest DNA input.

At the lowest DNA input, the allele frequency plots show a wider spread of allele frequencies, with more alleles falling into the ambiguous windows (Fig. S1); as the DNA input decreased, the ILB ratios drifted further from the 1:1 expected for heterozygous SNPs. However, the concordance rates were still high for these samples (using the OM) despite the wider variance in allele frequencies.

3.1.5. *Repeatability and reproducibility study*

There was little variance between sequencing runs for the control

Table 6
Average run quality metrics with standard errors for Sequencing Run 1 (6 sequencing runs, modelling the manufacturer's recommendations) and Sequencing Runs 2–4 (14 sequencing runs).

Quality Metric	Sequencing Run 1	Sequencing Runs 2–4
Density (K/mm^2)	1216 ± 54	1229 ± 71
Cluster Passing Filter (%)	86.88 ± 1.41	87.01 ± 1.23
Phasing (%)	0.129 ± 0.003	0.128 ± 0.010
Prephasing (%)	0.058 ± 0.009	0.060 ± 0.003
Dimers (%)	26.67 ± 1.74	19.36 ± 1.97
Coverage (Total Reads)	21,149,042	25,401,667

Table 7
Average sample performance with standard errors for control DNA NA24385 samples sequenced in Sequencing Run 1 (6 sequencing runs) compared to Sequencing Run 4 (4 sequencing runs).

Sample Performance	Sequencing Run 1	Sequencing Run 4
Call Rate (%)	99.67 ± 0.15	99.64 ± 0.12
Concordance Rate (%)	99.88 ± 0.04	99.91 ± 0.04
Heterozygosity (%)[1]	46.83 ± 0.05	46.85 ± 0.05
Coverage (Total Reads)	12,511,340	10,759,060

[1] True heterozygosity of NA24385 is 46.97%.

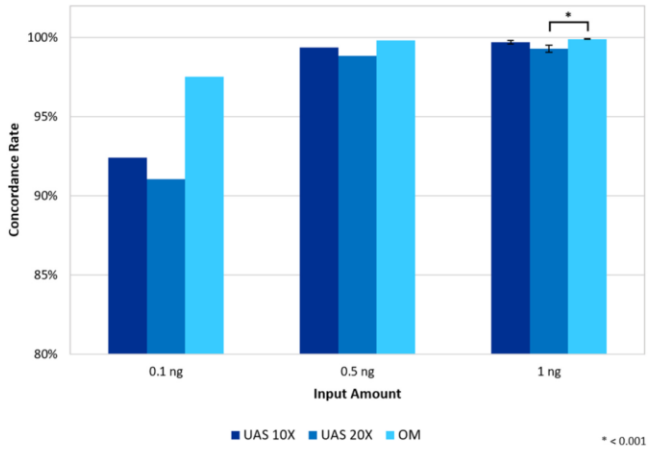
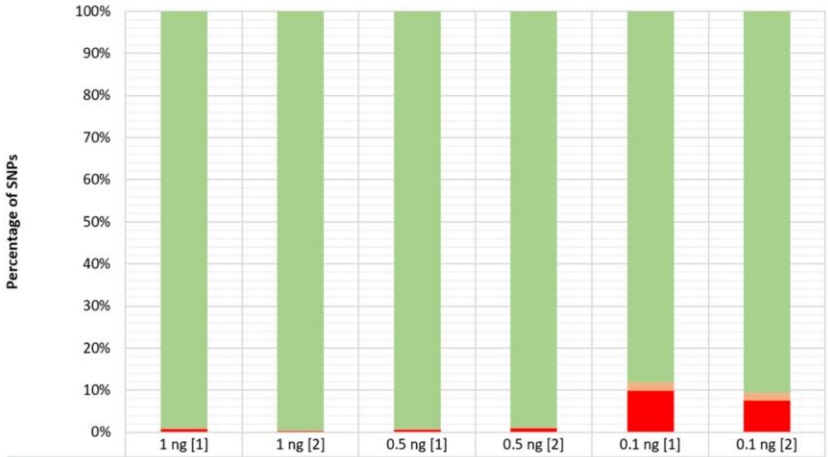
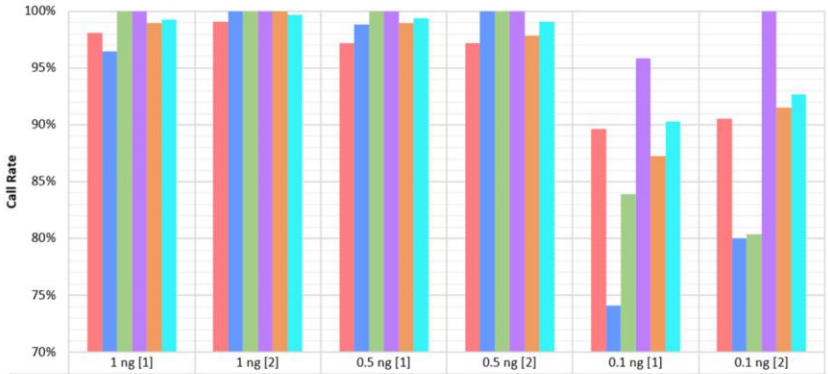


Fig. 6. Concordance rates for the control DNA NA24385 by DNA input amount with the default UAS 20X, UAS 10X and optimised analytical method (OM). Error bars represent standard error. Significant difference at a p value of 0.001 is indicated. Note: it was only possible to determine significance for the 1.0 ng DNA input amount (n = 8) but not for 0.5 ng (n = 2) and 0.1 ng (n = 3).



	1 ng [1]	1 ng [2]	0.5 ng [1]	0.5 ng [2]	0.1 ng [1]	0.1 ng [2]
Concordant	99.00%	99.60%	99.15%	98.85%	88.10%	90.67%
Allele Dropout	0.22%	0.08%	0.19%	0.23%	1.96%	1.71%
Allele Drop-In	0.00%	0.00%	0.00%	0.00%	0.04%	0.12%
Locus Dropout	0.77%	0.32%	0.66%	0.92%	9.90%	7.51%

Fig. 7. Percentage of SNPs that were concordant, non-concordant (allele dropout or allele drop-in) or not typed. Replicates of control DNA NA24385 are indicated by [1] and [2] for each DNA input amount (1.0 ng, 0.5 ng and 0.1 ng).



	1 ng [1]	1 ng [2]	0.5 ng [1]	0.5 ng [2]	0.1 ng [1]	0.1 ng [2]
X-SNPs	98.11%	99.06%	97.17%	97.17%	89.62%	90.57%
Y-SNPs	96.47%	100.00%	98.82%	100.00%	74.12%	80.00%
aiSNPs	100.00%	100.00%	100.00%	100.00%	83.93%	80.36%
piSNPs	100.00%	100.00%	100.00%	100.00%	95.83%	100.00%
iiSNPs	98.94%	100.00%	98.94%	97.87%	87.23%	91.49%
kiSNPs	99.26%	99.68%	99.36%	99.10%	90.29%	92.68%

Fig. 8. Call rate by SNP category. Replicates of control DNA NA24385 are indicated by [1] and [2] for each DNA input amount (1.0 ng, 0.5 ng and 0.1 ng).

DNA NA24385 samples. Across all batches, the call rate, accuracy and precision exceeded 99% within each group (Fig. 9). There was also little variance within and between the two operators, with no significant differences reported. The call rates (A: 99.58 ± 0.17%, B: 99.73 ± 0.10%), accuracy (A: 99.88 ± 0.04%, B: 99.91 ± 0.04%) and precision (A: 99.16 ± 0.12%, B: 99.41 ± 0.10%) exceeded 99% for all 10 samples (Fig. 10).

3.1.6. *Positive control performance*

There was wide variability in the total reads across all PC samples;

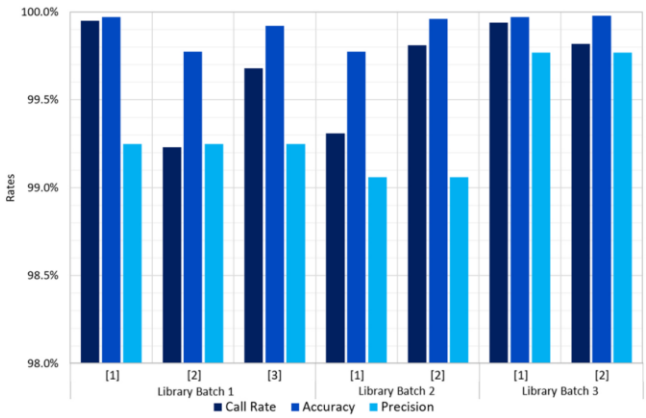


Fig. 9. Inter-run repeatability with 1.0 ng DNA inputs of control DNA NA24385 with the average call rate, accuracy and precision for each library batch. Replicates are indicated by [1], [2] and [3].

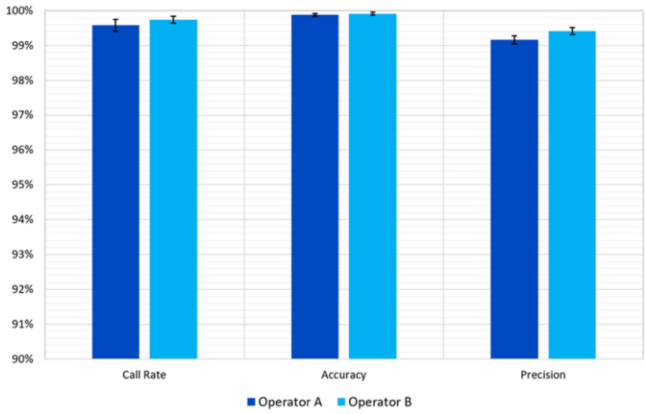


Fig. 10. Operator repeatability and reproducibility with 1.0 ng DNA inputs of control DNA NA24385 average call rate, accuracy and precision, each with standard error (five samples per operator).

however, the call rate, concordance rate and heterozygosity were high (Table 8). The coverage met the manufacturer's threshold of 15 million reads for two PCs (33%) but was less than six million reads for another two (33%). The call rates exceeded 99% and the concordance of the typed SNPs exceeded 99.7% in all PCs. The non-concordant genotypes were restricted to allele dropout, as indicated by the heterozygosity for the control DNA NA24385 being below the true value of 46.97%.

Table 8
The average performance of positive control (PC) samples (n = 6) with standard error.

Sample Metric	Sample Performance
Coverage (Total Reads)	12,511,341 ± 2,141,045
Call Rate (%)	99.67 ± 0.15
Concordance Rate (%)	99.88 ± 0.04
Heterozygosity (%)[1]	46.83 ± 0.05

[1]True heterozygosity of NA24385 is 46.97%.

3.1.7. Contamination study

For the seven NC samples analysed for contamination, a total of 65 SNPs, all single allele drop-in, were observed with an average of 9.3 ± 3.6 SNPs per profile. Only two NCs exceeded the 20 SNP limit for the baseline noise threshold (NC_5: 27 SNPs; NC_6: 23 SNPs), while the other NCs did not surpass seven SNPs typed (Fig. 11). With the exception of an II SNP typed in NC_6, all allele drop-ins were KI SNPs. The source of the majority of SNPs (97%) seen in the seven NCs could be attributed to the adjacent wells during library preparation as they shared one allele in common with the drop-in at that locus. The source of two SNPs in two NCs (NC_4 and NC_8) could not be determined. The total coverage ranged from 72 to 1489 total reads with the read depth for individual SNP calls only exceeding 100 reads for the NCs that had more than 20 SNPs called.

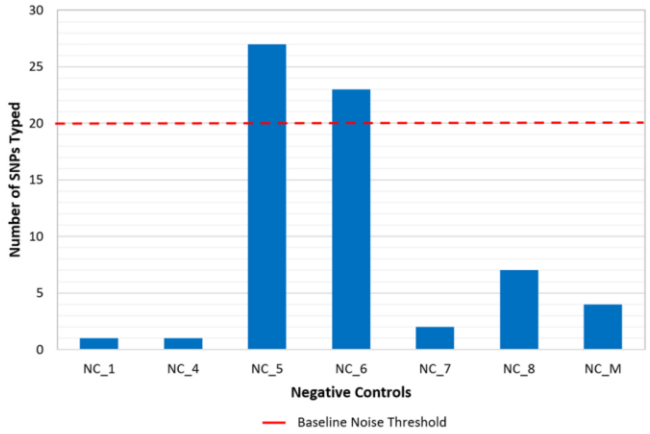


Fig. 11. Number of SNPs typed for each negative control (NC) sample. The baseline noise threshold is 20 typed SNPs.

Contamination was not observed in the control DNA NA24385 samples with a DNA input of 1.0 or 0.5 ng. For the control DNA NA24385 samples with a DNA input of 0.1 ng (n = 3), a total of 20 KI SNP loci were observed with allele drop-in and flagged for non-concordance as incorrectly being typed as heterozygous when the reference genotype at that locus was homozygous. The average read depth ranged from 13 to 52 reads per allele. From these observations, the risk of allele drop-in with a DNA input of 0.1 ng is 0.065%.

3.1.8. *Genotyping errors*

Genotyping errors, where an unexpected allele was detected and removed by the UAS, were observed in 40 samples (74%) and no correlation was found with sample type, DNA input or DI of the sample. There were 181 unexpected alleles typed, occurring at a rate of 0.033% or 3.35 alleles per sample. The majority of the unexpected alleles were adenine (52%) and thymine (44%). Of these 181 unexpected alleles, 157 alleles (87%) had allele frequencies below 0.05 and were removed from the profile when the OM was applied, having no effect on the final genotype (Fig. 12). Unexpected alleles were not reproduced in replicates of the control DNA NA24385, casework-type or reference-type samples sourced from the same DNA donor. Erroneous genotypes with allele frequencies greater than 0.05 resulted in the locus being removed from

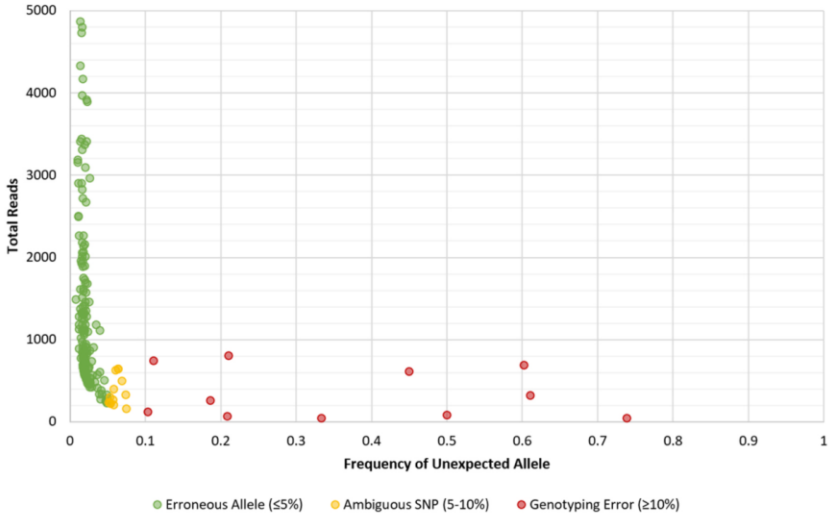


Fig. 12. Read depth at loci with unexpected alleles plotted against the allele frequency. Erroneous alleles (green: MAF $\leq$ 0.05) were not reported and had no impact on the final genotype. Ambiguous genotypes (orange: 0.05 $<$ MAF $<$ 0.10) and genotyping errors (red: MAF $\geq$ 0.10) resulted in the SNP being removed from the profile.

the profile. SNPs were removed as a result of genotyping error at a rate of 0.004%, or 0.44 SNPs per sample. Of the 11 genotyping errors with allele frequencies greater than 0.10, four were associated with KI SNP rs2280018. Of these, two were not replicated in the sample duplicates.

3.1.9. Locus and allele dropout

SNPs on the Y chromosome had a higher overall frequency of dropout than SNPs on any of the other chromosomes, with 48 Y SNPs (56%) not being typed in at least one sample. For the autosomal and X chromosomes, 23.2 ± 0.6% SNPs located on each chromosome were not typed in at least one sample. With the exception of Y SNPs, 27 ± 2% of SNPs in each SNP category were not typed in at least one sample.

The dropout percentage for each SNP was calculated as the proportion of samples where the SNP in question had not been typed. KI SNPs had a greater chance of locus dropout, with three SNPs exceeding 70% - rs4720273 at 73% and rs7951628 and rs509448 at 71% (Table 9). This higher chance of KI SNP dropout was consistent with the lower average coverage per SNP. Due to the large number of KI SNPs, higher rates of dropout can be tolerated than for non-KI SNPs. Poor performing non-KI SNPs with dropout proportions greater than 30% included: rs1413212 (II SNP, 71%), rs2920816 (II SNP, 31%), rs1207651 (X SNP, 39%), rs475025 (X SNP, 37%), rs2745708 (X SNP, 31%), rs35383156 (X SNP, 31%), rs9786169 (Y SNP, 53%), rs7067290 (Y SNP, 47%) and rs2032664 (Y SNP, 43%).

There were few trends in the allele dropout rates recorded in the control DNA NA24385 samples ranging from 0.1 to 1.0 ng of DNA input (n = 15). For the 1.0 ng input samples (n = 10), 107 occurrences of allele dropout were observed where a SNP was incorrectly typed as a homozygote, 73 (68%) of which were unique. The SNP with the highest allele dropout rate of 90% was the KI SNP rs4530357. The average number of SNPs with allele dropout increased with decreasing DNA input amounts (1.0 ng: 10.7 SNPs; 0.5 ng: 21.5 SNPs; 0.1 ng: 221.3 SNPs). Only at the lowest DNA input were non-KI SNPs affected by allele dropout; six II SNPs and one AI SNP were incorrectly typed as homozygotes. Similar to locus dropout, there were no noticeable trends by chromosome.

3.1.10. Casework-type and reference-type samples

The DNA input, DI and call rate for each sample tested is specified in Table 10. With the exception of the hair sample, all samples returned a call rate over 90% (n = 33). For the bone samples (n = 11), only three did not meet the target DNA input amount of 1.0 ng (Bone 2, Bone 3 and Bone 5). Despite this, all bone samples performed well regardless of the DNA input and DI (as high as 88 for Bone 5). The bone samples gave call rates over 90% for the majority of SNP classes; the only exceptions were the PI SNPs for Bone 5 (83%) and Y SNPs for Bone 2 (81%). The tooth samples (n = 2) performed well, and Tooth 1 produced a high call rate (93%) despite a DNA input of 0.166 ng. The nail (n = 2) and blood samples (n = 2) produced profiles with call rates exceeding 90%. The sequencing of the hair sample was unsuccessful, with only 14 SNPs being typed; however, this was not unexpected as nuclear DNA was not detected by the quantitation assay.

Of the reference-type samples (buccal swabs, n = 16), all had a high average call rate of 98.5 ± 0.5% despite Family 2 and Family 3 not meeting the recommended target DNA input amount of 1.0 ng. There were two samples where the call rates were lower than expected given the quantitation values were high. Family 9 was diluted to the target concentration of 0.04 ng/μL from 3.91 ng/μL and Blood 2 from 8.50 ng/μL, yet the call rates were 92.05% and 90.64% respectively.

There were differences observed between the sequencing results of the sample sequenced using the DNA input amounts derived from the SA target and the LA target (Table 11). Preparing the library using the maximum volume according to the LA target concentration resulted in a 10-fold increase in total reads, as well as increasing the number of SNPs typed by 497. The heterozygosity was lower for the SA target sample, indicating a higher degree of allele dropout. The LA target sample had a higher average read depth across all SNPs and fewer SNPs with allele frequencies falling below the MAF threshold and within the ambiguous windows. Furthermore, when the calculated LA target DNA input was

Table 9
SNP locus dropout by SNP category.

Dropout Percentage	X SNP (106)	Y SNP (85)	AI SNP (56)	PI SNP (24)	II SNP (94)	KI SNP (9867)
No Dropout Observed	76	37	40	19	63	7594
< 10%	22	36	10	4	26	1600
10 – 19%	4	6	4	1	3	365
20 – 29%	0	3	2	0	0	154
30 – 39%	4	0	0	0	1	81
40 – 49%	0	2	0	0	0	34
50 – 59%	0	1	0	0	0	26
60 – 69%	0	0	0	0	0	10
70 – 79%	0	0	0	0	1	3

Table 10
Casework-type and reference-type sample DNA input amounts, degradation indices (DI) and call rates.

Sample Name	DNA Input (ng)	Degradation Index (DI)	Call Rate (%)
Family 1	1.000	0.81	97.85
Family 2	0.150	3.33	95.37
Family 3	0.935	1.15	96.92
Family 4	1.000	1.64	99.86
Family 5	1.000	1.46	99.81
Family 6	1.000	2.32	99.76
Family 7	1.000	1.14	99.74
Family 8	1.000	1.16	99.76
Family 9	1.000	0.98	92.05
Family 10	1.000	0.66	99.79
Unrelated 1	1.000	0.89	99.18
Unrelated 2	1.000	0.65	98.07
Unrelated 3	1.000	0.96	99.68
Unrelated 4	1.000	1.30	99.37
Unrelated 5	1.000	0.96	99.70
Unrelated 6	1.000	0.66	99.86
Hair 1	0.000	N/A[1]	0.14
Blood 1	1.000	1.06	98.86
Blood 2	1.000	0.59	90.64
Tooth 1	0.166	5.79	93.22
Tooth 2	1.000	1.08	99.93
Bone 1	1.000	1.01	99.00
Bone 2	0.742	2.28	95.03
Bone 3	0.158	2.67	96.99
Bone 4	1.000	4.96	99.67
Bone 5	0.213	82.94	97.36
Bone 6	1.000	1.98	98.43
Bone 7	1.000	5.36	99.14
Bone 8	1.000	2.30	99.50
Bone 9	1.000	4.85	98.93
Bone 10	1.000	1.78	98.70
Bone 11	1.000	1.76	99.34
Nail 1	1.000	1.43	99.13
Nail 2	1.000	1.26	96.70

[1]DNA was not detected, therefore DI could not be calculated.

Table 11
Comparison of sequencing metrics using small autosomal (SA) and large autosomal (LA) target concentrations for determining the DNA input amount.

Sample	DNA Input (ng)	Total Reads	Call Rate	Heterozygosity	Concordance
Bone 5 (SA target)	1.000	2471,725	92.50%	45.57%	92.76%
Bone 5 (LA target)	0.213	24,595,917	97.36%	47.14%	

0.213 ng, the estimated calculated SA target DNA input was approximately 17 ng. This exceeded the recommended 1.0 ng of DNA for the assay, yet had no adverse effects on the profile. Following this result, the LA target was used to calculate DNA input for library preparation if the DI is > 1 to maximise DNA input.

Family 1 donor provided a fresh buccal swab (Family 1), a blood sample stored on FTA card at room temperature for 18 months (Blood 1) and a rootless hair cutting stored at room temperature for 38 years (Hair 1). There was high concordance between the buccal and blood genotypes (98%); however, only the PI SNPs were fully concordant (Table 12). Higher read depths were achieved in the blood sample and the ILB ratios were closer to the 1:1 expected for heterozygous SNPs. The hair sample did not generate a usable genotype and only single alleles were typed.

Family 2 donor also provided a fresh buccal swab (Family 2) and a primary tooth stored at room temperature for 30 years (Tooth 1). Both samples did not meet the recommended DNA input amount and were moderately degraded (Family 2: DI = 3.33; Tooth 1: DI = 5.79). Both samples had a high degree of concordance between the genotypes (94%, Table 12). Each SNP category had concordance rates exceeding 90%, with the highest concordance observed in the AI SNPs (98) and the lowest in the II SNPs (91%). The buccal swab had a greater average read depth across the typed loci, but more SNPs fell within the ambiguous window and were not reported.

3.2. Assessment of sex, externally visible characteristics and biogeographical ancestry predictions

3.2.1. Prediction of biological sex

Table S2 shows the comparison of the self-declared or reported biological sex of the DNA donor and the predictions made by the UAS and the validated pipeline. Hair 1 was excluded from this study as no SNPs from the X or Y chromosome were typed and predictions could not be generated. Of the 47 genotypes, the UAS was consistent with the true biological sex for 45 predictions (96%; Table 12). All 28 male genotypes were correctly identified, and 17 of the 19 female genotypes (89%) were correctly identified. Two buccal swabs produced inconclusive results. The buccal swabs from Family 2 and Family 9, who self-declared as biological females, both had two Y SNPs typed in their profile (29 ± 4 reads per SNP). Prior to library preparation, the quantification assay did not detect male DNA in the female samples to indicate a mixture (Table S1). Furthermore, the male to female ratio did not indicate a mixture in samples where male DNA was detected. The validated biological sex inference pipeline improved the accuracy to 100%.

3.2.2. Prediction of externally visible characteristics

Table 13 outlines the UAS predictions for hair colour for four phenotypes (red, blond, brown and black hair colour) compared to the volunteer's self-declared hair colour at 20 years of age. Of the 19 genotypes, 11 predictions (58%) were consistent with the self-declared information. This included the only blond-haired volunteer of the study, both black-haired volunteers, and eight of the 15 brown-haired volunteers (53%). The remaining eight self-declared brown-haired genotypes produced inconsistent predictions, or the phenotype was unable to be determined due to insufficient SNPs typed. All inconsistencies had largest p-values greater than 0.7, except for Family 8 which had a p-value of 0.35 for self-declared brown hair colour. Of the six inconsistencies observed (32%), the red-haired volunteer and two brown-haired volunteers were incorrectly predicted as blond, and three brown-haired volunteers were incorrectly predicted as black.

Table 13
Comparison of the hair colour predictions with the self-declared hair colour at 20 years old. Refer to Table S3 for all hair colour p-values.

Sample	Declaration	Largest p-value	Predicted Hair Colour	Comment
Family 1	Brown	0.90	Brown	Consistent
Family 2	Brown	-	Undetermined	-
Family 3	Brown	0.83	Brown	Consistent
Family 4	Dark brown	0.53	Brown	Consistent
Family 5	Brown	0.64	Brown	Consistent
Family 6	Brown	0.73	Blond	Inconsistent
Family 7	Red	0.78	Blond	Inconsistent
Family 8	Dark brown	0.52	Black	Inconsistent
Family 9	Dark brown	0.61	Brown	Consistent
Family 10	Dark brown	0.61	Brown	Consistent
Unrelated 1	Black	0.77	Black	Consistent
Unrelated 2	Dark brown	0.67	Brown	Consistent
Unrelated 3	Black	0.89	Black	Consistent
Unrelated 4	Dark brown	0.90	Black	Inconsistent
Unrelated 5	Dark brown	0.82	Black	Inconsistent
Unrelated 6	Brown	0.84	Blond	Inconsistent
Blood 1	Brown	0.90	Brown	Consistent
Tooth 1	Brown	-	Undetermined	-
Tooth 2	Dark Blond	0.88	Blond	Consistent

The prediction consistency of the UAS performed similarly for eye colour as it did for hair colour. Table 14 outlines the predictions for three possible phenotypes (brown, blue and intermediate eye colour) compared to the DNA donor's self-declared eye colour information. Of the 19 genotypes, 14 predictions (74%) were consistent with the declarations, three predictions (16%) were inconsistent, and two genotypes (10%) had too few SNPs to return a prediction. These inconsistent predictions corresponded to all five volunteers who self-declared intermediate eye colour, including hazel and green colours. All nine brown and five blue or grey eye colours were correctly predicted by the algorithm.

3.2.3. Prediction of biogeographical ancestry

Table S5 outlines the UAS predictions for BGA compared to the self-declared ancestries for themselves and their parents and grandparents. Of the 19 genotypes, 14 predictions (74%) were consistent with the included and excluded populations. There were two predictions (10%) that were partially consistent due to the inclusion of one of the population groups making up an individual's admixed ancestry. For these, the European population cluster was included for Unrelated 2 and Unrelated 6 who had self-declared European ancestry. However, they also declared Middle Eastern and Australian Aboriginal ancestry, respectively, which were unable to be detected by the UAS. There were three predictions (16%) where the BGA was inconclusive as all three population clusters were excluded. These genotypes included either admixed individuals (Family 8: admixed European and South East Asian) or individuals from

Table 12
Comparison of sequencing metrics using different biological samples from the same source.

Sample Source	Sample	DNA Input (ng)	Total Reads	Call Rate	Heterozygosity	Concordance
Family 1	Family 1	1.000	5785,007	97.87%	47.33%	97.91%
	Blood 1	1.000	14,871,976	98.88%	47.56%	
	Hair 1	0.000	622	0.14%	0.00%	N/A
Family 2	Family 2	0.150	3,583194	95.37%	46.44%	94.02%
	Tooth 1	0.166	3359,256	93.22%	45.82%	

Table 14
Comparison of the eye colour predictions with the self-declared eye colour. Refer to Table S4 for all eye colour p-values.

Sample	Declaration	Largest p-value	Predicted Eye Colour	Comment
Family 1	Brown	0.72	Brown	Consistent
Family 2	Hazel	-	Undetermined	-
Family 3	Grey	0.91	Blue	Consistent
Family 4	Grey	0.68	Blue	Consistent
Family 5	Brown	0.83	Brown	Consistent
Family 6	Green	0.89	Blue	Inconsistent
Family 7	Blue	0.94	Blue	Consistent
Family 8	Hazel	0.94	Brown	Inconsistent
Family 9	Brown	0.76	Brown	Consistent
Family 10	Green	0.72	Brown	Inconsistent
Unrelated 1	Brown	0.99	Brown	Consistent
Unrelated 2	Brown	0.66	Brown	Consistent
Unrelated 3	Brown	1.00	Brown	Consistent
Unrelated 4	Brown	1.00	Brown	Consistent
Unrelated 5	Brown	1.00	Brown	Consistent
Unrelated 6	Blue	0.92	Blue	Consistent
Blood 1	Brown	0.72	Brown	Consistent
Tooth 1	Hazel	-	Undetermined	-
Tooth 2	Blue	0.92	Blue	Consistent

population groups not represented in the PCA plot (Unrelated 1: Central Asian; Unrelated 5: Oceanian).

3.3. Assessment of kinship determination on GEDmatch PRO™

Kinship detection and prediction accuracy was assessed for the 10 family group volunteers, providing 39 true genetic relationships to test. The One-to-Many Kinship algorithm in GEDmatch PRO™ was reliable for detecting and classifying genetic relationships spanning from first to fifth degree (i.e. 2nd cousins). When comparing the default UAS 20X, UAS 10X and OM methods, the OM pipeline improved both detection rate and prediction accuracy (Table 15). Both the UAS 10X and OM methods had a detection rate of 79% for all relationships, which improved to 97% when considering only fifth degree relationships and closer. However, these methods were set apart by the accuracy of the predictions, with 90% of relationships correctly identified when the OM was applied. Only one relationship was detected beyond the fifth degree for all three analysis methods; a sixth degree relationship between Family 5 and Family 9 which was incorrectly estimated as a fifth degree relationship (Table 16). However, separating the more distant degrees of relationship can be challenging due to the overlap in the overall shared cM ranges (Fig. 13) [53]. Therefore, all detected relationships were predicted accurately within 1 degree of the true relationship.

For the additional biological samples tested, both the genotypes from the blood sample and primary tooth resulted in the same number of detected relatives and predicted the same degree of kinship as their corresponding buccal swab. This included identifying the buccal swab as "self" (i.e. originating from the same DNA donor).

Table 15
Detection and prediction accuracy rates for the One-to-Many Kinship algorithm for 39 relationships amongst the 10 volunteers of the family group. The methods include the UAS 10X, default UAS 20X and the optimised analytical method (OM).

Analysis Method	UAS 10X	UAS 20X	OM
Detected Relationships	31	29	31
Missed Matches (≤ Fifth Degree)	1	3	1
Missed Matches (> Fifth Degree)	7	7	7
Detection Rate	79.49%	74.36%	79.49%
Detection Rate (≤ Fifth Degree)	96.77%	90.32%	96.77%
Accuracy Rate (Exact)	80.65%	89.66%	90.32%
Accuracy Rate (± 1 degree)	100%	100%	100%
Average Overlapping SNPs	9743	9593	9615

The upload of the AncestryDNA® profile for Family 8 also resulted in the same relatives detected with the same degree of kinship prediction as the uploaded Kintelligence profile for the Family 8 donor. This included identifying the uploaded Kintelligence profile as "self". These matches were made despite the large difference in SNP densities, with the AncestryDNA® profile being comprised of over 500,000 SNPs compared to the 10,230 SNPs of the Kintelligence Kit. Furthermore, there were an average of 9623 SNPs overlapping between the Kintelligence Kit (analysed with OM) and AncestryDNA® profile.

4. Discussion

The Kintelligence Kit has been shown to be successful in typing a variety of forensic sample types to produce high quality genotypes and the internal validation has satisfied the relevant SWGDAM criteria. Our optimised analysis thresholds produce higher concordance rates than both the default UAS 20X analysis method and the lowered UAS 10X analysis method, especially at lower DNA inputs (0.1 ng). It has also been demonstrated that increasing the number of samples in the library preparation batch has not compromised sequencing. In a previous study, the sequencing plexity was increased from three to up to 32 samples, with no impact on profile quality; however, there was a significant reduction in SNPs typed which impacted on the kinship detection range [54,55]. The optimised protocol modifications developed in this validation have improved library preparation efficiency, while maintaining confidence in the final genotype with negligible loss of genotyped SNPs and concordance.

This assessment performed similarly to other evaluations of the Kintelligence Kit [56–58]. Their sensitivity studies support the conclusions of this study that, despite the recommended input of 1.0 ng of DNA, a DNA input of 0.1 ng still produces nearly complete profiles [56,57]. Peck et al. [56] sequenced DNA template as low as 0.05 ng and still retrieved over 90% of SNPs; however, the lower heterozygosity indicated an increased rate of allele dropout. Studies not included in the other evaluations of the kit included the occurrence of genotyping errors and locus and allele dropout. Unexpected alleles were likely caused by cytosine deamination, a base substitution of cytosine to thymine (C>T) or guanine to adenine (G>A) on the opposite DNA strand. Mutations resulting from cytosine deamination have been observed to largely contribute to the background noise of MPS, likely as a biological phenomenon or an artefact of thermocycling [59]. The trends observed in this study for locus dropout show that samples are less likely to lose highly informative SNPs such as those used to predict BGA and EVCs. As these inferences rely on high representation of a smaller number of SNPs than those used to infer kinship, it is critical that these SNPs have low rates of locus dropout.

Lower rates of contamination were observed in this study as opposed to others, with most NC samples having fewer than 10 SNPs typed and a maximum of 27 SNPs. This was substantially lower than other studies that observed 64 SNPs [57] and up to 92 SNPs [56]. However, none of the NCs from this study were indicative of widespread contamination as all called SNPs were single alleles with very low coverage (less than 0.02% of the average coverage per locus for other samples) and therefore would be eliminated by the MAF threshold during sample analysis.

A range of casework-type samples have been genotyped in various studies with varying success [56,58]; however, this study focused on sample types typical of unidentified and missing persons casework. Subsequently, mixtures were untested due to the expectation of single-source profiles; although, there have been varying rates of success in the deconvolution of mixed samples analysed with the Kintelligence Kit to date [56,60]. The UAS predicts contributor status for each sample based on a number of criteria related to X SNP and Y SNP detection, as

Table 16
Confusion matrix for kinship predictions using the optimised analytical method (OM).

		True Kinship Degree							
		1st	2nd	3rd	4th	5th	6th	7th	8th
Predicted Kinship Degree	1st	**6**							
	2nd		**4**						
	3rd			**6**					
	4th			1	**8**				
	5th				1	**4**	1		
	6th								
	7th								
	8th								
	Undetected					1	2	2	3
Totals		6	4	7	9	5	3	2	3

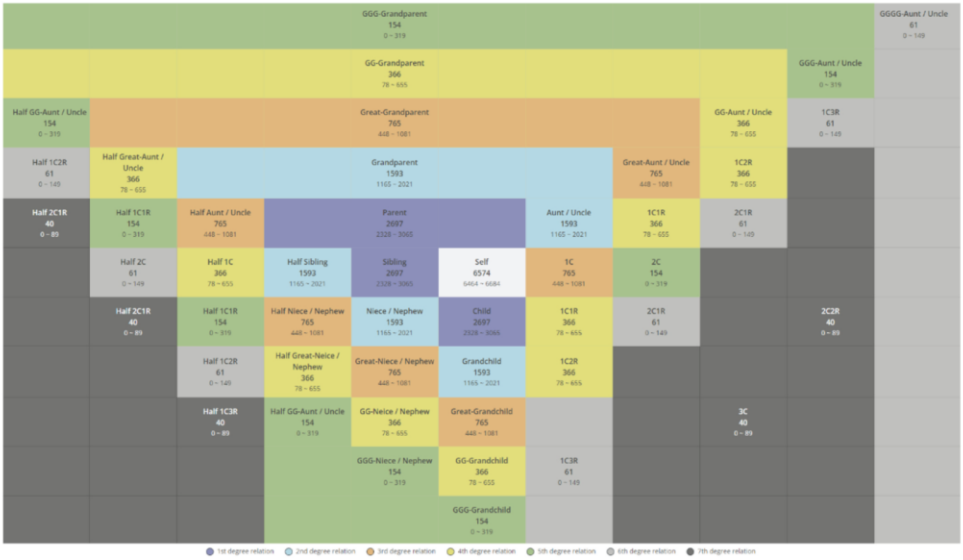


Fig. 13. Familial relationships with shared centimorgan (cM) ranges defined by GEDmatch PRO™ [26].

well as overall heterozygosity. In this study, the contributor status for two female volunteer samples was inconclusive due to the drop-in of two Y SNPs. All remaining samples were determined to be single-source. Of the 34 casework-type and reference-type samples, all 16 of the reference buccal swabs and 13 bone and teeth samples produced high quality profiles with call rates exceeding 95%. The remaining samples of different tissue types had call rates greater than 90% with the exception of the hair sample. Peck et al. [56] tested 11 bone samples ranging from 0.1 to 1.0 ng of input DNA, including from cremated and embalmed remains, and reported high recovery rates of over 96.9% for seven of these samples (63.6%). The lower DNA input samples were more susceptible to locus and allele dropout; however, most bone samples were able to meet the recommended input target. Staadig et al. [58] reported high concordance between bone samples sequenced with the Kintelligence Kit but observed lower call rates than other TAS panels.

While 74% of inferred BGAs were consistent with self-declared ancestries, the use of the PCA in the UAS is limited by the small number of population clusters represented in a two-dimensional PCA. This complicates the inference of BGA from individuals outside of European, Asian and African populations and the interpretation of admixture [61,62]. The predictions derived from the Kintelligence Kit could be improved by testing alternative BGA pipelines which have additional population groups and are able to infer admixture [63]. One such alternative is the Structure model-based likelihood estimator for inferring admixture proportions from ancestral population clusters [20,21].

Where predictions could be made by the UAS, all individuals with self-declared intermediate eye colour were incorrectly predicted as more likely to have brown or blue eye colour. Previous EVC studies have reported substantially lower success rates associated with the correct prediction of intermediate eye colours than that of predicting brown and blue eye colour predictions [64,65]. The majority of inconsistencies for predicting hair colour resulted from predicting brown as blond and dark brown as black. It is possible the incorrect hair colour predictions result from imperfect knowledge of the genetic pathways (and SNPs) associated with melanin expression and/or imprecise self-declaration of hair colour. The UAS also restricts the ability to make EVC predictions from a partial SNP profile.

The Kintelligence Kit genotypes analysed on GEDmatch PRO™

demonstrated that this panel is effective for detecting and predicting kinship out to the fifth degree. Furthermore, all predictions were within 1 degree of the true relationship. Snedecor et al. (2022) demonstrated that locating shared, uninterrupted haplotypes using kinship coefficients in "windows" across the genome was able to correctly identify first through third degree relatives using simulated Kintelligence Kit genotypes uploaded to GEDmatch PRO™, with reasonable performance observed for fourth degree relatives [66]. Kinship inference for a buccal swab was consistent with that for blood and primary tooth from the same donor. Furthermore, upload of the AncestryDNA® profile (containing over 500,000 SNPs), resulted in the same detected relatives with the same degree of kinship. There were no predicted relationships further than one degree from the true relationship and these inaccuracies did not affect first and second degree relationships.

5. Conclusions

The Kintelligence Kit has been validated and optimised for operational use in the AFP Program's accredited forensic laboratory as a new capability for extended kinship analysis, including FIGG, to aid select UHR cases where STR profiling (or other non-genetic identification methods) have not led to an identification. The kit is highly sensitive down to 0.1 ng of DNA input (albeit with some loss of SNP genotypes) and produces high quality SNP profiles from a range of forensic samples with sufficient call rates for bioinformatic analysis using pipelines within the UAS. Several modifications have been made to the throughput and analytical pipelines to improve efficiency in the laboratory and confidence in the resulting genotype. The BGA and EVC pipelines could be further evaluated and compared to existing pipelines outside of the UAS for improved prediction accuracy. The use of the Kintelligence Kit with GEDmatch PRO™ provides Australian law enforcement with an accessible tool for FIGG, with the capacity to detect and accurately predict kinship out to the fifth degree. Additional genetic information can be generated for uniparental lineage using Y SNPs and identity using II SNPs, and this will be the focus of future studies. The Kintelligence Kit presents the AFP Program with a fit for purpose solution for the in-house application of TAS technology for biological sex, BGA, EVCs and extended kinship analysis to advance unresolved unidentified and missing persons cases.

Funding

This research was funded by the Australian Federal Police (AFP) Innovation Fund and the AFP National DNA Program for Unidentified and Missing Persons. Jessica Watson is supported by an Australian Government Research Training Program (RTP) Scholarship.

CRediT authorship contribution statement

Jessica Watson, Jodie Ward, Dennis McNevin: Conceptualization. **Jessica Watson, Jodie Ward, Dennis McNevin, Kelly Grisedale**: Methodology. **Jessica Watson, Dennis McNevin**: Formal analysis. **Jessica Watson**: Writing – original draft presentation. **Jessica Watson, Jodie Ward, Dennis McNevin, Kelly Grisedale, Michelle Spiden, Shelley Seddon**: Writing – review & editing. **Jodie Ward, Dennis McNevin**: Supervision. **Jodie Ward**: Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank: the Australian Facility for Taphonomic Experimental Research (AFTER) donors from the University of Technology Sydney (UTS) Body Donation Program involved in this study for their invaluable contribution to forensic science; the volunteers from the Australian Federal Police (AFP), UTS and their families for providing their DNA and associated meta data for this study; Dr Paul Roffey (AFP), Michelle Peck (Signature Science), Rachel Oefelein (DNA Labs International), Daniela Cuenca (California Department of Justice), Daniel Hellwig (Intermountain Forensics), and Dr Daniel Kling (Oslo University Hospital) for sharing their expertise which guided aspects of this study; Verogen Inc. and QIAGEN for their technical assistance throughout this study; and the Australian law enforcement agencies that authorised the use of specific AFP Program casework samples for this study.

Institutional Review Board Statement

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of University of Technology Sydney (UTS) Human Research Ethics Committee (HREC) (UTS HREC REF NO. ETH21–5821, ETH21–6606, ETH23–8117 and ETH18–2999).

Informed Consent Statement

The cadavers were donated to the Australian Facility for Taphonomic Experimental Research (AFTER) through the UTS Body Donation Program. Written informed consent was obtained from each donor, or from their legal guardians. Written informed consent was obtained from all living sample donors involved in the study.

Supplementary Material

The following supporting information is available: Supplementary Material: Questionnaire – ETH21–5821 Genetic Identification of Unidentified Human Remains; Fig. S1: Allele frequency plots by DNA input amount. The minor allele frequency (MAF) window corresponding with sequencing error (0.00 – 0.05) is highlighted in red, and the ambiguous windows are in orange (0.05 – 0.10, 0.90 – 0.95). Replicates of control DNA NA24385 are indicated by [1] and [2] for each DNA input amount (1.0 ng, 0.5 ng and 0.1 ng).; Table S1: Quantification results and degradation indices (DI) for all casework-type and reference-type samples.; Table S2: Comparison of biological sex predictions from the Universal Analysis System (UAS) and validated pipeline with the self-declared or reported biological sex.; Table S3: The p-values generated by the Universal Analysis System (UAS) for each sample categorised by hair colour. If a prediction could not be generated, the UAS reported the results as inconclusive.; Table S4: The p-values generated by the Universal Analysis System (UAS) for each sample categorised by eye colour. If a prediction could not be generated, the UAS reported the results as inconclusive.; Table S5: Comparison of the biogeographical ancestry predictions with the self-declared ancestries of the donor's self (S), mother (M), father (F), maternal grandfather (MGF), maternal grandmother (MGM), paternal grandfather (PGF) and paternal grandmother (PGM). Refer to Table S6 for ancestry codes used for the declarations.; Table S6: Codes used to declare population and sub-population ancestry.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2023.102972.

References

1 A. Dahal, D. McNevin, M. Chikhani, J. Ward, An Interdisciplinary Forensic Approach for Human Remains Identification and Missing Persons Investigations, Wiley Interdisciplinary Reviews: Forensic Science,, 2023 e1484.
2 Australian Federal Police. National DNA Program for Unidentified and Missing Persons Canberra: Commonwealth of Australia; 2021. Available from: ⟨https://www.missingpersons.gov.au/support/national-dna-program-unidentified-and-missing-persons⟩.
3 J. Ward, A multifaceted, multijurisdictional, multiagency, and multidisciplinary approach to investigating unidentified and missing persons cases in Australia, Forensic Sci. Int Genet Suppl. Ser. 8 (2022) 53–55.
4 Scientific Working Group on DNA Analysis Methods. Recommendations from the SWGDAM Ad Hoc Working Group on Familial Searching. SWGDAM Communications, 2013.
5 M. Kayser, P. De Knijff, Improving human forensics through advances in genetics, genomics and molecular biology, Nat. Rev. Genet 12 (3) (2011) 179–192.
6 R.A. Wickenheiser, Expanding DNA database effectiveness, Forensic Sci. Int Synerg 4 (2022), 100226.
7 R.M. Mateen, M.F. Sabar, S. Hussain, R. Parveen, M. Hussain, Familial DNA analysis and criminal investigation: Usage, downsides and privacy concerns, Forensic Sci. Int 318 (2021), 110576.
8 S.B. Seo, J.L. King, D.H. Warshauer, C.P. Davis, J. Ge, B. Budowle, Single nucleotide polymorphism typing with massively parallel sequencing for human identification, Int J. Leg. Med 127 (6) (2013) 1079–1086.
9 K.M. Canturk, R. Emre, B. Buspinar, M. Ozen, Current status of the use of single-nucleotide polymorphisms in forensic practices, Genet Test. Mol. Biomark. 18 (7) (2014) 455–460.
10 Verogen. High-quality outcomes from low-quality samples with ForenSeq Kintelligence [Application Note]. ForenSeq Kintelligence Kit, 2021.
11 B. Budowle, A. Van Daal, Forensically relevant SNP classes, Biotechniques 44 (5) (2008) 603–610.
12 K.K. Kidd, A.J. Pakstis, W.C. Speed, E.L. Grigorenko, S.L.B. Kajuna, N.J. Karoma, et al., Developing a SNP panel for forensic identification of individuals, Forensic Sci. Int 164 (1) (2006) 20–32.
13 L. Chaitanya, I.Z. Pajnić, S. Walsh, J. Balažic, T. Zupanc, M. Kayser, Bringing colour back after 70 years: Predicting eye and hair colour from skeletal remains of World War II victims using the HIrisPlex system, Forensic Sci. Int Genet 26 (2017) 48–57.
14 L.A. Marano, C. Fridman, DNA phenotyping: current application in forensic science</p>. Res Rep Forensic, Med Sci. 9 (2019) 1–8.
15 S. Walsh, L. Chaitanya, L. Clarisse, L. Wirken, J. Draus-Barini, L. Kovatsi, et al., Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage, Forensic Sci. Int Genet 9 (2014) 150–161.
16 S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, et al., The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA, Forensic Sci. Int Genet 7 (1) (2013) 98–115.
17 J.R. Kidd, F.R. Friedlaender, W.C. Speed, A.J. Pakstis, F.M. De La Vega, K.K. Kidd, Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples, Invest. Genet 2 (1) (2011) 1.
18 J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (9) (2006) 1713–1724.
19 H.A. Gaspar, G. Breen, Probabilistic ancestry maps: a method to assess and visualize population substructures in genetics, BMC Bioinforma. 20 (1) (2019) 1–11.
20 J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2) (2000) 945–959.
21 L. Porras-Hurtado, Y. Ruiz, C. Santos, C. Phillips, A. Carracedo, M. Lareu, An overview of STRUCTURE: Applications, parameter settings, and supporting software, Fron Genet 4 (98) (2013) 1–13.
22 C. Morimoto, S. Manabe, S. Fujimoto, Y. Hamano, K. Tamaki, Discrimination of relationships with the same degree of kinship using chromosomal sharing patterns estimated from high-density SNPs, Forensic Sci. Int Genet 33 (2018) 10–16.
23 Tillmar A. Application of Investigative Genetic Genealogy in Sweden. Human Identification Solutions (2021): Labroots; 2021.
24 E.M. Greytak, C. Moore, S.L. Armentrout, Genetic genealogy for cold case and active investigations, Forensic Sci. Int 299 (2019) 103–113.
25 N. Scudder, J. Robertson, S.F. Kelty, S.J. Walsh, D. McNevin, An international consideration of a standards-based approach to forensic genetic genealogy, Forensic Sci. Int Genet Suppl. Ser. 7 (1) (2019) 512–514.
26 GEDmatch P.R.O. GEDmatch P.R.O. 2022. Available from: ⟨https://pro.gedmatch.com/⟩.
27 Kumar S.A. GEDmatch: A Platform for Data Driven Forensic Intelligence, 2020. Available from: ⟨https://promega.foleon.com/theishireport/november-2020/gedmatch-a-platform-for-data-driven-forensic-intelligence/⟩.
28 D. Kling, C. Phillips, D. Kennett, A. Tillmar, Investigative genetic genealogy: Current methods, knowledge and practice, Forensic Sci. Int Genet 52 (2021), 102474.
29 J. Byrnes, N. Myers, P. Underhill, Genetic genealogy in the genomic era, in: D. Primorac, M. Schanfield (Eds.), Forensic DNA Applications: An Interdisciplinary Perspective, CRC Press, 2014, p. 483.
30 D. Kling, On the use of dense sets of SNP markers and their potential in relationship inference, Forensic Sci. Int Genet 39 (2019) 19–31.
31 Verogen. MiSeq FGx Sequencing System Reference Guide. MiSeq FGx Sequencing System, 2021.
32 Verogen. Data Sheet: ForenSeq Kintelligence Kit. ForenSeq Kintelligence Kit, 2021.
33 Verogen. Data Sheet: ForenSeq DNA Signature Prep Kit. ForenSeq DNA Signature Prep Kit, 2018.
34 Verogen. S.N.P. Typing in Universal Analysis Software and Kinship Estimation with GEDmatch PRO [Technical Note]. ForenSeq Kintelligence Kit, 2021.
35 Verogen. Verogen and Gene by Gene Form Groundbreaking Partnership to Accelerate Adoption of Forensic Investigative Genetic Genealogy, 2022. Available from: ⟨https://verogen.com/verogen-and-gene-by-gene-form-groundbreaking-partnership-to-accelerate-adoption-of-forensic-investigative-genetic-genealogy/⟩.
36 QIAGEN. QIAamp® DNA Investigator Handbook, 2020.
37 QIAGEN. EZ1&2 DNA Investigator Kit, 2022.
38 QIAGEN. MinElute® Handbook, 2020.
39 O.M. Loreille, T.M. Diegoli, J.A. Irwin, M.D. Coble, T.J. Parsons, High efficiency DNA extraction from bone by total demineralization, Forensic Sci. Int Genet 1 (2) (2007) 191–195.
40 Sigma-Aldrich. Amicon® Ultra-15 Centrifugal Filters 30 kDa MWCO Millipore, 2023. Available from: ⟨https://www.sigmaaldrich.com/AU/en/product/mm/ufc9030⟩.
41 Applied Biosystems. Quantifiler™ HP and Trio DNA Quantification Kits User Guide [User Guide]. Thermo Fisher Scientific, 2017.
42 Promega. QuantiFluor® ONE dsDNA System, 2022.
43 Verogen. ForenSeq Kintelligence Kit Reference Guide, ForenSeq Kintelligence Kit, 2021.
44 Department of Genetic Identification of Erasmus MC HIrisPlex-S Eye, Hair and Skin Colour DNA Phenotyping Webtool, 2022. Available from: ⟨https://hirisplex.erasmuscmc.nl⟩.
45 Scientific Working Group on DNA Analysis Methods, Validation Guidelines for DNA Analysis Methods, 2016.
46 Scientific Working Group on DNA Analysis Methods. Addendum to "SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories" to Address Next Generation Sequencing, 2019.
47 H. Wickham, R. François, L. Henry, K. Müller, dplyr: A Grammar of Data Manipulation. R package version 1.0.2, 2020. URL: ⟨https://CRAN.R-project.org/package=dplyr⟩.
48 Wickham H. , Chang W. , Henry L. , Pederson T.L. , Takahashi K. , Wilke C. , et al. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics, 2023. URL: ⟨h⟩⟨ttps://cran.r-project.org/web/packages/ggplot2/index.html⟩.
49 R. Core Team, R: A language and environment for statistical computing, Vienna, Austria, 2020. URL: ⟨https://www.R-project.org/⟩.
50 X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, et al., pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinforma. 12 (1) (2011) 1–8.
51 Verogen. Universal Analysis Software v2.0, and above, Reference Guide, Univers. Anal. Softw. (2021).
52 Applied Biosystems. GlobalFiler PCR Amplification Kits. Thermo Fisher Scientific, 2019.
53 Bettinger B.T. The Shared cM Project 4.0 tool v4: DNA Painter; 2020. Available from: ⟨h⟩⟨ttps://dnapainter.com/tools/sharedcmv4⟩.
54 J. Antunes, S. Radecke, J. Snedecor, G. Padmabandu, K.M. Stephens, Application of a dense SNP PCR multiplex for high throughput kinship determination using Next Generation Sequencing (NGS), Forensic Sci. Int Genet Suppl. Ser. 8 (2022) 324–325.
55 Antunes J., Walichiewicz P., Forouzmand E., Barta R., Didier M., Han Y., et al. Developmental Validation of the ForenSeq® Kintelligence Kit, MiSeq Fgx® Sequencing System and ForenSeq Universal Analysis Software. Preprint (SSRN, June 6, 2023). Available from: ⟨https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4466078⟩ [Last accessed: September 25, 2023].
56 Peck M.A., Koeppel A.F., Gorden E.M., Bouchet J.L., Heaton M.C., Russell D.A., et al. Internal Validation of the ForenSeq Kintelligence Kit for Application to Forensic Genetic Genealogy. Forensic Genom, 2022.
57 Cuenca D. An improved method for forensic genetic genealogy in an operational crime lab. Xtalks: Verogen; 2021.
58 Staadig A., Sidstedt M., Tillmar A. ForenSeq® Kintelligence A Targeted DNA Sequencing Assay for Forensic Genetic Genealogy. Rättsmedicinalverket National Board of Forensic Medicine: Linköping University, Polisen Swedish Police; 2021.
59 G. Chen, S. Mosier, C.D. Gocke, M.-T. Lin, J.R. Eshleman, Cytosine deamination is a major cause of baseline noise in next-generation sequencing, Mol. Diagn. Ther. 18 (2014) 587–593.
60 R. Mitchell, S. Enke, K. Eskey, T. Ferguson, R. Just, A method to enable forensic genetic genealogy investigations from DNA mixtures, Forensic Sci. Int Genet Suppl. Ser. 8 (2022) 159–161.
61 E.Y. Cheung, M.E. Gahan, D. McNevin, Prediction of biogeographical ancestry in admixed individuals, Forensic Sci. Int Genet 36 (2018) 104–111.
62 E.Y. Cheung, M.E. Gahan, D. McNevin, Predictive DNA analysis for biogeographical ancestry, Aust. J. Forensic Sci. 50 (6) (2018) 651–658.
63 D. McNevin, Forensic Inference of Biogeographical Ancestry from Genotype: The Genetic Ancestry Lab, Wiley Interdisciplinary Reviews: Forensic Science,, 2020 e1356.
64 N.M. Salvo, K. Janssen, M.K. Kirsebom, O.S. Meyer, T. Berg, G.-H. Olsen, Predicting eye and hair colour in a Norwegian population using Verogen's ForenSeqTM DNA Signature Prep Kit, Forensic Sci. Int Genet (2021), 102620.
65 S. Walsh, F. Liu, K.N. Ballantyne, M. van Oven, O. Lao, M. Kayser, IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information, Forensic Sci. Int Genet 5 (3) (2011) 170–180.
66 J. Snedecor, T. Fennell, S. Stadick, N. Homer, J. Antunes, K. Stephens, et al., Fast and accurate kinship estimation using sparse SNPs in relatively large database searches, Forensic Sci. Int Genet 61 (2022), 102769.

4.1 Supplementary Material

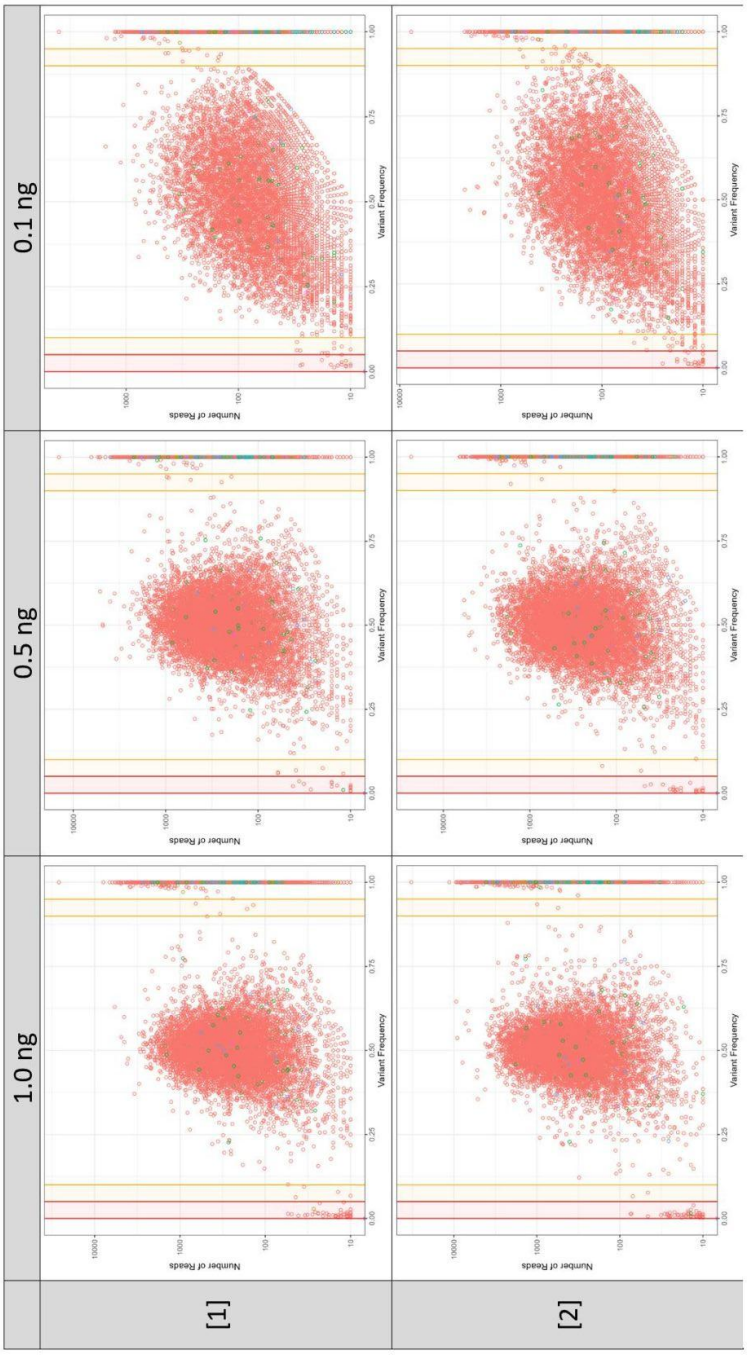


Figure S1. Allele frequency plots by DNA input amount. The minor allele frequency (MAF) window corresponding with sequencing error (0.00 – 0.05) is highlighted in red, and the ambiguous windows are in orange (0.05 – 0.10, 0.90 – 0.95). Replicates of control DNA NA24385 are indicated by [1] and [2] for each DNA input amount (1.0 ng, 0.5 ng and 0.1 ng).

Table S1. Quantification results and degradation indices (DI) for all casework-type and reference-type samples.

Sample	Small Autosomal Target Concentration (ng/μL)	Large Autosomal Target Concentration (ng/μL)	Male DNA Target Concentration (ng/μL)	Degradation Index (DI)
Family 1	2.895	3.588	0.000	0.81
Family 2	0.020	0.006	0.000	3.33
Family 3	0.043	0.037	0.043	1.15
Family 4	0.726	0.443	0.461	1.64
Family 5	0.297	0.203	0.293	1.46
Family 6	0.221	0.095	0.000	2.32
Family 7	0.066	0.058	0.072	1.14
Family 8	6.262	5.388	0.000	1.16
Family 9	3.907	3.996	0.000	0.98
Family 10	3.137	4.758	0.000	0.66
Unrelated 1	14.410	16.208	0.000	0.89
Unrelated 2	2.322	3.564	0.000	0.65
Unrelated 3	1.859	1.939	0.000	0.96
Unrelated 4	3.978	3.053	0.000	1.30
Unrelated 5	9.525	9.913	0.000	0.96
Unrelated 6	1.813	2.747	0.000	0.66
Hair 1	0.000	0.000	0.000	N/A[1]
Blood 1	0.312	0.293	0.000	1.06
Blood 2	8.500	14.400	0.000	0.59
Tooth 1	0.038	0.007	0.000	5.79
Tooth 2	0.408	0.379	0.000	1.08
Bone 1	0.327	0.322	0.286	1.01
Bone 2	0.037	0.016	0.038	2.28
Bone 3	0.016	0.006	N/A[2]	2.67
Bone 4	5.372	1.08	5.647	4.96
Bone 5	0.705	0.009	0.650	82.94
Bone 6	0.548	0.277	0.000	1.98
Bone 7	0.703	0.131	0.000	5.36
Bone 8	0.305	0.133	0.000	2.30
Bone 9	0.314	0.065	0.393	4.85
Bone 10	0.151	0.085	0.169	1.78
Bone 11	0.972	0.553	0.941	1.76
Nail 1	1.366	0.956	0.982	1.43
Nail 2	0.939	0.747	1.072	1.26

[1] DNA was not detected, therefore DI could not be calculated.

[2] Male DNA quantification data unavailable.

Table S2. Comparison of biological sex predictions from the Universal Analysis System (UAS) and validated pipeline with the self-declared or reported biological sex.

Sample	Biological Sex	UAS Pipeline		Validated Pipeline	
		Prediction	Comment	Prediction	Comment
PC 1	Male	Male	Consistent	Male	Consistent
PC 2	Male	Male	Consistent	Male	Consistent
PC 3	Male	Male	Consistent	Male	Consistent
PC 4	Male	Male	Consistent	Male	Consistent
PC 5	Male	Male	Consistent	Male	Consistent
PC 6	Male	Male	Consistent	Male	Consistent
PC 7	Male	Male	Consistent	Male	Consistent
PC 8	Male	Male	Consistent	Male	Consistent
PC 9	Male	Male	Consistent	Male	Consistent
PC 10	Male	Male	Consistent	Male	Consistent
NA24385 0.5ng (1)	Male	Male	Consistent	Male	Consistent
NA24385 0.5ng (2)	Male	Male	Consistent	Male	Consistent
NA24385 0.1ng (1)	Male	Male	Consistent	Male	Consistent
NA24385 0.1ng (2)	Male	Male	Consistent	Male	Consistent
NA24385 0.1ng (3)	Male	Male	Consistent	Male	Consistent
Family 1	Female	Female	Consistent	Female	Consistent
Family 2	Female	Inconclusive	-	Female	Consistent
Family 3	Male	Male	Consistent	Male	Consistent
Family 4	Male	Male	Consistent	Male	Consistent
Family 5	Male	Male	Consistent	Male	Consistent
Family 6	Female	Female	Consistent	Female	Consistent
Family 7	Male	Male	Consistent	Male	Consistent
Family 8	Female	Female	Consistent	Female	Consistent
Family 9	Female	Inconclusive	-	Female	Consistent
Family 10	Female	Female	Consistent	Female	Consistent
Unrelated 1	Female	Female	Consistent	Female	Consistent
Unrelated 2	Female	Female	Consistent	Female	Consistent
Unrelated 3	Female	Female	Consistent	Female	Consistent
Unrelated 4	Female	Female	Consistent	Female	Consistent
Unrelated 5	Female	Female	Consistent	Female	Consistent
Unrelated 6	Female	Female	Consistent	Female	Consistent
Blood 1	Female	Female	Consistent	Female	Consistent
Blood 2	Female	Female	Consistent	Female	Consistent
Tooth 1	Female	Female	Consistent	Female	Consistent
Tooth 2	Female	Female	Consistent	Female	Consistent
Bone 1	Male	Male	Consistent	Male	Consistent
Bone 2	Male	Male	Consistent	Male	Consistent
Bone 4	Male	Male	Consistent	Male	Consistent

Bone 5	Male	Male	Consistent	Male	Consistent
Bone 6	Female	Female	Consistent	Female	Consistent
Bone 7	Female	Female	Consistent	Female	Consistent
Bone 8	Female	Female	Consistent	Female	Consistent
Bone 9	Male	Male	Consistent	Male	Consistent
Bone 10	Male	Male	Consistent	Male	Consistent
Bone 11	Male	Male	Consistent	Male	Consistent
Nail 1	Male	Male	Consistent	Male	Consistent
Nail 2	Male	Male	Consistent	Male	Consistent

Table S3. The p-values generated by the Universal Analysis System (UAS) for each sample categorised by hair colour. If a prediction could not be generated, the UAS reported the results as inconclusive.

Sample	Brown	Red	Black	Blond
Family 1	0.90	0.04	0.05	0.01
Family 2	Inconclusive	Inconclusive	Inconclusive	Inconclusive
Family 3	0.83	0.09	0.02	0.06
Family 4	0.53	0.00	0.10	0.36
Family 5	0.64	0.01	0.09	0.27
Family 6	0.14	0.12	0.01	0.73
Family 7	0.10	0.11	0.00	0.78
Family 8	0.35	0.00	0.52	0.12
Family 9	0.61	0.00	0.12	0.27
Family 10	0.61	0.00	0.29	0.09
Unrelated 1	0.22	0.00	0.77	0.01
Unrelated 2	0.67	0.21	0.04	0.08
Unrelated 3	0.11	0.00	0.89	0.00
Unrelated 4	0.09	0.00	0.90	0.00
Unrelated 5	0.18	0.00	0.82	0.00
Unrelated 6	0.15	0.00	0.01	0.84
Blood 1	0.90	0.04	0.05	0.01
Tooth 1	Inconclusive	Inconclusive	Inconclusive	Inconclusive
Tooth 2	0.11	0.00	0.01	0.88

Table S4. The p-values generated by the Universal Analysis System (UAS) for each sample categorised by eye colour. If a prediction could not be generated, the UAS reported the results as inconclusive.

Sample	Intermediate	Brown	Blue
Family 1	0.21	0.63	0.16
Family 2	Inconclusive	Inconclusive	Inconclusive
Family 3	0.07	0.02	0.91
Family 4	0.11	0.21	0.68
Family 5	0.12	0.83	0.05
Family 6	0.07	0.04	0.89
Family 7	0.04	0.02	0.94
Family 8	0.05	0.94	0.01
Family 9	0.14	0.76	0.10
Family 10	0.19	0.72	0.09
Unrelated 1	0.01	0.99	0.00
Unrelated 2	0.22	0.66	0.12
Unrelated 3	0.00	1.00	0.00
Unrelated 4	0.00	1.00	0.00
Unrelated 5	0.00	1.00	0.00
Unrelated 6	0.05	0.03	0.92
Blood 1	0.19	0.72	0.09
Tooth 1	Inconclusive	Inconclusive	Inconclusive
Tooth 2	0.05	0.03	0.92

Table S5. Comparison of the biogeographical ancestry predictions with the self-declared ancestries of the donor's self (S), mother (M), father (F), maternal grandfather (MGF), maternal grandmother (MGM), paternal grandfather (PGF) and paternal grandmother (PGM). Refer to Table S6 for ancestry codes used for the declarations.

Sample	Excluded Populations	Included Populations	Declarations							Comments
			S	*M*	*F*	*MGF*	*MGM*	*PGF*	*PGM*	
Family 1	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	Consistent
Family 2	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	Consistent
Family 3	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	ANC1	AUS8	AUS8	Consistent
Family 4	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	ANC1	AUS8	AUS8	Consistent
Family 5	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	Consistent
Family 6	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	ANC1	AUS8	AUS8	Consistent
Family 7	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	Consistent
Family 8	Europe East Asia Africa		SEA4	AUS8	SEA4	ANC1	ANC1	SEA4	NWE2	Inconclusive
Family 9	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	Consistent
Family 10	East Asia Africa	Europe	AUS8		AUS8			SEU1	AUS8	Consistent
Unrelated 1	East Asia Africa Europe		CAS2	CAS2	CAS2	CAS2	CAS2	CAS2	CAS2	Inconclusive
Unrelated 2	East Asia Africa	Europe	AUS5	AUS8	MEA1	ANC3	ANC3	MEA1	SEU1	Partially consistent, the European contribution to admixture was predicted
Unrelated 3	Europe Africa	East Asia	NEA1	NEA1	NEA1	NEA1	NEA1	NEA1	NEA1	Consistent
Unrelated 4	Europe Africa	East Asia	SEA2	SEA2	SEA2	SEA2	SEA2	SEA2	SEA2	Consistent
Unrelated 5	East Asia Africa Europe		AUS14	OCE1	OCE1	OCE1	OCE1	OCE1	OCE1	Inconclusive
Unrelated 6	East Asia Africa	Europe	AUS14	AUS1	AUS8	AUS8	AUS1	ANC3	ANC1	Partially consistent, the European contribution to admixture was predicted
Blood 1	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	Consistent
Tooth 1	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	Consistent
Tooth 2	East Asia Africa	Europe	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	AUS8	Consistent

Table S6. Codes used to declare population and sub-population ancestry.

Population	Sub-Population	Code
Sub-Saharan African	South Africa	SSA1
	Zimbabwe	SSA2
	Mauritius	SSA3
	Kenya	SSA4
	Ethiopia	SSA5
	Nigeria	SSA6
	Somalia	SSA7
	Gambia	SSA8
	Ghana	SSA9
	Other (specify)	SSA10
North African	Egypt	NAF1
	Sudan	NAF2
	Libya	NAF3
	Morocco	NAF4
	Algeria	NAF5
	Tunisia	NAF6
	Other (specify)	NAF7
Middle Eastern	Lebanon	MEA1
	Iraq	MEA2
	Iran	MEA3
	Turkey	MEA4
	Syria	MEA5
	Saudi	MEA6
	Israel	MEA7
	United Arab Emirates	MEA8
	Libya	MEA9
	Other (specify)	MEA10
Southern European	Italy	SEU1
	Greece	SEU2
	Malta	SEU3
	Cyprus	SEU4
	Portugal	SEU5
	Spain	SEU6
	Gibraltar	SEU7
	Other (specify)	SEU8
Eastern European	Poland	EUE1
	Croatia	EUE2
	Macedonia	EUE3
	Bosnia/Herzegovina	EUE4
	Russia	EUE5
	Serbia	EUE6
	Hungary	EUE7
	Romania	EUE8
	Ukraine	EUE9
	Czech	EUE10
	Slovenia	EUE11
	Slovakia	EUE12
	Latvia	EUE13
	Bulgaria	EUE14
	Albania	EUE15
	Lithuania	EUE16
	Estonia	EUE17

	Belarus	EUE18
	Montenegro	EUE19
	Other (specify)	EUE20
Anglo/Celtic	England	ANC1
	Scotland	ANC2
	Ireland	ANC3
	Wales	ANC4
	Channel Islands	ANC5
	Jersey	ANC6
	Isle of Man	ANC7
	Other (specify)	ANC8
North/West European	Germany	NWE1
	Netherlands	NEW2
	France	NWE3
	Austria	NWE4
	Switzerland	NWE5
	Sweden	NWE6
	Denmark	NWE7
	Finland	NWE8
	Belgium	NWE9
	Norway	NWE10
	Other (specify)	NWE11
South Asian	India	SAS1
	Sri Lanka	SAS2
	Bangladesh	SAS3
	Maldives	SAS4
	Other (specify)	SAS5
Central Asian	Pakistan	CAS1
	Nepal	CAS2
	Afghanistan	CAS3
	Bhutan	CAS4
	Kazakhstan	CAS5
	Uzbekistan	CAS6
	Armenia	CAS7
	Azerbaijan	CAS8
	Georgia	CAS9
	Other (specify)	CAS10
North/East Asian	China/Taiwan	NEA1
	Korea	NEA2
	Japan	NEA3
	Mongolia	NEA4
	Other (specify)	NEA5
South East Asian	Philippines	SEA1
	Vietnam	SEA2
	Malaysia	SEA3
	Indonesia	SEA4
	Thailand	SEA5
	Singapore	SEA6
	Cambodia	SEA7
	Myanmar	SEA8
	Laos	SEA9
	Other (specify)	SEA10
Oceanian	Fiji	OCE1
	Papua New Guinea	OCE2

	Samoa	OCE3
	Tonga	OCE4
	Cook Islands	OCE5
	Solomon Islands	OCE6
	New Caledonia	OCE7
	Vanuatu	OCE8
	Māori	OCE9
	Other (specify)	OCE10
American	American European	AME1
	American African	AME2
	American Hispanic	AME3
	Brazil	AME4
	Chile	AME5
	Columbia	AME6
	Argentina	AME7
	El Salvador	AME8
	Peru	AME9
	Uruguay	AME10
	Venezuela	AME11
	Mexico	AME12
	Ecuador	AME13
	West Index	AME14
	Other (specify)	AME15
Australasian	Australian Aboriginal	AUS1
	Torres Strait Island	AUS2
	Australian African	AUS3
	Australian North African	AUS4
	Australian Middle Eastern	AUS5
	Australian Southern European	AUS6
	Australian Eastern European	AUS7
	Australian Anglo/Celtic	AUS8
	Australian North/West European	AUS9
	Australian South Asian	AUS10
	Australian Central Asian	AUS11
	Australian North/East Asian	AUS12
	Australian South East Asian	AUS13
	Australian Oceanian	AUS14
	Australian American	AUS15
	Other (specify)	AUS16

5. CHAPTER FIVE

CHARACTERISATION OF IDENTITY-INFORMATIVE GENETIC MARKERS IN THE AUSTRALIAN POPULATION WITH EUROPEAN ANCESTRY

Contributions of authors:

Watson, JL, Cho K, Grisedale K, Ward J and McNevin D. (2024) 'Characterisation of Identity-Informative Genetic Markers in the Australian Population with European Ancestry.' Forensic Science International: Genetics (2024): 103169. DOI: 10.1016/j.fsigen.2024.103169.

Jessica Watson (Candidate)

Conceptualised and assisted in the experimental design for the study. Carried out sample collection, sample preparation, DNA testing and data analysis. Prepared manuscript, edited manuscript following review by other co-authors, referenced manuscript, carried out manuscript's submission and incorporated feedback from peer review process.

Signed: Production Note: Signature removed prior to publication.

Kaymann Cho (AFP Biology staff)

Contributed to sample preparation and DNA testing. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Kelly Grisedale (AFP DNA Program staff)

Assisted in the experimental design for the study. Contributed to DNA testing and data analysis. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Jodie Ward (Co-supervisor)

Conceptualised and assisted in the experimental design for the study. Acquired funding for the study. Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

Dennis McNevin (Co-Supervisor)

Conceptualised and assisted in the experimental design for the study. Contributed to data analysis. Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

5.1 Characterisation of Identity-Informative Genetic Markers in the Australian Population with European Ancestry

Forensic Science International: Genetics 74 (2025) 103169

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Characterisation of identity-informative genetic markers in the Australian population with European ancestry

Jessica L. Watson [a,b,c,*], Kaymann Cho [c], Kelly Grisedale [a,c], Jodie Ward [a,b], Dennis McNevin [a,b]

[a] National DNA Program for Unidentified and Missing Persons, Australian Federal Police, Australia
[b] Centre for Forensic Science, School of Mathematical & Physical Science, Faculty of Science, University of Technology Sydney, Australia
[c] Biology, AFP Forensics, Australian Federal Police, Australia

ARTICLE INFO

Keywords:
Single nucleotide polymorphism
Allele frequency
Australia
Human identification
Massively parallel sequencing
Short tandem repeat

ABSTRACT

Identity-informative single nucleotide polymorphisms (iiSNPs) are valuable genetic markers for human identification and kinship testing in forensic casework, especially when the quality and quantity of DNA evidence is not suitable for routine short tandem repeat (STR) profiling. This study analysed 105 buccal samples representing the Australian population with European ancestry in order to assign allele frequencies and conduct population genetic analyses for 94 iiSNPs and 20 STRs. The markers were assessed by calculating relevant forensic statistics and testing for deviations from Hardy-Weinberg and linkage equilibrium. No linkage of statistical significance was observed between any of the pair-wise combinations of the combined 114 identity-informative markers and only one STR exhibited deviation from Hardy-Weinberg equilibrium (D8S1179). The probability of matching genotypes being observed within this population was of the order of 10^{-23} for STRs, 10^{-38} for iiSNPs and 10^{-60} for the combined identity-informative marker panel, improving the ability to discriminate between individuals when calculating likelihood ratios in direct or indirect matching scenarios. Further, the addition of iiSNPs will facilitate identifications when suboptimal STR profiles are recovered from compromised or challenging samples and aid comparisons to genetic relatives for familial or kinship testing.

1. Introduction

First introduced in the 1990s, short tandem repeats (STRs) are segments of repeated DNA motifs consisting of two to six bases dispersed throughout the genome [1]. STRs located on the autosomal chromosomes are the most common genetic marker currently targeted for forensic DNA profiling applications [2]. Differentiation between individuals is made possible by the combination of alleles inherited from each biological parent, with each allele defined by the number of times the DNA motif is repeated within an STR [3,4]. The combination of multiple STRs within one profile increases the discrimination power and uniqueness of the profile. Their highly polymorphic nature is due in some part to their high mutation rates of the order of 10^{-3} per meiosis [3].

STR profiling is the gold standard forensic genetic method for human identification and is typically used in criminal and coronial investigations to identify a person of interest [5]. In the majority of jurisdictions, STRs have been the only genetic marker acknowledged by the Court as a sound method of DNA profiling and able to be used as evidence [6]. As a result, law enforcement databases have been populated with evidentiary and reference STR profiles [7]. STRs are also used routinely for the identification of human remains in coronial investigations, incorporating missing persons and disaster victim identification (DVI) efforts [8,9].

Single nucleotide polymorphisms (SNPs) are single base genetic variants [10]. In the human genome, the average person will have approximately 5 million SNPs [2,11]. For the last 20 years, SNPs have been investigated as an alternative marker to STRs, but SNP genotyping is yet to become common practice in forensic genetic laboratories [12–14]. As biallelic genetic markers such as SNPs are less polymorphic, their discrimination power is significantly lower than that for STRs [9,15]. In order to produce profiles with a similar discrimination power, larger SNP panels are required [16]. Kidd et al. proposed that at least 45 SNPs would have the equivalent discrimination power of the 13 CODIS STRs [15]. However, the mutation rates for autosomal SNPs are approximately 100,000 times lower than in STRs (10^{-8}, compared to 10^{-3}), making SNP genotypes more stable than STRs across multiple generations and thereby reducing the risk of mutations confounding

* Corresponding author at: National DNA Program for Unidentified and Missing Persons, Australian Federal Police, Australia.
E-mail address: Jessica.Watson@student.uts.edu.au (J.L. Watson).

https://doi.org/10.1016/j.fsigen.2024.103169
Received 5 May 2024; Received in revised form 10 October 2024; Accepted 27 October 2024
Available online 28 October 2024
1872-4973/Crown Copyright © 2024 Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

typing [13,15]. SNPs are less prone to degradation due to their shorter amplicon sizes compared to STRs.

The most common genotyping technology currently available for STRs involves fragment length analysis. The targeted DNA regions are amplified with primers containing fluorescent dye labels. When the DNA fragments are separated by size during capillary electrophoresis (CE), fluorescent imaging generates an electropherogram consisting of fluorescent signals (peaks) representing the alleles within each dye channel [17,18]. Whereas CE only provides information on the size of DNA fragments, massively parallel sequencing (MPS) is capable of determining the actual DNA sequence [19]. MPS is a genotyping capability that sequences millions of DNA fragments from multiple samples in one sequencing run. This technology can be applied to both STR and SNP genotyping.

Identity-informative SNPs (iiSNPs) are a category of SNP that have characteristics most beneficial for individualisation of a genotype. iiSNPs require high heterozygosity and should also ideally have low allele frequency heterogeneity to minimise the difference in allele frequencies between populations [20]. SNP allele frequency databases have currently been developed for population and subpopulation groups in Europe, Asia, North America and South America [21–29].

Allele frequency databases are important to develop as these frequencies will tend to vary in different populations as a result of genetic drift [30]. The allele frequencies, reflecting the genetic diversity within a population, can be used to determine the random match probability (RMP) of a particular genotype in a forensic investigation [4]. This can be used to calculate a likelihood ratio (LR), the most commonly employed statistical method to compare an unknown and known DNA profile [31]. This is a ratio of two conditional probabilities for the same observations under alternative hypotheses [32]. It is also possible to calculate a combined LR by combining multiple DNA marker panels [33,34].

In order to be suitable for forensic use, identity-informative markers should: 1) be in Hardy-Weinberg equilibrium (HWE) to ensure that genotype frequencies can be inferred from allele frequencies (within locus independence); 2) be in linkage equilibrium (LE) to ensure that locus genotypes are independently inherited and that LRs from individual loci can be multiplied together (between locus independence); 3) have high heterozygosity to increase the discrimination power of the panel; and 4) the first three conditions should apply across subpopulations [35]. If a locus has significant deviation from HWE, it means a process is influencing the distribution of alleles and genotype frequencies within a population (e.g. inbreeding, hidden population structures, natural selection) [36]. LE tests assess the probability that the alleles of any two loci are inherited independently as a result of recombination and are usually influenced by the physical proximity between the pair [37].

Taylor et al. (2017) published STR allele frequencies in Australian and New Zealand populations for the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA, USA), including those whose members declared Aboriginal, European and Asian ancestries [30]. However, there has been no published study on the suitability of a SNP panel for population groups relevant to Australia or establishment of a SNP allele frequency database for this region to date. Furthermore, studies have primarily examined each class of identity marker separately and not assessed the LE between SNPs and STRs or the power of discrimination for a profile containing both marker types [21,23,33,38]. In this study, we examine the suitability of the 94 iiSNPs included in the ForenSeq® DNA Signature Prep Kit (Verogen, Inc., San Diego, CA, USA; now a QIAGEN company) and the ForenSeq® Kintelligence Kit (Verogen, Inc.) for use in the Australian population with European ancestry.

2. Methods

2.1. Ethics approval and sample procurement

Ethics approval for this research was granted by the University of Technology Sydney (UTS) Human Research Ethics Committee (HREC) (UTS HREC NO. ETH21-5821 and amendments ETH21–6606 and ETH23-8117 relate). All volunteers provided a buccal swab with informed consent. A questionnaire was completed to provide self-declared biogeographical ancestry (BGA) for each participant, as well as their parents and grandparents. A total of 105 volunteers with self-declared Australian European ancestry provided self-administered buccal swabs.

Additional casework-type samples (two teeth and eight bones) were sourced from the Australian Facility for Taphonomic Experimental Research with ethics approval (UTS HREC NO. ETH18–2999) and approved research samples submitted to the Australian Federal Police National DNA Program for Unidentified and Missing Persons.

2.2. Sample preparation

DNA from the buccal swabs was manually extracted using the EZ1® DNA Investigator Kit (QIAGEN, Hilden, Germany) [39]. For the bone and tooth samples, 500 mg of pulverised powder underwent total demineralisation lysis, concentration using the Amicon® 30 K Ultra Centrifugal Filter (Sigma-Aldrich, St. Louis, MO, USA) and extraction with the MinElute® PCR Purification Kit (QIAGEN). Samples were quantified with the Quantifiler™ Trio DNA Quantification Kit (Thermo Fisher Scientific) [40] on a QuantStudio™ 5 Real-time PCR System (Thermo Fisher Scientific) [41]. All protocols were performed according to the manufacturers' recommended protocols unless otherwise specified.

2.3. Library preparation and sequencing

There were two MPS panels utilised in this study; the ForenSeq® DNA Signature Prep Kit and the ForenSeq® Kintelligence Kit, each with the same 94 iiSNPs [42,43]. For both kits, the recommended DNA input for library preparation is 1.0 ng. For samples that required dilution and were degraded (with a degradation index (DI) greater than 1), the large autosomal (LA) target concentration was used to determine DNA input. If the samples were not degraded (with a DI equal to or less than 1), the small autosomal (SA) target concentration was used to avoid over-diluting the DNA extract.

For 99 buccal swab samples, libraries were prepared using the ForenSeq® DNA Signature Prep Kit with primer mix B according to either the manufacturer's recommended protocol (n = 42) [42] or an automated library preparation method utilising a quantitative polymerase chain reaction (qPCR) normalisation protocol (n = 57) [44]. Sequencing was performed on the MiSeq® FGx Sequencing System (Verogen, Inc.) using the MiSeq® FGx Reagent Kit (Verogen, Inc.) and standard flow cell according to the manufacturer's recommended protocol [45]. Each ForenSeq® DNA Signature Prep sequencing batch consisted of a positive control (2800 M), negative control and 14 samples. Different index combinations were used on successive sequencing runs to limit sample cross-contamination between batches. The profiles were analysed on the Universal Analysis Software v1.3 (UAS; Verogen, Inc.) using the default analytical and interpretation thresholds and the STR and SNP genotypes exported in Sample Details Reports [46].

For the remaining six buccal swabs and 10 casework-type samples, libraries were prepared following a modified protocol for the ForenSeq® Kintelligence Kit [47]. Sequencing was performed on the MiSeq® FGx Sequencing System using the MiSeq® FGx Reagent Kit and standard flow cell according to the manufacturer's recommended protocol [45]. The ForenSeq® Kintelligence Kit sequencing batches consisted of three samples per flow cell, with a positive control (NA24385) and negative control for each library preparation batch of 12 libraries. The SNP profiles were exported in Sample Reports from the UAS v2.5 and analysed according to the optimised thresholds published by Watson *et al.* (2023) [47,48].

2.4. STR profiling

For buccal swab samples (n = 6) and casework-type samples (n = 10) sequenced with the ForenSeq® Kintelligence Kit, STR profiles were generated using the GlobalFiler™ PCR Amplification Kit [49]. The buccal swab samples were amplified on the Veriti™ 96-Well Fast Thermal Cycler (Thermo Fisher Scientific) with a 29 cycle amplification using the LA target for DNA input. The casework-type samples were amplified with a 30 cycle method. Capillary electrophoresis was performed on the 3500xL Genetic Analyser (Thermo Fisher Scientific) according to the manufacturer's recommended protocols [50]. The resulting electropherogram profiles were analysed using the GeneMapper™ *ID-X* v1.6 software. To distinguish between background noise and the detected peaks in the DNA profile, the analytical threshold was set to 250 relative fluorescence units (RFU) for the buccal swab samples amplified at 29 cycles and 60 RFU for the casework-type samples amplified at 30 cycles. For both amplification methods, the homozygous threshold was set to 1000 RFU.

2.5. Statistical analysis

Statistical analysis was performed on the iiSNPs and a subset of STRs, consisting of the 20 STRs in common between the GlobalFiler™ and ForenSeq® DNA Signature Prep panels. This included D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, D10S1248, D1S1656, D12S391 and D2S1338. Fig. 1 shows the centimorgan (cM) positions of the STRs and SNPs on each chromosome.

Forensic efficiency parameters and allele frequencies for the iiSNP genotypes were generated using the STR Analysis for Forensics (STRAF) v2.1.5 program [51]. This included calculating the probability of matching (PM), polymorphism information content (PIC), expected heterozygosity (H_{exp}), observed heterozygosity (H_{obs}), power of exclusion (PE), typical paternity index (TPI) and power of discrimination (PD).

The number of possible genotypes (PG) for each locus was calculated based on the number of observed alleles (N) following:

$$PG = \frac{N(N+1)}{2}$$

LE and HWE tests were conducted for all iiSNPs and STRs with the Arlequin v3.5.2.2 software [52]. For both tests, a sequential Bonferroni correction was applied in order to account for false positives that would arise as a result of multiple comparisons [53].

In large populations with minimum inbreeding, the fixation index (F) is approximately equivalent to the coancestry coefficient (θ). θ is the probability that two homologous alleles, one drawn from each of the two individuals, are identical by descent (IBD) [55]. This was calculated for each locus following:

$$\theta \approx F = 1 - \frac{H_{obs}}{H_{exp}}$$

For each of the 10 casework-type samples, an LR was calculated comparing the single-source casework profile and a theoretical matching single source reference profile using the following propositions:

H_1: The DNA originated from the person of interest (POI).
H_2: The DNA originated from someone other than, and unrelated to, the POI in the Australian population with European ancestry.

The Balding-Nichols formulae (National Research Council (NRC) Recommendation 4.2 Equation 4.10a and 4.10b) were used to calculate the match probabilities [35]. The iiSNP allele frequencies determined in this study and the STR frequencies in the Australian European population published by Taylor et al. (2017) were used in a bespoke Excel-based workbook [30]. Single-source LRs were produced for the 20 STRs, 94 iiSNPs and the product of the two calculated for the combined LR for the 114 combined identity markers. A θ of 0.02 was used as recommended by Buckleton et al. (2016) and Taylor et al. (2017) to be consistent with current practice for STRs in the Australian population

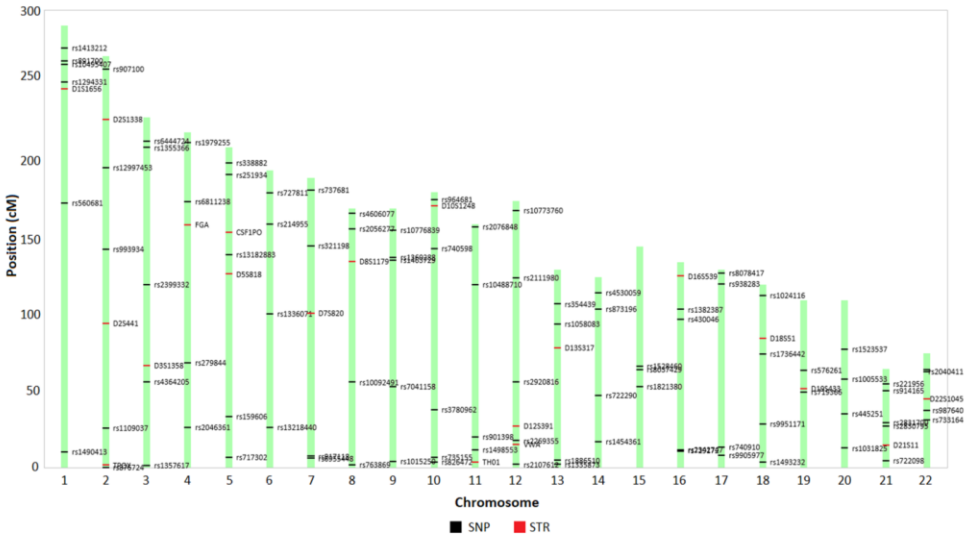


Fig. 1. Chromosome map with the positions of single nucleotide polymorphisms (SNPs; black) and short tandem repeats (STRs; red) in centimorgans (cM), as derived from HapMap [54].

with European ancestry [30,56].

3. Results

3.1. Observed alleles and possible genotypes

For the 105 individuals with self-declared European ancestry that were genotyped, no off-ladder microvariant alleles were observed in the STR profiles or tri-allelic genotypes in the STR and iiSNP profiles. Full STR and iiSNP profiles were obtained for the majority (90 %) of samples; locus dropout of one or two iiSNPs or STRs was observed for nine samples (9 %), and one sample produced a partial profile with combined call rate of 93 %.

As all iiSNPs targeted by the ForenSeq® Kintelligence and ForenSeq® DNA Signature Prep Kits are biallelic (N = 2), there are only three possible genotypes for each locus which indicates a low degree of polymorphism. When all 94 iiSNPs are taken into account, the number of possible genotypes is 7.07×10^{44} (assuming no linkage), enhancing the potential for individualising DNA profiles. However, STRs are highly polymorphic and will have multiple alleles observed for each locus. The total number of observed alleles across the 105 STR profiles was 184 corresponding with a total of 4.30×10^{30} possible genotypes (Table 1). The alleles observed ranged from five alleles for TH01 to 16 alleles for D1S1656.

3.2. Forensic efficiency parameters

The allele frequencies for the iiSNPs in the Australian population with European ancestry are in Table S1. The allele population frequencies for the STRs are in Table S2. The forensic efficiency parameters for all STRs and iiSNPs are detailed in Table S3 and Table S4, respectively. Summaries of the forensic efficiency parameters are provided in Fig. 2 and Table 2.

Although for individual loci the PM was always higher for iiSNPs than for STRs (Fig. 2a), the combined probability of matching (CPM) was calculated to be 1.71×10^{-23} for the 20 STRs and 7.77×10^{-38} for the 94 iiSNPs. Therefore, the probability that any two genotypes will match at all 114 loci within the population is negligible. For the PIC, PE, TPI and PD parameters calculated for all markers, SNPs produced lower values on average when compared to STRs (Fig. 2).

Table 1
Number of observed alleles and possible genotypes for short tandem repeats (STRs) in common between the GlobalFiler™ PCR Amplification Kit and the ForenSeq® DNA Signature Prep Kit.

STR Locus	Observed Alleles (N)	Possible Genotypes (PG)
D3S1358	8	28
vWA	7	21
D16S539	7	21
CSF1PO	8	28
TPOX	6	15
D8S1179	10	45
D21S11	11	55
D18S51	14	91
D2S441	10	45
D19S433	9	36
TH01	5	10
FGA	10	45
D22S1045	7	21
D5S818	7	21
D13S317	7	21
D7S820	9	36
D10S1248	7	21
D1S1656	16	120
D12S391	14	91
D2S1338	12	66
Total	**184**	**4.30×10^{30}**

3.3. Within locus (HWE) and between locus (LE) independence tests

To assess locus independence, HWE tests were conducted for all STRs and iiSNPs. Of the 94 iiSNPs, four loci (rs1357617, rs4374205, rs6955448 and rs1335873) returned p-values below 0.05 that indicated deviation from HWE. However, it is likely that approximately 5 % (0.05 $\times 94 \approx 4.7$) of the total tests will deviate from HWE by chance due to multiple comparisons. After a sequential Bonferroni correction was applied, no significant deviation was observed at any locus. For the STRs, only D8S1179 showed deviation from HWE ($p < 0.05$) and it remained statistically significant after applying the sequential Bonferroni correction, implying that the locus should be excluded when using the STR frequencies derived from this dataset.

LE tests were performed for inter-locus independence on all 6441 pairwise combinations of STRs and iiSNPs, of which 256 were syntenic pairs. There were 379 pairs (5.88 %) that were in disequilibrium ($p < 0.05$) consisting of 243 SNP/SNP pairs, 132 STR/SNP pairs and 4 STR/STR pairs. Of these, 14 pairs were located on the same chromosome with the distance between loci ranging from 7.59 cM (D22S1045/rs987640) to 212.76 cM (rs1355366/rs6444724). However, after a sequential Bonferroni correction was applied, all pairs were found to be in LE.

3.4. Accounting for population sub-structure

In a population with no inbreeding, F would equal 0, indicating that the number of observed and expected heterozygous genotypes are the same. In this study, F was calculated for each STR and iiSNP locus. The overall population had an average F of −0.002 across all loci, suggesting that there is little evidence of population sub-structure [55]. Some variation were observed, with F values ranging from −0.21 (rs1024116) to 0.35 (rs1357617; Fig. 3). However, this variation, with some values falling slightly above or below 0, is likely a result of sampling error. $F \approx \theta$ for large populations with minimal inbreeding, but nevertheless, we recommend a conservative θ correction factor of 0.02 for an Australian population with European ancestry to be consistent with the recommendations of Buckleton et al. (2016) and Taylor et al. (2017) for STRs [30,56].

3.5. Likelihood ratio calculations

The 10 casework samples yielded combined call rates ranging from 96.49 % to 100 %, with locus dropout only occurring for iiSNPs (Table 3). The logarithm of the LRs had an average of 23.70 ± 1.29 for the STR panel and 38.79 ± 1.43 for the iiSNP panel (Fig. 4). As these markers are in LE, inter-locus independence is ensured and the LRs can be combined into an overall LR. The average logarithm of the combined LR was 62.49 ± 1.65, meaning the likelihood of the combined genotype being shared with a randomly selected member of the Australian population with European ancestry is negligible.

4. Discussion

STR allele frequency data for the Australian population has previously been studied using the AmpFlSTR® Profiler Plus and GlobalFiler™ assays [30,57,58]. When compared to the latter and more widely adopted GlobalFiler™ STR study, Taylor *et al.* (2017) collected DNA from seven subpopulation groups in Australia and New Zealand with 2274 samples, of which 528 were identified as "Australian Caucasian" (assumed to be congruent with our DNA donors who declared European ancestry) [30]. Guidelines for genetic population data suggest a minimum of 500 individuals are required to generate reliable allele frequencies due to the high degree of polymorphism in STRs [59]. However, far fewer are required for bi-allelic SNPs which are far less polymorphic. There are dissimilarities between the STR analyses in this study and that by Taylor et al. (2017), likely due to the smaller samples size employed in this study [30]. As such, the Australian Caucasian STR

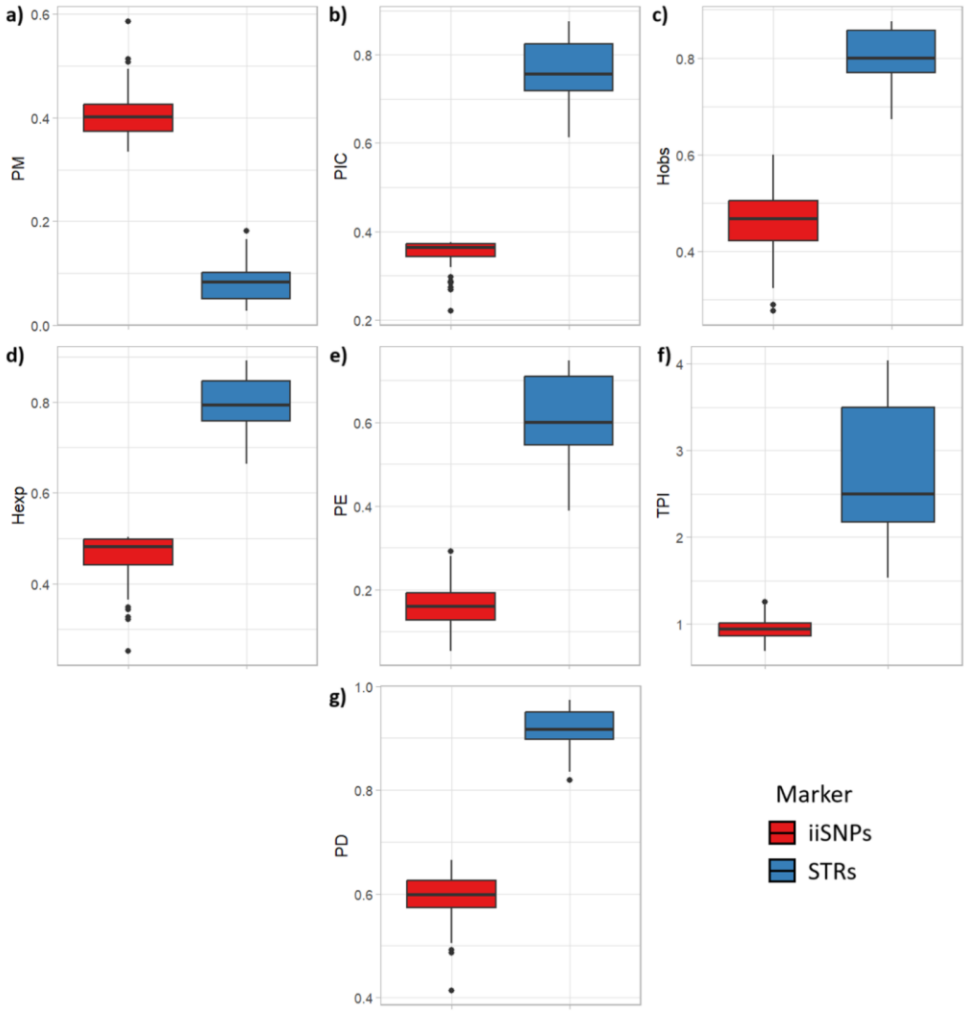


Fig. 2. Forensic efficiency parameters for all individual identity-informative single nucleotide polymorphism (iiSNP) and short tandem repeat (STR) markers: a) probability of matching (PM); b) polymorphism information content (PIC); c) observed heterozygosity (H_{obs}); d) expected heterozygosity (H_{exp}); e) power of exclusion (PE); f) typical paternity index (TPI); and g) power of discrimination (PD).

frequencies published by Taylor et al. (2017) were used for the calculation of LRs in this study (because of the larger sample size) and are recommended for forensic use for the Australian population with European ancestry [30].

The first difference between this study and the Australian Caucasian data by Taylor et al. (2017) was highlighted in the number of observed alleles; an additional 40 alleles were observed by Taylor et al. (2017) across 16 loci [30]. The largest differences were at each of the loci D21S11, D19S433 and FGA, where Taylor *et al.* found seven more alleles than in this study. An additional allele was also observed at D1S1656 (allele 19) in this study that was not observed in the larger published dataset by Taylor et al. (2017) [30]. The added alleles increased the number of possible genotypes to 2.57×10^{40} from 274 observed alleles, compared to the 184 alleles that were observed in this study.

This study demonstrated that established STR allele frequency data generated with CE could be combined with iiSNP data to improve the discriminatory power of a DNA profile. The majority of STR profiles produced in this study were sequenced using MPS technology, for which sequence-based allelic frequencies could have higher discrimination power than the length-based allelic frequencies determined with CE

Table 2
Summary of forensic efficiency parameters calculated for the individual identity-informative single nucleotide polymorphism (iiSNP) and short tandem repeat (STR) markers.

Forensic Efficiency Parameter	STR		iiSNP	
	Minimum	Maximum	Minimum	Maximum
PM	0.030 (D12S391)	0.181 (TPOX)	0.335 (rs1357617)	0.587 (rs938283)
PIC	0.613 (D1S1656)	0.876 (D1S1656 and D12S391)	0.220 (rs938283)	0.375 (13 iiSNPs)
H_{obs}	0.673 (D5S818)	0.876 (D18S51 and D12S391)	0.276 (rs938283)	0.600 (rs1024116)
H_{exp}	0.664 (TPOX)	0.891 (D1S1656 and D12S391)	0.253 (rs938283)	0.500 (13 iiSNPs)
PE	0.388 (D5S818)	0.747 (D18S51 and D12S391)	0.054 (rs938283)	0.291 (rs1024116)
TPI	1.529 (D5S818)	4.038 (D18S51 and D12S391)	0.691 (rs938283)	1.250 (rs1024116)
PD	0.819 (TPOX)	0.973 (D12S391)	0.413 (rs938283)	0.665 (rs1357617)

[60]. Using a panel such as the ForenSeq® DNA Signature Prep Kit allows for the generation of iiSNPs and STRs with sequence-based variation in alleles. However, the use of sequence-based allele frequencies for STRs is not commonly practiced in Australian laboratories at the time of publication.

SNP studies have been published for a number of population and subpopulation groups, but there has not yet been a study on the Australian population [21–29]. While a minimum of 500 individuals is recommended to adequately assess STR allele frequencies, substantially fewer are required for biallelic SNPs [23,27,28,59]. In this study, the most informative loci were rs717302 and rs1498553. The locus rs938283 was the lowest performing in the panel, producing the highest PM and the lowest PIC, H_{obs}, H_{exp}, PE, PD and TPI; these results were congruent with the findings for populations with European ancestry in studies in the United States, France and the United Kingdom [21,22,26].

When assessing intra- and inter-locus independence, all SNPs in this study were in HWE and LE for the Australian population with European ancestry. Furthermore, when combining these identity-informative SNP markers with the existing suite of STRs, pairwise tests showed all 114 loci were in LE. Only one STR, D8S1179, was found to be out of HWE. However, when compared to Taylor *et al.* (2017), there were variances in the p-values produced by up to 0.78 that were likely due to sampling error and D8S1179 was not out of equilibrium [30]. SE33 was not assessed in this study as only a few samples were profiled with the GlobalFiler™ PCR Amplification Kit and SE33 is not included in the ForenSeq® DNA Signature Prep panel. Similarly, D4S2408, D6S1043, D9S1122, D17S1301, D20S482, PentaD and PentaE were also not assessed as these markers were not included in the GlobalFiler™ panel. These STRs would require additional analysis to determine whether they are in LE with each other and with the 94 iiSNPs.

Table 3
Locus call rates (%) of casework-type samples for short tandem repeats (STRs) with the GlobalFiler™ PCR Amplification Kit and identity-informative single nucleotide polymorphisms (iiSNPs) with the ForenSeq® Kintelligence Kit. The call rates for the combined identity markers (20 STRs plus 94 iiSNPs) are also reported.

Sample	STRs (20)	iiSNPs (94)	Combined Identity Markers (114)
Tooth 1	100.0 %	95.7 %	96.5 %
Tooth 2	100.0 %	100.0 %	100.0 %
Bone 1	100.0 %	97.9 %	98.3 %
Bone 2	100.0 %	97.9 %	98.3 %
Bone 3	100.0 %	98.9 %	99.1 %
Bone 4	100.0 %	100.0 %	100.0 %
Bone 5	100.0 %	100.0 %	100.0 %
Bone 6	100.0 %	100.0 %	100.0 %
Bone 7	100.0 %	98.9 %	99.1 %
Bone 8	100.0 %	95.7 %	96.5 %

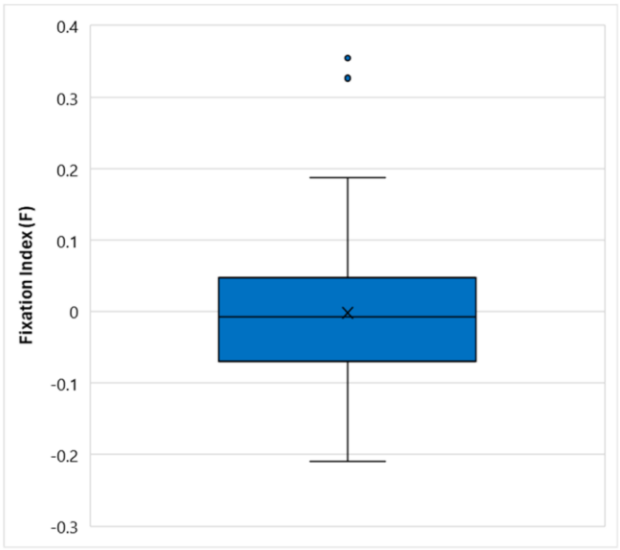


Fig. 3. The fixation index (F) for all short tandem repeat (STR) and identity-informative single nucleotide polymorphism (iiSNP) loci.

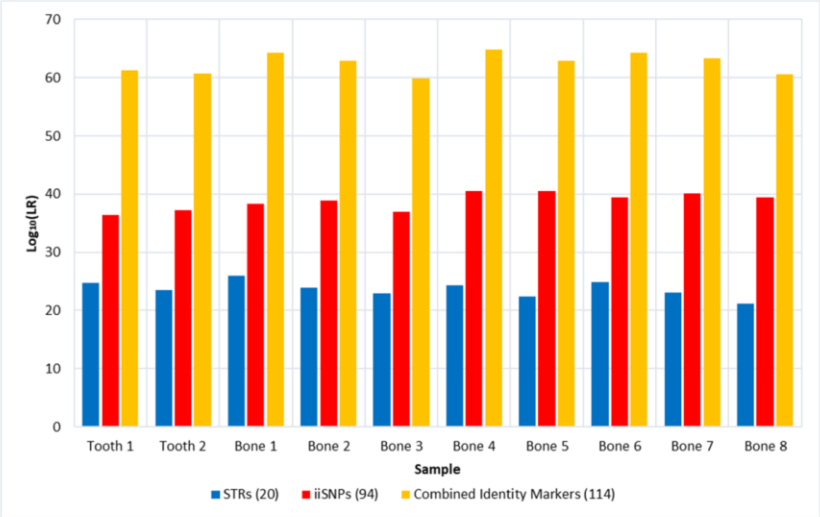


Fig. 4. Logarithm of the likelihood ratios (LR) generated for casework-type samples using short tandem repeats (STRs) with the GlobalFiler™ PCR Amplification Kit, identity-informative single nucleotide polymorphisms (iiSNPs) with the ForenSeq® Kintelligence Kit and the combined identity markers (20 STRs plus 94 iiSNPs).

The overall F for the Australian population with European ancestry was −0.002, indicating there is little evidence of population sub-structure [55]. Regardless, the NRC II recommendations specify the importance of accounting for population sub-structure by applying a θ correction factor between 0.01 (minimal inbreeding) and 0.03 (excess inbreeding) [35]. For STRs, Buckleton *et al.* (2016) and Taylor *et al.* (2017) recommended using a θ correction factor of 0.02 for an Australian population with European ancestry as a conservative measure when calculating the RMP [30,56]. In accounting for possible inbreeding, the θ correction factor raises the RMP and lowers the subsequent LR so as not to over-estimate the weight of the evidence [35]. The application of a conservative θ correction factor of 0.02 for the 94 iiSNPs is consistent with the use of this value for STRs.

The LRs calculated with the iiSNPs for single source profiles were orders of magnitude larger than those calculated with STRs, with a CPM of 10^{-38} for SNPs, compared to 10^{-23} for STRs. This value is similar to that observed by Kiesler *et al.* (2023), who produced a CPM of 10^{-39} for the population with European ancestry in the United States, and Davenport *et al.* (2023), who produced a CPM of 10^{-38} for the "White British" subpopulation in the United Kingdom [26]. Due to their independence, a combined LR can be calculated from the combination of STRs and iiSNPs with a CPM of 10^{-60} which would produce astronomical LRs for complete, matching single source profiles, far beyond the maximum LR reported in Australian forensic laboratories currently (100 billion) [61]. Similar CPMs for combined iiSNP and STR profiles were seen in population studies for French (10^{-69}) and Northeastern Peruvian Andes (10^{-66}) populations [21,23].

The combined LR may be beneficial for samples that produce partial STR profiles where iiSNPs could provide supplementary information to improve discrimination between individuals. The more powerful LRs could also impact kinship calculations, potentially extending the applicability of STRs beyond first order relationships (i.e. parent/offspring and full siblings) if their relatively high mutation rates can be accounted for [33,38].

5. Conclusions

This study has confirmed the forensic applicability of the 94 ForenSeq® iiSNPs in the Australian population with European ancestry, as well as the combined power of identity markers consisting of both iiSNPs and STRs for improved discrimination between individuals for forensic casework. By themselves, iiSNPs can produce LRs that exceed those produced with the established STRs, and their combined power may aid in identifying persons of interest through indirect matching to their genetic relatives or from challenging or compromised samples that have produced suboptimal partial profiles. This study has facilitated the creation of a SNP allele frequency database in Australia, starting with individuals of European ancestry. In order to expand the potential uses of iiSNP markers in routine casework, these loci should be evaluated in other Australian subpopulations including those with Aboriginal and Torres Strait Islander and Asian self-declared ancestries.

CRediT authorship contribution statement

Jodie Ward: Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Dennis McNevin:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Kaymann Cho:** Writing – review & editing, Methodology. **Jessica L. Watson:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Kelly Grisedale:** Writing – review & editing, Methodology, Formal analysis.

Informed Consent statement

Written informed consent was obtained from all volunteer sample donors involved in this study.

Institutional Review Board Statement

The study was conducted according to the guidelines of the

Declaration of Helsinki and approved by the Human Research Ethics Committee (HREC) of the University of Technology Sydney (UTS) (UTS HREC NO. ETH21–5821, ETH21–6606 and ETH23–8117).

Funding

This research was funded by the Australian Federal Police (AFP) National DNA Program for Unidentified and Missing Persons, AFP Innovation Fund and AFP Forensics. Jessica Watson is supported by an Australian Government Research Training Program (RTP) Scholarship.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the volunteer sample donors from the Australian Federal Police and the University of Technology Sydney for providing their DNA and associated meta data for this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2024.103169.

Data availability

Data are stored at the Australian Federal Police and may be made available to approved entities upon written request and subject to ethics and consent provisions.

References

[1] P. Gill, P.L. Ivanov, C. Kimpton, R. Piercy, N. Benson, G. Tully, et al., Identification of the remains of the Romanov family by DNA analysis, Nat. Genet. 6 (2) (1994) 130–135.
[2] Angers A., Kagkli D., Oliva L., Petrillo M., Raffael B. Study on DNA Profilng Technology for Its Implementation in Central Schengen Information System: EUR29766 Publications Office of the EU, Luxembourg; 2019. Available from: ⟨https://op.europa.eu/en/publication-detail/-/publication/7ae614b0-b274-11e9-9d01-01aa75ed71a1/language-en⟩.
[3] J.L. Weber, C. Wong, Mutation of human short tandem repeats, Hum. Mol. Genet. 2 (8) (1993) 1123–1128, https://doi.org/10.1093/hmg/2.8.1123.
[4] J.M. Butler, Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers, Elsevier, 2005.
[5] R. Puch-Solis, S. Pope, Interpretation of DNA data within the context of UK forensic science—evaluation, Emerg. Top. Life Sci. (2021), https://doi.org/10.1042/ETLS20200340.
[6] Australian Law Reform Commission. Report 96: Essentially yours - the protection of human genetic information in Australia. 2003.
[7] J.W. Hazel, E.W. Clayton, B.A. Malin, C. Slobogin, Is it time for a universal genetic forensic database? Science 362 (6417) (2018) 898–900, https://doi.org/10.1126/science.aav5475.
[8] C.H. Brenner, B.S. Weir, Issues and strategies in the DNA identification of World Trade Center victims, Theor. Popul Biol. 63 (3) (2003) 173–178, https://doi.org/10.1016/S0040-5809(03)00008-X.
[9] E.S.M. Iwamura, M.A. Guimarães, M.P. Evison, DNA methods to identify missing persons, in: S.J. Morewitz, C. Sturdy Colls (Eds.), Handbook of Missing Persons, Springer International Publishing, Cham, 2016, pp. 337–352.
[10] B. Budowle, A. Van Daal, Forensically relevant SNP classes, Biotechniques 44 (5) (2008) 603–610, https://doi.org/10.2144/000112806.
[11] R. Decorte, Genetic identification in the 21st century—current status and future developments, Forensic Sci. Int. 201 (1-3) (2010) 160–164.
[12] P. Gill, An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes, Int. J. Leg. Med. 114 (4-5) (2001) 204–210, https://doi.org/10.1007/s004149900117.
[13] J.M. Butler, M.D. Coble, P.M. Vallone, STRs vs. SNPs: thoughts on the future of forensic DNA testing, Forensic Sci., Med. Pathol. 3 (3) (2007) 200–205, https://doi.org/10.1007/s12024-007-0018-1.
[14] K.M. Canturk, R. Emre, B. Buşpinar, M. Ozen, Current status of the use of single-nucleotide polymorphisms in forensic practices, Genet. Test. Mol. Biomark. 18 (7) (2014) 455–460, https://doi.org/10.1089/gtmb.2013.0466.
[15] K.K. Kidd, A.J. Pakstis, W.C. Speed, E.L. Grigorenko, S.L.B. Kajuna, N.J. Karoma, et al., Developing a SNP panel for forensic identification of individuals, Forensic Sci. Int. 164 (1) (2006) 20–32, https://doi.org/10.1016/j.forsciint.2005.11.017.
[16] S.B. Seo, J.L. King, D.H. Warshauer, C.P. Davis, J. Ge, B. Budowle, Single nucleotide polymorphism typing with massively parallel sequencing for human identification, Int J. Leg. Med 127 (6) (2013) 1079–1086, https://doi.org/10.1007/s00414-013-0879-7.
[17] B.C. Durney, C.L. Crihfield, L.A. Holland, Capillary electrophoresis applied to DNA: determining and harnessing sequence and structure to advance bioanalyses (2009–2014), Anal. Bioanal. Chem. 407 (23) (2015) 6923–6938, https://doi.org/10.1007/s00216-015-8703-5.
[18] J.G. Shewale, L. Qi, L.M. Calandro, Principles, practice, and evolution of capillary electrophoresis as a tool for forensic DNA analysis, in: J.G. Shewale, Q. Liu (Eds.), Forensic DNA Analysis: Current Practices and Emerging Technologis, CRC Press, 2012, p. 131.
[19] Illumina. An Introduction to Next-Generation Sequencing Technology 2017. [Last accessed: 27 May]. Available from: ⟨https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf⟩.
[20] Y.-L. Wei, C.-X. Li, J. Jia, L. Hu, Y. Liu, Forensic Identification using a multiplex assay of 47 SNPs*, J. Forensic Sci. 57 (6) (2012) 1448–1456, https://doi.org/10.1111/j.1556-4029.2012.02154.x.
[21] A. Delest, D. Godfrin, Y. Chantrel, A. Ulus, J. Vannier, M. Faivre, et al., Sequenced-based French population data from 169 unrelated individuals with Verogen's ForenSeq DNA signature prep kit, Forensic Sci. Int. Genet. 47 (2020) 102304, https://doi.org/10.1016/j.fsigen.2020.102304.
[22] L. Davenport, L. Devesse, D.S. Court, D. Ballard, Forensic identity SNPs: Characterisation of flanking region variation using massively parallel sequencing, Forensic Sci. Int. Genet. 64 (2023) 102847, https://doi.org/10.1016/j.fsigen.2023.102847.
[23] E.K. Guevara, J.U. Palo, J.L. King, M.M. Buś, S. Guillen, B. Budowle, et al., Autosomal STR and SNP characterization of populations from the Northeastern Peruvian Andes with the ForenSeq™ DNA Signature Prep Kit, Forensic Sci. Int. Genet. 52 (2021) 102487, https://doi.org/10.1016/j.fsigen.2021.102487.
[24] R. Tao, X. Dong, X. Zhen, R. Xia, Y. Qu, S. Liu, et al., Population genetic analyses of Eastern Chinese Han nationality using ForenSeq™ DNA Signature Prep Kit, Mol. Genet. Genom. 299 (1) (2024) 9, https://doi.org/10.1007/s00438-024-02121-w.
[25] J.A. Aguilar-Velázquez, M.Á. Duran-Salazar, M.F. Córdoba-Mercado, C. E. Coronado-Avila, O. Salas-Salas, G. Martinez-Cortés, et al., Characterization of 58 STRs and 94 SNPs with the ForenSeq™ DNA signature prep kit in Mexican-Mestizos from the Monterrey city (Northeast, Mexico), Mol. Biol. Rep. 49 (8) (2022) 7601–7609, https://doi.org/10.1007/s11033-022-07575-y.
[26] K.M. Kiesler, L.A. Borsuk, C.R. Steffen, P.M. Vallone, K.B. Gettings, US population data for 94 identity-informative SNP loci, Genes 14 (2023) 1071, https://doi.org/10.3390/genes14051071.
[27] I. Grandell, R. Samara, A. Tillmar, A SNP panel for identity and kinship testing using massive parallel sequencing, Int. J. Leg. Med. 130 (2016) 905–914, https://doi.org/10.1007/s00414-016-1341-4.
[28] D. Augustinus, M.E. Gahan, D. McNevin, Development of a forensic identity SNP panel for Indonesia, Int J. Leg. Med. 129 (2015) 681–691, https://doi.org/10.1007/s00414-014-1055-4.
[29] F. Guo, Z. Liu, G. Long, B. Zhang, D. Liu, S. Yu, Performance and characterization of 94 identity-informative SNPs in Northern Han Chinese using ForenSeq™ DNA Signature Prep Kit, J. Forensic Leg. Med 103 (2024) 102678, https://doi.org/10.1016/j.flm.2024.102678.
[30] D. Taylor, J.-A. Bright, C. McGovern, S. Neville, D. Grover, Allele frequency database for GlobalFiler™ STR loci in Australian and New Zealand populations, Forensic Sci. Int. Genet. 28 (2017) e38–e40, https://doi.org/10.1016/j.fsigen.2017.02.012.
[31] P. Gill, T. Hicks, J.M. Butler, E. Connolly, L. Gusmão, B. Kokshoorn, et al., DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence-Guidelines highlighting the importance of propositions: Part I: evaluation of DNA profiling comparisons given (sub-) source propositions, Forensic Sci. Int. Genet. 36 (2018) 189–202, https://doi.org/10.1016/j.fsigen.2018.07.003.
[32] B.S. Weir, A.D. Anderson, A.B. Hepler, Genetic relatedness analysis: modern data and new challenges, Nat. Rev. Genet. 7 (10) (2006) 771–780, https://doi.org/10.1038/nrg1960.
[33] I.Z. Pajnič, T. Leskovar, M. Črešnar, Improving kinship probability in analysis of ancient skeletons using identity SNPs and MPS technology, Int. J. Leg. Med. 137 (2023) 1007–1015, https://doi.org/10.1007/s00414-023-03003-3.
[34] A. Angers, J. Drabek, M. Fabbri, M. Petrillo, M. Querci, whole genome sequencing and forensics genomics, Jt. Res. Cent. (2021), https://doi.org/10.2760/864087.
[35] National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence, National Academies Press (US), Washington (DC), 1996, https://doi.org/10.17226/5141.
[36] J.F. Crow, M. Kimura, An Introduction to Population Genetics Theory, Harper and Row, New York, 1970.
[37] A. Tillmar, C. Phillips, Evaluation of the impact of genetic linkage in forensic identity and relationship testing for expanded DNA marker sets, Forensic Sci. Int. Genet. 26 (2017) 58–65, https://doi.org/10.1016/j.fsigen.2016.10.007.
[38] C. Phillips, M. Fondevila, M. García-Magariños, A. Rodriguez, A. Salas, Á. Carracedo, et al., Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers, Forensic Sci. Int. Genet. 2 (3) (2008) 198–204, https://doi.org/10.1016/j.fsigen.2008.02.002.
[39] QIAGEN. EZ1&2 DNA Investigator Kit. 2022.

[40] Thermo Fisher Scientific. Quantifiler™ HP and Trio DNA Quantification Kits User Guide [User Guide]. 2017.
[41] Thermo Fisher Scientific, QuantStudio™ 5 Real-Time PCR Instrument(for Human Identification) [User Guide], Life Technol. (2017).
[42] Verogen. ForenSeq DNA Signature Prep Reference Guide. ForenSeq DNA Signature Prep Kit. 2020.
[43] Verogen. ForenSeq Kintelligence Kit Reference Guide. ForenSeq Kintelligence Kit. 2021.
[44] B. Mehta, S. Venables, P. Roffey, Comparison between magnetic bead and qPCR library normalisation methods for forensic MPS genotyping, Int. J. Leg. Med. 132 (2018) 125–132, https://doi.org/10.1007/s00414-017-1591-9.
[45] Verogen, MiSeq FGx Sequencing System Reference Guide, MiSeq FGx Seq. Syst. (2021).
[46] Verogen, Universal Analysis Software v1.3, Reference Guide, Universal Analysis Software, 2021.
[47] J.L. Watson, D. McNevin, K. Grisedale, M. Spiden, S. Seddon, J. Ward, Operationalisation of the ForenSeq® Kintelligence Kit for Australian Unidentified and Missing Persons Casework, Forensic Sci. Int. Genet. (2023) 102972, https://doi.org/10.1016/j.fsigen.2023.102972.
[48] Verogen. Universal Analysis Software v2.0 and above, Reference Guide. Universal Analysis Software. 2021.
[49] Thermo Fisher Scientific. GlobalFiler PCR Amplification Kits. 2019.
[50] Thermo Fisher Scientific. Applied Biosystems 3500/3500xL Genetic Analyzer User Guide. 2010.
[51] A. Gouy, M. Zieger, STRAF—a convenient online tool for STR data evaluation in forensic genetics, Forensic Sci. Int. Genet. 30 (2017) 148–151, https://doi.org/10.1016/j.fsigen.2017.07.007.
[52] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567, https://doi.org/10.1111/j.1755-0998.2010.02847.x.
[53] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc.: Ser. B (Methodol. 57 (1) (1995) 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
[54] The International HapMap Consortium, The International HapMap project, Nature 426 (2003) 789–796, https://doi.org/10.1038/nature02168.
[55] B.S. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population structure, Evolution 38 (6) (1984) 1358–1370, https://doi.org/10.2307/2408641.
[56] J. Buckleton, J. Curran, J. Goudet, D. Taylor, A. Thiery, B.S. Weir, F.S. T. Population-specific, values for the forensic STR markers: a worldwide survey, Forensic Sci. Int. Genet. 23 (2016) 91–100, https://doi.org/10.1016/j.fsigen.2016.03.004.
[57] S.J. Walsh, J.S. Buckleton, Autosomal microsatellite allele frequencies for a nationwide dataset from Australian Caucasian sub-population, Forensic Sci. Int. 168 (2-3) (2007) e47–e50, https://doi.org/10.1016/j.forsciint.2007.01.030.
[58] S.J. Walsh, J.S. Buckleton, Autosomal microsatellite allele frequencies for 15 regionally defined Aboriginal Australian population datasets, Forensic Sci. Int. 168 (2-3) (2007) e29–e42, https://doi.org/10.1016/j.forsciint.2007.01.029.
[59] Á. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, et al., New guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. 7 (2) (2013) 217–220, https://doi.org/10.1016/j.fsigen.2013.01.001.
[60] A. Alonso, P.A. Barrio, P. Müller, S. Köcher, B. Berger, P. Martin, et al., Current state-of-art of STR sequencing in forensic genetics, Electrophoresis 39 (21) (2018) 2655–2668, https://doi.org/10.1002/elps.201800030.
[61] Biology Specialist Advisory Group. Biology Specialist Advisory Group (BSAG) Statistics Scientific Working Group (StatSWG) recommendations for the interpretation of DNA, Version 3. BSAG StatSWG. 2016.

What follows is additional and not included in the published paper.

5.2 Introduction

Sewall Wright's fixation index relating sub-populations to the total population (F_{ST}) is a parameter that measures population structure by quantifying genetic variation within a subpopulation group compared to the total genetic variation across the whole population.[1] F_{ST} values within a population are utilised when calculating likelihood ratios (LRs) in forensic genetics to account for coancestry, allowing for uncertainty in the database used to calculate allele frequencies.[2, 3] F_{ST} values between sub-populations suggest genetic drift has separated them. Wright suggested the following interpretation:[4]

- 0.00 – 0.05: little genetic drift
- 0.05 – 0.15: moderate genetic drift
- 0.15 – 0.25: great genetic drift
- 0.25 and above: very great genetic drift

An inbreeding coefficient (F_{IS}) can also be defined as a measure of heterozygote deficiency due to inbreeding in each subpopulation.[5] This study compared the European Australian allele frequencies with data from European and North American populations having European ancestry for identity-informative single nucleotide polymorphisms (iiSNPs), as well as F_{ST} and F_{IS} in these populations.[6-8]

5.3 Methods

Data was gathered from the above study (Section 5.1) and three published population studies, and the European ancestry population data was extracted.[6-9] Table 5.1 summarises the population data. The data collated included the allele frequencies for each iiSNP, the observed heterozygosity (H_{obs}), the expected heterozygosity (H_{exp}) for each iiSNP and the population sizes.

Table 5.1 Population genetics studies of iiSNPs compared.

Metric	Watson et al (2024) [9]	Delest et al. (2020) [6]	Kiesler et al. (2023) [8]	Davenport et al. (2023) [7]
Population	Australian with European Ancestry	French	American Caucasian	White British
iiSNPs	94	92	94	94
Population Size	105	169	361	204

The expected heterozygosity of all populations (H_T) was calculated following:

$$H_T = 2 \times \overline{p} \times \overline{q}$$

where $\overline{p}$ is the frequency of the reference allele over all populations and $\overline{q}$ is the frequency of the alternative allele over all populations. $\overline{p}$ was calculated following:

$$\overline{p} = \frac{\sum_i p_i n_i}{\sum_i n_i}$$

where i is the population in question, p is the frequency of the reference allele in population i and n is the total number of observed alleles. As all iiSNPs were biallelic, $\overline{q}$ was calculated as $1 - \overline{p}$.

The locus-specific F_{ST} was calculated within each population following:

$$F_{\mathrm{ST}} = \frac{H_{\mathrm{T}} - H_{\mathrm{exp}}}{H_{\mathrm{T}}}$$

The locus-specific F_{IS} was calculated within each population following:

$$F_{\mathrm{IS}} = 1 - \frac{H_{\mathrm{obs}}}{H_{\mathrm{exp}}}$$

Statistical tests were used to determine significant differences by population for the locus-specific F_{ST} and F_{IS} values in R v4.2.1 using the Kruskal-Wallis H test.[10] The locus-specific F_{ST} and F_{IS} values were averaged across the population to generate the population-specific F_{ST} and F_{IS}.

5.4 Results

The locus-specific F_{ST} values were similar between the population groups and were not significantly different (x^2 = 0.84, p = 0.84; Table 5.2; Figure 5.1). The population-specific F_{ST} values were subsequently similar; the American Caucasian population had the highest F_{ST} at 0.050 and the Australian with European Ancestry population had the lowest at 0.041 (Table 5.2).

Similarly, the locus-specific F_{IS} values were not significantly different between population groups (x^2 = 0.85, p = 0.84; Table 5.3; Figure 5.2). The population-specific F_{IS} values were close to zero, ranging from -0.025 (French) to -0.001 (White British and Australian with European Ancestry). This indicates that all subpopulations exhibit negligible inbreeding.

Table 5.2 Locus-specific F_{ST} values for each population. The population-specific F_{ST} values are included in the bottom row as the average locus-specific F_{ST} values.

SNP	Australian with European Ancestry [9]	French [6]	Caucasian American [8]	White British [7]
rs1490413	0.021	0.010	0.004	0.039
rs560681	0.179	0.155	0.122	0.215
rs1294331	-0.008	-0.039	-0.034	-0.048
rs10495407	0.167	0.113	0.073	0.114
rs891700	-0.090	-0.074	-0.087	-0.086
rs1413212	-0.082	0.052	0.046	0.077
rs876724	0.055	0.004	0.099	0.152
rs1109037	-0.011	-0.031	-0.030	-0.032
rs993934	0.029	0.016	0.046	0.052
rs12997453	0.010	-0.003	0.018	0.031
rs907100	0.015	0.002	0.017	0.038
rs1357617	-0.015	0.244	0.219	0.112
rs4364205	-0.030	-0.018	-0.026	-0.026
rs2399332	0.118	0.033	0.066	0.117
rs1355366	0.042	0.011	-0.004	0.012
rs6444724	0.020	0.030	0.044	0.045
rs2046361	-0.050	0.136	0.120	0.127
rs279844	-0.002	0.000	0.000	-0.006
rs6811238	0.000	-0.005	0.004	-0.004
rs1979255	-0.177	0.054	-0.051	-0.079
rs717302	-0.045	-0.044	-0.042	-0.043
rs159606	0.068	0.109	0.131	0.050
rs13182883	-0.104	-0.112	-0.109	-0.125
rs251934	0.027	-0.021	0.014	0.017
rs338882	-0.020	-0.013	-0.022	-0.001
rs13218440	0.056	-0.013	0.042	0.008
rs1336071	-0.041	-0.015	-0.023	-0.020
rs214955	0.017	-0.004	0.005	0.012
rs727811	-0.013	0.035	-0.022	0.027
rs6955448	0.143	-	0.112	0.147
rs917118	0.046	0.055	0.072	0.017
rs321198	0.004	-0.029	-0.014	-0.030
rs737681	0.044	0.012	0.050	0.042
rs763869	-0.005	-0.004	0.000	-0.003
rs10092491	-0.082	-0.069	-0.070	-0.060
rs2056277	0.312	0.179	0.267	0.208
rs4606077	0.137	0.082	0.127	0.023
rs1015250	0.340	0.301	0.350	0.358
rs7041158	0.088	0.064	0.085	0.061

rs1463729	-0.111	-0.121	-0.073	-0.059
rs1360288	0.101	0.057	0.106	0.013
rs10776839	0.002	0.017	0.014	-0.001
rs826472	-0.044	-0.058	-0.083	-0.081
rs735155	0.002	-0.014	-0.015	-0.003
rs3780962	-0.040	-0.096	-0.098	-0.102
rs740598	0.010	0.028	0.080	0.040
rs964681	-0.017	-0.001	-0.013	0.029
rs1498553	-0.005	-0.004	-0.001	0.005
rs901398	0.085	-0.064	-0.083	-0.023
rs10488710	-0.004	0.007	0.095	0.119
rs2076848	-0.005	0.002	0.018	0.044
rs2107612	0.169	0.151	0.169	0.199
rs2269355	-0.004	0.011	0.014	0.025
rs2920816	0.035	0.031	0.014	0.027
rs2111980	-0.002	-0.009	-0.001	0.004
rs10773760	0.001	-0.003	0.011	0.012
rs1335873	-0.009	0.147	0.247	0.253
rs1886510	-0.022	-0.015	-0.019	-0.017
rs1058083	0.044	-0.013	-0.023	-0.018
rs354439	-0.043	-0.027	0.007	-0.015
rs1454361	-0.046	-0.030	-0.037	-0.035
rs722290	-0.007	-0.002	-0.004	-0.005
rs873196	0.052	0.017	0.010	-0.005
rs4530059	0.017	0.025	0.017	0.066
rs1821380	-0.021	-0.019	0.010	-0.014
rs8037429	0.007	-0.002	-0.001	-0.002
rs1528460	0.075	0.113	0.085	0.034
rs729172	0.037	0.053	0.008	0.034
rs2342747	0.092	0.169	0.130	0.169
rs430046	0.027	-0.014	0.010	0.017
rs1382387	0.172	0.098	0.171	0.102
rs9905977	0.117	0.125	0.133	0.124
rs740910	0.102	0.104	0.167	0.081
rs938283	0.451	0.355	0.402	0.462
rs8078417	0.130	0.089	0.241	0.198
rs1493232	0.055	0.151	0.027	0.039
rs9951171	-0.010	0.006	-0.004	-0.004
rs1736442	0.019	0.028	0.021	0.048
rs1024116	0.008	0.039	0.000	0.019
rs719366	0.028	0.117	0.056	0.029
rs576261	-0.005	-0.012	-0.017	-0.022
rs1031825	0.101	0.171	0.187	0.107
rs445251	-0.005	0.051	0.044	0.041

rs1005533	-0.002	0.011	0.000	0.002
rs1523537	-0.022	-0.009	-0.003	-0.017
rs722098	0.351	0.483	0.400	0.459
rs2830795	0.278	0.172	0.145	0.086
rs2831700	0.032	-	0.015	0.033
rs914165	0.025	0.006	-0.007	0.033
rs221956	0.099	0.113	0.132	0.114
rs733164	0.129	0.118	0.225	0.136
rs987640	-0.054	-0.038	-0.073	-0.061
rs2040411	0.007	0.013	0.017	0.005
rs1028528	0.271	0.116	0.194	0.185
Average F_{ST}	0.041	0.042	0.050	0.047

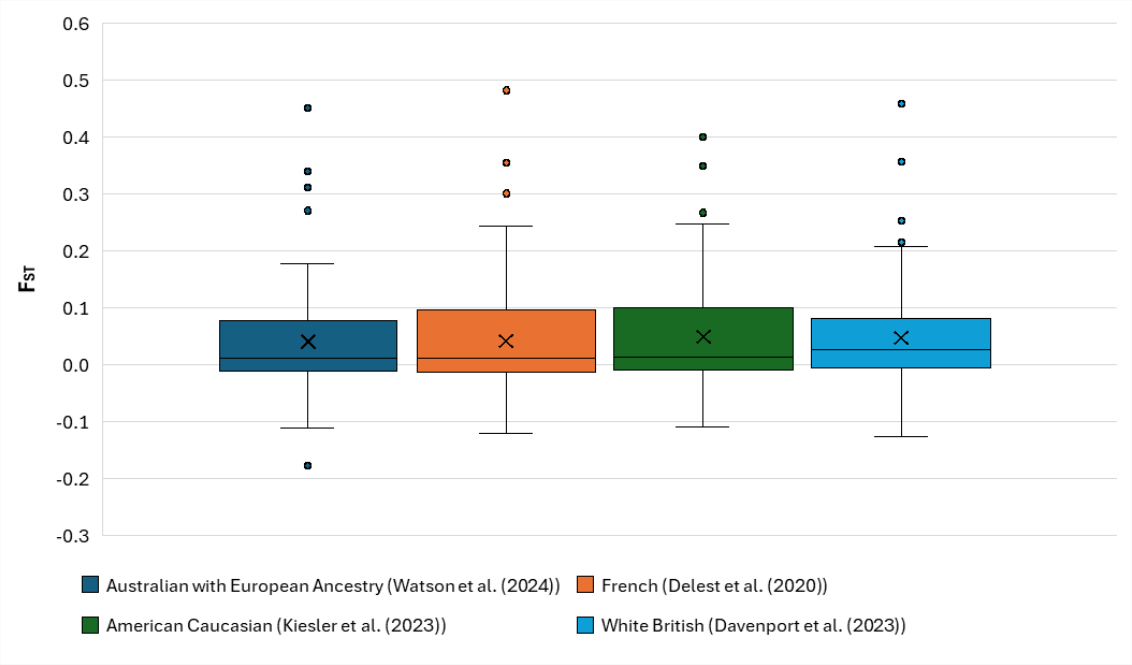


Figure 5.1 Locus-specific F_{ST} values ordered by population group.

Table 5.2 Locus-specific F_{IS} values for each population. The population-specific F_{IS} values are included in the bottom row as the average locus-specific F_{IS} values.

SNP	Australian with European Ancestry [9]	French [6]	Caucasian American [8]	White British [7]
rs1490413	-0.168	-0.005	0.082	-0.062
rs560681	0.039	-0.065	0.074	-0.024
rs1294331	-0.013	0.000	0.008	0.019
rs10495407	0.058	0.018	0.056	0.099
rs891700	0.014	-0.642	-0.007	-0.009
rs1413212	-0.011	0.227	-0.059	0.047
rs876724	-0.017	-0.052	0.057	-0.049

rs1109037	0.126	-0.256	0.036	-0.037
rs993934	0.188	0.027	-0.057	-0.067
rs12997453	-0.115	0.162	0.002	0.136
rs907100	0.071	0.027	-0.044	0.055
rs1357617	0.353	0.031	-0.002	0.010
rs4364205	-0.187	-0.052	-0.080	-0.088
rs2399332	-0.152	-0.015	-0.084	0.123
rs1355366	0.036	0.045	-0.009	0.074
rs6444724	-0.012	0.017	0.047	0.132
rs2046361	0.183	0.001	-0.086	-0.077
rs279844	0.132	0.050	0.051	-0.009
rs6811238	0.046	0.008	0.080	-0.053
rs1979255	0.182	-0.040	-0.022	0.029
rs717302	-0.082	0.051	-0.018	0.077
rs159606	-0.149	0.023	-0.022	0.040
rs13182883	0.056	-0.025	-0.046	-0.036
rs251934	0.044	-0.124	-0.086	0.015
rs338882	-0.096	-0.043	0.037	-0.052
rs13218440	0.015	-0.070	0.018	-0.137
rs1336071	0.052	0.027	0.080	-0.037
rs214955	0.126	-0.078	0.086	0.073
rs727811	0.130	0.019	-0.147	-0.062
rs6955448	0.327	-	-0.081	0.045
rs917118	0.005	-0.063	-0.049	-0.053
rs321198	-0.150	0.042	-0.081	-0.099
rs737681	0.084	-0.350	0.002	-0.034
rs763869	-0.100	0.056	-0.053	-0.087
rs10092491	0.127	-0.038	-0.016	-0.054
rs2056277	-0.052	-0.139	0.062	0.005
rs4606077	-0.063	0.033	0.004	-0.060
rs1015250	-0.018	-0.042	-0.009	-0.052
rs7041158	-0.191	0.112	0.001	-0.159
rs1463729	0.014	-0.061	-0.041	0.030
rs1360288	-0.011	0.021	-0.018	-0.027
rs10776839	0.004	0.033	-0.010	-0.047
rs826472	0.097	-0.013	0.045	-0.020
rs735155	-0.067	-0.278	0.012	-0.054
rs3780962	-0.070	0.007	-0.054	-0.054
rs740598	-0.074	0.133	-0.063	0.050
rs964681	0.042	0.082	-0.001	-0.146
rs1498553	0.070	0.009	-0.019	0.004
rs901398	0.060	0.060	-0.072	0.039
rs10488710	0.088	0.053	0.060	0.020
rs2076848	-0.082	0.039	-0.016	-0.005

rs2107612	0.010	0.052	-0.023	-0.007
rs2269355	0.014	0.037	-0.045	-0.102
rs2920816	-0.129	0.202	0.038	0.076
rs2111980	-0.014	0.112	-0.086	-0.020
rs10773760	0.100	-0.049	-0.023	-0.025
rs1335873	0.182	-0.201	0.051	0.075
rs1886510	-0.120	0.152	0.020	0.025
rs1058083	0.016	-0.201	0.034	0.016
rs354439	-0.084	0.158	0.016	-0.144
rs1454361	0.090	0.027	-0.030	0.004
rs722290	0.090	0.161	0.070	-0.076
rs873196	-0.037	-0.101	-0.064	0.103
rs4530059	-0.070	-0.223	-0.057	-0.066
rs1821380	-0.100	-0.249	0.084	0.105
rs8037429	-0.095	-0.125	0.097	-0.019
rs1528460	-0.099	-0.040	0.069	0.041
rs729172	-0.131	-0.128	-0.019	0.029
rs2342747	-0.125	0.089	-0.113	0.037
rs430046	-0.013	-0.026	0.035	0.039
rs1382387	0.071	-0.023	0.096	0.125
rs9905977	-0.062	-0.155	-0.063	0.098
rs740910	-0.011	-0.078	0.029	-0.044
rs938283	-0.091	-0.052	-0.075	0.111
rs8078417	-0.093	-0.250	-0.010	-0.119
rs1493232	0.120	-0.005	-0.051	0.130
rs9951171	0.038	-0.052	0.071	-0.057
rs1736442	-0.059	0.118	-0.069	0.071
rs1024116	-0.210	0.018	-0.048	-0.040
rs719366	-0.052	0.213	-0.019	0.075
rs576261	0.014	0.017	-0.077	-0.056
rs1031825	-0.071	0.175	0.058	-0.034
rs445251	0.012	-0.087	0.035	-0.005
rs1005533	-0.046	-0.094	-0.009	-0.042
rs1523537	0.026	-0.087	-0.040	0.037
rs722098	-0.006	0.014	0.022	0.184
rs2830795	-0.108	-0.005	-0.054	-0.004
rs2831700	-0.084	-	-0.125	-0.035
rs914165	0.011	-0.090	-0.062	-0.035
rs221956	-0.025	-0.142	-0.099	0.008
rs733164	0.044	0.091	0.078	0.097
rs987640	-0.057	-0.016	0.025	-0.031
rs2040411	0.154	-0.162	-0.092	0.013
rs1028528	0.046	-0.208	-0.005	-0.005
Average F_{IS}	-0.001	-0.025	-0.009	-0.001

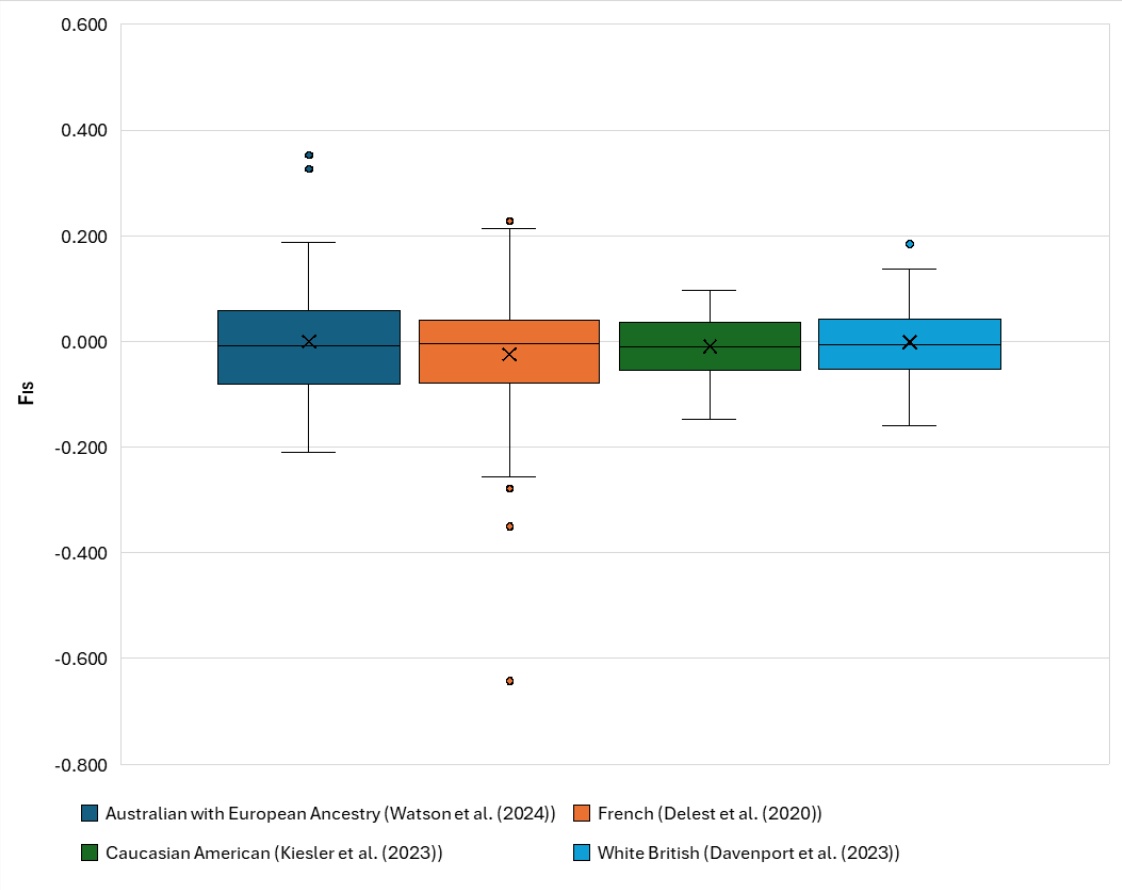


Figure 5.2 Locus-specific F_{IS} values ordered by population group.

5.5 Discussion and Conclusions

This analysis demonstrates that there is little difference between the F_{IS} values between the populations and that the means are close to zero, suggesting negligible inbreeding and that a θ correction for genotype frequencies close to zero is appropriate. There is also little difference between the F_{ST} values between the populations, but the means are in the range 0.04 to 0.05, suggesting some genetic drift between the populations. This means that allele frequency databases from one population with European ancestry will not necessarily be appropriate for another population with European ancestry for estimating LRs.

Further evaluation should be conducted of other ancestral groups such as African, South Asian, East Asian and Hispanic utilising available published data to determine the necessity for conducting population genetic studies within the relevant jurisdiction for populations with the same ancestry.[6-8, 11-16] Other population groups that have not been studied, including individuals with Polynesian, Aboriginal and Torres Strait Islander

ancestry, would require population genetic studies to be performed for the application of iiSNPs.

5.6 References

1. Wright S. The genetical structure of populations. Annals of eugenics. 1949;15(1):323-54. doi: 10.1111/j.1469-1809.1949.tb02451.x.
2. Steele CD, Court DS, Balding DJ. Worldwide Estimates Relative to Five Continental-Scale Populations. Ann Hum Genet. 2014;78(6):468-77. doi: 10.1111/ahg.12081.
3. Buckleton J, Curran J, Goudet J, Taylor D, Thiery A, Weir BS. Population-specific FST values for the forensic STR markers: A worldwide survey. Forensic Sci Int Genet. 2016;23:91-100. doi: 10.1016/j.fsigen.2016.03.004.
4. Wright S. Evolution and the genetics of populations, volume 4: variability within and among natural populations: University of Chicago press; 1978.
5. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 1984;38(6):1358-70. doi: 10.2307/2408641.
6. Delest A, Godfrin D, Chantrel Y, Ulus A, Vannier J, Faivre M, et al. Sequenced-based French population data from 169 unrelated individuals with Verogen's ForenSeq DNA signature prep kit. Forensic Sci Int Genet. 2020;47:102304. doi: 10.1016/j.fsigen.2020.102304.
7. Davenport L, Devesse L, Court DS, Ballard D. Forensic identity SNPs: Characterisation of flanking region variation using massively parallel sequencing. Forensic Sci Int Genet. 2023;64:102847. doi: 10.1016/j.fsigen.2023.102847.
8. Kiesler KM, Borsuk LA, Steffen CR, Vallone PM, Gettings KB. US Population Data for 94 Identity-Informative SNP Loci. Genes. 2023;14:1071. doi: 10.3390/genes14051071.
9. Watson JL, Grisedale K, Cho K, Ward J, McNevin D. Characterisation of Identity-Informative Genetic Markers in the Australian Population with European Ancestry [pre-print]. Available at SSRN 4854523. 2024.
10. R Core Team. R: A language and environment for statistical computing, Vienna, Austria. URL: https://www.R-project.org/. 2020.
11. Guevara EK, Palo JU, King JL, Buś MM, Guillen S, Budowle B, et al. Autosomal STR and SNP characterization of populations from the Northeastern Peruvian Andes with the ForenSeq™ DNA Signature Prep Kit. Forensic Sci Int Genet. 2021;52:102487. doi: 10.1016/j.fsigen.2021.102487.

12. Tao R, Dong X, Zhen X, Xia R, Qu Y, Liu S, et al. Population genetic analyses of Eastern Chinese Han nationality using ForenSeq™ DNA Signature Prep Kit. Molecular Genetics and Genomics. 2024;299(1):9. doi: 10.1007/s00438-024-02121-w.
13. Aguilar-Velázquez JA, Duran-Salazar MÁ, Córdoba-Mercado MF, Coronado-Avila CE, Salas-Salas O, Martinez-Cortés G, et al. Characterization of 58 STRs and 94 SNPs with the ForenSeq™ DNA signature prep kit in Mexican-Mestizos from the Monterrey city (Northeast, Mexico). Molecular Biology Reports. 2022;49(8):7601-9. doi: 10.1007/s11033-022-07575-y.
14. Grandell I, Samara R, Tillmar A. A SNP panel for identity and kinship testing using massive parallel sequencing. Int J Legal Med. 2016;130:905-14. doi: 10.1007/s00414-016-1341-4.
15. Augustinus D, Gahan ME, McNevin D. Development of a forensic identity SNP panel for Indonesia. Int J Legal Med. 2015;129:681-91. doi: 10.1007/s00414-014-1055-4.
16. Guo F, Liu Z, Long G, Zhang B, Liu D, Yu S. Performance and characterization of 94 identity-informative SNPs in Northern Han Chinese using ForenSeq™ DNA signature prep kit. J Forensic Legal Med. 2024;103:102678. doi: 10.1016/j.jflm.2024.102678.

5.7 Supplementary Material

Table S1. Allele population frequencies for the 94 identity-informative single nucleotide polymorphisms (iiSNPs) in the Australian population with European ancestry.

Locus	A	C	G	T	N
rs1490413	0.419	0.000	0.581	0.000	210
rs560681	0.719	0.000	0.281	0.000	210
rs1294331	0.000	0.643	0.000	0.357	210
rs10495407	0.291	0.000	0.710	0.000	210
rs891700	0.514	0.000	0.486	0.000	210
rs1413212	0.000	0.681	0.000	0.319	204
rs876724	0.000	0.700	0.000	0.300	210
rs1109037	0.576	0.000	0.424	0.000	210
rs993934	0.605	0.000	0.395	0.000	210
rs12997453	0.438	0.000	0.562	0.000	210
rs907100	0.000	0.429	0.571	0.000	210
rs1357617	0.524	0.000	0.000	0.476	210
rs4364205	0.000	0.000	0.552	0.448	210
rs2399332	0.000	0.000	0.700	0.300	210
rs1355366	0.000	0.381	0.000	0.619	210
rs6444724	0.000	0.419	0.000	0.581	210
rs2046361	0.519	0.000	0.000	0.481	210
rs279844	0.543	0.000	0.000	0.457	208
rs6811238	0.000	0.000	0.543	0.457	210
rs1979255	0.000	0.529	0.471	0.000	210
rs717302	0.500	0.000	0.500	0.000	210
rs159606	0.348	0.000	0.652	0.000	210
rs13182883	0.362	0.000	0.638	0.000	210
rs251934	0.610	0.000	0.391	0.000	210
rs338882	0.466	0.000	0.534	0.000	208
rs13218440	0.362	0.000	0.638	0.000	210
rs1336071	0.000	0.514	0.000	0.486	210
rs214955	0.000	0.576	0.000	0.424	210
rs727811	0.000	0.000	0.433	0.567	210
rs6955448	0.000	0.692	0.000	0.308	208
rs917118	0.000	0.724	0.000	0.276	210
rs321198	0.000	0.624	0.000	0.376	210
rs737681	0.000	0.610	0.000	0.391	210
rs763869	0.495	0.000	0.505	0.000	210
rs10092491	0.000	0.514	0.000	0.486	210
rs2056277	0.000	0.781	0.000	0.219	210
rs4606077	0.000	0.776	0.000	0.224	210
rs1015250	0.000	0.205	0.795	0.000	210
rs7041158	0.000	0.652	0.000	0.348	210
rs1463729	0.000	0.433	0.000	0.567	210

rs1360288	0.000	0.671	0.000	0.329	210
rs10776839	0.000	0.000	0.552	0.448	210
rs826472	0.000	0.638	0.000	0.362	210
rs735155	0.000	0.576	0.000	0.424	210
rs3780962	0.376	0.000	0.624	0.000	210
rs740598	0.586	0.000	0.414	0.000	210
rs964681	0.000	0.448	0.000	0.552	210
rs1498553	0.000	0.500	0.000	0.500	210
rs901398	0.000	0.238	0.000	0.762	210
rs10488710	0.000	0.524	0.476	0.000	210
rs2076848	0.519	0.000	0.000	0.481	210
rs2107612	0.710	0.000	0.291	0.000	210
rs2269355	0.000	0.514	0.486	0.000	210
rs2920816	0.602	0.000	0.398	0.000	206
rs2111980	0.000	0.548	0.000	0.452	210
rs10773760	0.548	0.000	0.452	0.000	210
rs1335873	0.581	0.000	0.000	0.419	210
rs1886510	0.510	0.000	0.491	0.000	210
rs1058083	0.305	0.000	0.695	0.000	210
rs354439	0.471	0.000	0.000	0.529	210
rs1454361	0.495	0.000	0.000	0.505	210
rs722290	0.000	0.514	0.486	0.000	210
rs873196	0.000	0.352	0.000	0.648	210
rs4530059	0.376	0.000	0.624	0.000	210
rs1821380	0.000	0.505	0.495	0.000	210
rs8037429	0.000	0.443	0.000	0.557	210
rs1528460	0.000	0.314	0.000	0.686	210
rs729172	0.000	0.000	0.605	0.395	210
rs2342747	0.324	0.000	0.676	0.000	210
rs430046	0.000	0.605	0.000	0.395	210
rs1382387	0.714	0.286	0.000	0.000	210
rs9905977	0.300	0.000	0.700	0.000	210
rs740910	0.671	0.000	0.329	0.000	210
rs938283	0.000	0.148	0.000	0.852	210
rs8078417	0.000	0.705	0.000	0.295	210
rs1493232	0.671	0.329	0.000	0.000	210
rs9951171	0.438	0.000	0.562	0.000	210
rs1736442	0.000	0.623	0.000	0.378	204
rs1024116	0.000	0.443	0.000	0.557	210
rs719366	0.605	0.000	0.395	0.000	210
rs576261	0.567	0.433	0.000	0.000	210
rs1031825	0.337	0.664	0.000	0.000	208
rs445251	0.000	0.476	0.524	0.000	210

rs1005533	0.471	0.000	0.529	0.000	210
rs1523537	0.000	0.462	0.000	0.538	210
rs722098	0.800	0.000	0.200	0.000	210
rs2830795	0.781	0.000	0.219	0.000	210
rs2831700	0.619	0.000	0.381	0.000	210
rs914165	0.376	0.000	0.624	0.000	210
rs221956	0.000	0.681	0.000	0.319	210
rs733164	0.310	0.000	0.691	0.000	210
rs987640	0.423	0.000	0.000	0.577	208
rs2040411	0.643	0.000	0.357	0.000	210
rs1028528	0.776	0.000	0.224	0.000	210

Table S2. Allele population frequencies for the 20 short tandem repeats (STRs) in the Australian population with European ancestry.

Locus	5	6	7	8	8.1	9	9.3	10	10.1	11	11.3	12	13	13.2
CSF1PO	0	0	0	0.014286	0	0.019048	0	0.228571	0	0.338095	0	0.280952	0.066667	0
D10S1248	0	0	0	0	0	0	0	0	0	0	0	0.038095	0.27619	0
D12S391	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D13S317	0	0	0	0.104762	0	0.057143	0	0.028571	0	0.328571	0	0.314286	0.128571	0
D16S539	0	0	0	0.009524	0	0.138095	0	0.07619	0	0.27619	0	0.271429	0.219048	0
D18S51	0	0	0	0	0	0	0	0.004762	0	0.009524	0	0.142857	0.133333	0
D19S433	0	0	0	0	0	0	0	0	0	0.004762	0	0.1	0.233333	0.014286
D1S1656	0	0	0	0	0	0	0	0	0	0.038462	0	0.158654	0.057692	0
D21S11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D22S1045	0	0	0	0	0	0	0	0	0	0.139423	0	0.004808	0	0
D2S1338	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D2S441	0	0	0	0	0	0.004762	0	0.180952	0.004762	0.380952	0.052381	0.033333	0.042857	0
D3S1358	0	0	0	0	0	0	0	0	0	0.004762	0	0	0.009524	0
D5S818	0	0	0	0.004808	0	0.009615	0	0.048077	0	0.418269	0	0.346154	0.163462	0
D7S820	0	0	0.042857	0.204762	0.004762	0.152381	0	0.219048	0	0.133333	0	0.185714	0.047619	0
D8S1179	0	0	0	0.028571	0	0.009524	0	0.090476	0	0.066667	0	0.157143	0.338095	0
FGA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TH01	0	0.190476	0.204762	0.090476	0	0.138095	0.37619	0	0	0	0	0	0	0
TPOX	0.004762	0	0	0.5	0	0.119048	0	0.057143	0	0.261905	0	0.057143	0	0
vWA	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Locus	14	14.2	14.3	15	15.2	15.3	16	16.3	17	17.1	17.3	18	18.3	19
CSF1PO	0.047619	0	0	0.004762	0	0	0	0	0	0	0	0	0	0
D10S1248	0.32381	0	0	0.180952	0	0	0.142857	0	0.028571	0	0	0	0	0.009524
D12S391	0	0	0	0.033333	0	0	0.014286	0	0.138095	0	0.014286	0.171429	0.014286	0.1
D13S317	0.038095	0	0	0	0	0	0	0	0	0	0	0	0	0
D16S539	0.009524	0	0	0	0	0	0	0	0	0	0	0	0	0
D18S51	0.17619	0	0	0.171429	0	0	0.114286	0	0.104762	0	0	0.066667	0	0.033333
D19S433	0.280952	0.028571	0	0.252381	0.052381	0	0.033333	0	0	0	0	0	0	0
D1S1656	0.067308	0	0.004808	0.125	0	0.043269	0.129808	0.057692	0.057692	0.004808	0.1875	0.004808	0.052885	0.004808
D21S11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D22S1045	0.028846	0	0	0.389423	0	0	0.331731	0	0.096154	0	0	0.009615	0	0
D2S1338	0	0	0	0	0	0	0.052381	0	0.195238	0	0	0.1	0	0.104762
D2S441	0.242857	0	0	0.052381	0	0	0.004762	0	0	0	0	0	0	0
D3S1358	0.128571	0	0	0.285714	0	0	0.2	0	0.233333	0	0	0.133333	0	0.004762
D5S818	0.009615	0	0	0	0	0	0	0	0	0	0	0	0	0
D7S820	0.009524	0	0	0	0	0	0	0	0	0	0	0	0	0
D8S1179	0.180952	0	0	0.104762	0	0	0.019048	0	0.004762	0	0	0	0	0
FGA	0	0	0	0	0	0	0	0	0	0	0	0.028571	0	0.085714
TH01	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TPOX	0	0	0	0	0	0	0	0	0	0	0	0	0	0
vWA	0.145631	0	0	0.087379	0	0	0.203883	0	0.218447	0	0	0.242718	0	0.082524

Locus	20	21	22	22.2	23	24	25	26	27	28	29	30
CSF1PO	0	0	0	0	0	0	0	0	0	0	0	0
D10S1248	0	0	0	0	0	0	0	0	0	0	0	0
D12S391	0.114286	0.128571	0.1	0	0.109524	0.02381	0.014286	0	0	0	0	0
D13S317	0	0	0	0	0	0	0	0	0	0	0	0
D16S539	0	0	0	0	0	0	0	0	0	0	0	0
D18S51	0.028571	0.004762	0.004762	0	0.004762	0	0	0	0	0	0	0
D19S433	0	0	0	0	0	0	0	0	0	0	0	0
D1S1656	0	0	0	0	0	0	0	0	0	0	0	0
D21S11	0	0	0	0	0	0	0	0	0.028846	0.149038	0.192308	0.283654
D22S1045	0	0	0	0	0	0	0	0	0	0	0	0
D2S1338	0.17619	0.042857	0.028571	0	0.109524	0.1	0.080952	0.004762	0.004762	0	0	0
D2S441	0	0	0	0	0	0	0	0	0	0	0	0
D3S1358	0	0	0	0	0	0	0	0	0	0	0	0
D5S818	0	0	0	0	0	0	0	0	0	0	0	0
D7S820	0	0	0	0	0	0	0	0	0	0	0	0
D8S1179	0	0	0	0	0	0	0	0	0	0	0	0
FGA	0.157143	0.190476	0.17619	0.014286	0.12381	0.119048	0.085714	0.019048	0	0	0	0
TH01	0	0	0	0	0	0	0	0	0	0	0	0
TPOX	0	0	0	0	0	0	0	0	0	0	0	0
vWA	0.019417	0	0	0	0	0	0	0	0	0	0	0

Locus	30.2	31	31.2	32	32.2	33.2	35.2	N
CSF1PO	0	0	0	0	0	0	0	210
D10S1248	0	0	0	0	0	0	0	210
D12S391	0	0	0	0	0	0	0	210
D13S317	0	0	0	0	0	0	0	210
D16S539	0	0	0	0	0	0	0	210
D18S51	0	0	0	0	0	0	0	210
D19S433	0	0	0	0	0	0	0	210
D1S1656	0	0	0	0	0	0	0	208
D21S11	0.033654	0.072115	0.096154	0.014423	0.072115	0.052885	0.004808	208
D22S1045	0	0	0	0	0	0	0	208
D2S1338	0	0	0	0	0	0	0	210
D2S441	0	0	0	0	0	0	0	210
D3S1358	0	0	0	0	0	0	0	210
D5S818	0	0	0	0	0	0	0	208
D7S820	0	0	0	0	0	0	0	210
D8S1179	0	0	0	0	0	0	0	210
FGA	0	0	0	0	0	0	0	210
TH01	0	0	0	0	0	0	0	210
TPOX	0	0	0	0	0	0	0	210
vWA	0	0	0	0	0	0	0	206

Table S3. Forensic parameters and Hardy-Weinberg equilibrium p-values for short tandem repeats (STRs).

Locus	Chr	Position (cM)	PM	PD	Hobs	Hexp	PIC	PE	TPI	p-value
D3S1358	3	67.049	0.084	0.916	0.838	0.793	0.757	0.672	3.088	0.455
vWA	12	15.412	0.060	0.940	0.767	0.820	0.790	0.539	2.146	0.187
D16S539	16	125.578	0.089	0.911	0.790	0.781	0.741	0.581	2.386	0.205
CSF1PO	5	154.386	0.108	0.892	0.676	0.751	0.705	0.392	1.544	0.193
TPOX	2	1.668	0.181	0.819	0.762	0.664	0.613	0.530	2.100	0.449
D8S1179	8	135.203	0.079	0.921	0.800	0.807	0.780	0.599	2.500	0.462
D21S11	21	14.636	0.054	0.946	0.846	0.840	0.817	0.687	3.250	0.323
D18S51	18	84.630	0.039	0.961	0.876	0.875	0.857	0.747	4.038	0.238
D2S441	2	94.709	0.102	0.898	0.771	0.758	0.720	0.547	2.188	0.147
D19S433	19	51.725	0.082	0.918	0.800	0.792	0.756	0.599	2.500	0.328
TH01	11	3.782	0.103	0.897	0.790	0.757	0.716	0.581	2.386	0.113
FGA	4	159.115	0.041	0.959	0.857	0.867	0.847	0.709	3.500	0.909
D22S1045	22	45.240	0.129	0.871	0.712	0.712	0.659	0.446	1.733	0.909
D5S818	5	126.909	0.165	0.835	0.673	0.679	0.617	0.388	1.529	0.825
D13S317	13	78.633	0.097	0.903	0.810	0.764	0.725	0.617	2.625	0.327
D7S820	7	101.330	0.059	0.941	0.857	0.834	0.808	0.709	3.500	0.286
D10S1248	10	171.638	0.101	0.899	0.781	0.767	0.726	0.564	2.283	0.575
D1S1656	1	248.508	0.030	0.970	0.875	0.891	0.876	0.745	4.000	0.790
D12S391	12	27.417	0.027	0.973	0.876	0.891	0.876	0.747	4.038	0.941
D2S1338	2	228.168	0.033	0.967	0.857	0.880	0.863	0.709	3.500	0.413

Chr: chromosome
cM: centimorgan
PM: probability of matching
PD: power of discrimination
Hobs: observed heterozygosity
Hexp: expected heterozygosity
PIC: polymorphism information content
PE: power of exclusion
TPI: typical paternity index

Table S4. Forensic parameters and Hardy-Weinberg equilibrium p-values for identity-informative single nucleotide polymorphisms (iiSNPs).

Locus	Chr	Position (cM)	PM	PD	Hobs	Hexp	PIC	PE	TPI	p-value
rs1490413	1	10.344	0.431	0.569	0.571	0.489	0.368	0.258	1.167	0.437
rs560681	1	173.519	0.434	0.566	0.390	0.406	0.322	0.108	0.820	1.000
rs1294331	1	252.689	0.401	0.599	0.467	0.461	0.354	0.160	0.938	0.607
rs10495407	1	264.510	0.426	0.574	0.390	0.414	0.327	0.108	0.820	1.000
rs891700	1	266.757	0.373	0.627	0.495	0.502	0.375	0.183	0.991	1.000
rs1413212	1	275.112	0.417	0.583	0.441	0.436	0.340	0.141	0.895	1.000
rs876724	2	0.054	0.427	0.573	0.429	0.422	0.332	0.132	0.875	1.000
rs1109037	2	25.846	0.359	0.641	0.429	0.491	0.369	0.132	0.875	0.175
rs993934	2	143.139	0.360	0.640	0.390	0.480	0.364	0.108	0.820	0.139
rs12997453	2	196.669	0.413	0.587	0.552	0.495	0.371	0.238	1.117	0.621
rs907100	2	261.368	0.367	0.633	0.457	0.492	0.370	0.153	0.921	0.449
rs1357617	3	1.267	0.335	0.665	0.324	0.501	0.374	0.074	0.739	0.004
rs4364205	3	56.460	0.438	0.562	0.590	0.497	0.372	0.280	1.221	0.041
rs2399332	3	120.167	0.448	0.552	0.486	0.422	0.332	0.175	0.972	0.562
rs1355366	3	209.799	0.385	0.615	0.457	0.474	0.360	0.153	0.921	0.454
rs6444724	3	214.028	0.386	0.614	0.495	0.489	0.368	0.183	0.991	0.464
rs2046361	4	26.496	0.343	0.657	0.410	0.502	0.375	0.120	0.847	0.458
rs279844	4	68.752	0.352	0.648	0.433	0.499	0.373	0.135	0.881	0.121
rs6811238	4	174.391	0.368	0.632	0.476	0.499	0.373	0.168	0.955	0.825
rs1979255	4	213.055	0.344	0.656	0.410	0.501	0.374	0.120	0.847	0.077
rs717302	5	6.712	0.399	0.601	0.543	0.502	0.375	0.228	1.094	0.461
rs159606	5	33.526	0.434	0.566	0.524	0.456	0.351	0.209	1.050	0.477
rs13182883	5	139.768	0.388	0.612	0.438	0.464	0.355	0.139	0.890	0.780
rs251934	5	191.986	0.380	0.620	0.457	0.478	0.363	0.153	0.921	0.807
rs338882	5	199.640	0.405	0.595	0.548	0.500	0.374	0.233	1.106	1.000
rs13218440	6	26.505	0.394	0.606	0.457	0.464	0.355	0.153	0.921	1.000
rs1336071	6	100.651	0.364	0.636	0.476	0.502	0.375	0.168	0.955	1.000
rs214955	6	159.848	0.359	0.641	0.429	0.491	0.369	0.132	0.875	1.000
rs727811	6	180.057	0.356	0.644	0.429	0.493	0.371	0.132	0.875	0.323
rs6955448	7	6.160	0.410	0.590	0.288	0.428	0.335	0.059	0.703	0.012
rs917118	7	7.494	0.440	0.560	0.400	0.402	0.320	0.114	0.833	1.000
rs321198	7	145.378	0.430	0.570	0.543	0.472	0.359	0.228	1.094	0.074
rs737681	7	181.920	0.374	0.626	0.438	0.478	0.363	0.139	0.890	1.000
rs763869	8	1.957	0.405	0.595	0.552	0.502	0.375	0.238	1.117	0.818
rs10092491	8	56.017	0.350	0.650	0.438	0.502	0.375	0.139	0.890	1.000
rs2056277	8	156.441	0.492	0.508	0.362	0.344	0.284	0.092	0.784	1.000
rs4606077	8	166.567	0.488	0.512	0.371	0.349	0.287	0.097	0.795	1.000
rs1015250	9	4.302	0.508	0.492	0.333	0.327	0.273	0.078	0.750	1.000
rs7041158	9	53.006	0.446	0.554	0.543	0.456	0.351	0.228	1.094	0.410
rs1463729	9	136.053	0.377	0.623	0.486	0.493	0.371	0.175	0.972	1.000

rs1360288	9	137.914	0.412	0.588	0.448	0.443	0.344	0.146	0.905	1.000
rs10776839	9	155.845	0.378	0.622	0.495	0.497	0.372	0.183	0.991	1.000
rs826472	10	3.569	0.382	0.618	0.419	0.464	0.355	0.126	0.861	0.129
rs735155	10	6.796	0.399	0.601	0.524	0.491	0.369	0.209	1.050	1.000
rs3780962	10	38.187	0.408	0.592	0.505	0.472	0.359	0.192	1.010	0.466
rs740598	10	143.729	0.402	0.598	0.524	0.488	0.368	0.209	1.050	0.466
rs964681	10	175.669	0.369	0.631	0.476	0.497	0.372	0.168	0.955	0.606
rs1498553	11	11.572	0.360	0.640	0.467	0.502	0.375	0.160	0.938	0.146
rs901398	11	20.235	0.471	0.529	0.343	0.365	0.297	0.083	0.761	0.725
rs10488710	11	119.996	0.357	0.643	0.457	0.501	0.374	0.153	0.921	0.633
rs2076848	11	157.844	0.400	0.600	0.543	0.502	0.375	0.228	1.094	0.318
rs2107612	12	2.140	0.430	0.570	0.410	0.414	0.327	0.120	0.847	0.745
rs2269355	12	17.707	0.373	0.627	0.495	0.502	0.375	0.183	0.991	1.000
rs2920816	12	56.272	0.420	0.580	0.544	0.482	0.364	0.229	1.096	1.000
rs2111980	12	124.518	0.382	0.618	0.505	0.498	0.373	0.192	1.010	0.819
rs10773760	12	168.442	0.357	0.643	0.448	0.498	0.373	0.146	0.905	0.157
rs1335873	13	2.118	0.353	0.647	0.400	0.489	0.368	0.114	0.833	0.024
rs1886510	13	4.799	0.412	0.588	0.562	0.502	0.375	0.248	1.141	0.137
rs1058083	13	94.111	0.421	0.579	0.419	0.426	0.334	0.126	0.861	1.000
rs354439	13	107.295	0.401	0.599	0.543	0.501	0.374	0.228	1.094	0.316
rs1454361	14	17.199	0.356	0.644	0.457	0.502	0.375	0.153	0.921	0.229
rs722290	14	47.503	0.357	0.643	0.457	0.502	0.375	0.153	0.921	0.627
rs873196	14	104.004	0.408	0.592	0.476	0.459	0.352	0.168	0.955	0.812
rs4530059	14	114.517	0.408	0.592	0.505	0.472	0.359	0.192	1.010	0.788
rs1821380	15	53.240	0.405	0.595	0.552	0.502	0.375	0.238	1.117	0.610
rs8037429	15	64.450	0.406	0.594	0.543	0.496	0.372	0.228	1.094	0.339
rs1528460	15	66.372	0.433	0.567	0.476	0.433	0.338	0.168	0.955	0.433
rs729172	16	11.313	0.421	0.579	0.543	0.480	0.364	0.228	1.094	0.792
rs2342747	16	11.861	0.435	0.565	0.495	0.440	0.342	0.183	0.991	0.447
rs430046	16	97.209	0.390	0.610	0.486	0.480	0.364	0.175	0.972	0.793
rs1382387	16	103.726	0.429	0.571	0.381	0.410	0.325	0.103	0.808	0.566
rs9905977	17	8.280	0.433	0.567	0.448	0.422	0.332	0.146	0.905	1.000
rs740910	17	13.409	0.412	0.588	0.448	0.443	0.344	0.146	0.905	1.000
rs938283	17	120.308	0.587	0.413	0.276	0.253	0.220	0.054	0.691	1.000
rs8078417	17	127.751	0.440	0.560	0.457	0.418	0.330	0.153	0.921	0.772
rs1493232	18	3.667	0.397	0.603	0.390	0.443	0.344	0.108	0.820	0.631
rs9951171	18	28.534	0.372	0.628	0.476	0.495	0.371	0.168	0.955	1.000
rs1736442	18	74.557	0.405	0.595	0.500	0.472	0.360	0.188	1.000	0.773
rs1024116	18	112.789	0.447	0.553	0.600	0.496	0.372	0.291	1.250	0.231
rs719366	19	49.407	0.399	0.601	0.505	0.480	0.364	0.192	1.010	1.000
rs576261	19	63.837	0.377	0.623	0.486	0.493	0.371	0.175	0.972	0.792
rs1031825	20	12.795	0.419	0.581	0.481	0.449	0.347	0.171	0.963	0.757
rs445251	20	35.366	0.374	0.626	0.495	0.501	0.374	0.183	0.991	0.600

rs1005533	20	58.015	0.389	0.611	0.524	0.501	0.374	0.209	1.050	1.000
rs1523537	20	77.584	0.371	0.629	0.486	0.499	0.374	0.175	0.972	0.469
rs722098	21	4.540	0.513	0.487	0.324	0.322	0.269	0.074	0.739	1.000
rs2830795	21	27.348	0.495	0.505	0.381	0.344	0.284	0.103	0.808	1.000
rs2831700	21	29.397	0.411	0.589	0.514	0.474	0.360	0.200	1.029	0.613
rs914165	21	50.554	0.391	0.609	0.467	0.472	0.359	0.160	0.938	0.613
rs221956	21	54.769	0.418	0.582	0.448	0.437	0.340	0.146	0.905	1.000
rs733164	22	31.366	0.415	0.585	0.410	0.429	0.336	0.120	0.847	0.777
rs987640	22	37.654	0.397	0.603	0.519	0.491	0.369	0.205	1.040	0.592
rs2040411	22	62.887	0.379	0.621	0.390	0.461	0.354	0.108	0.820	1.000
rs1028528	22	64.137	0.486	0.514	0.333	0.349	0.287	0.078	0.750	0.193

6. CHAPTER SIX

DNA INTELLIGENCE USING SEX-CHROMOSOME, PHENOTYPE-INFORMATIVE AND ANCESTRY-INFORMATIVE MARKERS IN AN AUSTRALIAN POPULATION

Contributions of authors:

Watson JL, Grisedale K, Ward J and McNevin D. (2025) 'DNA Intelligence Using Sex-Chromosome, Phenotype-Informative and Ancestry-Informative Markers in an Australian Population'. Australian Journal of Forensic Sciences, 1-18. DOI: 10.1080/00450618.2025.2491376.

Jessica Watson (Candidate)

Conceptualised and assisted in the experimental design for the study. Carried out sample collection, sample preparation, DNA testing and data analysis. Prepared manuscript, edited manuscript following review by other co-authors, referenced manuscript, carried out manuscript's submission and incorporated feedback from peer review process.

Signed: Production Note: Signature removed prior to publication.

Kelly Grisedale (AFP DNA Program staff)

Conceptualised and assisted in the experimental design for the study. Contributed to DNA testing and data analysis. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Jodie Ward (Co-supervisor)

Conceptualised and assisted in the experimental design for the study. Acquired funding for the study. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Dennis McNevin (Co-supervisor)

Conceptualised and assisted in the experimental design for the study, curated reference population data and designed the analysis pipelines used in this study. Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

AUSTRALIAN JOURNAL OF FORENSIC SCIENCES
https://doi.org/10.1080/00450618.2025.2491376

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

DNA intelligence using sex-chromosome, phenotype-informative and ancestry-informative markers in an Australian population

Jessica L Watson [a,b,c], Kelly Grisedale [a,b], Jodie Ward [b,c] and Dennis McNevin [b,c]

[a]Forensics Command, Australian Federal Police, Canberra, Australia; [b]National DNA Program for Unidentified and Missing Persons, Australian Federal Police, Canberra, Australia; [c]Centre for Forensic Science, School of Mathematical & Physical Science, Faculty of Science, University of Technology Sydney, Sydney, Australia

ABSTRACT

Single nucleotide polymorphism (SNP) data can be used to infer the biological sex, externally visible characteristics (EVCs) and biogeographical ancestry (BGA) of an unknown individual. There are multiple pipelines available that can be used to generate these inferences and provide investigative leads for law enforcement to pursue. It is important for inference pipelines to be evaluated within a population representative of the intended jurisdiction prior to casework implementation. This study presents the performance of several pipelines using an Australian study population with self-declared biological sex, eye colour, hair colour and recent ancestry. The proportion of consistent results for EVC inference was higher for the HIrisPlex online tool using published interpretation guidelines for eye colour (97%) and hair colour (80%) when compared to the MiSeq FGx® Universal Analysis Software (UAS) for eye colour (74%) and hair colour (69%). For inferring BGA, a principle coordinate analysis pipeline produced the most consistent results when compared to self-declared data (86%). This was improved to 90% when inconclusive results obtained from admixed individuals were analysed with Structure. This study highlights the strengths and limitations of multiple inference pipelines to assist in the development of interpretation and reporting guidelines for Australian applications.

ARTICLE HISTORY
Received 3 February 2025
Accepted 23 March 2025

KEYWORDS
Biogeographical ancestry; single nucleotide polymorphism; externally visible characteristics; biological sex; forensic DNA phenotyping

1. Introduction

Forensic genomics can be utilized to generate DNA intelligence when database searches fail to identify a DNA sample of an unknown individual, thus providing new investigative leads[1,2]. Single nucleotide polymorphisms (SNPs) are single base pair variants that can be used to infer the biological sex, externally visible characteristics (EVCs) and biogeographical ancestry (BGA) of an individual. Using massively parallel sequencing (MPS), it is possible to sequence millions of reads from multiple samples in a single run and derive

CONTACT Jessica L Watson Jessica.Watson@student.uts.edu.au
Supplemental data for this article can be accessed online at https://doi.org/10.1080/00450618.2025.2491376.
© 2025 Crown Copyright. Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

a single DNA genotype containing various classes of SNPs that can be analysed to generate various intelligence information[3].

For the generation of DNA intelligence, phenotype-informative SNPs (piSNPs) and ancestry-informative SNPs (aiSNPs) are increasingly considered to offer investigative leads for forensic casework applications. For example, by inferring an unknown person's biological sex, eye colour, hair colour and recent ancestry, physical traits such as pigmentation can be incorporated into a craniofacial reconstruction of unidentified human remains, narrowing the pool of potential candidates[4–6]. There are several forensic DNA panels available that assess multiple classes of SNPs, allowing for these inferences to be generated in a single workflow[7–11].

The biological sex of an unknown individual can be inferred through routine forensic DNA techniques such as DNA quantification and short tandem repeat (STR) typing. A DNA quantification workflow typically includes a male-specific target from the Y chromosome to indicate the presence of male DNA[12]. Autosomal STR typing kits incorporate biological sex inference through the amplification of an insertion/deletion (indel) in the Amelogenin gene, which has different variants on the X and Y chromosomes, as well as the inclusion of Y chromosome STRs (Y STRs)[13]. Many MPS panels also include a selection of Y STRs and/or X chromosome STRs (X STRs) to infer the biological sex of the DNA donor[8–10].

In order to infer an individual's BGA, ancestry-informative SNPs (aiSNPs) with low heterozygosity and high population heterogeneity are required[14]. Ideally, these SNPs have alleles that are shared by individuals within a population group but not with other population groups[14,15]. BGA inference requires a panel of these aiSNPs, a reference database consisting of populations with genotypes of individuals who have known ancestry and a prediction algorithm that compares an unknown person's genotype to the reference database. The Kidd Lab Panel of 55 aiSNPs is the most widely used set for BGA inference, and its utility has been demonstrated using a number of prediction algorithms[14,16].

Multidimensional scaling (MDS) or dimensionality reduction methods use eigenvalue decomposition to reduce the genotypes of a collection of individuals to two or three coordinates in a two- or three-dimensional (2D or 3D) space to explain the variance amongst genotypes[17]. Principle component analysis (PCA) requires a numerical representation of genotypes and is limited to biallelic genotypes to preserve the genetic distances between variants[18,19]. The resulting scatter plot visualizes the samples and their genetic distances from one another based on the derived coordinates. The Universal Analysis Software (UAS; Verogen) has an inbuilt PCA algorithm that generates a 2D plot[18]. In this plot, individuals with genetic similarity will cluster together, reflecting different population groups.

However, clusters that cannot be differentiated in a 2D plot, like that produced by the UAS, may become differentiated by plotting a third dimension. Principle coordinate analysis (PCoA) requires an input matrix of genetic distances which can account for tri- and tetra-allelic SNPs, allowing for more flexibility than the PCA method[20]. As for PCA, a 2D or 3D plot can then be generated and individuals with genetic similarity will cluster together.

Another category of BGA prediction algorithms are model-based likelihood estimators which are more appropriate for inferring admixture[17]. Structure is an algorithm that estimates the proportion of genetic contributions to a matrix of genotypes from

K multiple ancestral population groups[21,22]. Using a Bayesian updating framework, Markov chain Monte Carlo (MCMC) simulations adjust model parameters until the likelihood function for the genotype matrix is maximized[21,22]. The estimated ancestral population contributions for the questioned genotype are compared with those for reference genotypes with known ancestry in order to infer the BGA of the questioned genotype[17].

The Forensic Research/Reference on Genetics-knowledge base (FROG-kb) is a forensic application of the Allele FREquency Database (ALFRED) that calculates the probability of observing a genotype within each population in the database, commonly known as a random match probability (RMP)[23,24]. The populations are ranked by RMP, with genotypes more likely to be observed in populations exhibiting higher RMPs.

The inference of externally visible characteristics (EVCs) has similar requirements to BGA: a panel of piSNPs, a reference database consisting of genotypes of individuals with known phenotypes, and a prediction algorithm. For EVC inference, the HIrisPlex panel, which consists of 24 piSNPs, can be used to infer hair and eye colour[25–27]. EVCs are inferred using a multinomial logistic regression (MLR) model that associates categorical phenotype (i.e. hair colour, hair shade and eye colour) with reference genotypes that have known EVCs. The probability of each phenotype is reported as a p-value (not to be confused with statistical significance). The UAS has an inbuilt MLR algorithm with a fixed reference database of individuals with known EVCs and reports p-values for hair and eye colours[18]. Similarly, the HIrisPlex System also uses MLR and a private database to report p-values for hair and eye colours, as well as hair shades[25–27].

In this study, DNA from several individuals with varying eye colour, hair colour and ancestry was genotyped to evaluate the application of these inference pipelines in an Australian population. The inferences were compared against the self-declared information provided by the volunteers to assess the suitability of various inference algorithms for generating reliable and actionable DNA intelligence that could assist in identifying an unknown individual.

2. Methods

2.1. Ethics approval and sample procurement

This research was approved by the University of Technology Sydney (UTS) Human Research Ethics Committee (HREC; UTS HREC NO. ETH21–5821). Volunteers provided self-administered buccal swabs and completed a questionnaire to provide their self-declared biological sex, eye colour, hair colour (at 20 years old) and ancestry (of themselves, their parents and their grandparents; Supplementary Material 1). DNA was extracted following the manufacturer's recommended protocols with the EZ1® DNA Investigator Kit (QIAGEN) on the EZ1® Advanced XL (QIAGEN)[28]. Extracted DNA was subject to DNA quantification, STR profiling and SNP genotyping as indicated in Table S1.

2.2. DNA quantification

All samples were quantified using the Quantifiler™ Trio DNA Quantification Kit (Thermo Fisher Scientific) on the QuantStudio™ 5 Real-Time PCR System (Thermo Fisher Scientific) following the manufacturer's recommended protocols[12,29]. This kit quantified 80 bp small autosomal (SA), 214 bp large autosomal (LA) and 75 bp male-specific targets and a degradation index (DI) was calculated from the ratio of SA and LA target concentrations.

2.3. STR profiling

Some samples underwent STR profiling using the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific; $n = 25$)[13]. Amplification was performed on the Veriti™ 96-Well Fast Thermal Cycler (Thermo Fisher Scientific) with a 29 cycle protocol. The SA target concentration was used to calculate the required input volume of extracted DNA for a 1.0 ng DNA input template amount. Capillary electrophoresis was performed on the 3500×L Genetic Analyser (Thermo Fisher Scientific) and genotyped with GeneMapper™ *ID-X* v1.6[30,31]. The analytical threshold was 225 RFU and the homozygous threshold was 1000 RFU.

2.4. Library preparation and sequencing with the ForenSeq® DNA Signature Prep Kit

The ForenSeq® DNA Signature Prep Kit (QIAGEN) targets 24 piSNPs, 56 aiSNPs (of which two are also piSNPs), 24 Y STRs, 7 X STRs, 27 autosomal STRs, 94 identity-informative SNPs (iiSNPs) and Amelogenin[10]. Samples were diluted based on the SA target concentration to deliver 1.0 ng in 5 µL (0.2 ng/µL) and underwent manual library preparation following the manufacturer's recommended protocol with primer mix B ($n = 57$). Samples were prepared in batches of 14 samples with a positive control (PC; 2800 M Control DNA (Promega)) and a negative control (NC; nuclease-free water). Sequencing was performed on the MiSeq® FGx Sequencing System (QIAGEN) using the MiSeq® FGx Reagent Kit (QIAGEN) with a standard flow cell (SFC)[32,33]. The results were analysed on the UAS v1.3 using the default analytical and interpretation thresholds and exported in the Sample Details Report and Phenotype Estimation Report[34].

2.5. Library preparation and sequencing with the ForenSeq® Kintelligence Kit

The ForenSeq® Kintelligence Kit (QIAGEN) targets the same piSNPs, aiSNPs and iiSNPs as the ForenSeq® DNA Signature Prep Kit, as well as an additional 85 Y SNPs, 106 X SNPs and 9,687 kinship-informative SNPs[9,35]. Samples were diluted based on the SA target concentration to deliver 1.0 ng in 25 µL (0.04 ng/µL) and libraries were prepared following a modified protocol ($n = 16$)[36]. Samples were prepared in batches of up to 12 samples with a PC (NA24385 Control DNA (Coriell Institute)) and NC (nuclease-free water) and sequenced in batches of 3 samples on the MiSeq® FGx Sequencing System with the MiSeq® FGx Reagent Kit

Table 1. Criteria for inferring biological sex for each genotyping assay. The criteria for the ForenSeq® DNA signature prep kit and ForenSeq® kintelligence kit are as defined by the manufacturer[18,34].

Method	Male	Female	Inconclusive
Quantifiler™ Trio DNA Quantification Kit[12]	Y chromosome target concentration above detection threshold	Y chromosome target concentration below detection threshold	All targets below detection threshold
GlobalFiler™ PCR Amplification Kit[13]	Amelogenin typed XY and alleles typed at DYS391 and Y indel	Amelogenin typed XX and no alleles typed at DYS391 and Y indel	No alleles detected or the contributor status is a mixture
ForenSeq® DNA Signature Prep Kit[18]	≥3 X STRs and ≥ 3 Y STRs typed	≥3 X STRs and < 3 Y STRs typed	Negative control, the contributor status is a mixture, < 3 X STRs typed or < 3 Y STRs typed
ForenSeq® Kintelligence Kit[34]	≥10 Y SNPs typed	No Y SNPs typed and call rate ≥ 50%	Negative control, the contributor status is a mixture or male and female criteria are not met

and SFC. The results were exported from the UAS v2.5 in Sample Reports and analysed using published optimized thresholds with a Microsoft Excel macro[36]. The SNP profiles were manually edited on the UAS to be consistent with the genotypes generated with the optimized thresholds and the Phenotype and Ancestry Reports were exported.

2.6. Inference of biological sex

Biological sex was inferred from results obtained from quantification, STR profiling and SNP profiling. Table 1 defines the criteria for inferring whether the DNA donor was biologically male or female, or whether the biological sex was inconclusive.

2.7. Inference of hair and eye colour

EVCs were inferred for each sample using either the ForenSeq® DNA Signature Prep Kit or the ForenSeq® Kintelligence Kit as the same 24 piSNPs are targeted by both kits. The in-built UAS MLR pipeline was assessed using the exported Phenotype Estimation Report (UAS v1.3, ForenSeq® DNA Signature Prep Kit) and Phenotype & Ancestry Report (UAS v2.5, ForenSeq® Kintelligence Kit). The EVC inferences were made with maximum p-value for each category of eye colour (blue, intermediate and brown) and hair colour (red, blond, brown and black). P-values were unable to be generated by the UAS unless all 24 piSNPs were typed.

The piSNPs were uploaded to the HIrisPlex online tool to report p-values and area under the receiver operating characteristic curve (AUC) values to account for information loss in partial profiles[37]. As a result, samples that only yielded partial piSNP profiles were able to be analysed with the HIrisPlex online tool. Hair colour was inferred from the p-values for the hair colours (red, blond, brown and black) and hair shade (light and dark) using the Enhanced Model Version 1 Prediction Guide by Walsh et al.; the final inferences were either red, blond, blond or brown, brown, brown or black or black[27]. If maximum eye colour p-value exceeded 0.9, that eye colour was inferred. However, if the maximum eye colour p-value was less

Table 2. Counts of self-declared hair colours of volunteers.

Hair Colour	Count
Red	2
Blond	5
Dark Blond	11
Brown	32
Dark Brown	20
Black	3

Table 3. Counts of self-declared eye colours of volunteers.

Eye Colour	Count
Blue	30
Grey[a]	2
Green[b]	8
Hazel[b]	9
Brown	24

[a]Categorised as blue.
[b]Categorised as intermediate.

Table 4. Performance metrics used to assess EVC inferences. p = number of volunteers with a particular EVC (positive); N = number of volunteers without a particular EVC (negative); TP = number of P for which EVC was correctly inferred (true positive); TN = number of N for which EVC was correctly inferred (true negative); FN = number of P for which EVC was incorrectly inferred (false negative); FP = number of N for which EVC was incorrectly inferred (false positive).

Metric	Formula
Sensitivity or true positive rate (TPR)	$\frac{TP}{P}$
Specificity or true negative rate (TNR)	$\frac{TN}{N}$
Positive predictive value (PPV)	$\frac{TP}{TP+FP}$
Negative predictive value (NPV)	$\frac{TN}{TN+FN}$
Balanced accuracy	$\frac{1}{2}\left(\frac{TP}{P}+\frac{TN}{N}\right)$

than 0.9, it was inferred that the DNA donor could have that eye colour or an intermediate eye colour.

The EVC inferences were compared to the self-declared hair and eye colours of the volunteers for consistency (Tables 2 and 3). The UAS and HIrisPlex pipelines were assessed by calculating the performance metrics in Table 4.

2.8. Inference of biogeographical ancestry

BGA was inferred for each sample using either the ForenSeq® DNA Signature Prep Kit or ForenSeq® Kintelligence Kit as the same 56 aiSNPs are targeted by both kits.

Table 5. Population groups and inclusion criteria for each biogeographical ancestry (BGA) inference method.

Method	Population Groups	Inclusion Criteria	Inconclusive Criteria
Universal Analysis Software (UAS)	African, East Asian and European[a]	Questioned genotype is closer to the population cluster centroid than at least one other reference genotype from that population	Questioned genotype is further from the population cluster centroid than all other reference genotypes from that population
Principle Coordinate Analysis (PCoA)	Sub-Saharan African, Middle Eastern/North African, Oceanian, American, European, South Asian and East Asian	Questioned genotype is closer to the population cluster centroid than at least one other reference genotype from that population	Questioned genotype is further from the population cluster centroid than all other reference genotypes from that population
Structure	Sub-Saharan African, Middle Eastern/North African, Oceanian, American, European, South Asian and East Asian	A population is inferred as a major (> 50%) or a minor (10–50%) contributor	N/A
Forensic Resource Reference on Genetics – Knowledge Base (FROG-kb)	Sub-Saharan African, Middle Eastern/North African, Oceanian, American, European, South Asian, East Asian, Asian[b]	Populations included until the RMP decreases by at least a factor of 3	N/A
Optimised Pipeline	Sub-Saharan African, Middle Eastern/North African, Oceanian, American, European, South Asian and East Asian	If sample is inconclusive with PCoA, the sample is analysed with Structure	N/A

[a]Admixed American not considered in this study.
[b]The 161 populations on FROG-kb for the KiddLab 55 aiSNPs were organized into eight population groups based on geography and ethnicity. See Table S2 for population groupings.

Inferences were generated using four pipelines: UAS (PCA), PCoA, Structure and FROG-kb (Table 5). The in-built PCA algorithm on the UAS plots the sample in relation to three population groups (African, East Asian and European); however, the Admixed American group was determined not to be relevant to an Australian population and excluded from this study[18]. Two principle coordinates were considered.

For PCoA and Structure, reference population data was compiled from 2,262 individuals with known ancestries from 1000 Genomes, the HGDP-CEPH database and the Simons Genome Diversity Project[38–41]. PCoA was performed using the 'ape: Analyses of Phylogenetics and Evolution' package in R and three principle coordinates were considered[42,43]. The reference and questioned genotypes were imported into the Structure software and analysed with the following parameter settings: 10000 burnin repetitions 10,000 MCMC repetitions after burnin, Admixture Model, Allele Frequencies Correlated and computation of the probability of the data (for estimating K, the number of ancestral populations)[21,22,44]. When running the simulations, K was set to seven and with 10 iterations. Finally, the aiSNPs were uploaded to FROG-kb in the format for the 'KiddLab – Set of 55 AI SNPs' to generate RMP values for each population in the database[23,24].

The population groups and interpretation criteria for each BGA inference method are detailed in Table 4. An 'Optimised Pipeline' was derived where samples were analysed with PCoA and, if the results were inconclusive, the sample was analysed with Structure.

The inferences were compared to the self-declared ancestry of the volunteers and determined to be consistent if all and only self-declared population(s) were inferred, partially consistent if some of the self-declared population(s) were inferred or all self-declared population(s) and additional populations were inferred, inconsistent if no self-declared population(s) were included or inconclusive if an inference could not be generated.

3. Results

3.1. Biological sex

The biological sex inferred by the samples analysed with the Quantifiler™ Trio Quantification Kit ($n = 73$), GlobalFiler™ PCR Amplification Kit ($n = 25$) and ForenSeq® DNA Signature Prep Kit ($n = 57$) were consistent between methods and with the self-declared sex (Table S2). However, the inference was inconclusive for two (12.5%) of the samples analysed with the ForenSeq® Kintelligence Kit ($n = 16$) for which volunteers self-declared as biologically female. Both profiles had two Y SNPs called with the coverage ranging from 22 to 40 reads per SNP. The quantification and STR profiling of these samples did not indicate contamination or a mixture that would result in an inconclusive biological sex inference.

3.2. Eye colour

When analysed with the UAS, 74% of inferences were consistent with the self-declared eye colour of the volunteers, while 23% were inconsistent (Figure 1, Table S3). Of the inconsistencies, 94% were inferred as more likely to have brown or blue eye colour when the individuals self-declared as having intermediate eye colour (Figure 2). The remaining inconsistency was for an individual with self-declared brown eye colour and all of the p-values reported by the UAS were less than 0.5 with blue eye colour having the maximum p-value (blue eye colour $p = 0.41$, intermediate eye colour $p = 0.25$, brown eye colour $p = 0.34$; Figure 2). There were two samples with partial piSNP profiles (call rate of 80%) and the UAS was unable to generate p-values.

The HIrisPlex pipeline generated more consistent inferences than the UAS, with 97% of genotypes resulting in eye colour inferences consistent with the self-declared information (Figure 1, Table S3). Only two inferences were inconsistent: one sample was inferred as brown eye colour by both HIrisPlex and UAS ($p = 0.94$), despite the individual self-declaring intermediate eye colour (Figure 3); the second inconsistency involved an individual with self-declared brown eye colour, where the HIrisPlex pipeline inferred likely blue or intermediate eye colour and the UAS pipeline inferred likely blue eye colour (Figure 3). Notably, 88% of the samples that were inconsistent with the UAS pipeline produced consistent inferences with the HIrisPlex pipeline.

Table 6 shows the calculated sensitivity and specificity for each pipeline by eye colour. For all eye colours, these values were equal to or higher for HIrisPlex than for the UAS, with the exception of specificity for intermediate eye colour.

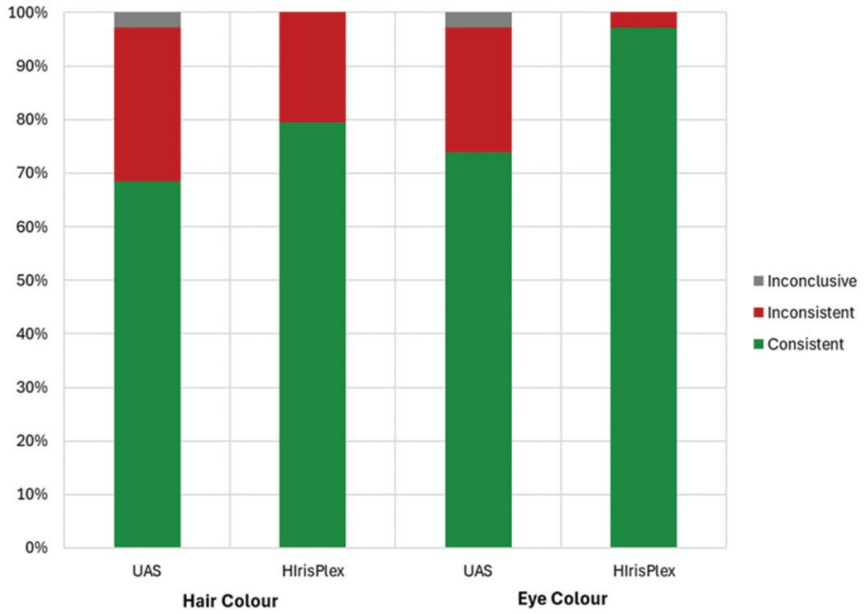


Figure 1. Externally visible characteristic (EVC) inference consistency with self-declared information of the volunteers by method: universal analysis software (UAS) and HIrisPlex.

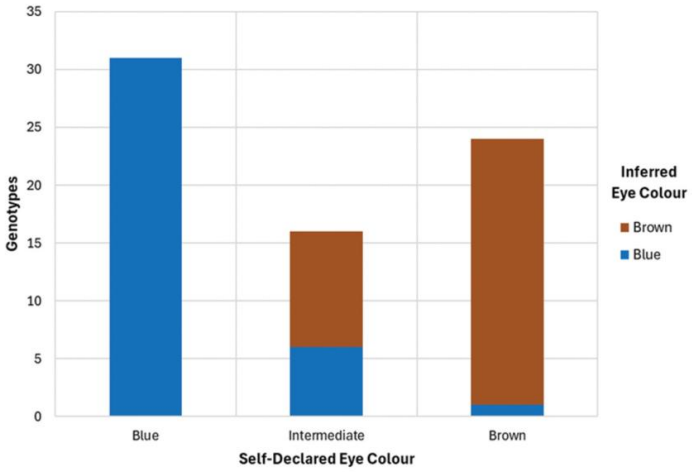


Figure 2. Eye colours inferred using the universal analysis software (UAS) by the self-declared eye colour of the volunteers.

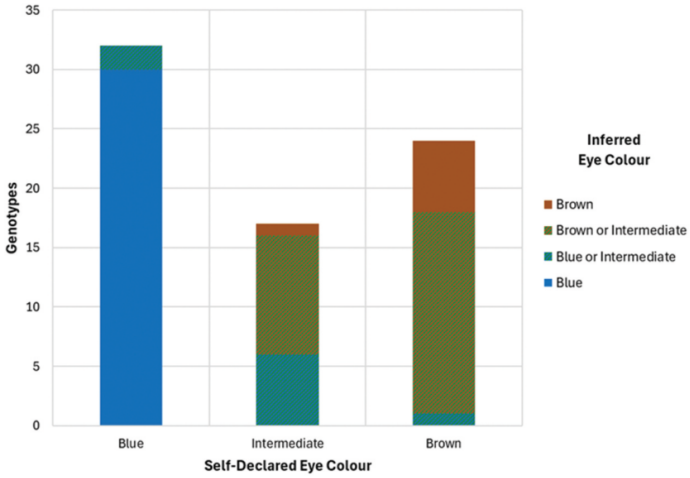


Figure 3. Eye colours inferred using the HIrisPlex by the self-declared eye colour of the volunteers.

Table 6. Performance metrics for universal analysis software (UAS) and HIrisPlex methods by eye colour category.

Eye Colour	Method	Sensitivity (TPR)	Specificity (TNR)	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Balanced Accuracy
Blue	UAS	1.00	0.83	0.82	1.00	0.91
	HIrisPlex	1.00	0.97	0.97	1.00	0.99
Intermediate	UAS	0.00	1.00	0.00	0.77	0.39
	HIrisPlex	0.97	0.97	0.97	0.97	0.97
Brown	UAS	0.86	0.79	0.70	0.97	0.84
	HIrisPlex	0.97	0.97	0.97	0.97	0.97

3.3. Hair colour

Hair colour inferences were less consistent than eye colour inferences for both pipelines (Figure 1, Table S4). With the UAS, 69% of inferences were consistent with the self-declared hair colour, while 80% were consistent when analysed with the HIrisPlex pipeline. The two samples with partial piSNP profiles that were unable to generate results in the UAS pipeline and were therefore inconclusive, produced consistent inferences with the HIrisPlex pipeline.

All inconsistent inferences generated by the HIrisPlex pipeline were also inconsistent for the UAS pipeline, of which 10 samples were from volunteers with self-declared brown or dark brown hair colour and five were from those with self-declared blond or dark blond hair colour (Figures 4, 5). An additional six samples had UAS inferences that were inconsistent with the self-declared hair colour, of which five had self-declared brown hair colour and one had self-declared red hair colour (Figure 4). For all hair colours, the sensitivity and specificity for the HIrisPlex method was equal to or higher than the UAS (Table 7).

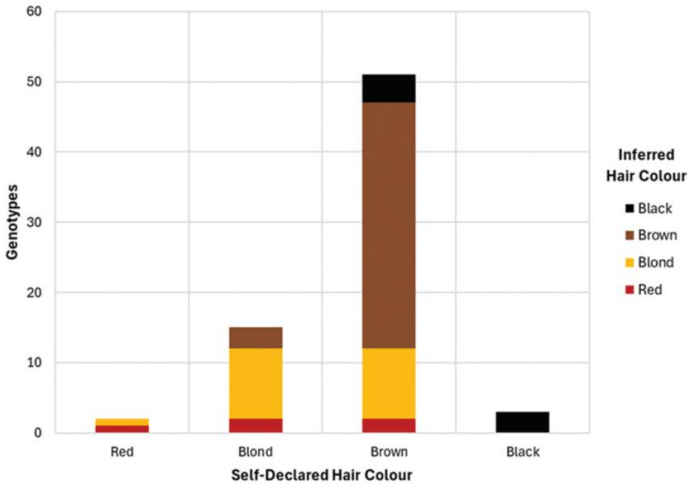


Figure 4. Hair colours inferred using the universal analysis software (UAS) by the self-declared hair colour of the volunteers.

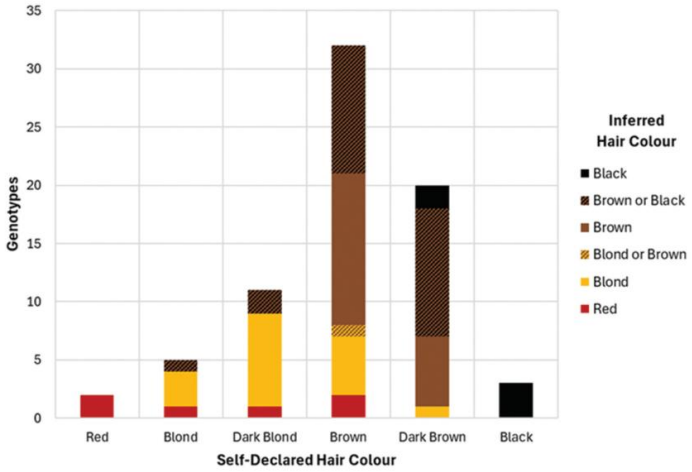


Figure 5. Hair colours inferred using HIrisPlex by the self-declared hair colour of the volunteers.

3.4. Biogeographical ancestry

The inferences generated using the PCoA pipeline had the highest consistency (86%) with the self-declared data, followed by the UAS pipeline (84%; Figure 6). The inferences that were partially consistent with the self-declared data (6% for PCoA, 8% for UAS) were due to only one of the admixed populations being detected; for the UAS pipeline, this was due to the population not being represented in the reference data (Figure 6, Table 8). Both MDS methods yielded inconclusive results for six samples each (8%), with four samples in

Table 7. Sensitivity and specificity for universal analysis software (UAS) and HIrisPlex methods by hair colour category.

Hair Colour	Method	Sensitivity (TPR)	Specificity (TNR)	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)	Balanced Accuracy
Red	UAS	0.50	0.94	0.20	0.98	0.59
	HIrisPlex	1.00	0.94	0.33	1.00	0.67
Blond	UAS	0.67	0.80	0.48	0.90	0.69
	HIrisPlex	0.71	0.89	0.67	0.91	0.79
Brown	UAS	0.69	0.85	0.92	0.52	0.72
	HIrisPlex	0.81	0.86	0.93	0.64	0.79
Black	UAS	1.00	0.96	0.50	1.00	0.75
	HIrisPlex	1.00	0.96	0.93	1.00	0.96

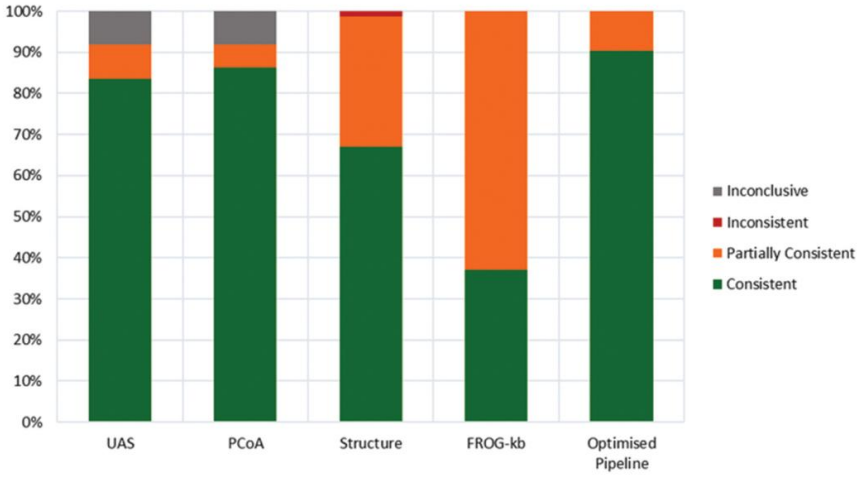


Figure 6. Biogeographical ancestry (BGA) inference consistency with self-declared ancestry of the volunteers by method: universal analysis software (UAS), principle coordinate analysis (PCoA), structure, forensic resource reference on genetics – knowledge base (FROG-kb) and the optimised pipeline.

Table 8. Categorization of biogeographical ancestry (BGA) results that were partially consistent with the self-declared ancestry of the volunteers by method: universal analysis software (UAS), principle coordinate analysis (PCoA), structure, forensic resource reference on genetics – knowledge base (FROG-kb) and the optimized pipeline.

Partially Consistent Inference	UAS	PCoA	Structure	FROG-kb	Optimised Pipeline
Some populations from self-declared admixed individuals detected and additional population(s) detected	0	0	1 (1.37%)	5 (6.85%)	1 (1.37%)
All populations from self-declared admixed individuals detected and additional population(s) detected	0	0	2 (2.74%)	3 (4.11%)	2 (2.74%)
Population from self-declared non-admixed individual detected and additional population(s) detected	0	0	18 (24.66%)	37 (50.68%)	0
Some populations from self-declared admixed individual detected	6 (8.22%)	4 (5.48%)	2 (2.74%)	1 (1.37%)	4 (5.48%)

common between the pipelines. All inconclusive results were derived from volunteers with self-declared admixture.

The Structure pipeline produced consistent BGA inferences for 67% of samples, with 32% being partially consistent and one sample exhibiting an inconsistent inference (Figure 6). In one instance, structure inferred the BGA as likely Middle Eastern ancestry with over 90% contribution, which was inconsistent with the self-declared European ancestry. FROG-kb was the least consistent, with only 37% of inferences being consistent with the self-declared ancestry and the remaining 62% being partially consistent. The majority of partially consistent inferences were samples with self-declared non-admixed ancestry, with the results inferring the self-declared population as well as additional populations (Table 8).

The optimized pipeline involved application of PCoA followed by Structure analysis for only those genotypes with inconclusive PCoA results, which improved the overall proportion of consistent results to 90%. The remaining samples were partially consistent with the self-declared ancestry and all inconclusive results were eliminated. Three samples that were inconclusive with PCoA produced consistent inferences for the self-declared admixed ancestry when analysed with Structure, while the other three samples were partially consistent (Table 8). Overall, the inferences were consistent across the five pipelines for 32% of the samples, with an additional 26% producing consistent inferences with UAS, PCoA, Structure and optimized pipelines.

3.5. Discussion

In agreement with previous studies of non-Australian populations, the HIrisPlex online tool and associated published interpretation guidelines provide flexibility for analysing partial profiles[25–27]. While the UAS uses the HIrisPlex model in an offline format and generates similar p-values to the online tool, its functionality for inferring EVCs is limited to samples with full piSNP profiles. Furthermore, additional information is provided by the online tool to assist with interpretation, such as AUC values for each category and p-values for hair shade[25–27].

Relatively high error rates for eye colour have been observed in other studies that rely on the maximum p-value approach[5,27,45,46]. These error rates were reduced following recommendations to only make inferences when the maximum p-value exceeds 0.7; however, this approach increases the likelihood of obtaining an inconclusive result[27,45]. In this study, uncertainty was incorporated by inferring both the highest p-value eye colour and intermediate eye colour if the highest p-value did not exceed 0.90, increasing the reliability of the inference pipeline. For reporting purposes, the conclusion that 'the eye colour of the individual is likely to be brown or intermediate' could alternatively be written as 'the eye colour of the individual is not likely to be blue'.

In previous studies, intermediate eye colours were the most difficult to infer as p-values tend to favour blue or brown eye colours[27,46]. In this study, 17 individuals self-declared green or hazel eye colours. The p-values generated by the HIrisPlex online tool for intermediate eye colour ranged from 0.05 to 0.34 and were never the maximum p-value produced. The sensitivity and specificity for inferring intermediate eye colour using the HIrisPlex online tool with the maximum p-value approach at the time of writing was 0.001 and 0.999, respectively[37]. However, after applying the additional thresholds in this study

to incorporate the possibility of intermediate eye colour in the inference, both sensitivity and specificity metrics were 0.97.

Inferring hair colour of an individual was more difficult than for eye colour and the consistency with self-declared data was lower. Prediction guidelines published by Walsh et al. incorporate a spectrum of hair colours[27]. Unlike the UAS, the HIrisPlex online tool generates p-values for inferring light and dark hair shades which can be used to further refine the hair colour inference with the published decision tree (i.e. dark blonde, light brown and dark brown hair colour inferences)[27]. These additional categories increased the consistency of the hair colour inferences generated by the HIrisPlex online tool compared to the in-built UAS tool.

The majority of volunteers in this study (93%) self-declared having blond or brown hair colour at 20 years old, reducing the impact of environmental and age-related hair colour changes. Both pipelines generated similar results, with the majority of inferences indicating blond or brown hair (UAS pipeline: 83%; HIrisPlex: 85%). Red hair colour had the lowest PPV for both pipelines (UAS: 0.20; HIrisPlex: 0.33), as four individuals with self-declared blond or brown hair colour were incorrectly inferred as likely having red hair. The proportion of individuals with self-declared blonde or brown hair colour who were inferred as having red hair was higher in this study than observed in previous studies[27,45,47,48].

Other studies have previously shown a correlation between pigmentation for hair, eye and skin colour and BGA[49,50]. The majority (81%) of volunteers in this study had self-declared European ancestry with no recent admixture. Other self-declared populations included East Asian (2.7%), South Asian (2.7%) and individuals with self-declared recent admixture (13%). The most important component of BGA inference is the suitability of the reference database for the jurisdiction. The UAS was designed primarily for American populations, with a 2D PCA plot of three super populations and one superimposed population representing admixed American individuals. This latter population was not deemed relevant for an Australian application of this pipeline.

The MDS approaches (UAS and PCoA) produced BGA inferences most consistent with the self-declared ancestries of the volunteers but were unable to infer admixture. The bespoke reference database used for PCoA represented a greater number of population groups and allowed for an additional PC, which was important for distinguishing between clusters[51]. BGA inferences for individuals with self-declared recent admixture, when analysed with PCoA, were either inconclusive (60%) or partially consistent (40%). The partially consistent inferences included the major ancestral contributor, where three grandparents were from the same population group and the fourth was from a different population group. The Structure and FROG-kb pipelines were more likely to infer multiple population contributions. When applying the same bespoke reference database as for PCoA, Structure helped to interpret the inconclusive results using PCoA in the optimized pipeline proposed in this study.

One limitation of BGA inferences is that they are currently restricted to continental population groups. Inference of subcontinental BGA will require the selection of DNA markers designed to reflect fine scale genetic distances within these populations[52]. Furthermore, Structure and FROG-kb often inferred both European and Middle Eastern ancestry for individuals with self-declared European ancestry only, likely due to the close geographic proximity of these regions. Use of these inferences in forensic casework will

require careful consideration of the risks associated with potential misinterpretation and misuse of the intelligence.

4. Conclusions

Where an identification cannot be achieved through direct comparison or database searching with STR profiles, inferences of biological sex, EVCs and BGA from SNP genotypes can assist law enforcement in reducing a pool of potential candidates for either a coronial or criminal investigation. The in-built UAS algorithms for hair and eye colour inference did not perform as well as the HIrisPlex online tool due to the limitation that all piSNPs must be typed in order to generate p-values. For BGA inferences, the PCoA method using our curated reference database was the highest performing pipeline but was unable to infer admixture. However, Structure was effective in refining BGA inferences for individuals with self-declared admixture that produced inconclusive results using PCoA, when integrated into an optimized analysis pipeline. By combining these inference methods, the interpretation and reporting of DNA intelligence can be improved for Australian jurisdictions.

Acknowledgments

The authors would like to thank the volunteers for providing their DNA and associated meta data for this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by the Australian Federal Police (AFP) Innovation Fund and the AFP National DNA Program for Unidentified and Missing Persons. Jessica L Watson is supported by an Australian Government Research Training Program Scholarship.Institutional Review Board Statement.This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of University of Technology Sydney (UTS) Human Research Ethics Committee (HREC) (UTS HREC NO. ETH21-5821).

ORCID

Jessica L Watson http://orcid.org/0000-0002-4649-3395
Kelly Grisedale http://orcid.org/0009-0000-3869-2725
Jodie Ward http://orcid.org/0000-0001-9359-9806
Dennis McNevin http://orcid.org/0000-0003-1665-3367

Author contributions

Conceptualization: J.L.W, D.M., K.G. and J.W.; Methodology: J.L.W., D.M., K.G. and J.W.; Formal Analysis: J.L.W., D.M. and K.G.; Writing – Original Draft Presentation: J.L.W.; Writing – Review & Editing: J.L.W., D.M., K.G. and J.W.; Supervision: D.M. and J.W.; Funding Acquisition: J.W.

Informed consent statement

Written informed consent was obtained from all volunteer sample donors involved in this study.

Data availability statement

Data are stored at the Australian Federal Police and may be made available to approved entities upon written request and subject to consent provisions.

References

1. Kayser M, Branicki W, Parson W, Phillips C. Recent advances in forensic DNA phenotyping of appearance, ancestry and age. Forensic Sci Int Genet. 2023;65:102870. doi:10.1016/j.fsigen.2023.102870
2. Kayser M, De Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet. 2011;12(3):179–192. doi:10.1038/nrg2952
3. Novroski NMM. The use of massively parallel sequencing in forensic DNA typing: what will the future hold? Forensic Sci Virtual Event Ser 2020. 2020. https://www.labroots.com/webinar/massively-parallel-sequencing-forensic-dna-typing-future-hold?srsltid=AfmBOopADFEeqo4piWqzUQ5ulPoEtq7Qbft9aav8ZpBoFEW_SJ6D4Fse
4. Chaitanya L, Iz P, Walsh S, J B, Zupanc T, Kayser M. Bringing colour back after 70 years: predicting eye and hair colour from skeletal remains of world war ii victims using the hirisplex system. Forensic Sci Int Genet. 2017;26:48–57. doi:10.1016/j.fsigen.2016.10.004
5. Draus-Barini J, Walsh S, Pośpiech E, Kupiec T, Głąb H, Branicki W, Kayser M. Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. Investig Genet. 2013;4(1):3. doi:10.1186/2041-2223-4-3
6. Watherston J, McNevin D, Gahan ME, Bruce D, Ward J. Current and emerging tools for the recovery of genetic information from post mortem samples: new directions for disaster victim identification. Forensic Sci Int Genet. 2018;37:270–282. doi:10.1016/j.fsigen.2018.08.016
7. Ion ampliseq ngs panels for targeted sequencing. 2024. [accessed 2024 14 March]. https://www.thermofisher.com/au/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-select-targets/ampliseq-target-selection.html
8. Tillmar A, Sturk-Andreaggi K, Daniels-Higginbotham J, Thomas JT, Marshall C. The force panel: an all-in-one snp marker set for confirming investigating genetic genealogy leads and for general forensic applications. Genes. 2021;12(12):1968. doi:10.1101/2021.11.30.470354
9. Antunes J, Walichiewicz P, Forouzmand E, Barta R, Didier M, Han Y, Perez JC, Snedecor J, Zlatkov C, Padmabandu G. Developmental validation of the forenseq® kintelligence kit, miseq fgx® sequencing system and forenseq universal analysis software. Forensic Sci Int Genet. 2024;71:103055. doi:10.1016/j.fsigen.2024.103055
10. Verogen. Forenseq DNA signature prep reference guide. ForenSeq DNA signature prep kit. 2020.
11. Verogen. Forenseq imagen kit. 2022.
12. Thermo Fisher Scientific. Quantifiler™ hp and trio DNA quantification kits user guide. Life Technologies. 2017.
13. Thermo Fisher Scientific. Globalfiler pcr amplification kits. Life Technologies. 2019.
14. Kidd JR, Friedlaender FR, Speed WC, Pakstis AJ, La Vega Fm D, Kidd KK. Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples. Investig Genet. 2011;2(1):1. doi:10.1186/2041-2223-2-1
15. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, et al.A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis. 2006;9(27):1713–1724. doi:10.1002/elps.200500671

16. Kidd KK, Speed WC, Pakstis AJ, Furtado MR, Fang R, Madbouly A, Maiers M, Middha M, Friedlaender FR, Kidd JR. Progress toward an efficient panel of SNPs for ancestry inference. Forensic Sci Int Genet. 2014;10:23–32. doi:10.1016/j.fsigen.2014.01.002
17. McNevin D. Forensic inference of biogeographical ancestry from genotype: the genetic ancestry lab. Wiley Interdiscip Rev. 2020;2(2):e1356. doi:10.1002/wfs2.1356
18. Verogen. 2021. Universal analysis software v2.0 and above, reference guide. Universal analysis software.
19. Gaspar HA, Breen G. Probabilistic ancestry maps: a method to assess and visualize population substructures in genetics. BMC Bioinf. 2019;20(1):1–11. doi:10.1186/s12859-019-2680-1
20. Paradis E, Claude J, Strimmer K. Ape: analyses of phylogenetics and evolution in r language. Bioinformatics. 2004;20(2):289–290. doi:10.1093/bioinformatics/btg412
21. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–959. doi:10.1093/genetics/155.2.945
22. Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo A, Lareu MAn overview of structure: applications, parameter settings, and supporting softwareFron Genet. 2013;4(98): 1–13. doi:10.3389/fgene.2013.00098
23. Kidd KK, Soundararajan U, Rajeevan H, Pakstis AJ, Moore KN, Ropero-Miller JD. The redesigned forensic research/reference on genetics-knowledge base, frog-kb. Forensic Sci Int Genet. 2018;33:33–37. doi:10.1016/j.fsigen.2017.11.009
24. Rajeevan H, Soundararajan U, Pakstis AJ, Kidd KK. Introducing the forensic research/reference on genetics knowledge base, frog-kb. Investig Genet. 2012;3(1):18. doi:10.1186/2041-2223-3-18
25. Chaitanya L, Breslin K, Zuñiga S, Wirken L, Pośpiech E, Kukla-Bartoszek M, Sijen T, Knijff PD, Liu F, Branicki W, et al. The hirisplex-s system for eye, hair and skin colour prediction from DNA: introduction and forensic developmental validation. Forensic Sci Int Genet. 2018;35:123–135. doi: 10.1016/j.fsigen.2018.04.004
26. Walsh S, Chaitanya L, Breslin K, Muralidharan C, Bronikowska A, Pospiech E, Koller J, Kovatsi L, Wollstein A, Branicki W. Global skin colour prediction from DNA. Hum Genet. 2017;136 (7):847–863. doi:10.1007/s00439-017-1808-5
27. Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L, Maeda H, Ishikawa T, Sijen T, De Knijff P, et al. Developmental validation of the hirisplex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet. 2014;9:150–161. doi:10.1016/j.fsigen.2013.12.006
28. QIAGEN. 2022. Ez1&2 DNA investigator kit.
29. Thermo Fisher Scientific. Quantstudio™ 5 real-time pcr instrument (for human identification) user guide. Life Technologies. 2017.
30. Thermo Fisher Scientific. Applied biosystems 3500/3500xl genetic analyzer user guide. Life Technologies. 2010.
31. Genemapper™ id-x software v1.6. 2024. [accessed 2024]. https://www.thermofisher.com/order/catalog/product/A39975.
32. Verogen miseq fgx sequencing reagents. 2024. [accessed 2024 15 August]. https://www.qiagen.com/us/products/human-id-and-forensics/nextgeneration-sequencing/verogen-miseq-fgx-sequencing-reagents
33. Verogen. 2021. Miseq fgx sequencing system reference guide. MiSeq FGx sequencing system.
34. Verogen. 2021. Universal analysis software v1.3, reference guide. Universal analysis software.
35. Verogen forenseq kintelligence kit. 2024. [accessed 2024 24 May]. https://www.qiagen.com/us/products/human-id-and-forensics/nextgeneration-sequencing/verogen-forenseq-kintelligence-kit.
36. Watson JL, McNevin D, Grisedale K, Spiden M, Seddon S, Ward JOperationalisation of the forenseq® kintelligence kit for Australian unidentified and missing persons casework. Forensic Sci Int Genet. 2023;68: 102972. doi: 10.1016/j.fsigen.2023.102972
37. Hirisplex-s eye, hair and skin colour DNA phenotyping webtool. 2022. [accessed 2023 19 April]. https://hirisplex.erasmuscmc.nl.

38. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz M H-Y, et al. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526(7571):75–81: doi: 10.1038/nature15394
39. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE. Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526(7571):68–74. doi:10.1038/nature15393
40. Cann HM, De Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A. A human genome diversity cell line panel. Science. 2002;296 (5566):261–262. doi:10.1126/science.296.5566.261b
41. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, et al. The Simons genome diversity project: 300 genomes from 142 diverse populations. Nature. 2016;538(7624):201–206. doi:10.1038/nature18964
42. R Core Team2020R: a language and environment for statistical computing, Vienna, Austria.
43. Paradis E, Bolker B, Claude J, Cuong HS, Desper R, Durand B, Dutheil J, Gascuel O, Heibl C, Lawson D, et al. Ape: analyses of phylogenetics and evolution. R package version 5.5. 2021.
44. Pritchard JK. Structure software 2.3.4. California, United: Pritchard Lab, Stanford University; 2012. https://web.stanford.edu/group/pritchardlab/structuresoftware/releaseversions/v2.3.4/html/structure.html
45. Salvo NM, Janssen K, Kirsebom MK, Meyer OS, Berg T, Olsen G-H. Predicting eye and hair colour in a Norwegian population using Verogen's ForenSeq™ DNA signature prep kit. Forensic Sci Int Genet. 2021;56:102620. doi:10.1016/j.fsigen.2021.102620
46. Walsh S, Liu F, Ballantyne KN, van Oven M, Lao O, Kayser M. Irisplex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet. 2011;5(3):170–180. doi:10.1016/j.fsigen.2010.02.004
47. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, Branicki W, Kayser M. The hirisplex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet. 2013;7(1):98–115. doi:10.1016/j.fsigen.2012.07.005
48. Carratto TMT, Marcorin L, Do Valle-Silva G, de Oliveira MLG, Donadi EA, Simões AL, Castelli EC, Mendes-Junior CT. Prediction of eye and hair pigmentation phenotypes using the hirisplex system in a Brazilian admixed population sample. Int J Legal Med. 2021;135(4):1329–1339. doi:10.1007/s00414-021-02554-7
49. Palmal S, Adhikari K, Mendoza-Revilla J, Fuentes-Guajardo M, de Cerqueira Ccs, Bonfante B, Chacón-Duque JC, Sohail A, Hurtado M, Villegas V. Prediction of eye, hair and skin colour in Latin americans. Forensic Sci Int Genet. 2021;53:102517. doi:10.1016/j.fsigen.2021.102517
50. CCSd C, Hünemeier T, Gomez-Valdés J, Ramallo V, Volasko-Krause CD, Barbosa AAL, Vargas-Pinilla P, Dornelles RC, Longo D, Rothhammer F, et al. Implications of the admixture process in skin color molecular assessment. PLoS One. 2014;9(5):e96886. doi:10.1371/journal.pone.0096886
51. Cheung EY, Gahan ME, McNevin D. Predictive DNA analysis for biogeographical ancestry. Aust J Forensic Sci. 2018;50(6):1–8. doi:10.1080/00450618.2017.1422021
52. Schneider PM, Prainsack B, Kayser M. The use of forensic DNA phenotyping in predicting appearance and biogeographic ancestry. Deutsches Ärzteblatt Int. 2019;116(51):873. doi:10.3238/arztebl.2019.0873

6.1 Supplementary Material

Supplementary Table 1: Quantification, STR profiling and SNP sequencing results for inferring biological sex

Sample ID	Self-Declared Biological Sex	Quantification Results (QuantiFiler Trio)					STR Profiling Results (GlobalFiler)	
		Small Autosomal Target (ng/µL)	Large Autosomal Target (ng/µL)	Y Target (ng/µL)	Degradation Index	Inferred Biological Sex	Amelogenin	DYS391
Individual 1	Female	0.291	0.359	0.000	0.810	Female	XX	Absent
Individual 2	Female	0.020	0.006	0.000	3.333	Female	XX	Absent
Individual 4	Male	0.043	0.037	0.043	1.150	Male	XY	Present
Individual 5	Female	0.099	0.053	0.000	1.864	Female	XX	Absent
Individual 6	Male	0.726	0.443	0.461	1.637	Male	XY	Present
Individual 8	Male	0.297	0.203	0.293	1.465	Male	XY	Present
Individual 11	Female	0.221	0.095	0.000	2.317	Female	XX	Absent
Individual 13	Male	0.066	0.058	0.072	1.144	Male	XY	Present
Individual 14	Male	20.449	15.095	16.548	1.355	Male	XY	Present
Individual 15	Female	6.262	5.388	0.000	1.162	Female	XX	Absent
Individual 16	Female	3.907	3.996	0.000	0.978	Female	XX	Absent
Individual 21	Female	4.903	2.962	0.000	1.655	Female	N/A	N/A
Individual 23	Female	3.137	4.758	0.000	0.659	Female	XX	Absent
Individual 40	Female	14.410	16.208	0.000	0.889	Female	XX	Absent
Individual 41	Female	6.774	7.607	0.000	0.891	Female	N/A	N/A
Individual 42	Male	7.617	6.902	7.938	1.104	Male	N/A	N/A
Individual 43	Female	21.095	31.369	0.000	0.672	Female	XX	Absent
Individual 44	Female	2.322	3.564	0.000	0.651	Female	XX	Absent
Individual 45	Female	1.859	1.938	0.000	0.959	Female	XX	Absent
Individual 46	Male	3.022	3.807	3.041	0.794	Male	N/A	N/A
Individual 47	Male	4.013	6.046	5.722	0.664	Male	N/A	N/A
Individual 48	Female	8.295	10.944	0.000	0.758	Female	N/A	N/A
Individual 49	Female	4.412	3.920	0.000	1.125	Female	N/A	N/A
Individual 50	Female	3.978	3.053	0.000	1.303	Female	XX	Absent
Individual 51	Male	3.508	3.367	3.287	1.042	Male	N/A	N/A
Individual 52	Female	14.156	14.547	0.000	0.973	Female	N/A	N/A
Individual 53	Female	9.525	9.913	0.000	0.961	Female	XX	Absent
Individual 55	Female	20.087	17.082	0.000	1.176	Female	N/A	N/A
Individual 56	Female	1.183	2.747	0.000	0.431	Female	XX	Absent
Individual 60	Female	1.165	1.332	0.000	0.874	Female	N/A	N/A
Individual 61	Female	1.339	0.811	0.000	1.652	Female	N/A	N/A
Individual 62	Female	1.887	2.289	0.000	0.824	Female	N/A	N/A
Individual 63	Female	5.520	7.049	0.000	0.783	Female	N/A	N/A
Individual 64	Male	1.633	1.426	1.406	1.145	Male	N/A	N/A
Individual 65	Female	1.052	1.260	0.000	0.836	Female	N/A	N/A
Individual 66	Female	1.925	0.657	0.000	2.931	Female	XX	Absent
Individual 67	Female	0.351	0.329	0.000	1.070	Female	XX	Absent
Individual 68	Female	1.102	1.070	0.000	1.030	Female	XX	Absent
Individual 70	Female	1.336	0.494	0.000	2.706	Female	XX	Absent
Individual 71	Female	6.399	7.852	0.000	0.815	Female	N/A	N/A
Individual 72	Female	5.838	6.511	0.000	0.897	Female	N/A	N/A
Individual 73	Female	4.973	7.137	0.000	0.697	Female	N/A	N/A
Individual 74	Female	8.651	11.737	0.000	0.737	Female	N/A	N/A
Individual 75	Female	6.840	11.933	0.000	0.573	Female	N/A	N/A
Individual 76	Female	1.914	2.111	0.000	0.907	Female	N/A	N/A
Individual 77	Female	2.976	3.470	0.000	0.858	Female	N/A	N/A
Individual 78	Female	4.691	5.048	0.000	0.929	Female	N/A	N/A
Individual 79	Male	3.474	5.331	3.262	0.652	Male	N/A	N/A
Individual 80	Female	6.975	8.368	0.000	0.834	Female	N/A	N/A
Individual 81	Male	7.894	12.112	7.813	0.652	Male	N/A	N/A
Individual 82	Male	2.519	2.823	2.569	0.892	Male	N/A	N/A
Individual 83	Female	5.844	9.555	0.000	0.612	Female	N/A	N/A
Individual 84	Female	7.726	9.007	0.000	0.858	Female	N/A	N/A
Individual 85	Female	5.510	6.886	0.000	0.800	Female	N/A	N/A
Individual 86	Male	5.380	4.258	7.499	1.263	Male	N/A	N/A
Individual 87	Female	2.457	2.593	0.000	0.947	Female	N/A	N/A
Individual 88	Female	8.221	13.237	0.000	0.621	Female	N/A	N/A
Individual 89	Female	8.348	12.979	0.000	0.643	Female	N/A	N/A
Individual 90	Female	4.619	6.629	0.000	0.697	Female	N/A	N/A
Individual 91	Female	9.841	14.308	0.000	0.688	Female	N/A	N/A
Individual 92	Female	8.980	8.298	0.000	1.082	Female	N/A	N/A
Individual 93	Male	10.384	10.998	10.833	0.944	Male	N/A	N/A
Individual 94	Male	0.028	0.002	0.028	11.387	Male	N/A	N/A
Individual 95	Female	2.029	1.460	0.000	1.390	Female	XX	Absent
Individual 96	Female	2.529	3.164	0.000	0.799	Female	N/A	N/A
Individual 97	Male	9.256	10.533	9.958	0.879	Male	N/A	N/A
Individual 98	Female	10.112	10.572	0.000	0.956	Female	N/A	N/A
Individual 99	Female	7.128	9.977	0.000	0.714	Female	N/A	N/A
Individual 100	Female	4.360	5.532	0.000	0.788	Female	XX	Absent
Individual 101	Female	9.606	12.608	0.000	0.762	Female	N/A	N/A
Individual 102	Male	5.874	6.086	5.341	0.965	Male	N/A	N/A
Individual 103	Female	4.440	7.161	0.000	0.620	Female	N/A	N/A
Individual 104	Male	4.211	5.022	4.508	0.839	Male	N/A	N/A

STR Profiling Results (GlobalFiler)		STR Profiling Results (DNA Signature Prep)				SNP Results (Kintelligence)		
Y Indel	Inferred Biological Sex	Amelogenin	X STRs Call Rate	Y STRs Call Rate	Inferred Biological Sex	X SNPs Call Rate	Y SNPs Call Rate	Inferred Biological Sex
Absent	Female	N/A	N/A	N/A	N/A	98.11%	0.00%	Female
Absent	Female	N/A	N/A	N/A	N/A	94.34%	2.34%	Inconclusive
Present	Male	N/A	N/A	N/A	N/A	96.23%	92.94%	Male
Absent	Female	XX	100.00%	0.00%	Female	N/A	N/A	N/A
Present	Male	N/A	N/A	N/A	N/A	100.00%	100.00%	Male
Present	Male	N/A	N/A	N/A	N/A	100.00%	98.82%	Male
Absent	Female	N/A	N/A	N/A	N/A	100.00%	0.00%	Female
Present	Male	N/A	N/A	N/A	N/A	99.06%	100.00%	Male
Present	Male	N/A	N/A	N/A	Male	100.00%	100.00%	Male
Absent	Female	N/A	N/A	N/A	N/A	100.00%	0.00%	Female
Absent	Female	N/A	N/A	N/A	N/A	94.34%	2.35%	Inconclusive
N/A	N/A	XX	100.00%	0.00%	Female	100.00%	0.00%	Female
Absent	Female	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Absent	Female	N/A	N/A	N/A	N/A	100.00%	0.00%	Female
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
Absent	Female	N/A	N/A	N/A	N/A	100.00%	0.00%	Female
Absent	Female	XX	100.00%	0.00%	Female	N/A	N/A	N/A
Absent	Female	N/A	N/A	N/A	N/A	100.00%	0.00%	Female
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
Absent	Female	N/A	N/A	N/A	N/A	100.00%	0.00%	Female
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
Absent	Female	N/A	N/A	N/A	N/A	100.00%	0.00%	Female
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
Absent	Female	N/A	N/A	N/A	N/A	100.00%	0.00%	Female
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
Absent	Female	N/A	N/A	N/A	Female	100.00%	0.00%	Female
Absent	Female	N/A	N/A	N/A	Female	98.11%	0.00%	Female
Absent	Female	N/A	N/A	N/A	Female	100.00%	0.00%	Female
Absent	Female	N/A	N/A	N/A	Female	99.06%	0.00%	Female
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XY	85.71%	54.17%	Male	N/A	N/A	N/A
Absent	Female	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
Absent	Female	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A
N/A	N/A	XX	100.00%	0.00%	Female	N/A	N/A	N/A
N/A	N/A	XY	100.00%	100.00%	Male	N/A	N/A	N/A

Supplementary Table 2. The population groups made based on geography and ethnicity of the 161 populations in the 'KiddLab – Set of 55 AISNPs' on Forensic Resource Reference on Genetics – Knowledge Base (FROG-kb).

Population Group	Population
Admixed	African American(ASW)
	African Americans
	Afro-Ecuadorian
Sub-Saharan Africa	Biaka
	Chagga
	Esan(ESN)
	Ethiopian Jews
	Gambian(GWD)
	Hausa
	Ibo (Af
	Lisongo
	Luhya(LWK)
	Masai
	Mbuti
	Mende(MSL)
	Sandawe
	Somali
	Somalis
	Yoruba
	Yoruba(YRI)
	Zaramo
Middle East / North Africa	Arabs from Northern Iraq
	Chaldeans
	Druze
	Hazaras
	Iranians
	Iranians
	Kairoun_Tunisia
	Kerkennah,Tunisia
	Kesra_Tunisia
	Kurds
	Kuwaiti
	Lybia
	Media_Tunisia
	Nebeur_Tunisia
	Palestinian Arabs
	Pathan
	Qatari
	Samaritans
	Saudi
	Shabaks
	Smar_South Tunisia
	Sousse_Tunisia
	Southern Tunisians
	Syriacs
	Turkish
	Turkmen
	Turks
	UAE_Arabs
	Yazidis
	Yemenite Jews

Population Group	Population
Europe	Adygei
	Ashkenazi Jews
	Basque
	British(GBR)
	Chuvash
	Danes
	Danes
	Finns
	Finns(FIN)
	Greek Cypriots
	Greeks
	Hungarians
	Iberian(IBS)
	Irish
	Khanty
	Komi-Zyrian
	Mixed Europeans
	Mixed Europeans(CEU)
	Norwegians
	Roman Jews
	Russians
	Russians,Archangel'sk
	Sardinian
	Toscani(TSI)
	Turkish Cypriots
Central Asia	Kazak
	Khazaks
	Kirghiz
	Mohannas
	Negroid Makrani
	Tajiksr
South Asia	Bengali(BEB)
	Gujarati(GIH)
	Kachari
	Keralite
	Punjabi(PJL)
	Sri Lankan Tamil(STU)
	Telugu
	Thoti
	Va(Burma)
East Asia	Ami (East
	Atayal
	Bai (East
	Baima Dee
	Cambodian
	Chengdu Tibetans
	Chinese Dai(CDX)
	Dai (East
	Dong(Guangxi)
	Hakka
	Hakka(Meizhou)
	Han Chinese(CHB)
	Han Chinese(CHS)
	Han(Guangdong)
	Han(Guangxi)
	Han(Guangzhou)

Population Group	Population
East Asia	Han(Shaanxi)
	Han(Shandong)
	Han(Yunnan)
	Hui(Ningxia)
	Inner Mongolians
	Japanese
	Japanese
	Jing(Guangxi)
	Khamba Tibetan
	Koreans
	Koreans
	Li (Asia
	Liangshan Tibetan
	Mainland Japanese
	Malaysians
	Miao(Guangxi)
	Miao(Guizhou)
	Mongols(Inner Mongolia)
	Okinawa Japanese
	Outer Mongolians
	Qiang
	Qinghai Tibetans
	San Francisco Chinese
	Taiwanese Han
	Tibetan(Tibet)
	Tibetans
	Tsaatan
	Tu (East
	Va(Yunnan)
	Vietnamese(KHV)
	Xibo (East
	Yi (East
	Yi(Sichuan)
Oceania	Aboriginals
	Micronesians
	Nasioi speakers
	Papuan New Guinean
	Samoans
America	Ecuadorian Mestizo
	Greenlanders
	Guihiba speakers
	Karitiana
	Kichwa
	Maya_Yucatec
	Mexican Pima
	Peruvian Quechuan
	Peruvian(PEL)
	Plains Amerindians
	Southwest Amerindians
	Surui,Rondonia
	Ticuna
	Yavapai

Supplementary Table 3: Results for inferring eye colour

Sample ID	Self-Declared Eye Colour	Universal Analysis Software				HIrisPlex			
		Blue p-value	Intermediate p-value	Brown p-value	Inferred Eye Colour	Blue p-value	Intermediate p-value	Brown p-value	Inferred Eye Colour
Individual 1	Brown	0.16	0.21	0.63	Brown	0.08	0.19	0.73	Brown or Intermediate
Individual 2	Hazel	Inconclusive	Inconclusive	Inconclusive	Inconclusive	0.21	0.23	0.56	Brown or Intermediate
Individual 4	Grey	0.91	0.07	0.02	Blue	0.90	0.07	0.02	Blue
Individual 5	Brown	0.14	0.16	0.70	Brown	0.14	0.15	0.71	Brown or Intermediate
Individual 6	Hazel	0.68	0.11	0.21	Blue	0.65	0.12	0.23	Blue or Intermediate
Individual 8	Brown	0.05	0.12	0.83	Brown	0.05	0.11	0.84	Brown or Intermediate
Individual 11	Green	0.89	0.07	0.04	Blue	0.88	0.07	0.04	Blue or Intermediate
Individual 13	Blue	0.94	0.04	0.02	Blue	0.93	0.05	0.02	Blue
Individual 14	Brown	0.01	0.05	0.94	Brown	0.01	0.05	0.94	Brown
Individual 15	Hazel	0.01	0.05	0.94	Brown	0.01	0.05	0.94	Brown
Individual 16	Brown	0.10	0.14	0.76	Brown	0.10	0.13	0.77	Brown or Intermediate
Individual 21	Blue	0.09	0.19	0.72	Brown	0.08	0.19	0.73	Brown or Intermediate
Individual 23	Green	0.92	0.05	0.03	Blue	0.91	0.06	0.03	Blue
Individual 40	Brown	0.00	0.01	0.99	Brown	0.00	0.01	0.99	Brown
Individual 41	Blue	0.93	0.05	0.02	Blue	0.93	0.06	0.02	Blue
Individual 42	Brown	0.08	0.14	0.79	Brown	0.07	0.13	0.79	Brown or Intermediate
Individual 43	Hazel	0.33	0.24	0.42	Brown	0.32	0.25	0.43	Brown or Intermediate
Individual 44	Brown	0.12	0.22	0.66	Brown	0.12	0.21	0.67	Brown or Intermediate
Individual 45	Brown	0.00	0.00	1.00	Brown	0.00	0.00	1.00	Brown
Individual 46	Blue	0.92	0.05	0.03	Blue	0.91	0.06	0.03	Blue
Individual 47	Blue	0.97	0.03	0.01	Blue	0.97	0.03	0.01	Blue
Individual 48	Brown	0.05	0.12	0.83	Brown	0.05	0.11	0.84	Brown or Intermediate
Individual 49	Brown	0.01	0.19	0.79	Brown	0.01	0.19	0.80	Brown or Intermediate
Individual 50	Brown	0.00	0.00	1.00	Brown	0.00	0.00	1.00	Brown
Individual 51	Brown	0.28	0.25	0.47	Brown	0.28	0.24	0.48	Brown or Intermediate
Individual 52	Hazel	0.81	0.08	0.12	Blue	0.78	0.09	0.13	Blue or Intermediate
Individual 53	Brown	0.00	0.00	1.00	Brown	0.00	0.00	1.00	Brown
Individual 55	Blue	0.94	0.04	0.02	Blue	0.93	0.05	0.02	Blue
Individual 56	Blue	0.92	0.05	0.03	Blue	0.91	0.06	0.03	Blue
Individual 60	Blue	0.95	0.03	0.02	Blue	0.95	0.04	0.02	Blue
Individual 61	Grey	0.95	0.04	0.01	Blue	0.94	0.05	0.01	Blue
Individual 62	Brown	0.16	0.21	0.63	Brown	0.16	0.21	0.63	Brown or Intermediate
Individual 63	Green	0.27	0.34	0.39	Brown	0.27	0.34	0.40	Brown or Intermediate
Individual 64	Blue	0.93	0.05	0.02	Blue	0.93	0.06	0.02	Blue
Individual 65	Blue	0.95	0.03	0.02	Blue	0.95	0.04	0.02	Blue
Individual 66	Brown	0.05	0.12	0.83	Brown	0.05	0.11	0.84	Brown or Intermediate
Individual 67	Green	0.21	0.32	0.46	Brown	0.21	0.32	0.47	Brown or Intermediate
Individual 68	Blue	0.93	0.05	0.02	Blue	0.93	0.06	0.02	Blue
Individual 70	Blue	0.95	0.04	0.01	Blue	0.94	0.05	0.01	Blue
Individual 71	Blue	0.88	0.08	0.03	Blue	0.88	0.09	0.03	Blue or Intermediate
Individual 72	Brown	0.05	0.12	0.83	Brown	0.05	0.11	0.84	Brown or Intermediate
Individual 73	Blue	0.92	0.05	0.03	Blue	0.91	0.06	0.03	Blue
Individual 74	Brown	0.05	0.12	0.83	Brown	0.05	0.11	0.84	Brown or Intermediate
Individual 75	Hazel	0.12	0.22	0.66	Brown	0.12	0.21	0.67	Brown or Intermediate
Individual 76	Blue	0.97	0.03	0.00	Blue	0.97	0.03	0.00	Blue
Individual 77	Brown	0.25	0.16	0.58	Brown	0.25	0.16	0.59	Brown or Intermediate
Individual 78	Blue	0.96	0.04	0.01	Blue	0.96	0.04	0.01	Blue
Individual 79	Green	0.15	0.23	0.62	Brown	0.14	0.23	0.63	Brown or Intermediate
Individual 80	Blue	0.98	0.02	0.00	Blue	0.98	0.02	0.00	Blue
Individual 81	Blue	0.92	0.05	0.03	Blue	0.91	0.06	0.03	Blue
Individual 82	Blue	0.95	0.03	0.02	Blue	0.95	0.04	0.02	Blue
Individual 83	Green	0.88	0.08	0.03	Blue	0.88	0.09	0.03	Blue or Intermediate
Individual 84	Hazel	0.88	0.08	0.03	Blue	0.88	0.09	0.03	Blue or Intermediate
Individual 85	Green	0.26	0.24	0.51	Brown	0.25	0.24	0.51	Brown or Intermediate
Individual 86	Blue	0.93	0.05	0.02	Blue	0.93	0.06	0.02	Blue
Individual 87	Blue	0.92	0.05	0.03	Blue	0.91	0.06	0.03	Blue
Individual 88	Blue	0.91	0.07	0.02	Blue	0.90	0.07	0.02	Blue
Individual 89	Blue	0.92	0.05	0.03	Blue	0.91	0.06	0.03	Blue
Individual 90	Green	0.27	0.22	0.51	Brown	0.27	0.22	0.51	Brown or Intermediate
Individual 91	Brown	0.41	0.25	0.34	Blue	0.41	0.25	0.35	Blue or Intermediate
Individual 92	Brown	0.05	0.12	0.83	Brown	0.05	0.11	0.84	Brown or Intermediate
Individual 93	Blue	0.95	0.04	0.01	Blue	0.94	0.05	0.01	Blue
Individual 94	Blue	Inconclusive	Inconclusive	Inconclusive	Inconclusive	0.93	0.05	0.01	Blue
Individual 95	Blue	0.92	0.05	0.03	Blue	0.91	0.06	0.03	Blue
Individual 96	Blue	0.94	0.04	0.02	Blue	0.93	0.05	0.02	Blue
Individual 97	Brown	0.09	0.14	0.77	Brown	0.08	0.14	0.78	Brown or Intermediate
Individual 98	Brown	0.10	0.14	0.76	Brown	0.10	0.13	0.77	Brown or Intermediate
Individual 99	Blue	0.97	0.02	0.00	Blue	0.97	0.03	0.00	Blue
Individual 100	Brown	0.00	0.01	0.99	Brown	0.00	0.01	0.99	Brown
Individual 101	Hazel	0.89	0.07	0.04	Blue	0.88	0.07	0.04	Blue or Intermediate
Individual 102	Hazel	0.21	0.32	0.46	Brown	0.21	0.32	0.47	Brown or Intermediate
Individual 103	Blue	0.86	0.08	0.06	Blue	0.85	0.09	0.06	Blue or Intermediate
Individual 104	Blue	0.94	0.05	0.01	Blue	0.93	0.06	0.01	Blue

Supplementary Table 4. Results for inferring hair colour

Sample ID	Self-Declared Hair Colour	Universal Analysis Software				
		Black p-value	Brown p-value	Blond p-value	Red p-value	Inferred Hair Colour
Individual 1	Brown	0.05	0.90	0.01	0.04	Brown
Individual 2	Brown	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive
Individual 4	Brown	0.02	0.83	0.06	0.09	Brown
Individual 5	Brown	0.13	0.55	0.31	0.00	Brown
Individual 6	Dark Brown	0.10	0.53	0.36	0.00	Brown
Individual 8	Brown	0.09	0.64	0.27	0.01	Brown
Individual 11	Brown	0.01	0.14	0.73	0.12	Blond
Individual 13	Red	0.00	0.10	0.78	0.11	Blond
Individual 14	Dark Brown	0.46	0.40	0.14	0.00	Black
Individual 15	Dark Brown	0.52	0.35	0.12	0.00	Black
Individual 16	Dark Brown	0.12	0.61	0.27	0.00	Brown
Individual 21	Dark Blond	0.29	0.61	0.09	0.00	Brown
Individual 23	Dark Brown	0.01	0.15	0.83	0.00	Blond
Individual 40	Black	0.77	0.22	0.01	0.00	Black
Individual 41	Brown	0.02	0.64	0.33	0.01	Brown
Individual 42	Brown	0.12	0.49	0.39	0.00	Brown
Individual 43	Dark Brown	0.09	0.44	0.47	0.00	Blond
Individual 44	Dark Brown	0.04	0.67	0.08	0.21	Brown
Individual 45	Black	0.89	0.11	0.00	0.00	Black
Individual 46	Dark Brown	0.02	0.16	0.82	0.00	Blond
Individual 47	Dark Brown	0.07	0.59	0.32	0.01	Brown
Individual 48	Dark Brown	0.02	0.37	0.34	0.26	Brown
Individual 49	Dark Brown	0.07	0.86	0.06	0.00	Brown
Individual 50	Dark Brown	0.90	0.09	0.00	0.00	Black
Individual 51	Dark Brown	0.03	0.57	0.30	0.10	Brown
Individual 52	Dark Brown	0.12	0.45	0.42	0.00	Brown
Individual 53	Dark Brown	0.82	0.18	0.00	0.00	Black
Individual 55	Dark Blond	0.01	0.17	0.70	0.12	Blond
Individual 56	Brown	0.01	0.15	0.84	0.00	Blond
Individual 60	Dark Blond	0.03	0.25	0.72	0.00	Blond
Individual 61	Brown	0.01	0.06	0.03	0.91	Red
Individual 62	Brown	0.08	0.69	0.17	0.07	Brown
Individual 63	Dark Brown	0.15	0.67	0.17	0.00	Brown
Individual 64	Brown	0.01	0.47	0.35	0.17	Brown
Individual 65	Dark Blond	0.01	0.19	0.30	0.50	Red
Individual 66	Brown	0.12	0.58	0.30	0.00	Brown
Individual 67	Brown	0.03	0.77	0.09	0.11	Brown
Individual 68	Brown	0.08	0.66	0.26	0.00	Brown
Individual 70	Blond	0.07	0.61	0.31	0.01	Brown
Individual 71	Brown	0.14	0.64	0.20	0.02	Brown
Individual 72	Brown	0.16	0.53	0.29	0.01	Brown
Individual 73	Dark Blond	0.01	0.18	0.63	0.18	Blond
Individual 74	Brown	0.03	0.47	0.31	0.18	Brown
Individual 75	Dark Brown	0.22	0.63	0.15	0.00	Brown
Individual 76	Brown	0.05	0.78	0.16	0.00	Brown
Individual 77	Dark Brown	0.07	0.63	0.28	0.02	Brown
Individual 78	Brown	0.14	0.44	0.41	0.01	Brown
Individual 79	Brown	0.15	0.41	0.44	0.00	Blond
Individual 80	Dark Blond	0.03	0.41	0.55	0.00	Blond
Individual 81	Brown	0.01	0.24	0.73	0.02	Blond
Individual 82	Blond	0.02	0.22	0.76	0.01	Blond
Individual 83	Brown	0.03	0.47	0.36	0.14	Brown
Individual 84	Dark Blond	0.10	0.53	0.35	0.02	Brown
Individual 85	Blond	0.00	0.00	0.00	1.00	Red
Individual 86	Brown	0.03	0.48	0.12	0.37	Brown
Individual 87	Dark Blond	0.01	0.17	0.82	0.00	Blond
Individual 88	Red	0.00	0.18	0.08	0.74	Red
Individual 89	Brown	0.04	0.33	0.62	0.01	Blond
Individual 90	Brown	0.09	0.67	0.23	0.00	Brown
Individual 91	Brown	0.12	0.43	0.44	0.01	Blond
Individual 92	Brown	0.09	0.51	0.39	0.01	Brown
Individual 93	Brown	0.06	0.41	0.47	0.06	Blond
Individual 94	Dark Blond	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive
Individual 95	Blond	0.01	0.13	0.86	0.00	Blond
Individual 96	Dark Blond	0.03	0.29	0.67	0.01	Blond
Individual 97	Dark Brown	0.41	0.42	0.17	0.00	Brown
Individual 98	Brown	0.00	0.09	0.15	0.75	Red
Individual 99	Brown	0.07	0.54	0.38	0.01	Brown
Individual 100	Black	0.74	0.26	0.00	0.00	Black
Individual 101	Brown	0.07	0.22	0.71	0.00	Blond
Individual 102	Dark brown	0.11	0.60	0.24	0.05	Brown
Individual 103	Blond	0.01	0.31	0.68	0.01	Blond
Individual 104	Dark Blond	0.02	0.68	0.03	0.26	Brown

HIrisPlex						
Black p-value	Brown p-value	Blond p-value	Red p-value	Light p-value	Dark p-value	Inferred Hair Colour
0.04	0.91	0.01	0.03	0.25	0.75	Brown/Dark Brown
0.21	0.69	0.10	0.00	0.32	0.68	Dark Brown/Black
0.02	0.83	0.05	0.10	0.77	0.23	Brown/Dark Brown
0.12	0.61	0.27	0.00	0.72	0.28	Dark Brown/Black
0.08	0.63	0.28	0.00	0.78	0.22	Dark Brown/Black
0.07	0.67	0.25	0.01	0.69	0.31	Dark Brown/Black
0.01	0.16	0.72	0.11	0.99	0.01	Blond
0.00	0.00	0.00	1.00	0.99	0.01	Red
0.45	0.47	0.08	0.00	0.36	0.64	Dark Brown/Black
0.42	0.47	0.11	0.00	0.23	0.77	Dark Brown/Black
0.11	0.64	0.24	0.00	0.68	0.32	Dark Brown/Black
0.25	0.66	0.09	0.00	0.23	0.77	Dark Brown/Black
0.02	0.22	0.76	0.00	0.99	0.01	Blond
0.55	0.44	0.01	0.00	0.01	0.99	Black
0.02	0.66	0.31	0.01	0.93	0.07	Brown/Dark Brown
0.11	0.55	0.34	0.00	0.73	0.27	Dark Brown/Black
0.08	0.49	0.42	0.00	0.86	0.14	Brown/Dark Brown
0.03	0.68	0.09	0.20	0.65	0.35	Dark Brown/Black
0.84	0.16	0.00	0.00	0.01	0.99	Black
0.02	0.21	0.77	0.00	0.98	0.02	Blond
0.06	0.60	0.32	0.01	0.79	0.21	Dark Brown/Black
0.02	0.38	0.33	0.27	0.93	0.07	Brown/Dark Brown
0.07	0.86	0.06	0.00	0.39	0.61	Brown/Dark Brown
0.78	0.22	0.00	0.00	0.00	1.00	Black
0.03	0.62	0.25	0.10	0.94	0.06	Brown/Dark Brown
0.14	0.57	0.29	0.00	0.80	0.20	Brown/Dark Brown
0.62	0.37	0.00	0.00	0.00	1.00	Black
0.01	0.21	0.65	0.14	0.99	0.01	Blond/Dark Blond
0.01	0.19	0.80	0.00	0.99	0.01	Blond
0.03	0.28	0.69	0.00	0.95	0.05	Blond/Dark Blond
0.00	0.03	0.02	0.94	0.87	0.13	Red
0.07	0.72	0.14	0.07	0.77	0.23	Brown/Dark Brown
0.14	0.70	0.16	0.00	0.51	0.49	Brown/Dark Brown
0.01	0.51	0.35	0.13	0.98	0.02	Brown/Dark Brown
0.00	0.16	0.27	0.56	0.98	0.02	Red
0.12	0.60	0.28	0.00	0.63	0.37	Dark Brown/Black
0.03	0.78	0.10	0.09	0.69	0.31	Brown/Dark Brown
0.07	0.66	0.26	0.00	0.73	0.27	Dark Brown/Black
0.05	0.63	0.30	0.01	0.78	0.22	Dark Brown/Black
0.11	0.66	0.21	0.02	0.48	0.52	Dark Brown/Black
0.14	0.58	0.27	0.01	0.65	0.35	Dark Brown/Black
0.01	0.19	0.61	0.19	0.99	0.01	Blond/Dark Blond
0.03	0.51	0.28	0.18	0.88	0.12	Brown/Dark Brown
0.17	0.68	0.14	0.00	0.39	0.61	Dark Brown/Black
0.06	0.78	0.16	0.00	0.70	0.30	Brown/Dark Brown
0.05	0.65	0.28	0.02	0.73	0.27	Dark Brown/Black
0.14	0.48	0.38	0.01	0.78	0.22	Dark Brown/Black
0.14	0.49	0.37	0.00	0.77	0.23	Dark Brown/Black
0.04	0.44	0.52	0.00	0.93	0.07	Blond/Dark Blond
0.01	0.27	0.70	0.02	0.99	0.01	Blond
0.02	0.26	0.72	0.01	0.98	0.02	Blond
0.02	0.45	0.41	0.12	0.91	0.09	Brown/Dark Brown
0.09	0.54	0.35	0.02	0.76	0.24	Dark Brown/Black
0.00	0.26	0.00	0.74	0.75	0.25	Red
0.03	0.44	0.13	0.40	0.83	0.17	Brown/Dark Brown
0.02	0.21	0.78	0.00	0.98	0.02	Blond
0.00	0.15	0.06	0.79	0.97	0.03	Red
0.04	0.39	0.57	0.01	0.95	0.05	Blond/Dark Blond
0.09	0.71	0.20	0.00	0.69	0.31	Brown/Dark Brown
0.09	0.45	0.44	0.01	0.72	0.28	Dark Brown/Black
0.09	0.56	0.34	0.01	0.83	0.17	Brown/Dark Brown
0.06	0.42	0.45	0.07	0.89	0.11	Dark Blond/Brown
0.00	0.09	0.75	0.16	1.00	0.00	Blond
0.01	0.19	0.80	0.00	0.99	0.01	Blond
0.03	0.35	0.61	0.01	0.96	0.04	Blond/Dark Blond
0.31	0.56	0.13	0.00	0.37	0.63	Dark Brown/Black
0.00	0.11	0.16	0.73	0.99	0.01	Red
0.06	0.55	0.38	0.01	0.81	0.19	Brown/Dark Brown
0.64	0.36	0.00	0.00	0.01	0.99	Black
0.07	0.26	0.67	0.00	0.93	0.07	Blond/Dark Blond
0.11	0.61	0.24	0.04	0.68	0.32	Dark Brown/Black
0.01	0.34	0.64	0.01	0.98	0.02	Blond/Dark Blond
0.02	0.59	0.04	0.35	0.55	0.45	Dark Brown/Black

Supplementary Table 5. Results for inferring biogeographical ancestry

Sample ID	Self-Declared Biogeographical Ancestry				Universal Analysis Software
	Maternal Grandfather	Maternal Grandmother	Paternal Grandfather	Paternal Grandmother	Included Populations
Individual 1	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 2	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 4	Australian Anglo Celtic	England	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 5	Australian Anglo Celtic	Australian Anglo Celtic	England	Australian Anglo Celtic	European
Individual 6	Australian Anglo Celtic	England	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 8	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 11	Australian Anglo Celtic	England	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 13	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 14	England	England	Indonesia	Netherlands	Inconclusive
Individual 15	England	England	Indonesia	Netherlands	Inconclusive
Individual 16	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 21	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Italy	European
Individual 23	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 40	Nepal	Nepal	Nepal	Nepal	Inconclusive
Individual 41	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 42	Australian Anglo Celtic	Australian Anglo Celtic	England	Australian Anglo Celtic	European
Individual 43	Switzerland	Australian Anglo Celtic	Italy	Australian Anglo Celtic	European
Individual 44	Ireland	Ireland	Lebanon	Italy	European
Individual 45	China	China	China	China	East Asian
Individual 46	Ireland	Australian Anglo Celtic	Australian Anglo Celtic	England	European
Individual 47	England	England	England	England	European
Individual 48	Hungary	West Indes	France	France	Inconclusive
Individual 49	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 50	Vietnam	Vietnam	Vietnam	Vietnam	East Asian
Individual 51	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 52	Italy	Italy	England	Australian Anglo Celtic	European
Individual 53	Fiji	Fiji	Fiji	Fiji	Inconclusive
Individual 55	Australian Anglo Celtic	Aboriginal Australian	Ireland	England	European
Individual 56	Aboriginal Australian	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 60	Brazil	Brazil	Brazil	Brazil	European
Individual 61	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 62	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 63	England	Ireland	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 64	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 65	England	Finland	Australian Anglo Celtic	Aboriginal Australian	European
Individual 66	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European

Principle Coordinate Analysis	Structure		Forensic Resource Reference on Genetics - Knowledge Base	Optimised Pipeline
Included Populations	Major Population (>50%)	Minor Population (10-50%)	Included Populations	Included Populations
European	European	Middle Eastern	European, Middle Eastern	European
European	European		European, Middle Eastern	European
European	European		European	European
European	European	Middle Eastern	European, Middle Eastern	European
European	Middle Eastern	European	European, Middle Eastern	European
European	Middle Eastern	European	European, Middle Eastern, Central Asian	European
European	European		European, Middle Eastern	European
European	European		European	European
Inconclusive		European, East Asian, Middle Eastern	European, Middle Eastern, South Asian, East Asian, Central Asian	European, Middle Eastern, East Asian
Inconclusive	European	East Asian	European, Middle Eastern, South Asian, East Asian, Central Asian	European, East Asian
European	Middle Eastern	European	European, Middle Eastern	European
European	Middle Eastern	European	European, Middle Eastern, Central Asian	European
European	European		European	European
South Asian	South Asian	Oceanian	South Asian, Middle Eastern, Central Asian	South Asian
European	European	Middle Eastern	European, Middle Eastern	European
European	European		European, Middle Eastern	European
European	Middle Eastern		European, Middle Eastern	European
European	Middle Eastern	European	European, Middle Eastern, Central Asian	European
East Asian	East Asian		East Asian	East Asian
European	European		European	European
European	European		European	European
Inconclusive	European	African, Middle Eastern	Middle Eastern	European, Middle Eastern, African
European	European		European, Middle Eastern	European
East Asian	East Asian		East Asian	East Asian
European	European		European, Middle Eastern	European
European	European	Middle Eastern	European, Middle Eastern	European
South Asian	South Asian		South Asian	South Asian
European	European		European, Middle Eastern	European
European	European		European, Middle Eastern	European
European	Middle Eastern	European	European, Middle Eastern	European
European	European		European	European
European	European	Middle Eastern	European	European
European	European		European	European
European	European		European, Middle Eastern	European
European	European		European, Middle Eastern	European
European	European	Middle Eastern	European, Middle Eastern	European

Sample ID	Self-Declared Biogeographical Ancestry				Universal Analysis Software
	Maternal Grandfather	Maternal Grandmother	Paternal Grandfather	Paternal Grandmother	Included Populations
Individual 67	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 68	Scotland	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 70	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 71	American White Caribbean	American European	Scotland	Australian Anglo Celtic	European
Individual 72	England	Malta	Egypt	Malta	European
Individual 73	South Africa/Netherlands	South Africa	Netherlands	Netherlands	European
Individual 74	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 75	Ireland	Ireland	England	England	European
Individual 76	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 77	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 78	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 79	South Africa	South Africa	South Africa	South Africa	European
Individual 80	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 81	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 82	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 83	England	England	England	England	European
Individual 84	American European	Germany	England	England	European
Individual 85	Scotland	Scotland	Netherlands	Netherlands	European
Individual 86	Australian Anglo Celtic	Australian Anglo Celtic	Poland	Poland	European
Individual 87	Australian North/West European	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 88	Australian Anglo Celtic	Australian Anglo Celtic	Scotland	Australian Anglo Celtic	European
Individual 89	Australian Anglo Celtic	Scotland	Australian Anglo Celtic	Ireland	European
Individual 90	Lithuania	Lithuania	England	England	European
Individual 91	Macedonia	Macedonia	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 92	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 93	China	China	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 94	Australian North/West European	Australian Southern European	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 95	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 96	Australian Anglo Celtic	Australian Anglo Celtic	Australian North/West European	Australian Anglo Celtic	European
Individual 97	Ukraine	Ukraine	Australian Anglo Celtic	Japan	European
Individual 98	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 99	Finland	Germany	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 100	China	China	England	Scotland	Inconclusive
Individual 101	Poland	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European
Individual 102	Australian Anglo Celtic	Australian Anglo Celtic	England	England	European
Individual 103	South Africa	South Africa	England	England	European
Individual 104	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	Australian Anglo Celtic	European

Principle Coordinate Analysis	Structure		Forensic Resource Reference on Genetics - Knowledge Base	Optimised Pipeline
Included Populations	Major Population (>50%)	Minor Population (10-50%)	Included Populations	Included Populations
European	European		European, Middle Eastern, Central Asian	European
European	European		European	European
European	European		European, Middle Eastern	European
European	European		European	European
European	Middle Eastern	European	European, Middle Eastern	European
European	European	Middle Eastern	European, Middle Eastern	European
European	European		European	European
European	European		European, Middle Eastern	European
European	European		European, Middle Eastern	European
European	European		European, Middle Eastern	European
European	European		European	European
European	European		European, Middle Eastern	European
European	European		European	European
European	European	Middle Eastern	European, Middle Eastern	European
European	European		European	European
European	European		European	European
European	European		European, Middle Eastern	European
European	European		European, Middle Eastern	European
European	European		European, Middle Eastern	European
European	European		European, Middle Eastern	European
European	European		European, Middle Eastern	European
European	European	Middle Eastern	European, Middle Eastern	European
European	European	Middle Eastern	European, Middle Eastern	European
European	European	Middle Eastern	European, Middle Eastern	European
European	European	Middle Eastern	European, Middle Eastern, Central Asian	European
Inconclusive	European	East Asian	European, Middle Eastern	European, East Asian
European	European		European	European
European	European		European	European
European	European	Middle Eastern	European, Middle Eastern	European
Inconclusive		Oceanian, Middle Eastern, European	European, Middle Eastern	European, Oceanian, Middle Eastern
European	European		European, Middle Eastern	European
European	European		European	European
Inconclusive		East Asian, Middle Eastern, European	East Asian, Central Asian	East Asian, European, Middle Eastern
European	European		European	European
European	European		European	European
European	European		European	European
European	European		European	European

7. CHAPTER SEVEN

EXTENDED KINSHIP INFERENCE. PART 1: EVALUATION OF SHORT TANDEM REPEATS AND SINGLE NUCLEOTIDE POLYMORPHISMS USING LIKELIHOOD RATIOS AND HAPLOTYPE MATCHING

Contributions of authors:

Watson JL, Grisedale K, Coakley L, McNevin D and Ward J. (2025) 'Extended Kinship Inference. Part 1: Evaluation of Short Tandem Repeats and Single Nucleotide Polymorphisms using Likelihood Ratios and Haplotype Matching' Forensic Genomics, 5(1), DOI: 10.1089/forensic.2025.0001.

Jessica Watson (Candidate)

Conceptualised and assisted in the experimental design for the study. Carried out sample collection, sample preparation, DNA testing and data analysis. Prepared manuscript, edited manuscript following review by other co-authors, referenced manuscript, carried out manuscript's submission and incorporated feedback from peer review process.

Signed: Production Note: Signature removed prior to publication.

Kelly Grisedale (AFP DNA Program staff)

Assisted in the experimental design for the study. Contributed to DNA testing and data analysis. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Louise Coakley (AFP DNA Program staff)

Assisted in the experimental design for the study. Contributed to data analysis. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Dennis McNevin (Co-supervisor)

Conceptualised and assisted in the experimental design for the study. Contributed to data analysis. Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

Jodie Ward (Co-supervisor)

Conceptualised and assisted in the experimental design for the study. Acquired funding for the study. Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

Forensic Genomics
Volume 5, Number 1, 2025
© Mary Ann Liebert, Inc.
DOI: 10.1089/forensic.2025.0001

Forensic Genomics

Open camera or QR reader and scan code to access this article and other resources online.

ORIGINAL ARTICLE

Extended Kinship Inference Part 1: Evaluation of Short Tandem Repeats and Single Nucleotide Polymorphisms Using Likelihood Ratios and Haplotype Matching

Jessica L. Watson, BForSc (Hons),[1–3,*] Kelly Grisedale, PhD,[2,3] Louise Coakley, MBA,[3,4] Dennis McNevin, PhD,[1,3] and Jodie Ward, PhD[1,3]

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

Abstract

Medium- and long-range kinship analysis using single nucleotide polymorphism (SNP) genotyping allows law enforcement to generate investigative leads by identifying an unknown individual through their close and distant genetic relatives. Short-range kinship inference can be conducted through calculation of likelihood ratios (LRs) using population allele frequency data and the degree of similarity between two DNA profiles but has historically been limited to short tandem repeat (STR) profiles. Alternatively, identical by descent (IBD) segment matching algorithms can be used to detect shared DNA haplotypes between two genetically related individuals that have been inherited from a common ancestor. The ForenSeq® Kintelligence Kit enables law enforcement and forensic laboratories to utilize medium-density SNP sequencing technology for extended kinship inference by analyzing 10,230 SNPs. In this study, DNA from two pedigrees was used to compare the ability of STR profiles, identity-informative SNP (iiSNP) profiles, Kintelligence profiles and Kintelligence and direct-to-consumer profiles available on public genetic genealogy databases to detect and classify genetic relationships. The DNA profiles were analyzed using DBLR™ software to calculate kinship LRs or uploaded to GEDmatch PRO™ for IBD segment matching with either database searching or one-to-one comparisons. The LRs calculated for STR and iiSNP profiles were able to correctly infer first degree relationships (i.e., parent, offspring, and full sibling), with the combined discrimination power able to distinguish between second degree relationships. LRs calculated for the Kintelligence profiles exceeded one million for 93% of full sibling to fifth degree relationships tested. IBD segment matching was effective for detecting first to fifth degree relatives when Kintelligence profiles were searched on the GEDmatch PRO™ database. The results of this study demonstrate that the Kintelligence Kit is a valuable tool for law enforcement and forensic investigators, offering an advanced method for medium-range kinship testing using either LRs or IBD segment matching.

Keywords: single nucleotide polymorphism, human identification, targeted amplicon sequencing, forensic investigative genetic genealogy, kinship, short tandem repeat

[1]*Faculty of Science, Centre for Forensic Science, School of Mathematical and Physical Sciences, University of Technology Sydney, New South Wales, Australia.*
[2]*Forensics Command, Australian Federal Police, Australian Capital Territory, Australia.*
[3]*National DNA Program for Unidentified and Missing Persons, Australian Federal Police, Australian Capital Territory, Australia.*
[4]*Genetic Genealogy Pty Ltd, Queensland, Australia.*

*Address correspondence to: *Jessica L. Watson, BForSc (Hons), Faculty of Science, Centre for Forensic Science, School of Mathematical and Physical Sciences, University of Technology Sydney, 15 Broadway, Ultimo, New South Wales, 2007, Australia,* Email: Jessica.Watson@student.uts.edu.au

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

Introduction

Kinship inference is a valuable identification tool to generate investigative leads from an unknown DNA sample where other forensic identification methods have been unsuccessful. This has been applied to the identification of persons of interest (PoIs) in criminal investigations and unidentified human remains in missing person investigations. The technique involves uploading an unknown DNA profile to a relevant database and collating a list of any DNA profiles that partially match the unknown DNA profile (a kinship match) that could be genetic relatives or directly comparing the unknown DNA profile with a known family reference sample.[1,2] Depending on the genetic marker panel, first (e.g., parent/offspring) to ninth (e.g., fourth cousins) relationships can be detected and classified as possible genetic relatives.

Alleles in common between two DNA profiles are identical by state (IBS). However, if the alleles are inherited from a common ancestor, they are also identical by descent (IBD).[3] Statistical support for relatedness between two individuals can be calculated using the likelihood ratio (LR) approach. This is the ratio of two conditional probabilities for observing IBS alleles in the DNA profiles under alternative hypotheses about the genetic relationship between the individuals.[4,5]

Short tandem repeat (STR) markers, the most commonly used marker type in forensic genetics and accepted in many jurisdictions as a scientific method for identification, have facilitated short-range kinship testing since 2002.[4,6,7] However, STR profiles have only been routinely applied to first degree relationships (parents, offsprings, and full siblings), as more distant genetic relatives are difficult to distinguish from unrelated individuals with the limited number of markers targeted by STR panels.[8] In contrast, single nucleotide polymorphism (SNP) panels target more markers and can therefore expand the reach of kinship inference to detect more distant relationships.[3,9]

It is possible that two alleles are IBS not because they are IBD but because of mutations occurring in different lineages independently. This can be accounted for with medium- to high-density SNP panels that target SNPs in close proximity to each other on the genome. These markers are less likely to be separated during recombination and therefore co-inherited, forming a haplotype or IBD segment.[10,11] Algorithms can compare dense SNP profiles to detect shared IBD segments and calculate the amount of shared DNA in centimorgans (cM). The number, size, and total length of these IBD segments can then be used to infer the type of relationship.[11]

IBD segment matching algorithms have been implemented by public genetic genealogy databases where individuals can upload their DNA profiles to find their relatives.[12] These databases are populated with direct-to-consumer (DTC) profiles with hundreds of thousands of SNPs generated by private genetic testing companies.[13] There are two such databases accessible to law enforcement for forensic investigative genetic genealogy (FIGG). These are GEDmatch PRO™, the law enforcement accessible portal for the GEDmatch™ database, maintained by QIAGEN and Bode Technology, and FamilyTreeDNA, maintained by Othram Inc., both of which have over two million profiles.[14,15] However, these profiles are not accessible to law enforcement for database searching unless the owner of the profile has opted in to law enforcement searching.[16–18]

The ForenSeq® Kintelligence Kit is a targeted amplicon sequencing (TAS) panel developed by Verogen Inc. (a QIAGEN subsidiary) to allow in-house extended kinship analysis of forensic samples.[19,20] It targets 10,230 SNPs across the human genome: 106 X chromosome SNPs, 85 Y chromosome SNPs, 56 ancestry-informative SNPs, 24 phenotype-informative SNPs, 94 identity-informative SNPs (iiSNPs), and 9,867 kinship-informative SNPs (kiSNPs). Kintelligence profiles are compatible with upload to GEDmatch PRO™ for IBD segment matching with either the one-to-one comparison with known profiles or using the one-to-many database searching function.[17] These tools have a windowed kinship algorithm that locates shared IBD segments using modified PC-AiR and PC-Relate tools that calculate kinship coefficients for the segments rather than identifying stretches of identical SNP alleles on homologous chromosomes.[21] The Kintelligence Kit has been validated and employed for casework by a number of law enforcement, academic, and private laboratories.[19,22–26]

Current FIGG guidelines require that once a putative identification has been made by law enforcement, additional genetic and nongenetic testing is required to confirm the results.[27] With extended kinship inference, LR calculations could provide statistical support for such identifications, especially when suitable direct or close family reference samples are unavailable. This study sampled two pedigrees to investigate the application of STRs and SNPs for extended kinship inference using LRs and IBD segment matching. This paper presents part one of the study, which focuses on assessing the power of medium-density SNP profiling for extending the capabilities of kinship inference beyond the use of STRs.

Methods

Ethics approval and sample procurement

Ethics approval for this research was granted by the University of Technology (UTS) Human Research Ethics

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

Committee (HREC) (UTS HREC NO. ETH21-5821). Volunteers from Pedigree Group 1 (Individuals 1–12) provided self-administered buccal swabs with informed consent for DNA profiling with the Kintelligence Kit. Profiles were subsequently uploaded to GEDmatch PRO™ (Fig. 1). Volunteers from Pedigree Group 2 (Individuals A– L) provided the kit numbers of pre-existing DTC kits uploaded to GEDmatch™ with informed consent to match against Kintelligence profiles uploaded to GEDmatch PRO™ (Fig. 2; Table 1).[a] The central individual for Pedigree Group 2 (Individual A) also provided a self-administered buccal swab for sequencing with the Kintelligence Kit. All volunteers completed a form mapping their genetic relationships with the other volunteers in the study.

DNA extraction and quantification

DNA was extracted from the buccal swabs manually using the QIAamp DNA Investigator® Kit (QIAGEN) following the manufacturer's protocols for Omni Swabs with an elution volume of 100 μL.[28] Quantification was performed using the Quantifiler™ Trio DNA Quantification Kit (Thermo Fisher Scientific) on a QuantStudio™ 5 Real-Time PCR System (Thermo Fisher Scientific) following the manufacturer's protocols.[29,30] A degradation index (DI) was calculated for each sample as the ratio of concentrations of the small autosomal (SA) target and the large autosomal (LA) target.

STR profiling

STR profiles were generated for Pedigree Group 1 samples using the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific).[31] The samples were amplified for 29 cycles on the Veriti™ 96-Well Fast Thermal Cycler (Thermo Fisher Scientific). The LA target was used to dilute the extracted DNA to the required input amount of 1.0 ng. Capillary electrophoresis was performed on the 3500xL Genetic Analyzer (Thermo Fisher Scientific) following the manufacturer's protocols.[32] The electropherograms were analyzed with GeneMapper™ ID-X v1.6 with an analytical threshold of 225 relative fluorescence units (RFU) and homozygous threshold of 1,000 RFU.[33]

Library preparation and sequencing

Libraries were prepared manually with the Kintelligence Kit following a modified protocol.[25,34] DNA input was calculated using the LA target concentration if the DI was >1 to avoid overdiluting degraded samples, otherwise the samples were diluted using the SA target concentration. Library preparation was performed in batches of 12 libraries including a positive control (NA24385; provided with the Kintelligence Kit) and a negative control (nuclease-free water). The libraries were normalized to 0.75 ng/μL using the QuantiFluor® ONE dsDNA System (Promega) on the Quantus™ Fluorometer (Promega) and pooled in batches of three libraries for sequencing.[35,36] Sequencing was performed on the MiSeq FGx® Sequencing System (QIAGEN) using a standard flow cell.[37,38]

The Kintelligence Sample Reports were exported from the Universal Analysis Software (UAS) v2.5 (QIAGEN) and analyzed according to optimized thresholds in a Microsoft Excel macro described by.[25,39] This included a total read threshold of 20 reads and relative minor allele frequency thresholds for homozygotes (0.95–1.00), heterozygotes (0.10–0.90), sequencing error (<0.05), and ambiguous variants (with relative frequencies in ranges 0.05–0.10 and 0.90–0.95).

Extended kinship analysis

STR, iiSNP (included in the Kintelligence Kit), combined identity marker (STR and iiSNP), and entire Kintelligence profiles were generated for all individuals in Pedigree Group 1. Profiles were compared using kinship LR calculation and IBD segment matching on GEDmatch PRO™. A Kintelligence profile was generated for Individual A of Pedigree Group 2. This profile was compared with DTC profiles of individuals in Pedigree Group 2 using IBD segment matching on GEDmatch PRO™.

LR calculations. To calculate LRs, two alternative hypotheses were tested for each pair-wise combination of individuals in Pedigree Group 1. For unrelated pairs of individuals (19 pairs), the first hypothesis (H_1) proposed a relationship ranging from first (parent/offspring or full sibling) to eighth degree relationships and the alternative hypothesis (H_2) proposed that the individuals were unrelated. For the genetically related pairs of individuals, H_1 proposed the true relationship of the individuals and H_2 proposed that they were unrelated. For example:

H_1: Individual 1 and Individual 2 are parent/offspring.

H_2: Individual 1 and Individual 2 are unrelated members of the European population.

For each pair-wise genotype comparison, G_xG_y, LRs were calculated using the Kinship module in DBLR™ v1.3 (STRmix™)[40] according to:

[a]*The use of DTC data for research was in compliance with the terms and conditions at the time the respective data were downloaded by the volunteers. All volunteers were notified at the conclusion of the study so they could opt out of law enforcement searching or remove their DNA data from GEDmatch™ to ensure that future use of their data is in compliance with their preferences and relevant terms and conditions.*

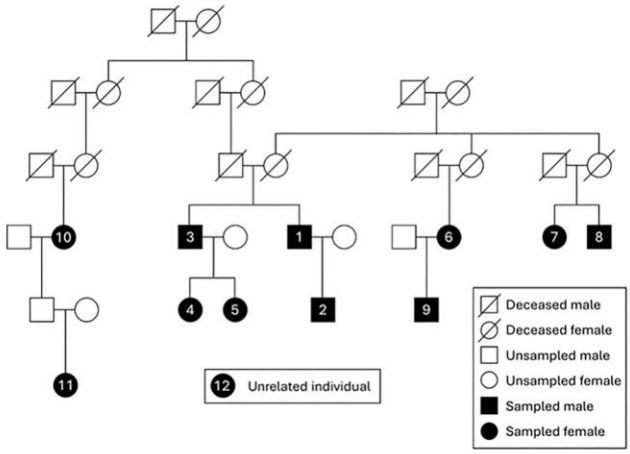


FIG. 1. Pedigree Group 1, with numbers corresponding to Individuals 1 through 12. Individuals included in the study are filled in (black), with living relatives linking the family members not filled (white). Deceased relatives are crossed out. Individual 12 is unrelated to all other individuals in the group.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

$$\mathrm{LR} = \frac{\mathrm{P}(G_x G_y | H_1)}{\mathrm{P}(G_x G_y | H_2)}$$

DBLR™ does not report $\log_{10}$ LRs that exceed 300 ($\mathrm{LR} > 10^{300}$) and return the $\log_{10}$ LR as infinite (∞).[40] When the relationship is impossible ($\mathrm{LR} = 0$), the $\log_{10}$ LR results are given as negative infinity ($-\infty$).[40] The verbal scale for statistical support provided by $\log_{10}$ LR results is provided in Table 2.

As all individuals in this study self-identified as Australian with European ancestry, population allele frequency data for each genetic marker panel were obtained from the sources in Table 3. A Kintelligence Kit linkage map was constructed using the Map Interpolator of the Rutger's Map v.3 with sex-averaged cM positions and imported into DBLR™ to account for linkage.[41]

IBD segment matching. The Kintelligence profiles were manually edited on the UAS to be consistent with the genotypes generated with the additional thresholds in the Microsoft Excel macro. GEDmatch PRO Reports were exported from the UAS and uploaded to GEDmatch PRO™ as laboratory validation samples. Database searching was conducted using the One-to-Many Kinship tool and matches were assessed in the high confidence and expanded match lists (Table 4). For samples that were not returned in either match list, the One-to-One Kintelligence DNA Comparison tool was used to directly compare the pair of samples with their unique kit numbers.

The estimated shared cM values, average IBD segment length, number of IBD segments, and number of overlapping SNPs between kits were analyzed. The theoretical values were estimated using the equations of Speed & Balding (2015), modified for total length of DNA in cM, rather than Mb (Supplementary Data).[42] This determined the expected values for shared cM, average IBD segment length, and number of IBD segments for all relationship levels tested in this study (Table 5). It should be recognized, however, that there is not a perfect correlation between cM and Mb and so the expected cM values are only approximate.

Results

DNA profiles

Table 6 outlines the call rates for the different profiles obtained for Pedigree Group 1. The average call rates exceeded 98% (STRs: 98.4 ± 1.2; iiSNPs: 98.3 ± 0.6; combined identity markers: 98.7 ± 0.5; Kintelligence: 98.4 ± 0.7). The Kintelligence profile for Individual A of Pedigree Group 2 had a call rate of 99.4%.

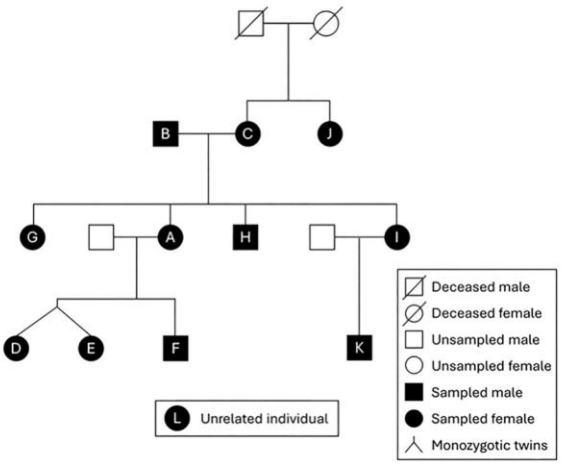


FIG. 2. Pedigree Group 2, with letters corresponding to Individuals A through L. Individuals included in the study are filled in (black), with living relatives linking the family members not filled (white). Deceased relatives are crossed out. Individual L is unrelated to all other individuals in the group.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

LR calculations with identity-informative genetic markers (Pedigree Group 1)

STR profiles. Figure 3 shows the $\log_{10}$ LRs for the true related and unrelated pairwise comparisons in Pedigree Group 1. For the first degree relationships, the parent/offspring $\log_{10}$ LRs ranged from 6.0 to 9.1 and the full sibling $\log_{10}$ LRs from 2.6 to 8.5, all providing very strong statistical support for the true relationship (H_1) as opposed to the pairs being unrelated (H_2). $\log_{10}$ LRs for the second degree relationships ranged from 0.3 to 4.4, providing limited statistical support H_1 for an avuncular relationship (Individual 1 and Individual 4, $\log_{10}$ LR = 0.3) and grandparent/grandchild (Individual 10 and Individual 11, $\log_{10}$ LR = 0.9).

The $\log_{10}$ LRs generated for the unrelated pairs provided strong statistical support for the pairs being unrelated as opposed to being first or second degree relatives. The range of true full sibling $\log_{10}$ LRs was 2.5 to 8.5 (Fig. 3). There was a false positive in the unrelated pairs with H_1 proposing a full sibling relationship (Individual 1 and Individual 12, $\log_{10}$ LR = 2.1), indicating it was approximately 120 times more likely to observe the STR profiles if the individuals were full siblings than if they were unrelated. Parent/offspring $\log_{10}$ LRs for the unrelated pairs were unable to be calculated because $P(G_xG_y \mid H_1) = 0$ as

Table 1. Direct-to-consumer Kits Available for Individuals A Through L in Group 2

Sample ID	*AncestryDNA*	*FamilyTreeDNA*	*23andMe*	*Living DNA*
Individual A	✓	✓	✓	✓
Individual B	✓	✓	✓	✓
Individual C	✓	✓	✓	✓
Individual D	✗	✓	✗	✗
Individual E	✓	✓	✗	✗
Individual F	✗	✓	✗	✗
Individual G	✓	✓	✗	✗
Individual H	✓	✓	✗	✗
Individual I	✓	✓	✗	✗
Individual J	✗	✓	✗	✗
Individual K	✗	✓	✗	✗
Individual L	✓	✗	✗	✗

Table 2. Verbal Scale for $\log_{10}$ LR Results and the Statistical Support for Each Hypothesis[41]

$\log_{10}$ LR	*Interpretation*
≤−5	Very strong statistical support for H_2
−4	Strong statistical support for H_2
−3	Moderately strong statistical support for H_2
−2	Moderate statistical support for H_2
−1	Limited statistical support for H_2
0	Uninformative
1	Limited statistical support for H_1
2	Moderate statistical support for H_1
3	Moderately strong statistical support for H_1
4	Strong statistical support for H_1
≥5	Very strong statistical support for H_1

LR, likelihood ratio.

Table 3. Population Allele Frequency Data Utilized for Each Marker Set: Short Tandem Repeat, Identity-Informative Single Nucleotide Polymorphism, Combined Identity Markers (STRs and iiSNPs), and All Kintelligence SNPs

Market set	*Panel*	*Markers*	*Population data*	*Source*
STRs	GlobalFiler™ PCR Amplification Kit[31]	21	Australian Caucasian	Taylor et al. (2017)[57]
iiSNPs	ForenSeq® Kintelligence Kit[34]	94	Australian with European Ancestry	Watson et al. (2024)[46]
Combined Identity Markers	GlobalFiler™ PCR Amplification Kit and ForenSeq® Kintelligence Kit[31,34]	114[a]	Australian Caucasian (STR) Australian with European Ancestry (iiSNP)	Taylor et al. (2017) and Watson et al. (2024)[46,57]
Kintelligence SNPs	ForenSeq® Kintelligence Kit[34]	10,039[b]	European	1000 Genomes[58]

[a]SE33 was excluded as linkage equilibrium tests could not be performed between SE33 and iiSNPs.[44]
[b]X and Y SNPs were excluded from the Kintelligence Kit for likelihood ratio calculations.
STR, short tandem repeat; iiSNP, identity-informative single nucleotide polymorphism.

there were no alleles in common for at least one STR, implying LR = 0 ($\log_{10}$ LR = $-\infty$) and therefore, the relationship is impossible.

For third degree relationships and beyond, the $\log_{10}$ LRs were insufficient for inferring relatedness. While some of the third degree relationships generated moderate statistical support for relatedness, the majority of $\log_{10}$ LRs fell below 1 and some even below 0. The average $\log_{10}$ LR generated for the pairs of true third to eighth degree relatives was 0.21 ± 0.08 and uninformative for determining if the pair were more likely to be related or unrelated. The $\log_{10}$ LRs calculated for the unrelated pairs proposing third to eighth degree relationships were in similar ranges to the truly related pairs at these degrees, with an average $\log_{10}$ LR of -0.06 ± 0.02, and could not be differentiated from the results produced for pairs of true relatives. Of these unrelated pairs testing third to eighth degree relationships, 44% produced $\log_{10}$ LRs above 0.

iiSNP profiles. The $\log_{10}$ LRs calculated with the 94 iiSNPs were less variable by degree than STRs for the true related pairs. However, the unrelated pairs had more variable $\log_{10}$ LRs that were typically higher than the $\log_{10}$ LRs calculated with the corresponding STR profiles. The $\log_{10}$ LRs were informative for true first and second degree relationships, with $\log_{10}$ LRs ranging from 4.6 to 9.4 for parent/offspring pairs, 4.8 to 7.2 for full sibling pairs, and 1.0 to 1.9 for second degree relationships (Fig. 4). One parent/offspring pair (Individual 3 and Individual 5) resulted in a $\log_{10}$ LR of $-\infty$ due to no alleles being in common at the iiSNP rs1493232 (Individual 3 typed CC from 27 reads; Individual 5 typed AA from 58 reads).

Table 4. Shared Centimorgan Thresholds Applied in the GEDmatch PRO™ One-to-Many Kinship Tool Match Lists Based on the Number of Overlapping SNPs

	Shared cM threshold	
Overlapping SNPs	*High confidence match list*	*Expanded match list*
9,000	170	120
8,000	190	140
6,000	200	160

For all threshold levels, the longest identical-by-descent (IBD) segment needs to be at least 30 cM.[17]

The unrelated pair observed as a false positive in the full sibling test with STRs (Individual 1 and Individual 12) also appeared as a false positive with iiSNP profiles, producing a higher $\log_{10}$ LR of 2.3. This provided moderate support for H_1, that these individuals were more likely related as full siblings than unrelated. There was an overlap in the $\log_{10}$ LRs calculated for true and false second degree relationships, with the $\log_{10}$ LRs for unrelated pairs ranging from −2.5 to 1.5.

As observed with the STR profiles, the $\log_{10}$ LRs did not provide statistical support for relatedness for third degree relationships and beyond, with an average $\log_{10}$ LR of 0.04 ± 0.04. This was not distinguishable from the $\log_{10}$ LRs calculated for the unrelated pairs, also producing an average $\log_{10}$ LR of 0.03 ± 0.02 across the third to eighth degree relationships.

Combined identity markers. The combined identity marker $\log_{10}$ LRs were calculated for both STRs (excluding SE33) and iiSNPs. This resulted in an increased statistical power compared with the individual panels and extended the informativeness of the kinship $\log_{10}$ LRs to include second degree relationships. The $\log_{10}$ LRs ranged from 11.0 to 18.5 for true parent/offspring pairs, 7.8 to 15.7 for full siblings, and 1.4 to 6.0 for second degree relationships (Fig. 5). Due to no typed allele in common at rs1493232, as previously discussed in the iiSNP profiles, the $\log_{10}$ LR was $-\infty$ for a parent/offspring relationship (Individual 3 and Individual 5).

The false positive observed in the unrelated STR and iiSNP profile tests for full siblings provided strong statistical support for H_1 (Individual 1 and Individual 12, $\log_{10}$ LR = 4.7), inferring it is 23,410 times more likely to observe the combined STR/iiSNP profiles if the individuals were full siblings than if they were unrelated. This was lower than the smallest $\log_{10}$ LR observed for a true full sibling relationship ($\log_{10}$ LR = 7.8). Of the 10

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

Table 5. Expected Values for Total Shared Centimorgan, Average Identical by Descent (IBD) Segment Length, and Number of IBD Segments Calculated for Different Relationships

Kinship degree	*Relationship*	*Total shared cM*	*Average IBD segment length (cM)*	*Number of IBD segments*
1	Parent / offspring	3,300	61.3	53.8
1	Full sibling	3,300	38.6	85.6
2	Avuncular	1,650	28.1	58.7
2	Grandparent / grandchild	1,650	38.6	42.8
3	First cousin	825	22.1	37.3
4	First cousin once removed	412.5	18.2	22.6
5	Second cousin	206.3	15.5	13.3
6	Second cousin once removed	103.1	13.5	7.6
7	Second cousin twice removed	51.6	11.9	4.3
8	Third cousin once removed	25.8	10.7	2.4
N/A	Unrelated	0	0	0

Values were calculated with the equations in Supplementary Data.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

true third degree relationships tested, half of the calculated $\log_{10}$ LRs provided only limited statistical support for H_1, whereas 2 $\log_{10}$ LRs provided limited statistical support for H_2. The average $\log_{10}$ LRs for testing third to eighth degree relationships was 0.25 ± 0.08 for the true related pairs and -0.03 ± 0.03 for the unrelated pairs.

Extended kinship analysis between kintelligence profiles (Pedigree Group 1)

LR calculations. When calculating $\log_{10}$ LRs using 10,039 autosomal SNPs available in the Kintelligence Kit, there was no overlap in $\log_{10}$ LRs for the related and unrelated pairs (Fig. 6; Table 7). All pairwise comparisons with H_1 proposing a parent/offspring relationship produced $\log_{10}$ LR = $-\infty$. Across the four pairs of true parent/offspring pairs, a total of 58 loci had no alleles in common and 97% were kiSNPs. All of these SNPs were typed homozygous in both profiles and the number of loci without a common allele ranged from 6 (Individual 6 and Individual 9) to 30 (individual 3 and Individual 5). The $\log_{10}$ LRs for the true full sibling relationships exceeded 300 (LR > 10^{300}), resulting in DBLR™ representing the $\log_{10}$ LR as infinite. For unrelated pairs tested with the full sibling hypothesis, the average $\log_{10}$ LR was -58.6 ± 0.5, providing strong statistical support for H_2 (Table 7).

Table 6. Call Rates for the DNA Profiles Generated for Each Sample in Pedigree Group 1; Calculated for Short Tandem Repeats (21), Identity-Informative SNPs (94), Combined Identity Markers (114), and Autosomal Kintelligence SNPs (10,039)

Sample ID	*STRs (%)*	*iiSNPs (%)*	*Combined identity markers (114)*	*Kintelligence (%)*
Individual 1	100.00	100.00	100.00	99.86
Individual 2	100.00	100.00	100.00	99.82
Individual 3	100.00	96.81	98.25	96.96
Individual 4	100.00	95.74	96.49	97.85
Individual 5	85.71	97.87	97.80	96.18
Individual 6	100.00	100.00	100.00	99.76
Individual 7	100.00	98.94	99.12	99.76
Individual 8	100.00	100.00	100.00	99.95
Individual 9	100.00	98.94	99.12	99.74
Individual 10	95.24	92.55	93.86	92.02
Individual 11	100.00	100.00	100.00	99.79
Individual 12	100.00	98.94	99.12	99.41

The smallest $\log_{10}$ LRs produced for second to fifth degree relationships was 10.32 for a fifth degree relationship (Individual 3 and Individual 10). Where H_1 tested second to fifth degree relationships, the $\log_{10}$ LRs produced for true relationships were distinguishable from those produced when testing the unrelated pairs (Table 7). Of the related pairs spanning sixth to eighth degree relationships, 63% produced $\log_{10}$ LRs exceeding 4.0, providing strong statistical support for H_1. These values were highly variable, with some related pairs providing limited probative value: Individual 2 and Individual 11 with a $\log_{10}$ LR of 1.9 (eighth degree relatives), Individual 5 and Individual 10 with a $\log_{10}$ LR of 1.3 (sixth degree relatives), and Individual 5 and Individual 11 with a $\log_{10}$ LR of 0.05 (eighth degree relatives).

The $\log_{10}$ LRs produced for the unrelated pairs did not provide support for H_1 until proposing seventh degree relationships, with 16% of seventh degree and 42% of eighth degree relationship tests exceeding a $\log_{10}$ LR of 0 (Fig. 6). These tests were uninformative, producing an average $\log_{10}$ LR of 0.08 ± 0.02. There was an outlier that was an exception to this (Individual 10 and Individual 12, $\log_{10}$ LR = 1.1), providing limited statistical support where H_1 proposed an eighth degree relationship. The false positive observed for the STR and iiSNP tests (Individual 1 and Individual 12) produced a $\log_{10}$ LR of -60.0 when evaluating the Kintelligence SNP profiles, showing that STRs and iiSNPs on their own are not always reliable for first degree relationship testing.

IBD segment matching. The database search using the One-to-Many Kinship tool detected 81% of the true relationships (n = 47) between the individuals in Pedigree Group 1 in either the high confidence or expanded match lists. The majority of these were fifth degree relationships

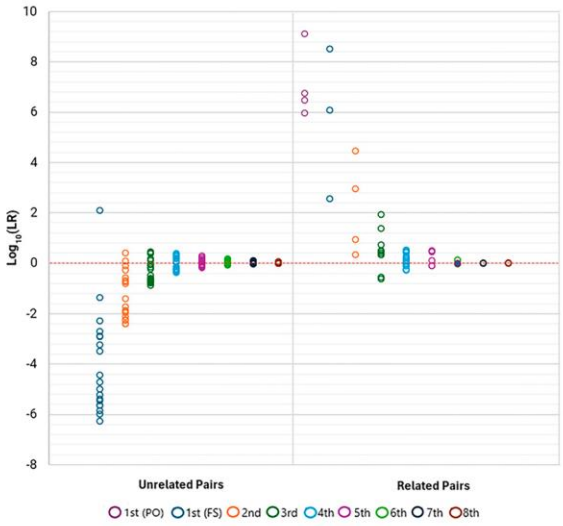


FIG. 3. Log_{10} LR generated using the short tandem repeats (STRs) in the GlobalFiler™ PCR Amplification Kit for each relationship degree for true related pairs ($n = 47$) and all relationship degrees for unrelated pairs ($n = 152$) in Pedigree Group 1. The dotted red line marks uninformative results (log_{10} LR = 0). Unrelated pairs have not been plotted for parent/offspring tests (log_{10} LR = -∞). PO: parent/offspring; FS: full sibling.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

or closer, of which 95% of possible first to fifth degree relationships were detected. A fourth degree relationship (Individual 2 and Individual 8) and a fifth degree relationship (Individual 4 and Individual 9) fell below the database searching thresholds (Table 4). Of the eight relationships greater than the fifth degree, only a sixth degree relationship (Individual 2 and Individual 10) was detected in the database search. The unrelated pairs in Pedigree Group 1 did not appear in the high confidence or expanded match lists as possible genetic relatives. The undetected related pairs were directly compared in the One-to-One Kintelligence DNA Comparison tool and the fourth to seventh degree relationships could be distinguished from the true unrelated pairs. However, the eighth degree relationships ($n = 3$) could not be distinguished from the 19 unrelated pairs tested based on the estimated shared cM values and the number and length of IBD segments detected. Between all Kintelligence profiles compared, there was an average overlap of 9,647 ± 43 SNPs.

The estimated shared cM decreased with increased genetic distance between the pairs, with 89% of the unrelated pairs having an estimated shared cM of 0 (Fig. 7). The two unrelated pairs with a single detected IBD segment were Individual 9 and Individual 11 (32 cM) and Individual 1 and Individual 12 (25 cM). Although full sibling relationships were expected to share approximately 3,300 cM in total, the three full sibling relationships tested produced much lower estimated shared cM values (2,295–2,569 cM). Full sibling and parent/offspring relationships, both first degree relationships, could be differentiated as the parent/offspring estimated shared cM values were closer to the expected shared cM (3,033–3,194 cM). The difference between the observed and expected total shared cM values decreased with increasing genetic distance between the samples. For relationship degrees beyond third order, the expected total shared cM values were within the ranges of the observed estimated shared cM values.

The types of second degree relationships (avuncular and grandparent/grandchild) could not be differentiated from the estimated shared cM value; however, they could be distinguished by examining the number of IBD segments and their average length (Fig. 8, Fig. 9). The avuncular relationships had an average of 28.3 IBD segments with an overall average length of 51.4 cM. This is

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

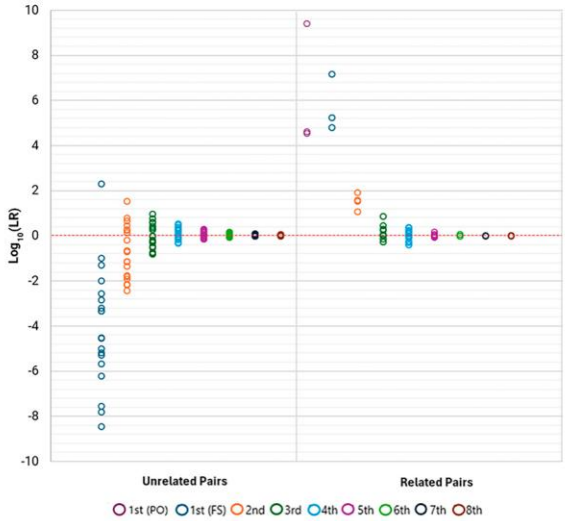


FIG. 4. Log_{10} LR generated using the identity-informative single nucleotide polymorphisms (iiSNPs) in the ForenSeq® Kintelligence Kit for each relationship degree for true related pairs ($n = 47$) and all relationship degrees for unrelated pairs ($n = 152$) in Pedigree Group 1. The dotted red line marks uninformative results (log_{10} LR = 0). Unrelated pairs have not been plotted for parent/offspring tests (log_{10} LR = $-\infty$). PO: parent/offspring; FS: full sibling.

a greater number of IBD segments than detected in the grandparent/grandchild relationship (Individual 10 and Individual 11) with 19 IBD segments detected. However, the IBD segments were longer for the grandparent/grandchild results with an average length of 74.9 cM. While these observed values were smaller and larger than the expected values for average length and number of IBD segments, respectively, they were consistent with regard to grandparent/grandchild relationships having fewer IBD segments of greater length (Table 5).

The average length of the IBD segments detected by the GEDmatch PRO™ algorithms decreased with increased genetic distance (Fig. 8). For the pairs of seventh and eighth degree relatives, and the unrelated pairs where only one IBD segment was detected, the length of that segment was recorded. The observed average length was greater than the expected average length for all relationships, with the exception of an eighth degree relationship (Individual 5 and Individual 11) where no IBD segments were detected. The largest difference between the observed and expected values was for parent/offspring relationships, with the average length for the four tested pairs ranging from 95 to 112 cM, compared with an expected average length of 61 cM.

The number of IBD segments detected between all pairs is shown in Figure 9, where there was low variability observed within the relationship degrees. The IBD segment count decreased with increased genetic distance between the individuals. The observed number of IBD segments was lower than the expected number, but this difference decreased with increased genetic distance. The largest difference between the observed and expected IBD segment count was for the full siblings, where the results ranged from 29 to 38 IBD segments as opposed to the expected 86 IBD segments.

Extended kinship analysis between kintelligence and DTC kits (Pedigree Group 2)

The individuals in Group 2 had up to four DTC kits each that were available for comparison to the Kintelligence profile for Individual A. All DTC kits for the known

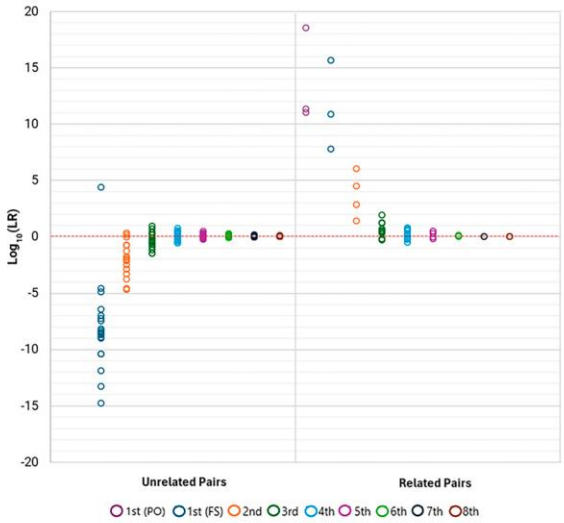


FIG. 5. Log_{10} LR generated using the combined identity markers in the GlobalFiler™ PCR Amplification Kit (short tandem repeats; STRs) and ForenSeq® Kintelligence Kit (identity informative single nucleotide polymorphisms; iiSNPs) for each relationship degree for true related pairs ($n = 47$) and all relationship degrees for unrelated pairs ($n = 152$) in Pedigree Group 1. The dotted red line marks uninformative results (log_{10} LR = 0). Unrelated pairs have not been plotted for parent/offspring tests (log_{10} LR = $-\infty$). PO: parent/offspring; FS: full sibling.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

relatives of Individual A were detected in the database search on GEDmatch PRO™ in the high confidence match list. There was little variability in the results produced between the comparison of the Kintelligence profile and any of the DTC kits. Figure 10 shows the estimated shared cM by individual for each DTC kit and similar results were obtained for the parent/offspring, full sibling, second degree relatives and the unrelated pairs when compared with the Kintelligence results for Pedigree Group 1. The FamilyTreeDNA profiles produced slightly lower estimated shared cM values than Ancestry for 71% of individuals with both kits ($n = 7$); there was no variation observed for DTC kits of Individual A (self; 4 kits) and Individual B (parent/offspring; 4 kits). As observed in the Pedigree Group 1 study, the estimated shared cM values were lower than the expected total shared cM values calculated in Table 4.

There was little difference among DTC kits in the average length and number of IBD segments identified by the GEDmatch PRO™ algorithm (Fig. 11, Fig. 12). As observed with Pedigree Group 1, the average length of the IBD segments were higher and the number of IBD segments were lower than the expected values calculated (Table 4). All four DTC kits for Individual A had an average IBD segment length of 151.9 cM (Fig. 11). The difference in IBD segment lengths across DTC kits for other individuals ranged from 1.5 cM (Individual G) to 8.6 cM (Individual C). The number of IBD segments detected for matches varied by only one segment between the DTC kits for an individual, showing low variability by DTC kit type (Fig. 12). There were three individuals whose DTC kits varied by two IBD segments; the Living DNA profiles for Individual B and Individual C (parent/offspring) and the Ancestry and FamilyTreeDNA profiles for Individual G (full sibling).

There was no variability in the comparisons with the DTC kits for Individual A except for the number of overlapping SNPs; the 23andMe profile had 7,879 SNPs available for comparison with the Kintelligence profile, whereas the other three profiles had a minimum of 9,700 SNPs. Excluding this outlier (the 23andMe kit), the average number of SNPs overlapping between the Individual A Kintelligence profile and the 24 DTC kits was 9687 ± 13 SNPs.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

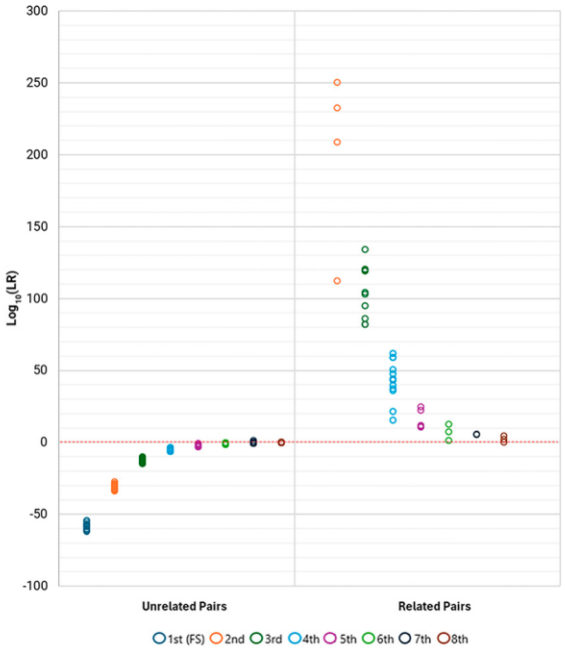


FIG. 6. Log_{10} LR generated using the 10,039 autosomal single nucleotide polymorphisms (SNPs) in the ForenSeq® Kintelligence Kit for each relationship degree for true related pairs ($n = 47$) and all relationship degrees for unrelated pairs ($n = 152$) in Pedigree Group 1. The dotted red line marks uninformative results (Log_{10} LR = 0). Related parent/offspring pairs and unrelated pairs testing parent/offspring relationships have not been plotted (log_{10} LR = $-\infty$). Related full sibling pairs have not been plotted (Log_{10} LR = ∞). FS: full sibling.

Discussion

The Kintelligence Kit provides an in-house solution for forensic laboratories with massively parallel sequencing technology to expand their kinship inference capabilities.

Table 7. Average Log_{10} LR Calculated for Each Relationship Degree for the True Related and Unrelated Pairs of Group 1

Relationship degree	*Related pairs*	*Unrelated pairs*
1st (full siblings)	>300	−58.61 ± 0.47
2nd	112.21–250.26	−31.02 ± 0.40
3rd	104.45 ± 5.49	−12.58 ± 0.29
4th	42.85 ± 3.65	−5.07 ± 0.20
5th	15.88 ± 2.77	−2.12 ± 0.16
6th	1.31–12.32	−0.91 ± 0.10
7th	5.16–5.71	−0.31 ± 0.10
8th	0.05–4.55	−0.12 ± 0.04

The range of log_{10} LRs has been given when there were fewer than five pairs tested. Parent/offspring tests were excluded (log_{10} LR = $-\infty$).

Whereas STR profiles have been historically limited to short-range kinship analysis for first degree relationships, sequencing several thousand SNPs enables medium-range kinship analysis to reliably distinguish between related and unrelated pairs out to the fifth degree (e.g., second cousins) for both IBD segment matching and LR calculations. The medium-density genotypes generated with the Kintelligence Kit can be analyzed using the traditional LR approach, providing statistical support for relatedness in accordance with international recommendations.[43] Alternatively, IBD segment matching can be applied using a windowed kinship algorithm to facilitate FIGG by estimating shared haplotype distributions.[17,21] Furthermore, this analytical method allows profiles derived from forensic samples to be compared with DTC profiles uploaded by members of the public that

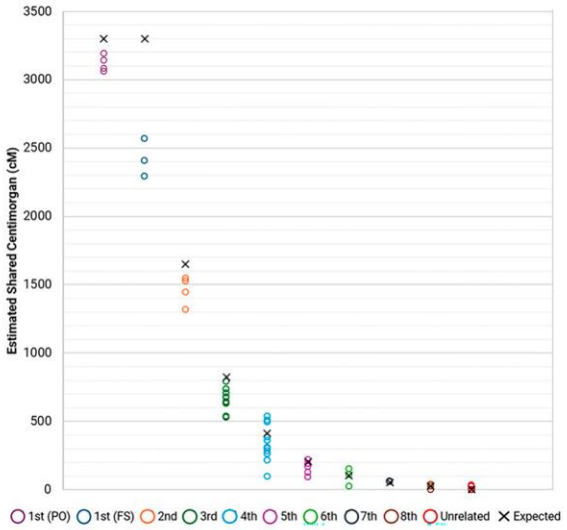


FIG. 7. Estimated shared centimorgan (cM) values between the Kintelligence profiles generated by GEDmatch PRO™ for the related and unrelated pairs in Pedigree Group 1. The expected values were calculated in Table 5. PO: parent/offspring; FS: full sibling.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

have given permission for them to be searched by law enforcement.

The current guidelines for kinship analysis recommend calculating LRs for first degree relatives (i.e., parent, offspring, and full sibling) using routine STR profiling.[1,2] The basis for this approach is that first degree relatives will share more alleles than unrelated individuals.[5,8] Searching law enforcement databases for indirect matches allows testing of the propositions that a pair of DNA donors are parent/offspring or unrelated, or that the pair are full siblings or unrelated.[1,5]

Several studies have recommended prioritizing parent/offspring familial reference samples over full siblings as the direct inheritance of chromosomes ensures at least one allele in common at every locus.[44,45] This recommendation is supported by the STR analysis in this study, which produced strong $\log_{10}$ LRs supporting relatedness for the parent/offspring pairs. However, the false positive results observed when testing a full sibling relationship between unrelated individuals demonstrated that STRs and iiSNPs alone are not always reliable for this relationship. Kinship analysis using the 94 iiSNPs performed similarly to STRs, with the resulting kinship LRs only slightly more powerful with the iiSNPs. This is consistent with a study that assessed the discriminating power of STRs and iiSNPs for direct comparisons of casework-type samples, where the average $\log_{10}$ LR was 23 for STRs and 38 for the 94 iiSNPs.[46]

For both STRs and iiSNPs, there were too few markers to differentiate more distant genetic relatives from unrelated individuals in this study. The unrelated pairs typically produced higher $\log_{10}$ LRs when comparing iiSNP profiles as opposed to STR profiles; this was not unexpected due to the biallelic nature of the 94 iiSNPs in this study, resulting in only three possible genotypes at a locus and increasing the probability of alleles being identical by chance.[47] Pajnič et al. (2023) proposed combining the LRs of STRs and SNPs to improve kinship probabilities for partial profiles obtained in an ancient DNA case study.[47] Similarly, this study showed a substantial increase in the $\log_{10}$ LRs when combining identity markers. The range of kinship analysis could be expanded to include second degree relationships when applying the combined identity marker panel; however, this still produced a false positive result for the full sibling tests between an unrelated pair. This approach can be applied to short-range kinship analysis to test first

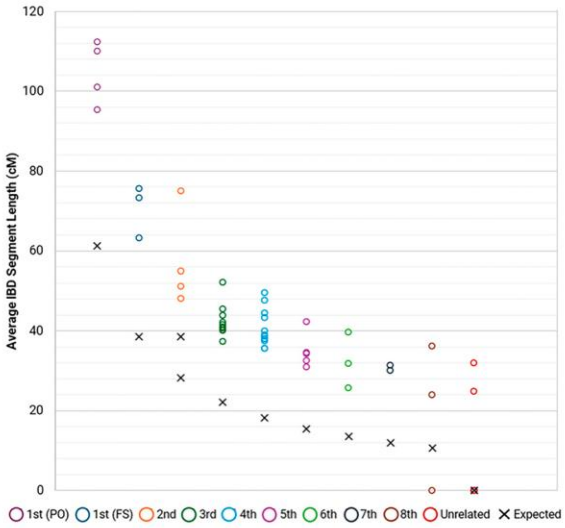


FIG. 8. Average identical by descent (IBD) segment lengths in centimorgans (cM) in the Kintelligence profiles that were detected by GEDmatch PRO™ for the related and unrelated pairs in Pedigree Group 1. The expected values were calculated in Table 5. The two expected values for second degree relationships are different for avuncular (28.1) and grandparent/grandchild (38.6) relationships. PO: parent/offspring; FS: full sibling.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

and second degree relationships and improve the statistical support for relatedness when comparing partial profiles.

When LRs were calculated with 10,039 autosomal SNPs targeted by the Kintelligence Kit, powerful statistical support allowed for true full siblings to fifth degree relatives (e.g., second cousins) to be differentiated from unrelated individuals. Log_{10} LRs could not be calculated for the parent/offspring relationships with the Kintelligence genotypes because at least one locus had no shared alleles between the profiles. This is because DBLR™ applies a binary approach to test parent/offspring relationships, requiring a shared allele at every locus. Genotyping errors, resulting in allele dropout, are expected with a panel of this size, with up to 30 SNPs (0.3% of the profile) showing no shared alleles between true parent/offspring pairs in this study.[48] It may be possible to apply a probabilistic model that accounts for genotyping error when calculating LRs, such as that published by.[48]

The majority of the relationships spanning full sibling to seventh degree (93%) produced log_{10} LRs that exceeded the extremely strong support threshold (log_{10} LR = 6) and 98% exceeding the very strong support threshold (log_{10} LR = 5).[43] These results are concordant with the findings of simulations of medium-density SNP panels to test kinship LR calculations.[3,9,49] Gettings et al. (2024) simulated Kintelligence data for first to fifth degree relationships and produced similar results, calculating median log_{10} LRs of 1,300 for first degree, 300 for second degree, 120 for third degree, 50 for fourth degree, and 20 for fifth degree relationships. Notably, 95% of fifth degree relationship log_{10} LRs exceeded 4, providing strong support for relatedness.[49]

A study simulating 6,600 SNPs for second to sixth degree relatives with direct lineage or one or two common ancestors concluded that while denser marker sets can produce more powerful LRs, they also increase the number of false positive relationships.[3] In this study, the unrelated pairwise comparisons with the Kintelligence profiles did not produce log_{10} LRs that falsely supported a relationship, except for one test hypothesizing a seventh degree relationship. However, this provided only limited support for relatedness and could be differentiated from

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

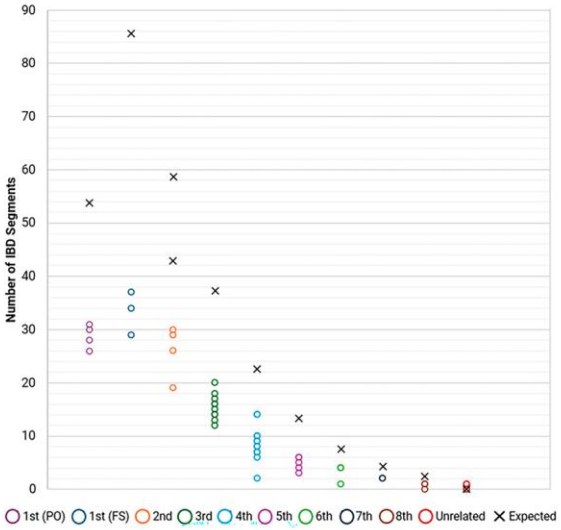


FIG. 9. Number of IBD segments in the Kintelligence profiles detected by GEDmatch PRO™ for the related and unrelated pairs in Pedigree Group 1. The expected values were calculated in Table 5. The two expected values for second degree relationships are different for avuncular (58.7) and grandparent/grandchild (42.8) relationships. PO: parent/offspring; FS: full sibling.

the $\log_{10}$ LRs generated for the true seventh degree relationships. Furthermore, the false positive result supporting a full sibling relationship between an unrelated pair with STRs and iiSNPs was eliminated when using Kintelligence genotypes.

IBD segment matching algorithms have become increasingly common tools for kinship analysis in forensics, particularly with the use of FIGG for long-range kinship testing.[12] The windowed kinship algorithm on GEDmatch PRO™ infers IBD segments by first assessing stretches of SNPs with at least one allele in common, followed by a second pass to identify SNPs within those segments that share both alleles.[21] While the algorithm was originally developed and validated for first to fourth degree relationships, this study found it was also capable of detecting fifth degree relatives in database searches, as well as facilitating direct comparisons to known profiles for sixth and seventh degree relatives. Furthermore, the number and average length of the detected IBD segments can be taken into account to differentiate relationships within the same degree, as demonstrated in this study with second degree relationships (avuncular and grandparent/grandchild). These variations are attributed to the number of meioses separating the queried individuals, with additional meiosis leading to smaller IBD segments being shared between them.[11,42]

When developing the windowed kinship algorithm, it was concluded that the estimated shared cM values produced by this algorithm would be, on average, 7% lower than those produced by the GEDmatch™ algorithm, which was developed for high-density SNP data analysis.[21] In this study, the estimated shared cM values were, on average, 16% lower than the expected values. It was not unexpected that the observed shared cM, average length of IBD segments and number of IBD segments were different to the expected values. This is because the calculations were based on the entire genome of approximately three billion nucleotides, whereas the Kintelligence Kit targets only 10,230 SNPs or approximately 0.00034% of the human genome. While the kiSNPs were selected to target linked SNPs along the chromosomes that would correlate with IBD segments, this is still an estimation, and the available data are more limited compared with microarray or WGS approaches. Microarrays, which are commonly employed by DTC companies for their genetic testing, target anywhere

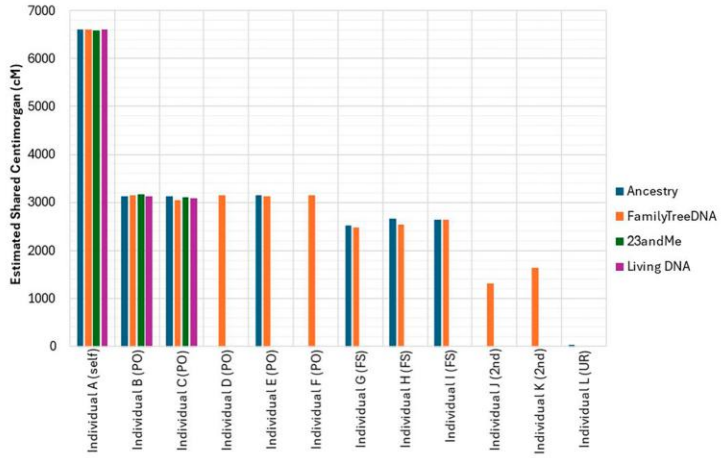


FIG. 10. Estimated shared centimorgan (cM) values generated by GEDmatch PRO™ between the Kintelligence profile for Individual A and the direct-to-consumer kits from Pedigree Group 2. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; UR: unrelated.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

from 500,000 to 2,500,000 nucleotides, while WGS aims to sequence the entire human genome of three billion nucleotides.[12]

The theoretical values calculated with the modified equations by Speed and Balding were lower than those published on the Shared cM Project 4.0 Tool v4 with the exception of full siblings.[42,50,51] The values published on this tool are based on data uploaded to the website by the public as opposed to calculations with sex-averaged recombination numbers per meiosis. The observed shared cM values for the full sibling relationships in this study were closer to those published on the Shared cM Project 4.0 than the expected shared cM values calculated. It is also important to note that there is not a perfect correlation between cM and Mb and so the expected cM values calculated by modifications to the equations of Speed and Balding (in Mb) are only approximations but should be reasonable approximations for average values.[42] The expected number of IBD segments, however, should not depend on the unit chosen for IBD segment length.

GEDmatch PRO™ provides expected cM ranges and the average cM in the Generation Chart as part of the One-to-Many Kinship tool, which are generally lower than the values published on the Shared cM Project.[17,52,53] However, the Generation Chart groups first degree relationships together, giving the expected range of 2,328 to 3,065 cM with an average of 2,697 cM. The Shared cM Project separates parent/offspring (range of 2,376–3,720 cM, average of 3,485 cM) and full sibling (range of 1,613–3,488 cM, average of 2,613 cM) values and, while there is an overlap in the expected ranges for these relationships, the parent/offspring relationships are typically several hundred cM higher.[17,52,53] The estimated shared cM observed in this study for the first degree relationships fell within the broad ranges observed by the Shared cM Project.[52,53] However, some results did not fall within the ranges provided by the Generation Chart on GEDmatch PRO™.[17] This study supports differentiation of the parent/offspring and full sibling expected shared cM in the Generation Chart for classification of first degree relationships.

The results obtained when Kintelligence profiles were compared with either other Kintelligence profiles (Pedigree Group 1) or DTC kits (Pedigree Group 2) were consistent, with no observable impact from the company providing the DTC testing service. Other studies assessing Kintelligence and microarray results have observed minimal variation in the estimated shared cM across different sequencing methods, confirming the Kintelligence Kit is reliable for detecting and classifying first to fourth degree relationships.[52,53] This lack of variation is due to the Kintelligence Kit being specifically designed to target SNPs common to various DTC kits and those used by

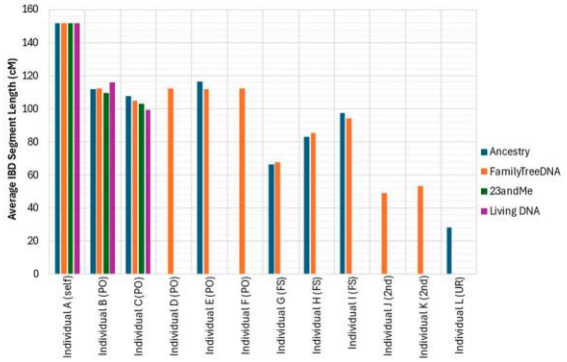


FIG. 11. Average identical by descent (IBD) segment lengths in centimorgans (cM) detected by GEDmatch PRO™ between the Kintelligence profile for Individual A and the direct-to-consumer kits from Pedigree Group 2. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; UR: unrelated.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

the existing GEDmatch™ algorithm for high-density SNP data.[19,21] This allows for public databases and profiles uploaded by the public to be leveraged for extended kinship analysis, eliminating the need for the creation of a new law enforcement database containing reference SNP profiles.

Kling & Tillmar (2019) observed that the IBD segment matching approach for kinship analysis was more effective at excluding unrelated individuals, while LR calculations were able to classify related individuals more accurately in simulations using high-density SNP profiles.[9] However, in this study, the expanded match list

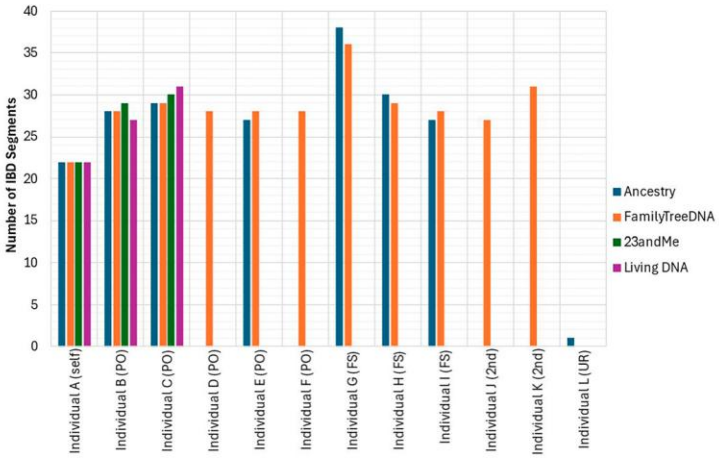


FIG. 12. Number of identical-by-descent (IBD) segments detected by GEDmatch PRO™ between the Kintelligence profile for Individual A and the direct-to-consumer kits from Pedigree Group 2. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; UR: unrelated.

generated following a database search on GEDmatch PRO™ often included individuals who were not known genetic relatives of the volunteers. Previous studies have concluded that, while absolute determination of a relationship is not possible, it is crucial to assess which relationship is most probable.[54] While the LR approach requires testing multiple sets of hypotheses to infer the most likely relationship between individuals, the degree of relatedness can be inferred from the total amount of DNA shared, the average shared IBD segment length, and the number of shared IBD segments using the IBD segment matching approach.

The Kintelligence Kit has been successfully used to resolve both criminal and unidentified human remains cases around the world, relying on genetic genealogy databases and IBD segment matching algorithms. However, FIGG requires confirmatory testing of the proposed identification using either a direct match with a PoI or partial matches and kinship LR calculations with STR profiles from first degree relatives of the PoI.[55,56] This study demonstrates that applying the LR approach to Kintelligence profiles is a suitable alternative when these preferred family reference samples are unavailable, allowing for confirmation of identity using more distant relatives with statistical support. Ideally, it would be preferable to provide support for genetic relationships on more than one lineage of a pedigree. However, forensic casework samples are often of poor quality and further evaluation is required to determine how extended kinship analysis should be approached with partial profiles. Part two of this study, published separately, assesses how genotyping error and information loss impact the ability to calculate LRs and detect IBD segments.[59]

Conclusions

This study demonstrates that the Kintelligence Kit can extend kinship inference beyond short-range familial searches using STR profiles. While identity-informative markers such as STRs and iiSNPs are suitable for first degree relationships, the Kintelligence Kit can be applied for kinship analysis of first to fifth degree relationships using kinship LR calculations and IBD segment matching. The SNP genotypes can be analyzed with kinship LRs to provide statistical support for hypotheses of relatedness or uploaded to law enforcement accessible genetic genealogy databases for analysis using the windowed kinship algorithm to infer IBD segments. Part two of this study builds on the assessment of the Kintelligence Kit and various kinship analysis methods, exploring the impact of suboptimal Kintelligence profiles, locus dropout, and allele dropout on kinship analysis.

Acknowledgments

The authors would like to thank the volunteers and their families for providing their DNA and associated metadata for this study; and Dr. Maarten Kruijver, Bjorn Sutherland, and Dr. Kevin Cheng from the Institute of Environmental Science and Research Limited for their assistance with using DBLR™.

Authors' Contributions

J.L.W.: Conceptualization (equal); methodology (lead); formal analysis (lead); writing—original draft (lead); writing—review and editing (equal). K.G.: Methodology (supporting); formal analysis (supporting); writing—review and editing (equal). L.C.: Methodology (supporting); writing—review and editing (equal). D.M.: Conceptualization (equal); methodology (supporting); formal analysis (supporting); writing—review and editing (equal); supervision (equal). J.W.: Conceptualization (equal); methodology (supporting); writing—review and editing (equal); supervision (equal); funding acquisition (lead).

Institutional Review Board Statement

This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of University of Technology Sydney (UTS) Human Research Ethics Committee (HREC) (UTS HREC NO. ETH21-5821).

Informed Consent Statement

Written informed consent was obtained from all volunteer sample donors involved in this study.

Data Availability Statement

Data are stored at the Australian Federal Police and may be made available to approved entities upon written request and subject to consent provisions.

Author Disclosure Statement

The authors declare no conflict of interest.

Funding Information

This research was funded by the Australian Federal Police (AFP) Innovation Fund and the AFP National DNA Program for Unidentified and Missing Persons. J.L.W. is supported by an Australian Government Research Training Program Scholarship.

Supplementary Material

Supplementary Data

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

References

1. National Institute of Justice. Familial DNA Searching: Current Approaches. National Institute of Justice; 2015. Available from: https://www.ojp.gov/library/publications/familial-dna-searching-current-approaches [Last accessed: November 18, 2024].
2. Scientific Working Group on DNA Analysis Methods. Recommendations from the SWGDAM Ad Hoc Working Group on Familial Searching. SWGDAM Communications; 2013. Available from: https://www.swgdam.org/publications [Last accessed: January 2, 2025].
3. Kling D. On the use of dense sets of SNP markers and their potential in relationship inference. Forensic Sci Int Genet 2019;39:19–31; doi: 10.1016/j.fsigen.2018.11.022
4. Maguire CN, McCallum LA, Storey C, et al. Familial searching: A specialist forensic DNA profiling service utilising the National DNA Database® to identify unknown offenders via their relatives—The UK experience. Forensic Sci Int Genet 2014;8(1):1–9; doi: 10.1016/j.fsigen.2013.07.004
5. Ge J, Chakraborty R, Eisenberg A, et al. Comparisons of familial DNA database searching strategies. J Forensic Sci 2011;56(6):1448–1456; doi: 10.1111/j.1556-4029.2011.01867.x
6. Schwab AP, Luu HS, Wang J, et al. Genomic privacy. Clin Chem 2018; 64(12):1696–1703; doi: 10.1373/clinchem.2018.289512
7. Kennett D. Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes. Forensic Sci Int 2019; 301:107–117; doi: 10.1016/j.forsciint.2019.05.016
8. Ge J, Budowle B, Chakraborty R. Choosing relatives for DNA identification of missing persons. J Forensic Sci 2011;56 Suppl 1:S23–S28; doi: 10.1111/j.1556-4029.2010.01631.x
9. Kling D, Tillmar A. Forensic genealogy—A comparison of methods to infer distant relationships based on dense SNP data. Forensic Sci Int Genet 2019;42:113–124; doi: 10.1016/j.fsigen.2019.06.019
10. Ge J, Budowle B. Forensic investigation approaches of searching relatives in DNA databases. J Forensic Sci 2021;66(2):430–443; doi: 10.1111/1556-4029.14615
11. Morimoto C, Manabe S, Fujimoto S, et al. Discrimination of relationships with the same degree of kinship using chromosomal sharing patterns estimated from high-density SNPs. Forensic Sci Int Genet 2018;33:10–16; doi: 10.1016/j.fsigen.2017.11.010
12. Kling D, Phillips C, Kennett D, et al. Investigative genetic genealogy: Current methods, knowledge and practice. Forensic Sci Int Genet 2021;52: 102474; doi: 10.1016/j.fsigen.2021.102474
13. Glynn CL. Bridging disciplines to form a new one: The emergence of forensic genetic genealogy. Genes (Basel) 2022;13(8):1381; doi: 10.3390/genes13081381
14. GEDmatch. Find DNA Matches for Free | Analyze Your DNA. GEDmatch; 2024. Available from: https://www.gedmatch.com/ [Last accessed: November 25, 2024].
15. FamilyTreeDNA. DNA Testing for Ancestry & Genealogy. FamilyTreeDNA; 2024. Available from: https://www.familytreedna.com/ [Last accessed: November 25, 2024].
16. GEDmatch. GEDmatch.com Terms of Service and Privacy Policy. 2024. Available from: https://www.gedmatch.com/terms-of-service-privacy-policy [Last accessed: November 25, 2024].
17. GEDmatch PRO. GEDmatch PRO. 2024. Available from: https://pro.gedmatch.com/ [Last accessed: November 25, 2024].
18. FamilyTreeDNA. FamilyTreeDNA Law Enforcement Guide; 2024. Available from: https://www.familytreedna.com/legal/law-enforcement-guide [Last accessed: November 25, 2024].
19. Antunes J, Walichiewicz P, Forouzmand E, et al. Developmental validation of the ForenSeq® kintelligence kit, MiSeq Fgx® sequencing system and ForenSeq universal analysis software. Forensic Sci Int Genet 2024; 71:103055; doi: 10.1016/j.fsigen.2024.103055
20. QIAGEN. Verogen ForenSeq Kintelligence Kit. 2024. Available from: https://www.qiagen.com/us/products/human-id-and-forensics/nextgeneration-sequencing/verogen-forenseq-kintelligence-kit [Last accessed: May 24, 2024].
21. Snedecor J, Fennell T, Stadick S, et al. Fast and accurate kinship estimation using sparse SNPs in relatively large database searches. Forensic Sci Int Genet 2022;61:102769; doi: 10.1016/j.fsigen.2022.102769
22. Arenson J. DNA Labs International is the First Accredited Lab to Use the ForenSeq® Kintelligence System to Aid in the Identification of Unidentified Remains. PRWeb: Deerfield Beach, FLA; 2022. Available from: https://biotechnetworks.org/2022/02/23/dna-labs-international-is-the-first-accredited-lab-to-use-the-forenseq-kintelligence-system-to-aid-in-the-identification-of-unidentified-remains/ [Last accessed: January 10, 2025].
23. Signature Science. SigSci's Center for Advanced Genomics Validating the Verogen ForenSeq® Kintelligence Workflow for Forensic Genetic Genealogy. Austin, TX; 2021. Available from: https://www.signaturescience.com/press/ngs_fgg/ [Last accessed: March 31, 2022].
24. Peck MA, Koeppel AF, Gorden EM, et al. Internal validation of the ForenSeq kintelligence kit for application to forensic genetic genealogy. Forensic Genom 2022;2(4):103–114; doi: 10.1089/forensic.2022.0014
25. Watson JL, McNevin D, Grisedale K, et al. Operationalisation of the ForenSeq® kintelligence kit for Australian unidentified and missing persons casework. Forensic Sci Int Genet 2024;68:102972; doi: 10.1016/j.fsigen.2023.102972
26. Jordan S. CHI becomes Nation's First Public Lab to earn Accreditation to Perform Forensic Genetic Genealogy. 2023. Available from: https://www.unthsc.edu/newsroom/story/chi-becomes-nations-first-public-lab-to-earn-accreditation-to-perform-forensic-genetic-genealogy/ [Last accessed: February 5, 2024].
27. Department of Justice. Forensic Genetic Genealogical DNA Analysis and Searching. Interim Policy; 2019. Available from: https://www.justice.gov/olp/page/file/1204386/dl [Last accessed: November 20, 2024].
28. QIAGEN. QIAamp® DNA Investigator Handbook; 2020. Available from: https://www.qiagen.com/us/products/human-id-and-forensics/investigator-solutions/qiaamp-dna-investigator-kit [Last accessed: January 5, 2024].
29. Thermo Fisher Scientific. Quantifiler™ HP and Trio DNA Quantification Kits User Guide. 2017. Available from: https://www.thermofisher.com/order/catalog/product/4482910 [Last accessed: January 5, 2024].
30. Thermo Fisher Scientific. QuantStudio™ 5 Real-Time PCR Instrument for Human Identification User Guide. Life Technologies; 2017. Available from: https://www.thermofisher.com/order/catalog/product/A34322 [Last accessed: January 5, 2024].
31. Thermo Fisher Scientific. GlobalFiler PCR Amplification Kits. 2019. Available from: https://www.thermofisher.com/order/catalog/product/4476135 [Last accessed: January 5, 2024].
32. Thermo Fisher Scientific. Applied Biosystems 3500/3500xL Genetic Analyzer User Guide. 2010. Available from: https://www.thermofisher.com/order/catalog/product/4406016 [Last accessed: January 5, 2024].
33. Thermo Fisher Scientific. GeneMapper™ ID-X Software v1.6. 2024. Available from: https://www.thermofisher.com/order/catalog/product/A39975 [Last accessed: January 5, 2024].
34. Verogen. ForenSeq Kintelligence Kit Reference Guide. 2021. Available from: https://verogen.com/products/forenseq-kintelligence-kit/ [Last accessed: February 19, 2024].
35. Promega. Quantus™ Fluorometer. 2022. Available from: https://www.promega.com.au/products/microplate-readers-fluorometers-luminometers/fluorometers/quantus-fluorometer/?catNum=E6150 [Last accessed: March 29, 2024].
36. Promega. QuantiFluor® ONE dsDNA System. 2022. Available from: https://www.promega.com.au/products/rna-analysis/dna-and-rna-quantitation/quantifluor-dsdna-system/?catNum=E2670 [Last accessed: March 29, 2024].
37. Verogen. MiSeq FGx Sequencing System Reference Guide. 2021. Available from: https://verogen.com/products/miseq-fgx-sequencing-system/ [Last accessed: February 19, 2024].
38. QIAGEN. Verogen MiSeq FGx Sequencing Reagents. 2024. Available from: https://www.qiagen.com/us/products/human-id-and-forensics/nextgeneration-sequencing/verogen-miseq-fgx-sequencing-reagents [Last accessed: August 15, 2024].
39. Verogen. Universal Analysis Software v2.0 and above Reference Guide. 2021. Available from: https://verogen.com/products/universal-analysis-software/ [Last accessed: February 19, 2024].
40. STRmix™. DBLR™. 2023. Available from: strmix.com/dblr / [Last accessed: January 29, 2024].
41. University R. Map Interpolator of the Rutgers Map. 2014. Available from: http://compgen.rutgers.edu/map_interpolator.shtml [Last accessed: October 20, 2024].
42. Speed D, Balding DJ. Relatedness in the post-genomic era: Is it still useful? Nat Rev Genet 2015;16(1):33–44; doi: 10.1038/nrg3821
43. Scientific Working Group on DNA Analysis Methods. Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Resulted Reported as Likelihood Ratios. 2016. Available from: https://www.swgdam.org/publications [Last accessed: January 2, 2025].
44. Zvénigorosky V, Sabbagh A, Gonzalez A, et al. The limitations of kinship determinations using STR data in ill-defined populations. Int J Legal Med 2020;134(6):1981–1990; doi: 10.1007/s00414-020-02298-w

45. Ge J, Budowle B. How many familial relationship testing results could be wrong? PLoS Genet 2020;16(8):e1008929; doi: 10.1371/journal.pgen.1008929
46. Watson JL, Grisedale K, Cho K, et al. Characterisation of identity-informative genetic markers in the Australian population with European Ancestry. Forensic Sci Int Genet 2025;74:103169; doi: 10.1016/j.fsigen.2024.103169
47. Pajnič IZ, Leskovar T, Črešnar M. Improving kinship probability in analysis of ancient skeletons using identity SNPs and MPS technology. Int J Legal Med 2023;137(4):1007–1015; doi: 10.1007/s00414-023-03003-3
48. Andersen MM, Kampmann M-L, Jepsen AH, et al. Shotgun DNA sequencing for human identification: Dynamic SNP selection and likelihood ratio calculations accounting for errors. Forensic Sci Int Genet 2025;74:103146; doi: 10.1016/j.fsigen.2024.103146
49. Gettings KB, Tillmar A, Sturk-Andreaggi K, et al. Review of SNP assays for disaster victim identification: Cost, time, and performance information for decision-makers. J Forensic Sci 2024;69(5):1546–1557; doi: 10.1111/1556-4029.15585
50. Bettinger B. The shared cM project: A demonstration of the power of citizen science. J Genet Geneal 2016;8(1):38–42. https://jogg.info/wp-content/uploads/2021/09/81.006.pdf
51. Bettinger BT. The Shared cM Project 4.0 tool v4. 2020. Available from: https://dnapainter.com/tools/sharedcmv4 [Last accessed: November 25, 2024].
52. Hymus CM, Tay JW, Coghlan ML, et al. Tree vs tree–comparing kinship prediction outcomes for three generations of a family using the Verogen Kintelligence and Infinium Global Screening Array-24 Kit. Aust J Forensic Sci 2024;56(Suppl 1):119–122; doi: 10.1080/00450618.2024.2324750
53. Daniel R, Raymond J, Sears A, et al. It's all relative: A multi-generational study using ForenSeq™ Kintelligence. Forensic Sci Int 2024;364:112208; doi: 10.1016/j.forsciint.2024.112208
54. Weir BS, Anderson AD, Hepler AB. Genetic relatedness analysis: Modern data and new challenges. Nat Rev Genet 2006;7(10):771–780; doi: 10.1038/nrg1960
55. Greytak EM, Moore C, Armentrout SL. Genetic genealogy for cold case and active investigations. Forensic Sci Int 2019;299:103–113; doi: 10.1016/j.forsciint.2019.03.039
56. Scientific Working Group on DNA Analysis Methods. Overview of Investigative Genetic Genealogy. SWGDAM Communications; 2020. Available from: https://www.swgdam.org/publications [Last accessed: January 2, 2025].
57. Taylor D, Bright J-A, McGovern C, et al. Allele frequency database for GlobalFiler™ STR loci in Australian and New Zealand populations. Forensic Sci Int Genet 2017;28:e38–e40; doi: 10.1016/j.fsigen.2017.02.012
58. The International Genome Sample Resource. 1000 Genomes Project. 2021. Available from: https://www.internationalgenome.org/ [Last accessed: October 17, 2023].
59. Watson JL, Grisedale K, Coakley L, et al. Extended kinship inference. Part 2: Evaluation of the impact of information loss on likelihood ratios and haplotype matching. Forensic Genom 2025;5(1):32–49; doi: 10.1089/forensic.2025.0002

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

7.1 Supplementary Material

Supplementary Information S1: Calculations for expected shared centimorgan (cM), average identical by descent (IBD) segment length and number of IBD segments. Equations are modified from those published by Speed & Balding (2015) for total length of DNA in cM.[1]

The average IBD segment length that two individuals share is:

$$\text{Average IBD Segment Length} = \frac{3300}{22 + (40.7 + 22.9) \times \frac{G_1 + G_2}{2}} \quad (1)$$

where the numerator is the total length of DNA in cM; A is the number of most recent common ancestors shared between the individuals; G_1 and G_2 are the number of generations between each individual and the most recent common ancestor(s); 22 is the number of autosomal chromosomes; and $(40.7 + 22.9)/2$ is the sex-averaged number of recombinations per meiosis.

The total cM that two individuals share is:

$$\text{Expected Shared cM} = A \times \frac{1}{2^{G_1+G_2-1}} \times 3300 \quad (2)$$

The number of IBD segments that two individuals share is:

$$\text{Number of IBD Segments} = A \times \frac{22 + (40.7 + 22.9) \times \frac{G_1 + G_2}{2}}{2^{G_1+G_2-1}} \quad (3)$$

Reference

1. Speed, D., Balding, D. Relatedness in the post-genomic era: is it still useful? Nat Rev Genet 2015;16:33–44; doi: 10.1038/nrg3821.

8. CHAPTER EIGHT

EXTENDED KINSHIP INFERENCE. PART 2: EVALUATION OF THE IMPACT OF INFORMATION LOSS ON LIKELIHOOD RATIOS AND HAPLOTYPE MATCHING

Contributions of authors:

Watson JL, Grisedale K, Coakley L, McNevin D and Ward J. (2025) 'Extended Kinship Inference. Part 2: Evaluation of the Impact of Information Loss on Likelihood Ratios and Haplotype Matching', Forensic Genomics, 5(1), DOI: 10.1089/forensics.2025.0002.

Jessica Watson (Candidate)

Conceptualised and assisted in the experimental design for the study. Carried out sample collection, sample preparation, DNA testing and data analysis. Prepared manuscript, edited manuscript following review by other co-authors, referenced manuscript, carried out manuscript's submission and incorporated feedback from peer review process.

Signed: Production Note: Signature removed prior to publication.

Kelly Grisedale (AFP DNA Program staff)

Assisted in the experimental design for the study. Contributed to DNA testing and data analysis. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Louise Coakley (AFP DNA Program staff)

Assisted in the experimental design for the study. Contributed to data analysis. Contributed to manuscript review and editing.

Signed: Production Note: Signature removed prior to publication.

Dennis McNevin (Co-supervisor)

Conceptualised and assisted in the experimental design for the study. Contributed to data analysis. Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

Jodie Ward (Co-supervisor)

Conceptualised and assisted in the experimental design for the study. Acquired funding for the study. Contributed to manuscript review and editing.

Signed: Production Note:
Signature removed prior to publication.

Forensic Genomics
Volume 5, Number 1, 2025
© Mary Ann Liebert, Inc.
DOI: 10.1089/forensic.2025.0002

Forensic Genomics

Open camera or QR reader and scan code to access this article and other resources online.

ORIGINAL ARTICLE

Extended Kinship Inference Part 2: Evaluation of the Impact of Information Loss on Likelihood Ratios and Haplotype Matching

Jessica L. Watson, BForSc (Hons),[1–3,*] Kelly Grisedale, PhD,[2,3] Louise Coakley, MBA,[3,4] Dennis McNevin, PhD,[1,3] and Jodie Ward, PhD[1,3]

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

Abstract

Medium-density single nucleotide polymorphism (SNP) profiles enable law enforcement to infer close and distant genetic relationships. Part one of this study demonstrated that the ForenSeq® Kintelligence Kit, which targets 10,230 SNPs, can facilitate extended kinship inference through kinship likelihood ratio (LR) and identical by descent (IBD) segment matching methods. However, if SNPs are not detected, or are incorrectly called, the ability to detect genetic relatives and accurately classify the relationship may be compromised. The Kintelligence profiles for the central individuals of two pedigrees described in Part One were edited to simulate information loss through locus and allele dropout. LRs were calculated with DBLR™ and SNP profiles were uploaded to GEDmatch PRO™ for database searching or direct comparison. The LRs decreased with increasing information loss but still provided strong statistical support for relatedness. LRs exceeded 100,000 for all full sibling to fifth degree relationships for up to 30% locus and allele dropout. Locus dropout did not significantly impact the ability to infer first to fifth degree relationships with IBD segment matching. Allele dropout had a greater impact, with 30% allele dropout impairing the ability to classify relationships to their correct degree. When allele dropout was greater than 10%, the fifth degree relative was no longer detected in the database search. This study highlights the robustness of LR calculations and the GEDmatch PRO™ IBD segment matching algorithms and the suitability of the Kintelligence Kit for medium-range kinship inference, with the algorithm maintaining the ability to infer relationships despite increasing information loss.

Keywords: single nucleotide polymorphism, human identification, targeted amplicon sequencing, forensic investigative genetic genealogy, kinship

Introduction

Extended kinship inference can be applied to forensic casework using single nucleotide polymorphism (SNP) profiles for the identification of persons of interest in criminal investigations and unidentified human remains in missing persons investigations.[1–3] However, forensic samples collected during an investigation may be degraded, inhibited or have low quantities of DNA. This

[1]Faculty of Science, Centre for Forensic Science, School of Mathematical and Physical Sciences, University of Technology Sydney, New South Wales, Australia.
[2]Forensics Command, Australian Federal Police, Australian Capital Territory, Australia.
[3]National DNA Program for Unidentified and Missing Persons, Australian Federal Police, Australian Capital Territory, Australia.
[4]Genetic Genealogy Pty Ltd, Queensland, Australia.

*Address correspondence to: Jessica L. Watson, BForSc (Hons), Faculty of Science, Centre for Forensic Science, School of Mathematical and Physical Sciences, University of Technology Sydney, 15 Broadway, Ultimo, New South Wales, 2007, Australia, Email: Jessica.Watson@student.uts.edu.au

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

can result in suboptimal DNA profiles being produced, limiting the ability to find direct or indirect matches (for short-range kinship testing) in a database or generate genetic intelligence, such as inferred biogeographical ancestry, externally visible characteristics, or distant genetic relatives of the DNA donor.[4–7]

There are two common methods used to assess genetic relationships between individuals: kinship likelihood ratio (LR) calculations and identical by descent (IBD) segment matching. LRs provide statistical support for alternative propositions about the genetic relationship between two individuals based on the probabilities of observing alleles identical by state (IBS) in the DNA profiles.[2] IBD segment matching algorithms compare medium- and high-density SNP profiles to infer haplotypes or IBD segments that are shared between genetic relatives and co-inherited from a common ancestor.[1,8]

The ForenSeq® Kintelligence Kit (Verogen Inc., QIAGEN) targets 10,230 SNPs across the human genome in a medium-density targeted amplicon sequencing panel.[6,9] The size of this panel facilitates kinship analysis using the LR calculation method or IBD segment matching using a database search or direct comparison.[8,10,11] Kintelligence profiles are compatible with upload to the law enforcement accessible genetic genealogy databases GEDmatch PRO™ and FamilyTreeDNA, allowing for comparison with over two million profiles uploaded by consenting members of the public on each database.[12,13] Part one of this study demonstrated Kintelligence profiles yielded LRs that provided very strong statistical support ($\log_{10}$ LR > 5) for relatedness between full siblings and fifth degree relatives.[14] The IBD segment matching tools on GEDmatch PRO™ could also detect all known first to fifth degree relatives available in the database.

Some studies have evaluated how locus dropout impacts kinship analysis for different SNP panels; however, further work is required to assess suboptimal profiles yielded from compromised forensic samples.[2,8,11] Partial profiles could appear as locus dropout, where loci are not detected during sequencing or fall below the analytical thresholds, or allele dropout, where a heterozygote is incorrectly called as homozygote. This paper presents part two of the study, which focuses on evaluating the Kintelligence profiles for central individuals of two pedigrees with simulated locus and allele dropout. This study aims to guide the interpretation of extended kinship inference results using LR calculations and IBD segment matching for partial profiles that could result from compromised forensic samples.

Methods

Ethics approval for this research was granted by the University of Technology (UTS) Human Research Ethics Committee (HREC) (UTS HREC NO. ETH21-5821 and amendment ETH21-6606). The methods for sample procurement, DNA extraction, quantification, library preparation, sequencing, LR calculation, and IBD segment matching are described in part one of this study.[14]

For the calculation of LRs for unrelated pairs of individuals (19 pairs), the first hypothesis (H_1) proposed a relationship ranging from first (parent/offspring or full sibling) to eighth degree relationships and the alternative hypothesis (H_2) proposed that the individuals were unrelated. For the genetically related pairs of individuals, H_1 proposed the true relationship of the individuals and H_2 proposed that they were unrelated. The statistical support for the hypotheses is presented using the verbal scale based on the $\log_{10}$ LR results (Table 1). Pedigree Group 1 (Individuals 1–12) provided buccal swabs for DNA profiling with the Kintelligence Kit (Fig. 1). Pedigree Group 2 (Individuals A–L) provided the kit numbers for pre-existing direct-to-consumer (DTC) data uploaded to GEDmatch™ and the central individual (Individual A) provided a buccal swab for DNA profiling with the Kintelligence Kit (Fig. 2; Table 2).[a] All Kintelligence profiles were uploaded to GEDmatch PRO™ as laboratory validation samples.

For both pedigrees, the Kintelligence profiles of the central individual (Individual 1 from Pedigree Group 1; Individual A from Pedigree Group 2) were manually edited using the Universal Analysis Software (Verogen, Inc.) to simulate varying degrees of locus and allele dropout. Table 3 outlines the call rates and autosomal heterozygosity for the generated Kintelligence profiles. Call rates were calculated as the number of typed SNPs out of a possible 10,230 SNPs for Individual 1 (male) and 10,145 SNPs for Individual A (female). Autosomal heterozygosity was calculated as the proportion of typed autosomal SNPs (up to 10,039 SNPs) that were heterozygous.

Locus dropout was assessed in increments of 5% with a minimum call rate of 70% generated for the profiles. SNPs were removed based on the frequency of locus dropout as observed by Watson et al. (2023).[10] After this list of SNPs was exhausted, SNPs were removed by increasing the total coverage threshold required to call a SNP. Allele dropout was calculated as the proportion of heterozygous autosomal SNPs that were called as heterozygous and changed to homozygous, across the testing range of 5% to 30% allele dropout. The minor allele of each heterozygous SNP was removed from the final genotype.

[a]*The use of DTC data for research was in compliance with the terms and conditions at the time the respective data were downloaded by the volunteers. All volunteers were notified at the conclusion of the study so they could opt out of law enforcement searching or remove their DNA data from GEDmatch™ to ensure that future use of their data is in compliance with their preferences and relevant terms and conditions.*

Table 1. Verbal scale for $\log_{10}$ LR results and the statistical support for each hypothesis[15]

$\log_{10}$ LR	*Interpretation*
≤−5	Very strong statistical support for H_2
−4	Strong statistical support for H_2
−3	Moderately strong statistical support for H_2
−2	Moderate statistical support for H_2
−1	Limited statistical support for H_2
0	Uninformative
1	Limited statistical support for H_1
2	Moderate statistical support for H_1
3	Moderately strong statistical support for H_1
4	Strong statistical support for H_1
≥5	Very strong statistical support for H_1

Results

Impact of locus dropout

LR calculations between Kintelligence profiles (Pedigree Group 1). Increasing locus dropout in the Kintelligence profile of Individual 1 resulted in the $\log_{10}$ LRs decreasing for all relationships greater than the first degree tested within Pedigree Group 1 (Fig. 3). The results for the first degree relationships were unaffected by locus dropout. As noted in part one, when a relationship is impossible (LR = 0), the $\log_{10}$ LR results are given as negative infinity ($-\infty$).[15] Alternatively, DBLR™ does not report $\log_{10}$ LRs that exceed 300 (LR > 10^{300}) and return the $\log_{10}$ LR as infinite (∞).

The parent/offspring relationship (Individual 2) produced a $\log_{10}$ LR of $-\infty$, inferring the relationship was impossible, for all levels of dropout due to the profiles failing to have at least one allele in common at every SNP. The full sibling relationship (Individual 3) was not observably impacted by locus dropout, with the LRs exceeding 10^{300} and remaining incomputable by DBLR™, returning a $\log_{10}$ LR of ∞ and inferring H_2 was impossible. The greatest decrease was observed in the second degree relationships (Individual 4 and Individual 5). For locus dropout ranging from 0% to 30%, the $\log_{10}$ LR decreased from 232 to 160 for Individual 4 and from 250 to 176 for Individual 5. This represented an average decrease of 12 in the $\log_{10}$ LR for every 5% of SNPs lost. However, these values all provided very strong statistical support for H_1 and the decreases in $\log_{10}$ LRs with locus dropout do not have any practical significance.

For the third degree relationships (Individual 6, Individual 7, and Individual 8), the $\log_{10}$ LR decreased steadily by approximately 7.2 for every 5% interval of locus dropout tested (Fig. 3). The exception was Individual 8, which saw the $\log_{10}$ LR increase from 86 to 88 when the call rate of Individual 1 was first reduced to 95%.

For the fourth and fifth degree relationships (Individual 9 and Individual 10, respectively), a small decrease

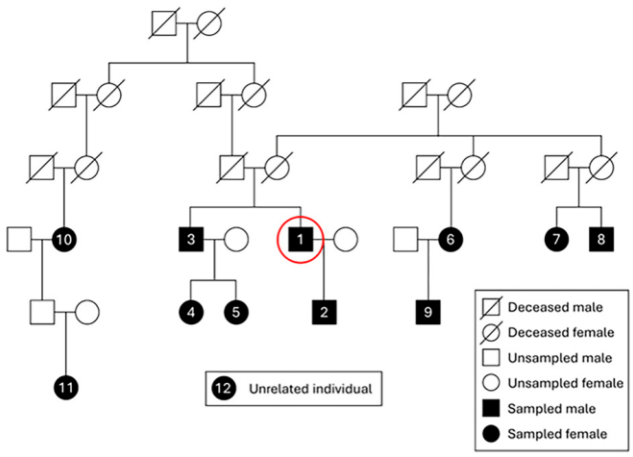


FIG. 1. Pedigree Group 1, with numbers corresponding to Individuals 1 through 12.[14] Individuals included in the study are filled in (black), with living relatives linking the family members not filled (white). Deceased relatives are crossed out. Individual 12 is unrelated to all individuals in the group. The central individual (Individual 1) is circled in red.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

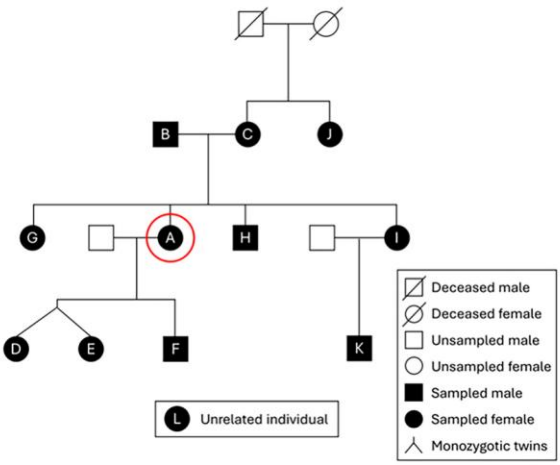


FIG. 2. Pedigree Group 2, with letters corresponding to Individuals A through L.[14] Individuals included in the study are filled in (black), with living relatives linking the family members not filled (white). Deceased relatives are crossed out. Individual L is unrelated to all individuals in the group. The central individual (Individual A) is circled in red.

in the $\log_{10}$ LR was observed with increasing locus dropout. For the fourth degree relationship, the $\log_{10}$ LR decreased from 43 to 28, with an average loss of 2.5 per 5% increment of locus dropout. The $\log_{10}$ LRs decreased by an average of 1.4 per 5% increment for the fifth degree relationship, from 22 to 17, still providing very strong statistical support for relatedness.

The seventh degree relationship (Individual 11) provided very strong statistical support for the proposed relationship when there was no locus dropout ($\log_{10}$ LR = 5.2). However, locus dropout resulted in a substantial drop in statistical support for relatedness when the call rate dropped to 70%, producing a $\log_{10}$ LR of 0.92. This means that, at this level of dropout, it is only 8.35 times more likely to observe these Kintelligence profiles if the individuals are seventh degree relatives than if they are unrelated. This was not distinguishable from the $\log_{10}$ LRs obtained from testing unrelated pairs of individuals.

Table 2. Direct-to-consumer kits available for Individuals A through L in Group 2[14]

Sample ID	*AncestryDNA*	*FamilyTreeDNA*	*23andMe*	*Living DNA*
Individual A	✓	✓	✓	✓
Individual B	✓	✓	✓	✓
Individual C	✓	✓	✓	✓
Individual D	✗	✓	✗	✗
Individual E	✓	✓	✗	✗
Individual F	✗	✓	✗	✗
Individual G	✓	✓	✗	✗
Individual H	✓	✓	✗	✗
Individual I	✓	✓	✗	✗
Individual J	✗	✓	✗	✗
Individual K	✗	✓	✗	✗
Individual L	✓	✗	✗	✗

When calculating the kinship LRs between the Individual 1 profiles and the unrelated individual (Individual 12), the $\log_{10}$ LRs increased as the rate of locus dropout increased (Fig. 4). However, this increase plateaued for all relationship degrees at 20% locus dropout. When testing H_1 for first to fourth degree relationships, there was very strong statistical support for H_2, consistent with the true unrelated status of this pair. When testing H_1 for fifth to seventh degree relationships, only limited to moderate statistical support was provided for H_2. Positive $\log_{10}$ LRs that statistically supported H_1 were produced for fifth degree at 20% locus dropout and greater, sixth degree at 15% locus dropout and greater, and seventh degree at 10% locus dropout and greater. The strongest statistical support for an incorrect inference of relatedness occurred at 20% locus dropout, where H_1 proposed a sixth degree relationship with a $\log_{10}$ LR of 1.5. The $\log_{10}$ LRs, when testing H_1 for fifth to seventh degree relationships, could not be distinguished from the results obtained from true seventh degree relationships when locus dropout was greater than 10%.

Table 3. Call rate and autosomal heterozygosity for Individual 1 (Pedigree Group 1) and Individual A (Pedigree Group 2) for the full Kintelligence profile, profiles generated with locus dropout ranging from 5% to 30% and profiles generated with allele dropout ranging from 5% to 30%

	Individual 1		*Individual A*	
Profile	*Call rate (%)*	*Autosomal heterozygosity (%)*	*Call rate (%)*	*Autosomal heterozygosity (%)*
Full profile	99.85	48.04	99.41	47.51
5% locus dropout	95.00	48.19	95.00	47.52
10% locus dropout	90.00	48.26	90.00	47.94
15% locus dropout	85.00	48.55	85.00	47.90
20% locus dropout	80.00	48.28	80.00	47.68
25% locus dropout	75.00	48.36	75.00	47.56
30% locus dropout	70.00	48.47	70.00	47.42
5% allele dropout	99.85	45.64	99.41	45.14
10% allele dropout	99.85	43.24	99.41	42.77
15% allele dropout	99.85	40.84	99.41	40.39
20% allele dropout	99.85	38.44	99.41	38.02
25% allele dropout	99.85	36.03	99.41	35.63
30% allele dropout	99.85	33.64	99.41	33.26

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

IBD segment matching between Kintelligence profiles (Pedigree Group 1). For the first to fourth degree relationships, all appeared in either the high confidence or expanded match lists for all levels of locus dropout tested. For Individual 1 profiles with call rates from 70% to 85%, the fifth degree relationship (Individual 10) results did not appear in the high confidence match list and only appeared in the expanded match list. The seventh degree relationship (Individual 11) and the unrelated individual (Individual 12) were not detected in the database search and were directly compared with Individual 1 using the One-to-One Kintelligence DNA Comparison tool.

As locus dropout increased, additional matches were identified by the One-to-Many Kinship tool that were not known genetic relatives of Individual 1 (Fig. 5). Very few of these matches were listed in the high confidence match list and those that were had an estimated shared cM over 200 (average: 205 ± 1 cM). The additional matches in the expanded match list had an average estimated shared cM of 160 ± 5 cM. No additional matches appeared in the high confidence match list when locus dropout was less than 15% or at 25% locus dropout. With the exception of the 25% locus dropout test, the number of additional matches in the extended match list increased as locus dropout increased. Furthermore, these additional matches tended to appear in the match list for the next level of locus dropout tested.

The Individual 1 profiles with lower call rates were able to match to the Individual 1 Kintelligence profiles with varying levels of locus dropout (0–30%) in the One-to-Many Kinship tool as likely "self," regardless of the number of SNPs being compared. All 21 pairwise comparisons resulted in high estimated shared cM values, with an average of 6,559 ± 4 cM. These values all fell within the range provided by GEDmatch PRO™ as likely to be self (6464–6684 cM).

There was minimal variability in the estimated shared cM for the relationships tested in Group 1 as locus dropout increased (Fig. 6). The standard deviations for the estimated shared cM values were calculated across the locus dropout tests for each individual and averaged to calculate the overall variance. The overall variance for the estimated shared cM values was 29.3 cM.

The average IBD segment length decreased with increasing locus dropout for first degree relationships (Individual 2 and Individual 3) and increased for the second degree relationships (Individual 4 and Individual 5; Supplementary Fig. S1). There was no discernible impact of locus dropout on the average IBD segment lengths for the other relationships. The overall variance in the average segment length was 2.9 cM.

The number of IBD segments detected by the algorithm increased with locus dropout for Individual 2, Individual 3, Individual 8, Individual 9, and Individual 10 but decreased for Individual 5 (Supplementary Fig. S2). The greatest impact of locus dropout on the number of IBD segments detected occurred for the parent/offspring relationship (Individual 2), where five additional segments were identified across the locus dropout range, and for the second degree relationship (Individual 5), which lost five segments (Supplementary Fig. S2). The overall variance in the number of IBD segments detected was 1.1 segments.

The unrelated individual (Individual 12) could not be distinguished from the seventh degree relationship (Individual 11) when considering the estimated shared cM and the average length and number of IBD segments detected. A single IBD segment was detected between Individual 1 and Individual 12 on chromosome 4 (24.5 cM) that dropped out at 15% locus dropout. A different IBD segment was detected on chromosome 13 at 20% locus dropout (24.1 cM) that increased in size for 25% and 30% locus dropout (43.1 cM), and a second

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

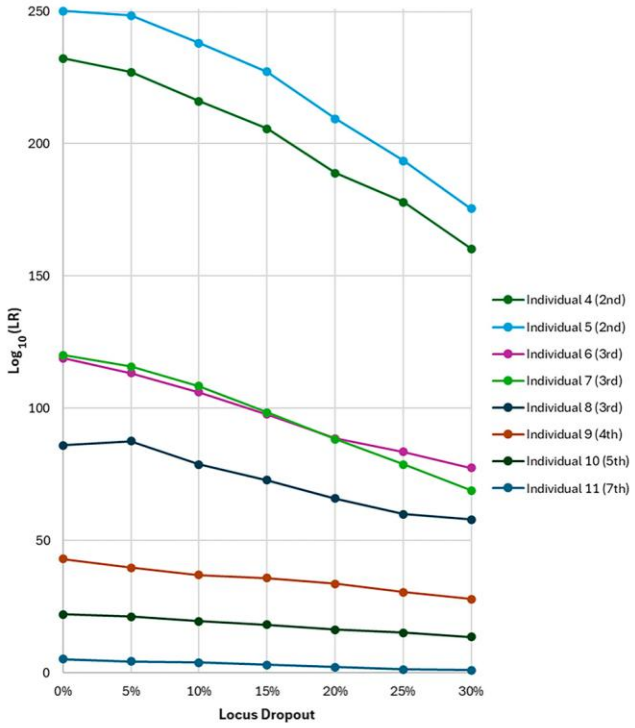


FIG. 3. Log_{10} LRs generated with locus dropout of Individual 1 ranging from 0% to 30% for all true relationships in Pedigree Group 1. The H_1 proposition was that the pair were related according to the true relationship degree. The H_2 proposition was that the pair were unrelated. The parent/offspring (Individual 2) was not plotted (log_{10} LR = $-\infty$). The full sibling (Individual 3) has not been plotted (log_{10} LR = ∞). 2nd: second degree relative; 3rd: third degree relative; 4th: fourth degree relative; 5th: fifth degree relative; 7th: seventh degree relative.

IBD segment only appeared when testing 25% locus dropout on chromosome 14 (22.5 cM).

IBD segment matching between Kintelligence and DTC kits (Pedigree Group 2). As observed with Pedigree Group 1, there was minimal variability in the results when comparing Individual A with varying degrees of locus dropout and the DTC profiles from Pedigree Group 2. The majority of DTC kits from genetically related individuals were included in the high confidence or expanded match lists. Only the Individual A 23andMe kit did not appear in the One-to-Many Kinship results at 30% locus dropout, as fewer than 6,000 overlapping SNPs were detected, failing to meet the threshold for inclusion in the database match list. Despite this, the direct comparison of this kit to the Individual A Kintelligence profile produced results that were consistent with all other DTC kits for Individual A.

All IBD segment matching results were consistent between the different types of relationships tested, as well as between the DTC kits for each individual. Locus dropout did not significantly impact the estimated shared cM (Fig. 7). The average IBD segment length and number of IBD segments detected decreased with increasing locus dropout for some individuals while increasing for others (Supplementary Fig. S3 and Supplementary Fig. S4,

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

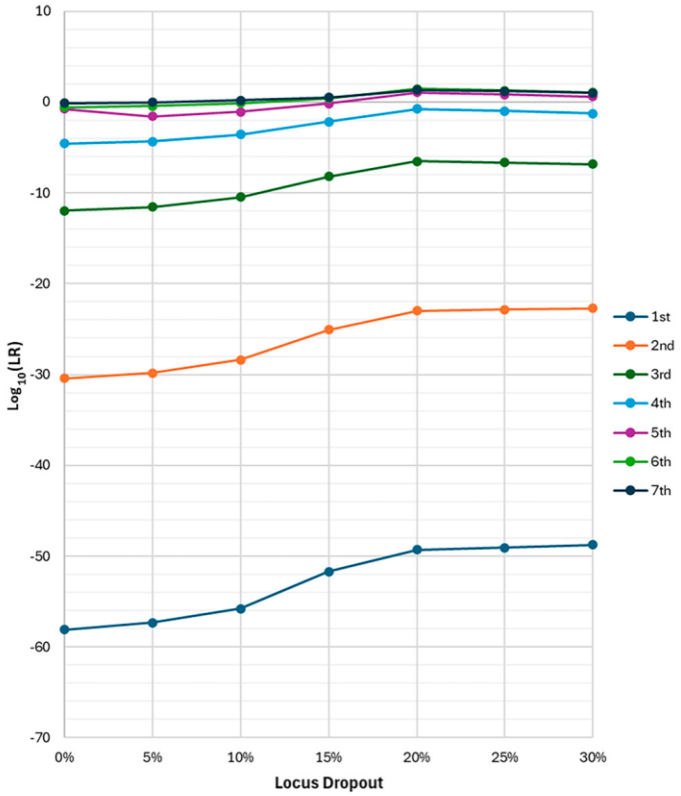


FIG. 4. Log_{10} LRs generated with locus dropout of Individual 1 ranging from 0% to 30% and a true unrelated individual (Individual 12). The H_1 proposition was that the pair were related according to a specific relationship degree. The H_2 proposition was that the pair were unrelated (true). Parent/offspring tests have not been included (log_{10} LR = $-\infty$). 1st: first degree relative; 2nd: second degree relative; 3rd: third degree relative; 4th: fourth degree relative; 5th: fifth degree relative; 6th: sixth degree relative; 7th: seventh degree relative.

respectively). The overall variances across locus dropout tests were 25.2 cM for the estimated shared cM, 3.4 cM for the average IBD segment lengths, and 1.3 for the number of IBD segments.

The greatest degree of variability in results due to increasing locus dropout was observed in the number of IBD segments detected (Supplementary Fig. S4). For the parent/offspring relationship (Individual C), several IBD segments were lost, with the AncestryDNA kit decreasing by five segments and the FamilyTreeDNA kit by only two segments. IBD segments were lost for both kits for one full sibling relationship (Individual G) after additional IBD segments were detected at 15% locus dropout for the AncestryDNA kit and 5% for the FamilyTreeDNA kit; by 30% locus dropout, the number of IBD segments detected for the AncestryDNA kit decreased by five segments, whereas the FamilyTreeDNA kit decreased by only two segments. In contrast, for the full sibling relationship (Individual I), additional IBD segments were detected as locus dropout increased, with the

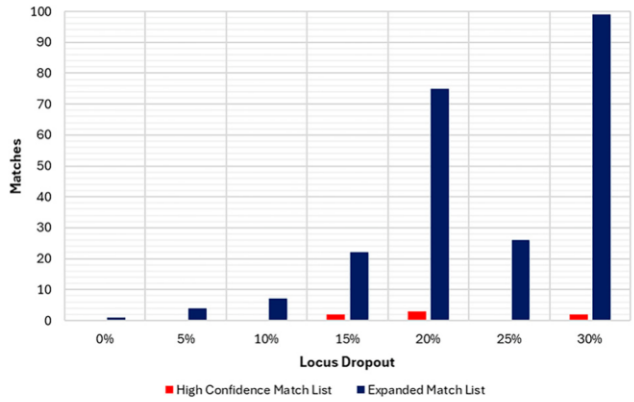


FIG. 5. Additional matches included in the high confidence and expanded match lists on GEDmatch PRO™ that were not known genetic relatives of Individual 1 at various degrees of locus dropout.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

AncestryDNA kit increasing by six segments and the FamilyTreeDNA kit by five segments.

The unrelated individual (Individual L) had only a single IBD segment detected (28.2 cM) when compared with the full Individual A Kintelligence profile. However, this segment was not detected when comparing Individual L with any of the profiles with locus dropout.

Impact of allele dropout

LR calculations between Kintelligence profiles (Pedigree Group 1). The $\log_{10}$ LR decreased with increasing allele dropout of the Individual 1 profile (Fig. 8). As observed in the locus dropout study, the impact of allele dropout on the $\log_{10}$ LRs for the first degree relationships was unable to be evaluated. The $\log_{10}$ LRs remained at ∞ for the full sibling relationship (Individual 3) as $\log_{10}$ LRs exceeding 300 are not reported by DBLR™. As there was already allele dropout observed in the comparison of full profiles for the parent/offspring relationship (Individual 2), the $\log_{10}$ LRs were represented as $-\infty$, inferring the relationship was impossible. The second degree relationships showed the greatest rate of decrease in $\log_{10}$ LRs at an average of 22.5 between each 5% increment of allele dropout tested, with those from Individual 4 decreasing from 232 to 111 and those from Individual 5 decreasing from 250 to 176.

The $\log_{10}$ LRs decreased in a similar pattern for the third degree relationships (Individual 6, Individual 7, and Individual 8), with minimal change between 15% and 20% allele dropout and between 25% and 30% allele dropout. Although Individual 6 and Individual 7 produced similar values, the $\log_{10}$ LRs for Individual 8 were lower but still provided very strong statistical support for relatedness. The $\log_{10}$ LR for the fourth degree relationship (Individual 9) decreased from 43 to 11, with an average reduction of 5.3 per 5% increment of allele dropout. The $\log_{10}$ LR for the fifth degree relationship (Individual 10) decreased from 22 to 4.7 at 30% allele dropout but still provided strong statistical support for H_1. Finally, the $\log_{10}$ LRs calculated for the seventh degree relationship (Individual 11) decreased from 5.2 at 0% allele dropout to being uninformative, where a $\log_{10}$ LR is at 0 and neither hypothesis is supported, at 15% allele dropout ($\log_{10}$ LR = 0.1) and provided limited statistical support for H_2 at 30% allele dropout ($\log_{10}$ = –0.5).

For the unrelated individual (Individual 12), allele dropout resulted in a decrease in the $\log_{10}$ LR (Fig. 9). The highest $\log_{10}$ LR across all tests was the H_1 proposition for a seventh degree relationship with no allele dropout ($\log_{10} LR = -0.1$), which provided limited statistical support for H_2. All first to fourth degree relationship tests provided very strong statistical support for H_2, with only limited to moderate statistical support for H_2 for fifth to seventh degree relationships.

IBD segment matching between Kintelligence profiles (Pedigree Group 1). In contrast to the locus dropout study, a gradual decrease was observed in almost all results obtained from the IBD segment matching algorithm with increasing allele dropout. Allele dropout

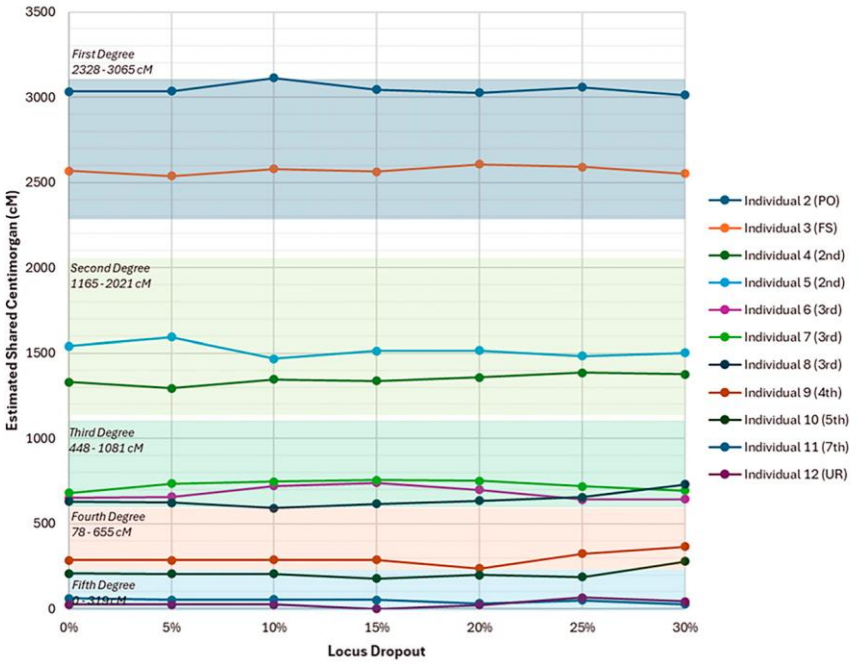


FIG. 6. Estimated shared centimorgan (cM) values shared between the Kintelligence profile of Individual 1 and other Kintelligence profiles in Pedigree Group 1 with locus dropout ranging from 0% to 30%. The ranges indicated for kinship degrees are as included in the generation chart on GEDmatch PRO™.[12] PO: parent/offspring; FS: full sibling; 2nd: second degree relative; 3rd: third degree relative; 4th: fourth degree relative; 5th: fifth degree relative; 7th: seventh degree relative; UR: unrelated.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

resulted in the more distant genetic relatives not being included in the match lists when searching the database. The fifth degree relationship (Individual 10) did not appear in the high confidence or expanded match lists using the One-to-Many Kinship tool when allele dropout exceeded 10%. As a result, the remaining comparisons for the fifth degree relationship were performed using the One-to-One Kintelligence DNA Comparison tool. Similarly, the comparisons for the seventh degree relationship (Individual 11) and unrelated individual (Individual 12), across all degrees of allele dropout tested, were conducted using the One-to-One Kintelligence DNA Comparison tool.

No additional matches were included in the high confidence match list that were not already known genetic relatives of Individual 1. However, several additional matches were observed in the expanded match list, with fewer than 10 matches for each version of the profile with varying allele dropout (Fig. 10). The average estimated shared cM for these matches was 139 ± 4 cM. The number of additional matches did not correlate with the allele dropout rate and the matches did not appear in any other match lists for different levels of allele dropout tested.

When comparing the full Kintelligence profile to those with increasing amounts of allele dropout, the estimated shared cM values showed a sharp decline after 5% allele dropout (Table 4). At 5% allele dropout, the estimated shared cM remained within the range provided by GEDmatch PRO™ for self (6,464–6,684 cM) but decreased and plateaued around 3400 cM after 10% allele dropout. This suggests that, beyond 5% dropout, the profiles are more likely to represent first degree relatives rather than originating from the same individual.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

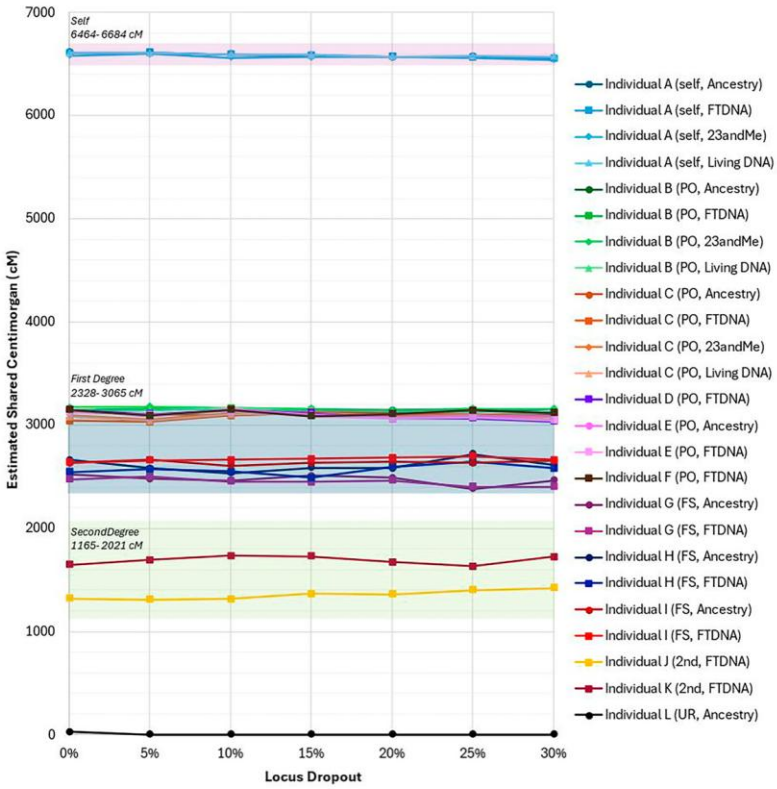


FIG. 7. Estimated shared centimorgan (cM) shared between the Kintelligence profile of Individual A and direct-to-consumer profiles in Pedigree Group 2 with locus dropout ranging from 0% to 30%. The ranges indicated for kinship degrees are as included in the generation chart on GEDmatch PRO™.[12] PO: parent/offspring; FS: full sibling; 2nd: second degree relative; UR: unrelated; FTDNA: FamilyTreeDNA.

Despite this decrease in shared cM, the number of IBD segments (22) and average IBD segment length (152 cM) remained consistent across the comparisons and allows for differentiation between self with allele dropout and true parent/offspring relationships. However, there were increasing numbers of SNPs where both alleles were not in common between the profiles, resulting in the decreased estimated shared cM. For all 21 pairwise comparisons of Individual 1 profiles, the estimated shared cM values were within the range for self if the allele dropout rates were within 5% (Table 4).

Figure 11 illustrates the gradual decrease in the estimated shared cM between Individual 1 and the individuals in Pedigree Group 1 as allele dropout increased. There was no observable impact on the parent/offspring relationship (Individual 2), with a variance of 35.6 cM. In contrast, the full sibling relationship (Individual 3) showed a decrease in the estimated shared cM from an average of 2,530 ± 40 cM (0–15% allele dropout) to 2298 cM at 20%, dropping below the range for first degree relationships (2,328–3,065 cM). The estimated shared cM decreased for second degree relationships

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

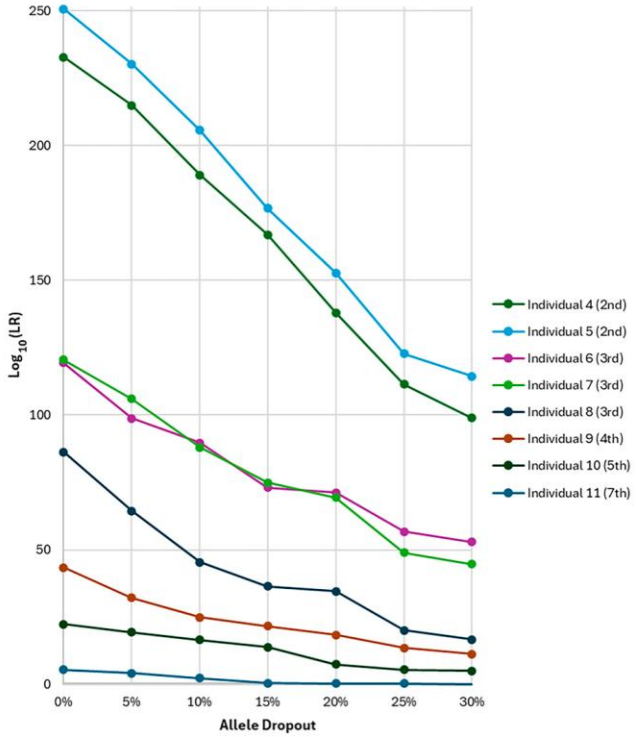


FIG. 8. Log_{10} LRs generated with allele dropout of Individual 1 ranging from 0% to 30% for all true relationships in Pedigree Group 1. The H_1 proposition was that the pair were related according to the true relationship degree. The H_2 proposition was that the pair were unrelated. The parent/offspring was not plotted (log_{10} LR = $-\infty$). The full sibling has not been plotted (log_{10} LR = ∞). 2nd: second degree relative; 3rd: third degree relative; 4th: fourth degree relative; 5th: fifth degree relative; 7th: seventh degree relative.

(Individual 4 and Individual 5) after 20% allele dropout, falling within the range for third degree relationships (448–1,081 cM) at 30% allele dropout.

There was greater variability observed in the third degree relationships. The estimated shared cM for Individual 7 remained stable until 30% allele dropout, while the estimated shared cM values for Individual 6 decreased from an average of 686 ± 30 (0–15% allele dropout) to 522 ± 6 cM (20–30% allele dropout). However, Individual 8 produced consistent results up to 20% allele dropout (average = 600 ± 29 cM) before dropping sharply to 143 cM, the greatest decrease observed across all allele dropout tests.

Overall, allele dropout had minimal impact on the estimated shared cM values for relationships of fourth degree and greater, except when no IBD segments were detected, i.e., no estimated shared cM (Fig. 11). The estimated shard cM for the fourth degree relationship (Individual 9) steadily increased from 285 cM (0% allele dropout) to a peak of 357 cM (15% allele dropout), before decreasing to 169 cM (30% allele dropout). Despite not appearing in the match lists, there was minimal variability observed in the estimated shared cM for the fifth degree relationship (Individual 10, variance = 46.6 cM) and seventh degree relationship (Individual 11,

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

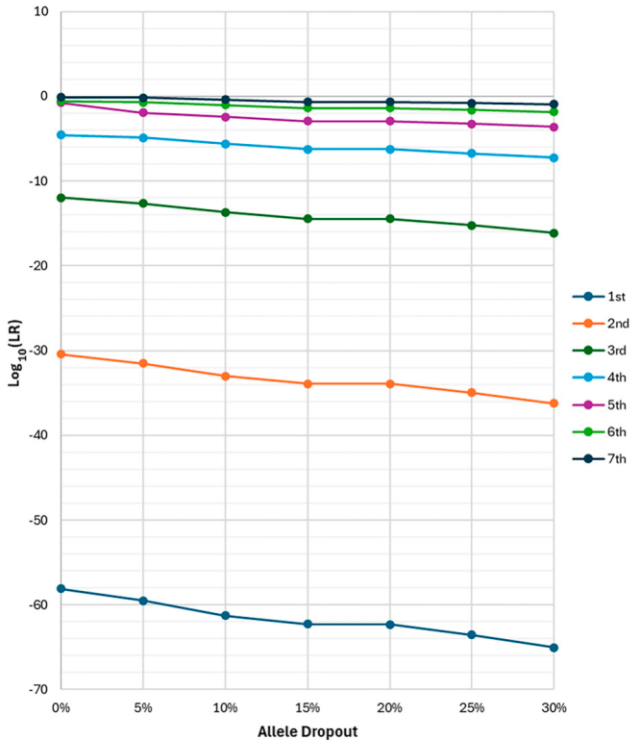


FIG. 9. Log_{10} LRs generated with allele dropout of Individual 1 ranging from 0% to 30% and a true unrelated individual (Individual 12). The H_1 proposition was that the pair were related according to a specific relationship degree. The H_2 proposition was that the pair were unrelated (true). Parent/offspring tests have not been included as (log_{10} LR = $-\infty$). 1st: first degree relative; 2nd: second degree relative; 3rd: third degree relative; 4th: fourth degree relative; 5th: fifth degree relative; 6th: sixth degree relative; 7th: seventh degree relative.

variance = 16.8 cM). However, at 30% allele dropout, no IBD segments were detected for the seventh degree relationship.

The single IBD segment detected between the Individual 1 and the unrelated individual (Individual 12) on chromosome 4 (24.5 cM) was identifiable when testing the Kintelligence profile with 0% to 10% allele dropout. At higher levels of allele dropout, no IBD segments were detected and the estimated shared cM was 0.

The average length of the IBD segments typically increased with increasing allele dropout for several relationships (Supplementary Fig. S5). This trend was observed in the full sibling (Individual 3), a second degree (Individual 4), some third degree (Individual 6 and Individual 7), and the fourth degree (Individual 9) relationships. However, there was minimal variability observed in the parent/offspring (Individual 2), other second degree (Individual 5), and seventh degree (Individual 11) relationships, until no IBD segments were detected at 30% allele dropout for the latter. An exception was Individual 8, which showed an increase in the average IBD segment length up to 20% allele dropout before dropping at 25%. The overall variance in the average IBD segment length across all relationships tested was 5.4 cM.

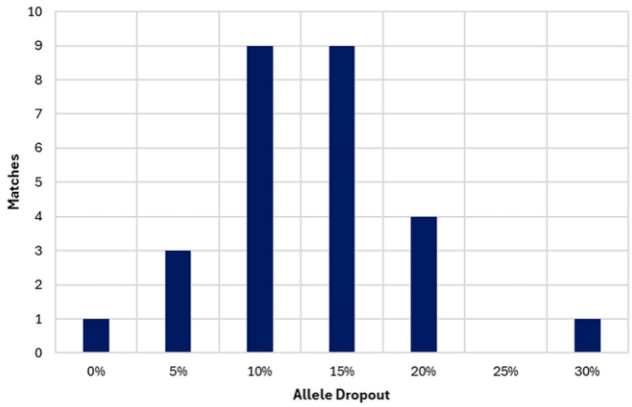


FIG. 10. Additional matches included in the expanded match list on GEDmatch PRO™ that were not known genetic relatives of Individual 1 at various degrees of allele dropout.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

Increasing allele dropout generally resulted in fewer IBD segments shared between relatives being detected (Supplementary Fig. S6). However, the parent/offspring (Individual 2) and fourth degree (Individual 9) relationships showed a greater number of IBD segments being detected with increasing allele dropout. For the former, an additional six IBD segments were detected across the allele dropout range tested. The latter saw a large increase of nine IBD segments from 0% to 20% allele dropout, before dropping to only four IBD segments at 30% allele dropout. Excluding these individuals, the overall variance in the number of IBD segments detected was 2.6 segments.

IBD segment matching between Kintelligence and DTC kits (Pedigree Group 2). Similar trends to those observed in Pedigree Group 1 were seen when comparing the Kintelligence profile of Individual A with varying amounts of allele dropout and the DTC kits from Pedigree Group 2. There was a greater impact of allele dropout on the results than for locus dropout. All DTC kits from the true genetic relatives were detected in the high confidence or expanded match lists using the One-to-Many Kinship tool; only the unrelated individual (Individual L) was not detected in the database search and required direct comparison using the One-to-One Kintelligence DNA Comparison tool.

When comparing the Kintelligence profile to the DTC kits for Individual A, there was no observable impact of 5% allele dropout on the estimated shared cM (Fig. 12). However, after this, the estimated shared cM decreased substantially until plateauing at approximately 3,400 cM when allele dropout was 15% or higher. This result was similar to those of the Individual 1 Kintelligence profiles, indicating that the relationship between the Individual A profiles was more likely to be of the first degree rather than self. Despite the impact on the estimated shared cM between Individual A profiles, no variation was observed in the average length of IBD segments (152 cM; Supplementary Fig. S7) or the number of IBD segments (22 segments;

Table 4. Estimated shared centimorgan (cM) results between the Individual 1 Kintelligence profiles with allele dropout ranging from 0% to 30%

Allele dropout	*0%*	*5%*	*10%*	*15%*	*20%*	*25%*	*30%*
0%		6,631.3	5,061.6	3,490.7	3,341.5	3,341.5	3,341.5
5%	6,631.3		6,661.7	5,194.3	3,493.7	3,341.5	3,341.5
10%	5,061.6	6,661.7		6,681.4	5,024.3	3,493.5	3,341.5
15%	3,490.7	5,194.3	6,681.4		6,683	5,223.8	3,598.5
20%	3,341.5	3,493.7	5,024.3	6,683		6,683	4,900.7
25%	3,341.5	3,341.5	3,493.5	5,223.8	6,683		6,683
30%	3,341.5	3,341.5	3,341.5	3,598.5	4,900.7	6,683	

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

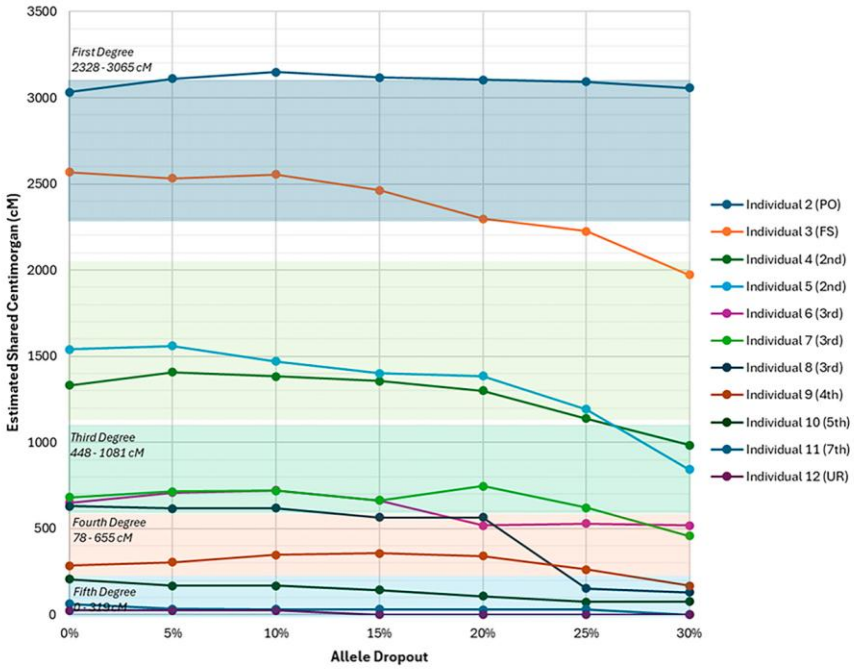


FIG. 11. Estimated shared centimorgan (cM) values shared between the Kintelligence profile of Individual 1 and other Kintelligence profiles in Pedigree Group 1 with allele dropout ranging from 0% to 30%. The ranges indicated for kinship degrees are as included in the generation chart on GEDmatch PRO™.[12] PO: parent/offspring; FS: full sibling; 2nd: second degree relative; 3rd: third degree relative; 4th: fourth degree relative; 5th: fifth degree relative; 7th: seventh degree relative; UR: unrelated.

Supplementary Fig. S8) with increasing allele dropout. Both the average length and number of IBD segment results support that the relationship is self with allele dropout as opposed to a parent/offspring relationship.

There was a decrease in the estimated shared cM when comparing the Individual A Kintelligence profiles to the DTC kits of the first and second degree relatives with increasing allele dropout (Fig. 12). The relationship with the least variability across the range of allele dropout was a parent/offspring relationship (Individual B, variance = 30.4 cM), with no differences observed between the different DTC kits. The other parent/offspring relationships decreased by approximately 200 cM once allele dropout exceeded 25%, bringing the overall variance for parent/offspring relationships to 92.7 cM. There was a steady decrease in the estimated shared cM values for full sibling relationships with increasing allele dropout. Individual G had the greatest reduction, with an average loss of 138 cM for every 5% interval of allele dropout tested. There was no observable impact for one of the second degree relationships (Individual J); however, the estimated shared cM for the other second degree relationship (Individual K) steadily decreased by 127 cM per 5% interval of allele dropout, before dropping by 303 cM between 25% and 30% allele dropout.

There was generally low variability in the relationships for the average length of IBD segments with an overall variance of 3.3 cM (Supplementary Fig. S7). The number of IBD segments showed a general decrease with increasing allele dropout and had an overall variance of 1.6 segments (Supplementary Fig. S8). Furthermore, no difference were observed between the DTC kits

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

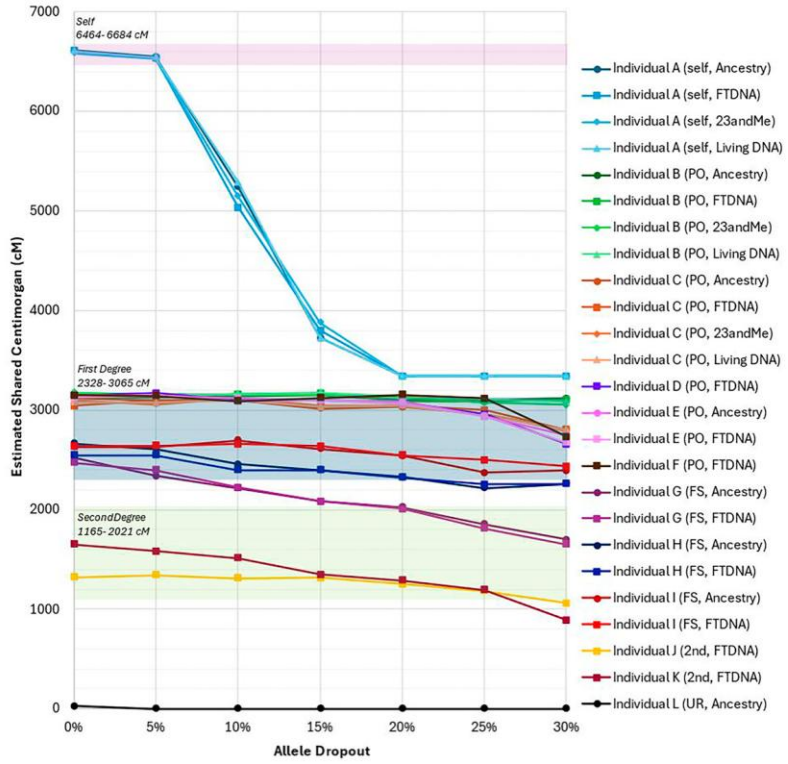


FIG. 12. Estimated shared centimorgan (cM) shared between the Kintelligence profile of Individual A and direct-to-consumer profiles in Pedigree Group 2 with allele dropout ranging from 0% to 30%. The ranges indicated for kinship degrees are as included in the generation chart on GEDmatch PRO™.[12] PO: parent/offspring; FS: full sibling; 2nd: second degree relative; UR: unrelated; FTDNA: FamilyTreeDNA.

for each individual. The greatest variability in the average length of IBD segments was observed in the parent/offspring relationships, whereas there was a slight increase in the average segment length for full sibling relationships and a second degree relationship (Individual K; Supplementary Fig. S7).

There were two individuals where allele dropout impacted the number of IBD segments detected (Supplementary Fig. S8). The results for both DTC kits for Individual G varied; the first 5% of allele dropout caused the loss of six IBD segments for the AncestryDNA kit that were not lost for the FamilyTreeDNA kit and up to four IBD segments were lost for each 5% interval. Individual K also showed a loss of IBD segments with increasing allele dropout, losing up to four IBD segments per 5% increment and a total of 16 IBD segments by 30% allele dropout.

Discussion

The Kintelligence Kit is a robust panel capable of producing extended kinship inferences with both LR calculations and IBD segment matching algorithms, even with suboptimal partial profiles. When testing locus and allele dropout, both kinship inference methods produced sufficient statistical support for relatedness, extending to fifth

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

degree relationships (e.g., second cousins). Both methods allow for direct comparison of DNA profiles between putative genetic relatives and the windowed kinship algorithm on GEDmatch PRO™ facilitates IBD segment matching for medium-density SNP profiles, enabling the identification of possible genetic relatives present in the database. In cases where other forensic techniques are unable to provide an identification, extended kinship inference can provide law enforcement with investigative leads. This method is particularly useful beyond the short-range kinship inference offered by short tandem repeat profiles, which are typically limited to first degree relationships.

The LR calculations were negatively impacted by both forms of information loss tested in this study but still provided at least very strong statistical support ($\log_{10}$ LR > 5) for all full sibling to fifth degree relationships.[16] For both the locus and allele dropout studies, the parent/offspring relationship could not be calculated, as there had to be at least one allele in common at every SNP. This was also the case in part one of this study, and the application of a probabilistic model is recommended to account for genotyping error due to the large size of the Kintelligence panel.[14] The ability to infer the full sibling relationship was not practically impacted by locus or allele dropout. As the $\log_{10}$ LRs exceeded 300, DBLR™ returned the $\log_{10}$ LR at ∞ for all full sibling tests. The first to fifth degree relationships could be differentiated from the unrelated individual for all levels of locus and allele dropout tested in this study. However, the seventh degree relationship could not be distinguished from the unrelated individual once locus dropout exceeded 20% or allele dropout exceeded 10%.

Partial medium-density SNP profiles have been evaluated for the impact of locus dropout on LR calculations in previous studies.[2,11] Gettings et al. (2024) observed a median $\log_{10}$ LR of approximately 780 for first degree, 200 for second degree, 70 for third degree, 30 for fourth degree, and 5 for fifth degree relationships using simulated partial Kintelligence profiles.[11] The results generated in this study with Pedigree Group 1 supported these findings for second to fourth degree relationships but produced $\log_{10}$ LRs exceeding 300 for full siblings and $\log_{10}$ LRs exceeding 13 for fifth degree relationships. This indicates very strong statistical support for all full sibling to fifth degree relationships tested with up to 30% locus dropout. In contrast, 97% of fourth degree and 81% of fifth degree relationships produced LRs that provided strong statistical support ($\log_{10}$ LR > 4) in the study by Gettings et al. (2024).[11,16]

Locus dropout was also evaluated by Kling (2019) when constructing a 6,600 SNP panel for extended kinship inference.[2] By reducing the original panel of approximately 10,000 SNPs to 6,600, the $\log_{10}$ LRs decreased for real-world third and fifth degree relatives by approximately half. This magnitude of loss in the statistical support for relatedness was similar to the findings in this study when testing the loss of 30% of available SNPs in the Kintelligence Kit. However, both of these other studies only assessed locus dropout as a factor influencing extended kinship inference.[2,11] In contrast, this study observed that allele dropout had a greater impact on the LR calculations and it is likely that Kintelligence profiles obtained from compromised forensic samples will likely contain a combination of both locus and allele dropout. Despite 30% of heterozygous SNPs being incorrectly called as homozygous, all $\log_{10}$ LRs for full sibling to fifth degree relationships exceeded 5, providing very strong statistical support for relatedness.[16]

The results obtained from IBD segment matching were less impacted by information loss than the LR calculations. Locus dropout had no observable impact on the ability to infer relatedness with up to 30% loss of the profile and first to fifth degree relatives could be differentiated from unrelated individuals at all levels of locus dropout tested. When the call rate of a Kintelligence profile is less than 70%, the profile is no longer compatible with upload to GEDmatch PRO™.[12] This limit was recommended by Snedecor et al. (2022) in the development of the windowed kinship algorithm, which tested locus dropout at intervals of 2,000 SNPs.[8] Their simulations found that locus dropout could not exceed 20% to reliably infer fifth degree relationships, 40% for fourth degree relationships and 60% for third degree relationships.[8] This study observed that the fifth degree relationship was still detectable at 30% locus dropout, but there was a greater number of potential matches listed in the database search that were not known genetic relatives.

Similar to the LR calculations, allele dropout had a greater impact on IBD segment matching but did not significantly impair the ability to detect first to fourth degree relationships and differentiate them from unrelated individuals, with only the fifth degree relative failing to appear in the database match list after 10% allele dropout. Allele dropout also complicated the ability to distinguish profiles originating from the same person from those of their first degree relatives; however, the average length and number of IBD segments can assist with differentiating self from a parent/offspring relationship as these values did not vary with allele dropout. Snedecor et al. (2022) simulated 5% to 100% allele dropout on the windowed kinship algorithm and concluded that the results were sufficient to infer first to third degree relationships up to 20% allele dropout and first to second degree relationships up to 40%.[8] However, the specificity

of extended kinship inference was not impacted and false positives were not introduced by reducing profile heterozygosity.[8] These findings were supported by the present study, where <10 matches were returned that were not known genetic relatives of the individual.

Furthermore, there were consistent findings between Pedigree Group 1 and Pedigree Group 2 for the first and second degree relationships, as well as comparisons with known unrelated individuals. This indicates that comparing Kintelligence profiles with either DTC kits uploaded by the public or other Kintelligence profiles does not impact the results.

Conclusions

This study demonstrated that partial medium-density SNP profiles generated with the Kintelligence Kit are suitable for extended kinship inference using both LR calculation or IBD segment matching. While both methods were impacted by information loss, allele dropout had a greater effect than locus dropout. Despite this, first to fifth degree relationships could still be detected and differentiated from unrelated individuals with the maximum dropout levels applied. LR calculations provided strong statistical support for full siblings to fifth degree relatives, excluding parent/offspring relationships where an allele must be in common at every SNP. The performance of the windowed kinship algorithm on GEDmatch PRO™ was maintained across all levels of locus dropout tested, whereas the accuracy of IBD segment matching results decreased with increasing allele dropout. Both extended kinship inference methods can therefore be employed for forensic samples that produce partial Kintelligence profiles.

Acknowledgments

The authors would like to thank the volunteers and their families for providing their DNA and associated metadata for this study; and Dr. Maarten Kruijver, Bjorn Sutherland, and Dr. Kevin Cheng from the Institute of Environmental Science and Research Limited for their assistance with using DBLR™.

Authors' Contributions

J.L.W.: Conceptualization (equal); methodology (lead); formal analysis (lead); writing—original draft (lead); writing—review and editing (equal). K.G.: Methodology (supporting); formal analysis (supporting); writing—review and editing (equal). L.C.: Methodology (supporting); writing—review and editing (equal). D.M.: Conceptualization (equal); methodology (supporting); formal analysis (supporting); writing—review and editing (equal); supervision (equal). J.W.: Conceptualization (equal); methodology (supporting); writing—review and editing (equal); supervision (equal); funding acquisition (lead).

Institutional Review Board Statement

This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of University of Technology Sydney (UTS) Human Research Ethics Committee (HREC) (UTS HREC NO. ETH21-5821).

Informed Consent Statement

Written informed consent was obtained from all volunteer sample donors involved in this study.

Data Availability Statement

Data are stored at the Australian Federal Police and may be made available to approved entities upon written request and subject to consent provisions.

Author Disclosure Statement

The authors declare no conflict of interest.

Funding Information

This research was funded by the Australian Federal Police (AFP) Innovation Fund and the AFP National DNA Program for Unidentified and Missing Persons. J.L.W. is supported by an Australian Government Research Training Program Scholarship.

Supplementary Material

Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Figure S5
Supplementary Figure S6
Supplementary Figure S7
Supplementary Figure S8

References

1. Kling D, Phillips C, Kennett D, et al. Investigative genetic genealogy: Current methods, knowledge and practice. Forensic Sci Int Genet 2021;52:102474; doi: 10.1016/j.fsigen.2021.102474
2. Kling D. On the use of dense sets of SNP markers and their potential in relationship inference. Forensic Sci Int Genet 2019;39:19–31; doi: 10.1016/j.fsigen.2018.11.022
3. Watson JL, McNevin D, Ward J. Genetic kinship testing techniques for human remains identification and missing persons investigations. Forensic Genom 2024;4(1):4–23; doi: 10.1089/forensic.2023.0018
4. Gill P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med 2001;114(4–5):204–210; doi: 10.1007/s004149900117
5. Seo SB, King JL, Warshauer DH, et al. Single nucleotide polymorphism typing with massively parallel sequencing for human identification. Int J Legal Med 2013;127(6):1079–1086; doi: 10.1007/s00414-013-0879-7
6. Antunes J, Walichiewicz P, Forouzmand E, et al. Developmental validation of the ForenSeq® Kintelligence Kit, MiSeq Fgx® sequencing system

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

and ForenSeq® universal analysis software. Forensic Sci Int Genet 2024; 71:103055; doi: 10.1016/j.fsigen.2024.103055
7. Scientific Working Group on DNA Analysis Methods. Recommendations from the SWGDAM Ad Hoc Working Group on Familial Searching. SWGDAM Communications; 2013. Available from: https://www.swgdam.org/publications [Last accessed: January 2, 2025].
8. Snedecor J, Fennell T, Stadick S, et al. Fast and accurate kinship estimation using sparse SNPs in relatively large database searches. Forensic Sci Int Genet 2022;61:102769; doi: 10.1016/j.fsigen.2022.102769
9. QIAGEN. Verogen ForenSeq Kintelligence Kit. 2024. Available from: https://www.qiagen.com/us/products/human-id-and-forensics/nextgeneration-sequencing/verogen-forenseq-kintelligence-kit [Last accessed: May 24, 2024].
10. Watson JL, McNevin D, Grisedale K, et al. Operationalisation of the ForenSeq® Kintelligence kit for Australian unidentified and missing persons casework. Forensic Sci Int Genet 2024;68:102972; doi: 10.1016/j.fsigen.2023.102972
11. Gettings KB, Tillmar A, Sturk-Andreaggi K, et al. Review of SNP assays for disaster victim identification: Cost, time, and performance information for decision-makers. J Forensic Sci 2024;69(5):1546–1557; doi: 10.1111/1556-4029.15585
12. GEDmatch PRO. GEDmatch PRO. 2024. Available from: https://pro.gedmatch.com/ [Last accessed: November 25, 2024].
13. FamilyTreeDNA. FamilyTreeDNA Law Enforcement Guide. 2024. Available from: https://www.familytreedna.com/legal/law-enforcement-guide [Last accessed: November 25, 2024].
14. Watson JL, Grisedale K, Coakley L, et al. Extended kinship inference. Part 1: Evaluation of short tandem repeats and single nucleotide polymorphisms using likelihood ratios and haplotype matching. Forensic Genom 2025;5(1):13–31; doi: 10.1089/forensic.2025.0001
15. STRmix™. DBLR™. 2023. Available from: strmix.com/dblr/ [Last accessed: 29 January].
16. Scientific Working Group on DNA Analysis Methods. Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Resulted Reported as Likelihood Ratios. 2016.

Downloaded by Jodie Ward from www.liebertpub.com at 03/22/25. For personal use only.

8.1 Supplementary Material

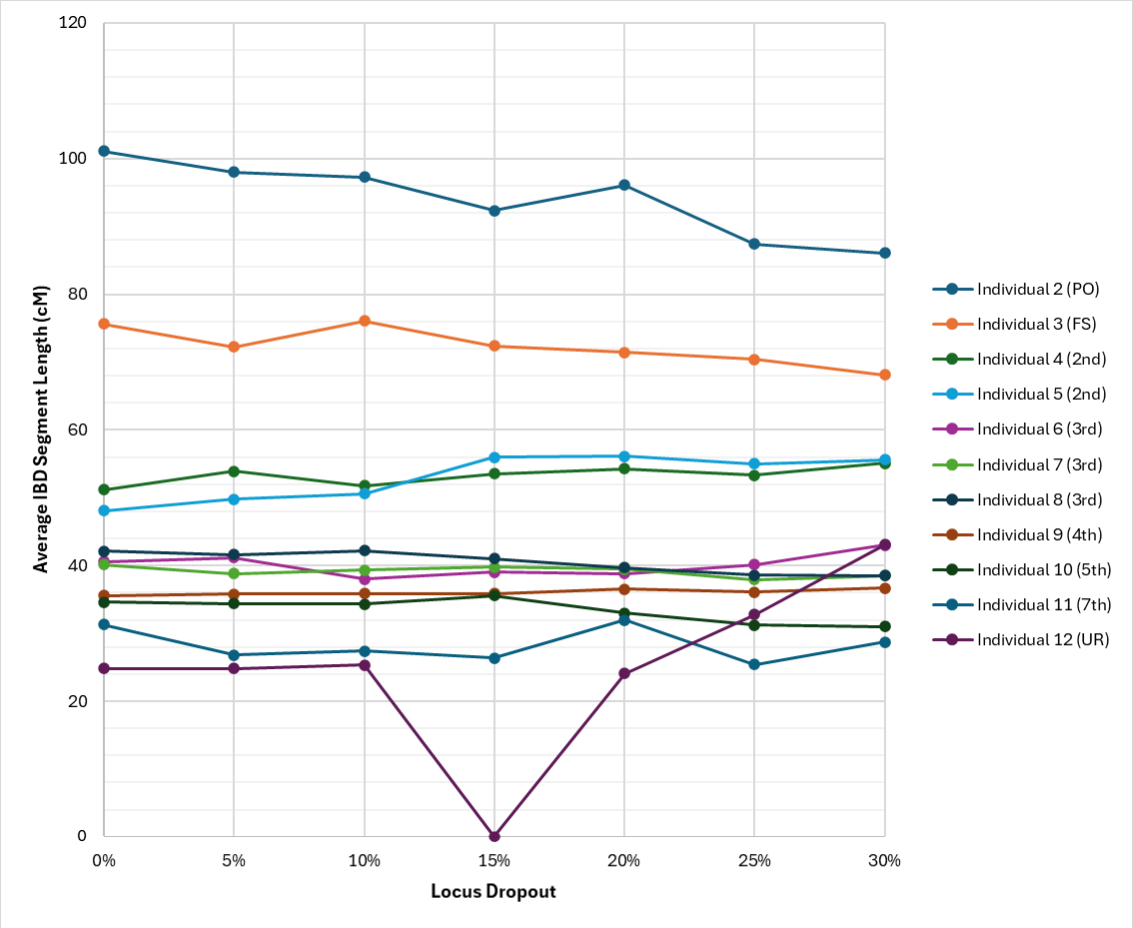


Figure S1. Average identical-by-descent (IBD) segment lengths in centimorgans (cM) shared between the Kintelligence profile of Individual 1 and other Kintelligence profiles in Pedigree Group 1 with locus dropout ranging from 0 to 30%. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; 3rd: third degree relative; 4th: fourth degree relative; 5th: fifth degree relative; 7th: seventh degree relative; UR: unrelated.

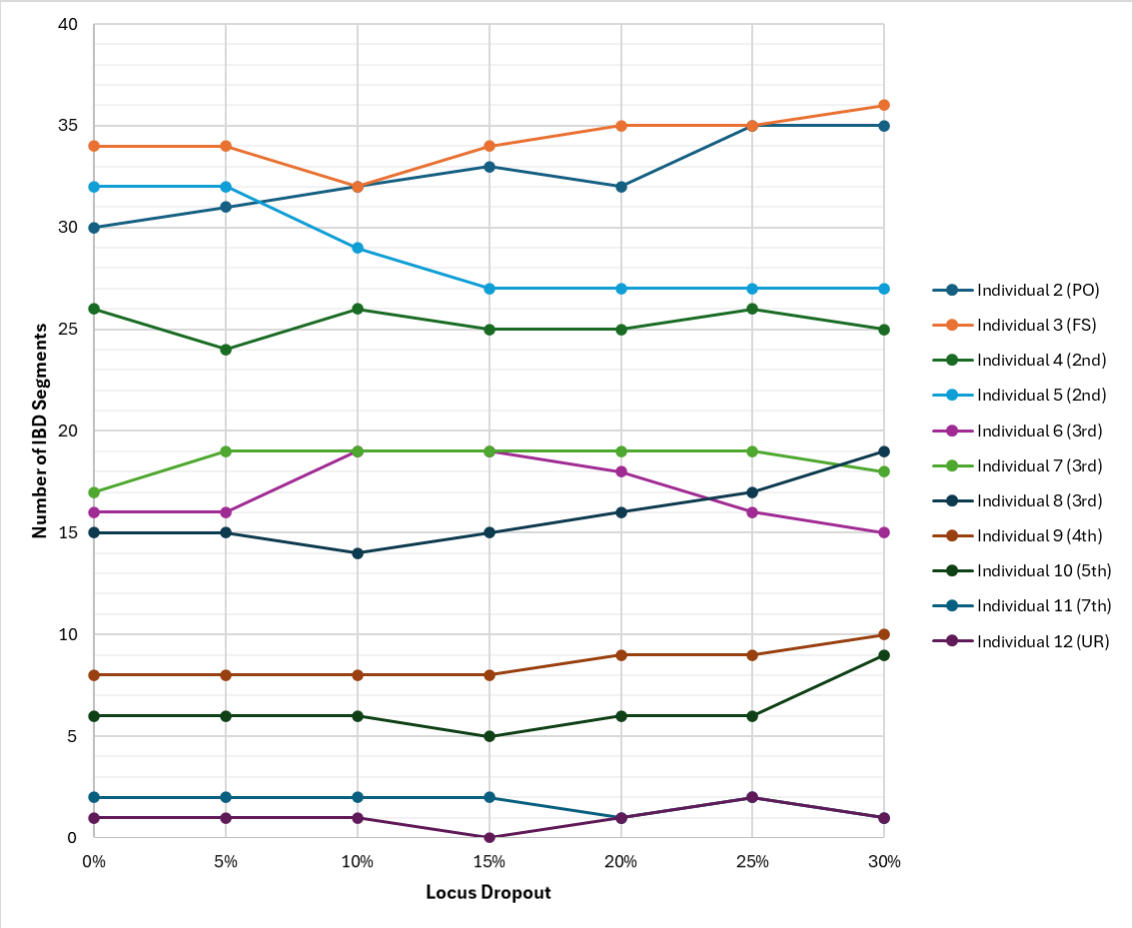


Figure S2. Number of identical-by-descent (IBD) segments detected by GEDmatch PRO™ shared between the Kintelligence profile of Individual 1 and other Kintelligence profiles in Pedigree Group 1 with locus dropout ranging from 0 to 30%. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; 3rd: third degree relative; 4th: fourth degree relative; 5th: fifth degree relative; 7th: seventh degree relative; UR: unrelated.

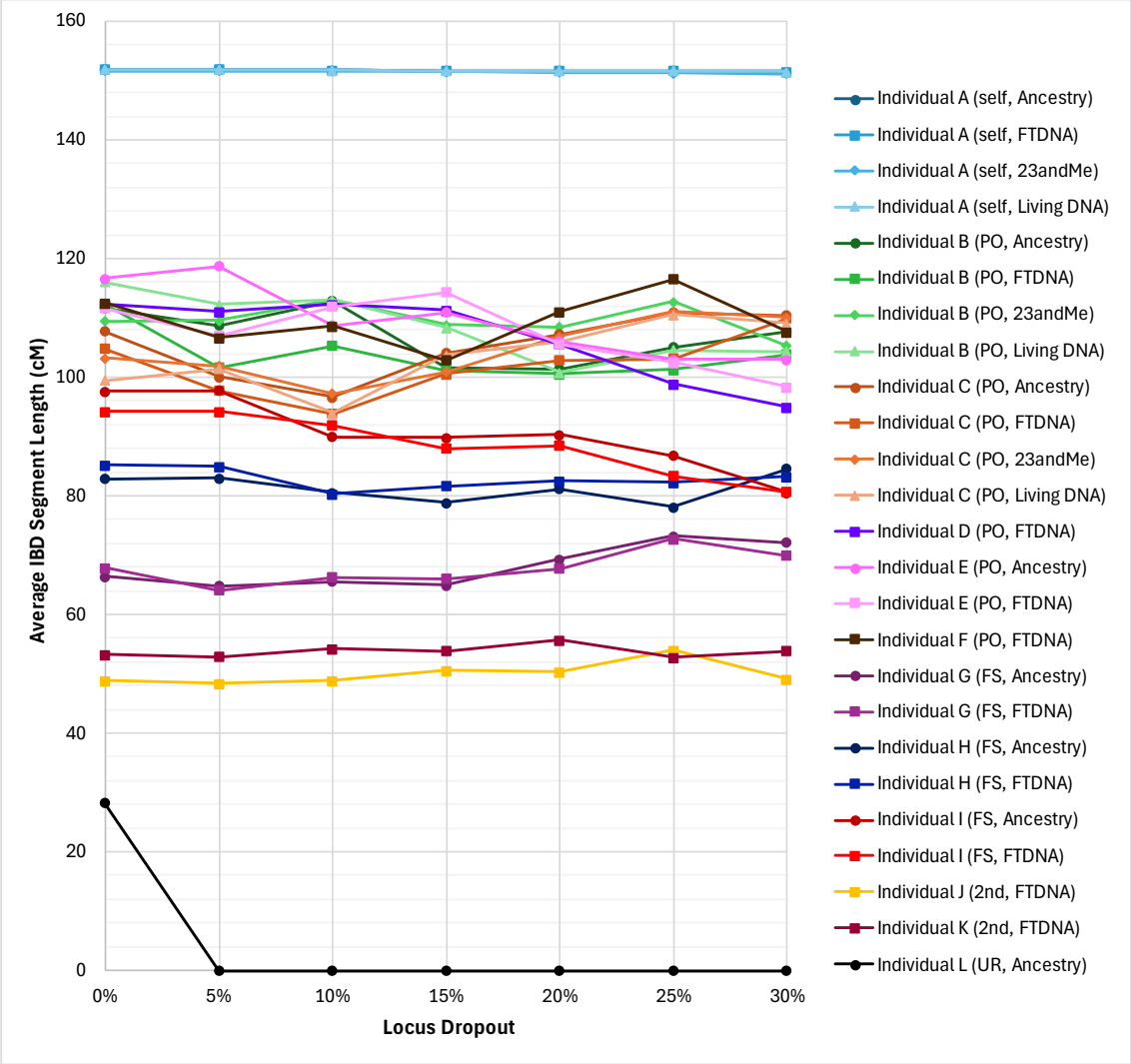


Figure S3. Average identical-by-descent (IBD) segment lengths in centimorgans (cM) shared between the Kintelligence profile of Individual A and direct-to-consumer profiles in Pedigree Group 2 with locus dropout ranging from 0 to 30%. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; UR: unrelated; FTDNA: FamilyTreeDNA.

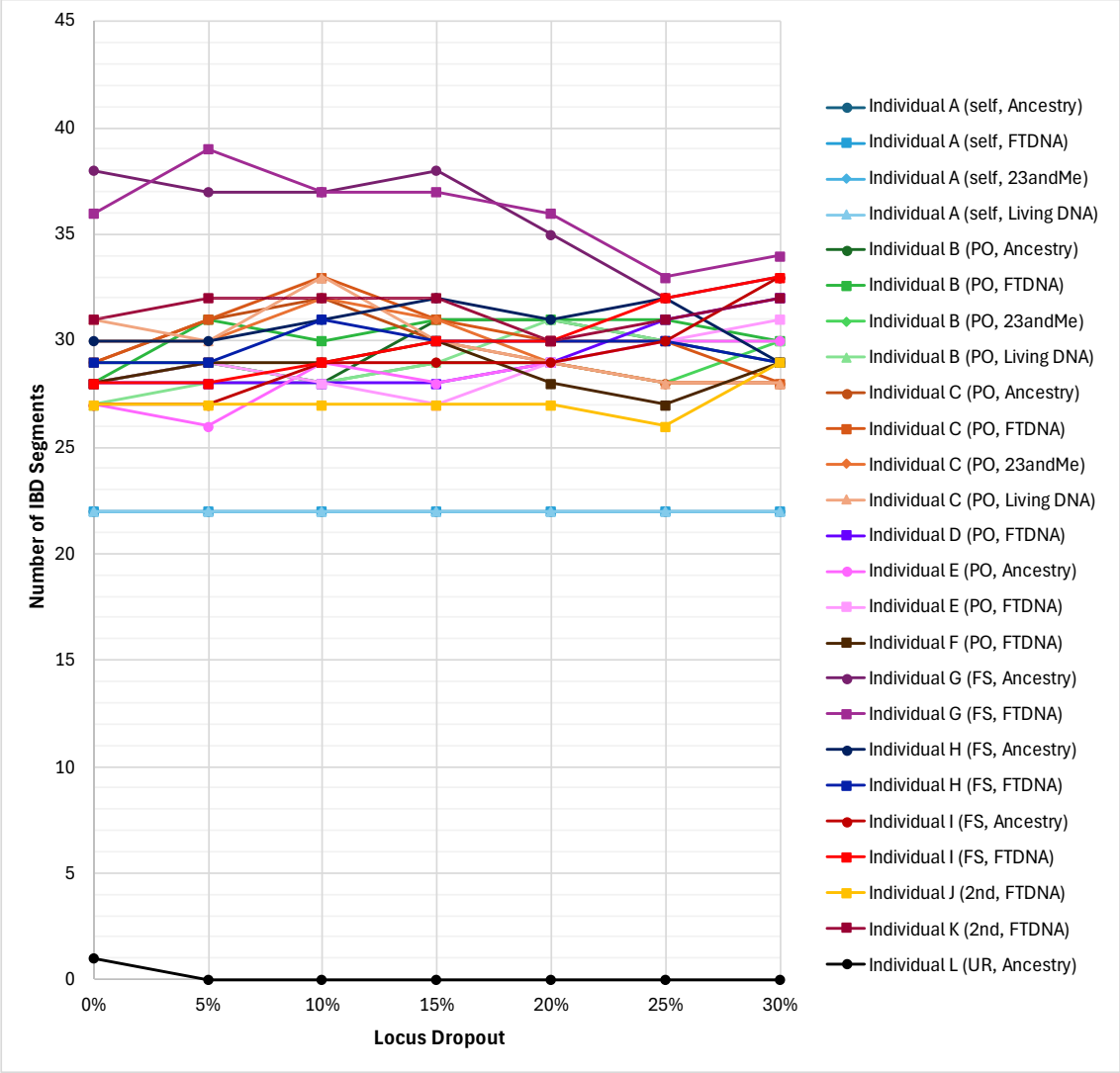


Figure S4. Number of identical-by-descent (IBD) segments detected by GEDmatch PRO™ shared between the Kintelligence profile of Individual A and direct-to-consumer profiles in Pedigree Group 2 with locus dropout ranging from 0 to 30%. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; UR: unrelated; FTDNA: FamilyTreeDNA.

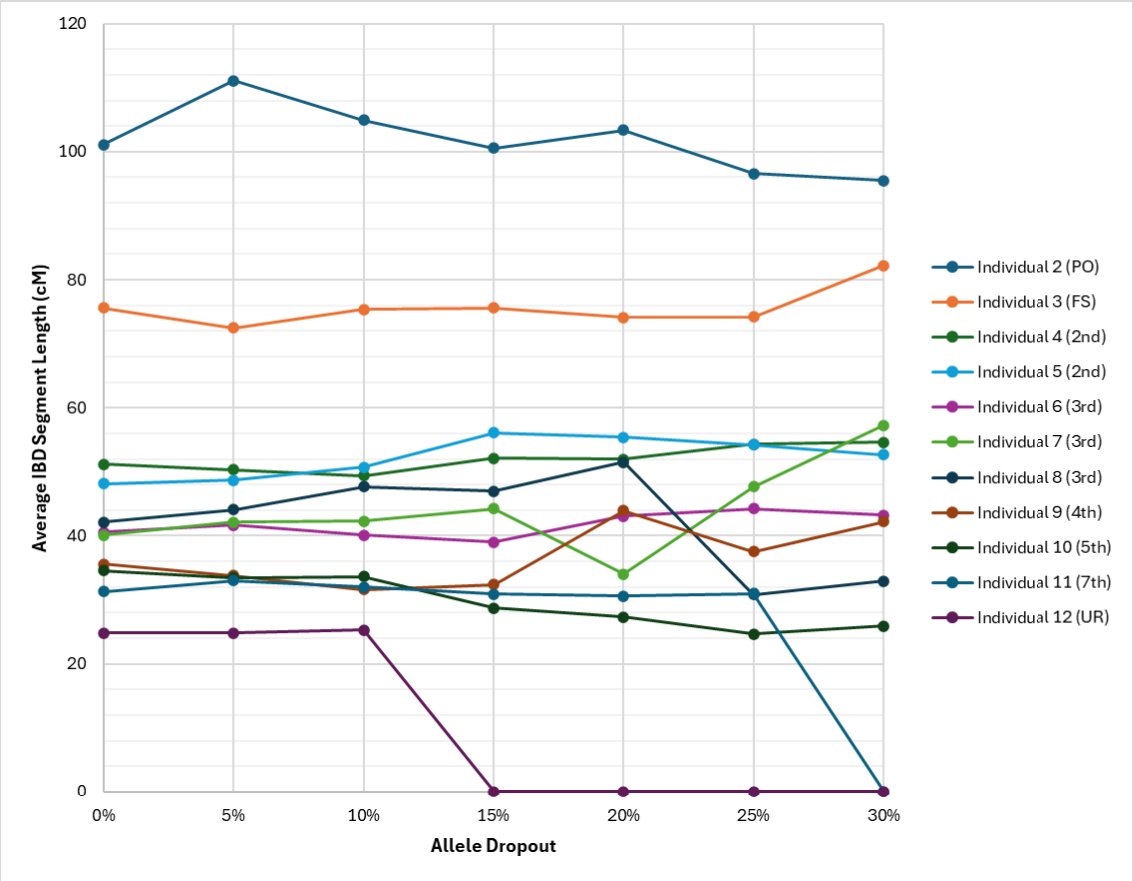


Figure S5. Average identical-by-descent (IBD) segment lengths in centimorgans (cM) shared between the Kintelligence profile of Individual 1 and other Kintelligence profiles in Pedigree Group 1 with allele dropout ranging from 0 to 30%. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; 3rd: third degree relative; 4th: fourth degree relative; 5th: fifth degree relative; 7th: seventh degree relative; UR: unrelated.

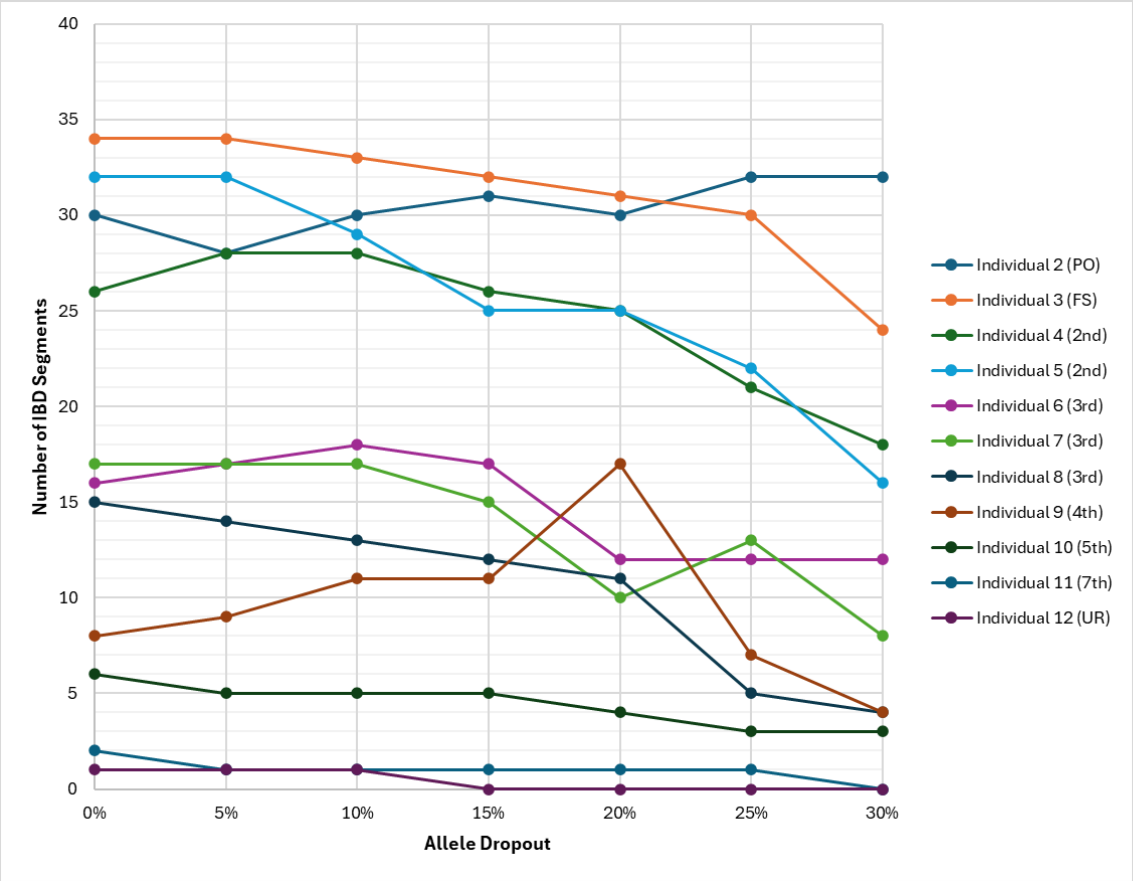


Figure S6. Number of identical-by-descent (IBD) segments detected by GEDmatch PRO™ shared between the Kintelligence profile of Individual 1 and other Kintelligence profiles in Pedigree Group 1 with allele dropout ranging from 0 to 30%. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; 3rd: third degree relative; 4th: fourth degree relative; 5th: fifth degree relative; 7th: seventh degree relative; UR: unrelated.

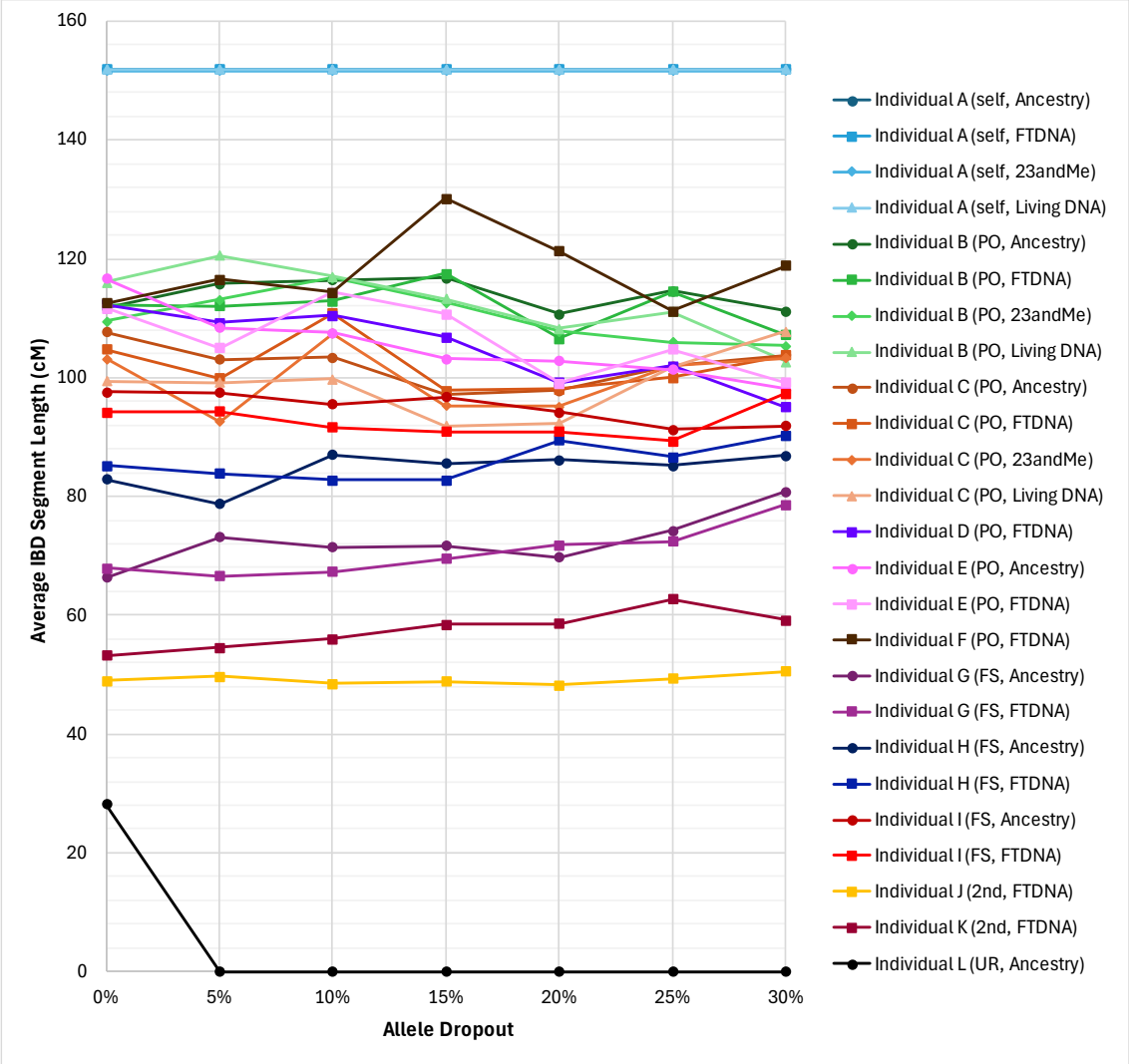


Figure S7. Average identical-by-descent (IBD) segment lengths in centimorgans (cM) shared between the Kintelligence profile of Individual A and direct-to-consumer profiles in Pedigree Group 2 with allele dropout ranging from 0 to 30%. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; UR: unrelated; FTDNA: FamilyTreeDNA.

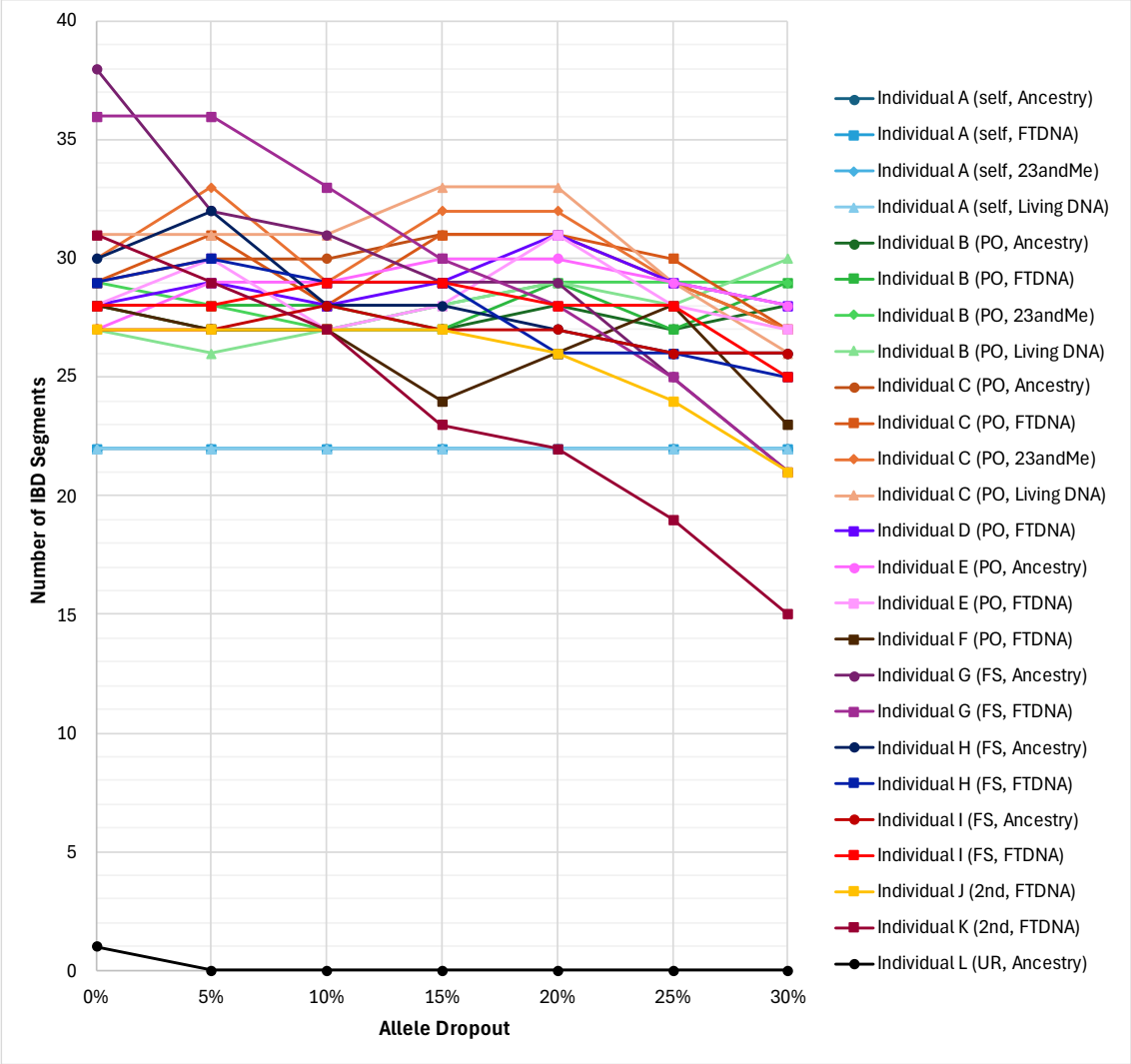


Figure S8. Number of identical-by-descent (IBD) segments detected by GEDmatch PRO™ shared between the Kintelligence profile of Individual A and direct-to-consumer profiles in Pedigree Group 2 with allele dropout ranging from 0 to 30%. PO: parent/offspring; FS: full sibling; 2nd: second degree relative; UR: unrelated; FTDNA: FamilyTreeDNA.

9. CHAPTER NINE

CONCLUSIONS

This research aimed to evaluate the emerging forensic genomics capabilities that could provide resolution or new investigative leads for UHR cases in Australia when routine forensic identification methods have been exhausted. Casework samples submitted to the AFP's DNA Program and casework-type samples obtained from AFTER allowed testing of authentic compromised samples typically encountered in UHR investigations to ensure the genotyping technologies would be robust for samples of varying quality and quantity. By studying reference-type samples with self-declared physical features, ancestry and relationships, the accuracy of genetic intelligence pipelines could be assessed and optimised with respect to an Australian population.

While WGS technology is not readily accessible to law enforcement forensic laboratories due to high costs and throughput requirements, TAS provides an effective solution using benchtop instruments and kits that target forensically relevant SNPs.[1] It has been demonstrated in this research that both the ForenSeq® Kintelligence Kit and FORCE panel produced genotypes that could be used for identification purposes, as well as inferring EVCs, estimating BGA and performing extended kinship analysis. Within the AFP DNA Program laboratory, the Kintelligence workflow produced higher quality SNP profiles than the FORCE panel for the range of compromised samples tested.

There have been several validations and evaluations of individual TAS panels that have been published, but few studies that have directly compared workflows.[2-5] McNevin et al. (2025) evaluated and compared the ForenSeq® Kintelligence Kit and various Ion AmpliSeq™ panels (Thermo Fisher Scientific), including the Precision ID Identity Panel, Precision ID Ancestry Panel, DNA Phenotyping Panel and HID Y-SNP Research Panel v1.[6] The study concluded that all of these panels were suitable for processing compromised samples encountered in UHR casework; however, the ForenSeq® Kintelligence Kit provided a streamlined process because all SNPs are incorporated into a single workflow. Additionally, the Ion AmpliSeq™ panels have not yet been evaluated for kinship applications (and are not likely to be suitable beyond first order kinship) whereas the ForenSeq® Kintelligence Kit enables extended kinship analysis.[7-9]

Since this research commenced, technology has undergone continual and rapid development to provide more sensitive and accurate sequencing results. To enhance applicability, small-density SNP panels designed for genetic intelligence (ForenSeq®

Imagen Kit) and higher throughput solutions of existing panels (ForenSeq® Kintelligence HT Kit) have been developed to allow laboratories to choose kits best suited to their casework and reporting requirements.[10-13] Furthermore, the MiSeq i100 and MiSeq i100 Plus Sequencing Systems were released in 2024 by Illumina 2024 and generate a higher sequencing output in less time than the MiSeq Sequencing System (Illumina).[14] While current applications have been limited to microbial, pathogen and cancer sequencing, future workflows could provide more powerful sequencing solutions for forensic casework.

Based on the conclusions of McNevin et al. (2025) and the research in this thesis, the ForenSeq® Kintelligence Kit was pursued in this research for implementation by the AFP DNA Program.[6] The internal validation allowed for the in-house, end-to-end workflow to be optimised and operationalised to support UHR and LTMP investigations with forensic genomics. As this panel combines multiple SNP classes for genetic identification and intelligence purposes, it has since been implemented in multiple private and law enforcement forensic laboratories in Australia, the US and Europe, with more undergoing evaluations or internal validations.[15-18] During the course of this PhD candidature, SNP profiles generated with the ForenSeq® Kintelligence Kit have provided investigative leads for multiple criminal and coronial cases that have resulted in identifications.[18, 19]

Optimisation of the library preparation throughput from three to 12 samples in this research occurred prior to the release of the ForenSeq® Kintelligence HT Kit which allows up to 96 libraries to be prepared simultaneously.[11, 12] While the sequencing throughput has been increased to allow up to 12 PM samples or 36 AM samples on a single flow cell, this decreases the coverage and, subsequently, the call rate. However, as the ForenSeq® Kintelligence HT Kit targets the same SNPs with the same library preparation chemistry as the ForenSeq® Kintelligence Kit, reducing the sequencing throughput will improve the number of SNPs called. This is important because a minimum of 70% of the available SNPs are required to be compatible for upload to GEDmatch PRO™ and to utilise their extended kinship analysis tools.

At the time this thesis was submitted, the UAS was only capable of determining if the ForenSeq® Kintelligence SNP profile is likely to be single-source or a mixture but had not been validated for deconvolution and mixture interpretation. Biological samples obtained during criminal investigations are more likely to be from more than one DNA contributor, leading to current research in how to apply probabilistic genotyping to deconvolute mixed SNP profiles. Mixture Deconvolution in R (MixDeR) is an open-source R package and

Shiny app under development to utilise EuroForMix to deconvolute two-person mixed ForenSeq® Kintelligence profiles.[20] With further research, probabilistic genotyping tools specifically designed for SNP deconvolution will enable TAS panels such as the ForenSeq® Kintelligence Kit to provide genetic intelligence support for criminal investigations.

Presentation of forensic evidence in criminal and/or coronial Courts requires reporting the LR to convey statistical weight, calculated using allele frequencies for each locus in the profile.[21] Within the ForenSeq® Kintelligence Kit, there are 94 iiSNPs that provide an alternative identity-informative marker type to STRs for the identification of an unknown individual. While iiSNPs are selected for their low population heterogeneity to minimise differences in allele population frequencies among different population groups, relevant allele population frequencies need to be used.[22, 23] This research presented the first characterisation of these iiSNPs using Australian-specific data and facilitated the establishment of a SNP allele frequency database for Australians with European ancestry. The CPM calculated for the 94 iiSNPs within this population was consistent with previous research assessing populations of individuals with European ancestry in the US and UK.[24, 25]

Additional research needs to be conducted to establish the allele population frequencies of other Australian population groups commonly utilised when reporting STRs, including individuals with Asian and Aboriginal and Torres Strait Islander ancestries.[26] This research has provided the framework for further investigation of iiSNPs for Australian casework applications and addresses some of the challenges faced with presentation of SNP evidence in Courts. SNP to SNP comparison for identification purposes was accepted in Court for the first time in California, US, in 2024. A pretrial admissibility hearing resulted in the Court deeming the MPS technology used for generating the ForenSeq® DNA Signature Prep Kit profile as reliable and that it had gained acceptance in the scientific community.[27, 28]

When an identification cannot be made through a direct comparison to an AM reference sample of a LTMP, characteristics of the unknown individual can still be inferred from their DNA to generate new investigative leads. In UHR casework, inferences of the biological sex, EVCs, BGA and genetic relatives of remains have been used to nominate candidates for missing persons to scientifically link using confirmatory testing. There are multiple algorithms available for generating DNA intelligence that should be evaluated for their suitability with respect to a jurisdiction’s population. However, there are several

factors that complicate EVC inferences such as incorrect declaration of phenotypes and environmental exposure resulting in hair colour changes.

Multiple algorithms are available for interpretating aiSNP genotypes to infer the BGA of an unknown individual. This research demonstrated that PCoA achieved the greatest success rates when compared to three other algorithms. PCoA produced BGA inference results that were consistent with self-declared data for the majority of volunteers and inferred one of the self-declared population groups for approximately half of individuals with admixture. Admixture is one of the greatest complicating factors in BGA inference and modern populations are likely to be admixed due to colonialism and migration.[29] Structure, which estimates the proportion of genetic contributions from ancestral contribution groups, was found to be more likely to infer admixture for non-admixed individuals. However, the majority of self-declared admixed individuals were inferred correctly. The proposed optimised pipeline utilises the high success rates from PCoA and the ability to infer admixture from Structure to improve the BGA inference pipeline with the same reference population database. As the application of DNA intelligence continues to be tested in forensic casework, pipelines must be further developed to refine the reporting and communication of inferences to law enforcement in order to minimise the risk of bias and misinterpretation of results.

The application of extended kinship analysis to generate investigative leads using possible genetic relatives of an unknown person on public genetic genealogy databases has significantly increased since the first case resolutions in 2019.[30-32] This technique has been applied to numerous cases to identify human remains, living 'Doe' individuals and suspects of serious crimes.[18, 33, 34] While IBD segment matching has been utilised for genetic genealogy research by the public, it is not suitable for identification and confirmatory testing using established forensic identification methods is required for this purpose.[35] In jurisdictions where short-range kinship analysis is permitted for familial or kinship searching of databases, a kinship LR is required.[36, 37]

There have been few studies comparing kinship LR calculations with IBD segment matching results to infer relatedness between genetically related individuals.[38] Several studies have confirmed that the ForenSeq® Kintelligence Kit can accurately detect relationships out to the fifth degree.[8, 9, 13, 39] The research presented in this thesis demonstrated that both kinship LR calculations and IBD segment matching pipelines are valuable and reliable tools for conducting medium-range kinship analysis using profiles generated with the ForenSeq® Kintelligence Kit. However, fewer studies have evaluated

the impact of information loss on the kinship algorithms. Specifically, simulated Kintelligence profiles have been tested to evaluate the impact of locus dropout on kinship LR calculations by Gettings et al. (2024) and both locus and allele dropout on IBD segment matching by Snedecor et al. (2022).[8, 13] In order to evaluate the combined robustness of extended kinship pipelines and the ForenSeq® Kintelligence Kit, this research conducted the first comprehensive evaluation of suboptimal Kintelligence profiles by simulating locus and allele dropout in profiles and compared to known first to eighth degree relatives. It demonstrated that the ForenSeq® Kintelligence Kit can produce powerful kinship inference results for first to fifth degree relationships using both the kinship LR and IBD segment matching approaches. By characterising the types of results obtained and comparing kinship LR calculations with IBD segment matching, this research will assist in presenting alternative extended kinship analysis pipelines as reliable methods for generating actionable investigative leads.

This research has demonstrated the applicability of emerging forensic genomics techniques to unidentified and missing persons investigations and provided new investigative avenues to scientifically link UHR to LTMP. Establishing a centralised and dedicated facility which offers standardised identification procedures alongside emerging forensic capabilities provides the most effective approach to advancing challenging investigations, identifying human remains and resolving LTMP cases.

The AFP DNA Program included established and emerging forensic capabilities, including the operationalisation of SNP genotyping techniques for generating DNA intelligence for law enforcement agencies. From January 2021 to March 2024, 99 UHR cases were submitted to the AFP DNA Program and a range of forensic techniques were used, resulting in 17 identifications of coronial significance and 32 UHR deemed to be of no coronial significance (e.g. ancestral Aboriginal or historical remains).[40]

As this research was supported by, and contributed to, the AFP DNA Program, several of the capabilities developed and validated as a component of this PhD research have now been applied to active UHR investigations in Australia. Since the operationalisation of the ForenSeq® Kintelligence Kit workflow in September 2023, the AFP has applied this capability to 32 UHR cases.[41] Of these, two case relied on BGA inferences to resolve the ancestry of historical remains and 30 cases underwent FIGG by utilising law enforcement accessible genetic genealogy databases. At the time this thesis was submitted, six candidates had been nominated to local law enforcement agencies for further investigation. The ForenSeq® Kintelligence Kit provides law enforcement with an

in-house, end-to-end pipeline that can be used to generate multiple types of DNA intelligence and assist in resolving UHR cases within Australia and in other jurisdictions around the world.

9.1 References

1. Kling D, Phillips C, Kennett D, Tillmar A. Investigative genetic genealogy: Current methods, knowledge and practice. Forensic Sci Int Genet. 2021;52:102474. doi: 10.1016/j.fsigen.2021.102474.
2. Tillmar A, Sturk-Andreaggi K, Daniels-Higginbotham J, Thomas JT, Marshall C. The FORCE panel: An all-in-one SNP marker set for confirming investigating genetic genealogy leads and for general forensic applications. bioRxiv. 2021. doi: 10.1101/2021.11.30.470354.
3. Staadig A, Hedman J, Tillmar A. Applying Unique Molecular Indices with an Extensive All-in-One Forensic SNP Panel for Improved Genotype Accuracy and Sensitivity. Genes. 2023;14(4):818. doi: 10.3390/genes14040818.
4. Churchill JD, Schmedes SE, King JL, Budowle B. Evaluation of the Illumina ® Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling. Forensic Sci Int Genet. 2016;20:20-9. doi: 10.1016/j.fsigen.2015.09.009.
5. Peck M, Idrizbegovic S, Bittner F, Parsons TJ. Optimization and performance of a very large MPS SNP panel for missing persons. 7th QIAGEN Investigator Forum: International Commission on Missing Persons; 2018.
6. McNevin D, Watson J, Grisedale K, Dahal A, Goodwin C, Ward J. Comparison of commercial targeted amplicon sequencing assays for human remains identification casework. Int J Legal Med. 2025;139:49-60. doi: 10.1007/s00414-024-03335-8.
7. Antunes J, Walichiewicz P, Forouzmand E, Barta R, Didier M, Han Y, et al. Developmental Validation of the ForenSeq® Kintelligence Kit, MiSeq Fgx® Sequencing System and ForenSeq Universal Analysis Software. Forensic Sci Int Genet. 2024;71:103055. doi: 10.1016/j.fsigen.2024.103055.
8. Snedecor J, Fennell T, Stadick S, Homer N, Antunes J, Stephens K, et al. Fast and accurate kinship estimation using sparse SNPs in relatively large database searches. Forensic Sci Int Genet. 2022;61:102769. doi: 10.1016/j.fsigen.2022.102769.

9. Daniel R, Raymond J, Sears A, Stock A, Scudder N, Padmabandu G, et al. It's all relative: A multi-generational study using ForenSeq™ Kintelligence. Forensic Sci Int. 2024;364:112208. doi: 10.1016/j.forsciint.2024.112208.

10. Verogen. ForenSeq Imagen Kit. 2022.

11. QIAGEN. ForenSeq® Kintelligence HT Kit Technical Information 2024. Available from: https://www.qiagen.com/br/resources/download.aspx?id=63849dbe-ab1f-4dbe-968f-e468055637a6&lang=en.

12. QIAGEN. Verogen ForenSeq Kintelligence HT Kit 2024. [Last accessed: 15 August]. Available from: https://www.qiagen.com/us/products/human-id-and-forensics/nextgeneration-sequencing/verogen-forenseq-kintelligence-ht-kit.

13. Gettings KB, Tillmar A, Sturk-Andreaggi K, Marshall C. Review of SNP assays for disaster victim identification: Cost, time, and performance information for decision-makers. J Forensic Sci. 2024;69(5):1546-57. doi: 10.1111/1556-4029.15585.

14. Illumina. MiSeq i100 Series 2025. [Last accessed: 10 February]. Available from: https://www.illumina.com/systems/sequencing-platforms/miseq-i100.html.

15. Signature Science. SigSci's Center for Advanced Genomics Validating the Verogen ForenSeq® Kintelligence Workflow for Forensic Genetic Genealogy Austin, TX2021. [Last accessed: 31/03/2022]. Available from: https://www.signaturescience.com/press/ngs_fgg/.

16. DNA Labs International. Genealogy 2021. [Last accessed: 31/03/2022]. Available from: https://dnalabsinternational.com/genealogy/.

17. Stoljarova-Bibb M, Järving R, Kohtla M, Puss F, Sajantila A, Sadam M. Getting intelligence with Kintelligence: analysis of known pedigrees and progress report on ongoing cases [oral presentation]. Congress of the International Society for Forensic Genetics. 2024.

18. Ward J, Coakley L, Grisedale K, Seddon S, Spiden M, Watson JL, et al. Operationalisation of the National DNA Program for Unidentified and Missing Persons' forensic investigative genetic genealogy capability for human remains identification in Australia [in press, accepted 27 January 2024]. Forensic Genom. 2024.

19. QIAGEN. DNA Labs International solves significant cold cases with QIAGEN workflow for forensic genetic genealogy [press release] 2023. [Last accessed: 19

November]. Available from: https://corporate.qiagen.com/English/newsroom/press-releases/press-release-details/2023/DNA-Labs-International-solves-significant-cold-cases-with-QIAGEN-workflow-for-forensic-genetic-genealogy/default.aspx.

20. Mitchell R, Peck M, Gorden E, Just R. MixDeR: a SNP mixture deconvolution workflow for forensic genetic genealogy. Forensic Sci Int Genet. 2025:103224. doi: 10.1016/j.fsigen.2025.103224.

21. Scientific Working Group on DNA Analysis Methods. Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Resulted Reported as Likelihood Ratios. 2016.

22. Wei Y-L, Li C-X, Jia J, Hu L, Liu Y. Forensic Identification Using a Multiplex Assay of 47 SNPs*. J Forensic Sci. 2012;57(6):1448-56. doi: 10.1111/j.1556-4029.2012.02154.x.

23. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SLB, Karoma NJ, et al. Developing a SNP panel for forensic identification of individuals. Forensic Sci Int. 2006;164(1):20-32. doi: 10.1016/j.forsciint.2005.11.017.

24. Davenport L, Devesse L, Court DS, Ballard D. Forensic identity SNPs: Characterisation of flanking region variation using massively parallel sequencing. Forensic Sci Int Genet. 2023;64:102847. doi: 10.1016/j.fsigen.2023.102847.

25. Kiesler KM, Borsuk LA, Steffen CR, Vallone PM, Gettings KB. US Population Data for 94 Identity-Informative SNP Loci. Genes. 2023;14:1071. doi: 10.3390/genes14051071.

26. Taylor D, Bright J-A, McGovern C, Neville S, Grover D. Allele frequency database for GlobalFiler™ STR loci in Australian and New Zealand populations. Forensic Sci Int Genet. 2017;28:e38-e40. doi: 10.1016/j.fsigen.2017.02.012.

27. QIAGEN. First US admissibility hearing using next-generation sequencing leading to the conviction in a double homicide 2024. [Last accessed: 10 February]. Available from: https://www.qiagen.com/us/knowledge-and-support/knowledge-hub/science-matters/human-id-and-forensics/first-us-admissibility-hearing-using-ngs-leads-to-conviction-in-a-double-homicide.

28. Taylor M. Next Generation Sequencing Accepted in Court for First Time: Forensic On the Scene and In the Lab; 2024. [Last accessed: 10 February]. Available from:

https://www.forensicmag.com/610513-Next-Generation-Sequencing-Deemed-Admissible-in-Court-for-First-Time/#:~:text=This%20case%20officially%20marks%20the,acceptance%20in%20the%20scientific%20field.

29. Cheung EY, Gahan ME, McNevin D. Prediction of biogeographical ancestry in admixed individuals. Forensic Sci Int Genet. 2018;36:104-11.

30. Arango T. The Cold Case That Inspired the 'Golden State Killer Detective to Try Genealogy. The New York Times. 2018 May 4, 2018;Sect. A.

31. Edge M, Coop G. How lucky was the genetic investigation in the Golden State Killer case? 2019.

32. Augenstein S. 'Buck Skin Girl' case break is success of new DNA Doe Project. Forensic Mag. 2018.

33. Forensic Genetic Genealogy Project v. 2022 [Internet]. Mendeley Data. 2023 [cited 17 August 2023]. Available from: data.mendeley.com/datasets/jcycgvhm96/1.

34. Dowdeswell TL. Forensic genetic genealogy: A profile of cases solved. Forensic Sci Int Genet. 2022;58:102679. doi: 10.1016/j.fsigen.2022.102679.

35. Department of Justice. Forensic Genetic Genealogical DNA Analysis and Searching. Interim Policy. 2019.

36. Scientific Working Group on DNA Analysis Methods. Overview of investigative genetic genealogy. SWGDAM Communications. 2020.

37. Scientific Working Group on DNA Analysis Methods. Recommendations from the SWGDAM Ad Hoc Working Group on Familial Searching. SWGDAM Communications. 2013.

38. Kling D, Tillmar A. Forensic genealogy—A comparison of methods to infer distant relationships based on dense SNP data. Forensic Sci Int Genet. 2019;42:113-24. doi: 10.1016/j.fsigen.2019.06.019.

39. Peck MA, Koeppel AF, Gorden EM, Bouchet JL, Heaton MC, Russell DA, et al. Internal Validation of the ForenSeq Kintelligence Kit for Application to Forensic Genetic Genealogy. Forensic Genom. 2022. doi: 10.1089/forensic.2022.0014.

40. Ward J. The operationalization of a multifaceted, multijurisdictional, multiagency and multidisciplinary capability for unidentified and missing persons investigations

in Australia. Aust J Forensic Sci. 2024;56(sup1):65-70. doi: 10.1080/00450618.2024.2324737.

41. Grisedale K. Operationalisation of a Forensic Investigative Genetic Genealogy Workflow for Unidentified and Missing Persons Casework: Consideration for Validation, Implementation, and Sustainability [oral presentation]. Forensic Investigative Genetic Genealogy Summit. 2024.

10. APPENDICES

Appendix 1. Ethics approval from UTS HREC under project number ETH21-5821

Human Ethics Application

Application ID :	ETH21-5821
Application Title :	Genetic identification of unidentified human remains
Date of Submission :	31/03/2021
Primary Investigator :	Prof Dennis Blair McNevin (Chief Investigator)
Other Personnel :	A/Prof Jodie Ward (Chief Investigator)
	Miss Jessica Lee Watson (5Research Student)

Section 1: Ethics Portal

Select your application type

What type of application are you looking for?
Please **do not** change your application type without first consulting with the Ethics Secretariat (9514 9772).*

- (●) New application (including scope-checking for nil/negligible risk research)
- ( ) Ratification of existing approval
- ( ) Transfer of existing approval
- ( ) Evaluation of teaching and learning activities
- ( ) Amendment to existing approval
- ( ) Program approval

You have selected "new application (including scope checking for nil/negligible risk research)". This option allows you to create a new form. The system will check if your application can be approved by the Faculty or whether it requires full ethics approval by the HREC. Please click "save" before continuing.

What should I know before I start?

Would you like more information on:

- ( ) This system
- ( ) The ethics process
- (●) Purpose of the ethics review process

Purpose of the Human Research Ethics Review Process

The ethical review process is valuable as it:

- Provides the opportunity to reflect on the research methodology;
- Ensures the research integrity of the work and
- Increases the trust levels of the research participants and the publishers about the output of the research project.

The online ethics application simplifies the process:

- The form is intuitive so you won't need to answer every question just those that are applicable to your research
- Some parts of the form are auto-populated
- You can save the form and return to it later to complete and submit it
- Guidelines and instructions are incorporated into the form
- Reduced paper handling.

Importance of ethics procedures
Regardless of the level of risk, all staff and students are expected to abide by the standards outlined in the National Statement, the Australian Code, and guidelines established by the UTS Human Research Ethics Committee (HREC). It is the responsibility of researchers, both staff and students, to familiarise themselves with these.

Section 1A: Risk evaluation

Risk A

Determining the level of risk and review

- Please answer each question carefully **and consecutively**.
- For assistance with answering these questions please refer to the National Statement on Ethical Conduct in Human Research as per the chapters listed below.
- If you need to contact the Research Ethics Officer you can call (02) 9514 9772
- Click on the help buttons (?) for more information
- You can save your application at any time by clicking on the save button on the left hand side in the toolbar. For further information and help in completing your application go to our website.

Does your research involve:

Projects involving covert observation, active concealment, or planned deception of participants

e.g. covert observation of the hand-washing behaviour of hospital employees, undisclosed role-playing by a researcher, etc. Does NOT include observation in a public place WITHOUT the use of photographs, images, video or audio footage (Chapter 2.3, p.19)

*

- ( ) Yes
- (●) No

Targeted recruitment or analysis of data(?) from any of the groups listed below (or where any of these groups are likely to be significantly over-represented in the group being studied)

- Women who are pregnant and the human fetus (Chapter 4.1, p. 61)
- Children and young people (under 18 years) (Chapter 4.2, p. 65)
- People in dependent or unequal relationships (e.g. lecturer/student [except T&L], doctor/patient, employer/employee) (Chapter 4.3, p.68)
- People highly dependent on medical care who may be unable to give consent Chapter 4.4, p.68)
- People with a cognitive impairment, an intellectual disability, or a mental illness (may include the disadvantaged/homeless) (Chapter 4.5, p. 70)
- People who may be involved in illegal activities (including those affected e.g. victims of domestic violence) (Chapter 4.6, p.73)
- Aboriginal and Torres Strait Islander Peoples (Chapter 4.7, p.77)

*

○ Yes
◉ No

Targeted recruitment of people in / from countries that score <50 on the Corruption Perception Index (CPI) (check here)
This includes any cohorts from these countries, i.e. it is not restricted to marginalised groups within these countries*

○ Yes
◉ No

Collection, use or disclosure of personal information without consent of the participant(?)

- **a record which may include your name, address and other details about the participant (e.g. date of birth, financial information etc.)**
- **photographs, images, video or audio footage**
- **fingerprints, blood or DNA samples**

*

○ Yes
◉ No

Collection, use or disclosure of health information(?)

- personal information that is information or an opinion about
  - the physical or mental health or a disability (at any time) of an individual; or
  - an individual's expressed wishes about the future provision of health services to him or her, or
  - a health service provided, or to be provided, to an individual or
- other personal information collected to provide, or in providing, a health service, or
- other personal information about an individual collected in connection with the donation, or intended donation, of a individual's body parts, organs, body substances, or
- other personal information that is genetic information about an individual arising from a health service provided to the individual in a form that is or could be predictive of the health (at any time) of the individual or of a genetic relative of the individual, or
- healthcare identifiers

N.B Includes information collected through physiological testing or assessment. Examples include but are not limited to EEG, EMG, BMI, blood pressure, DEXA, etc.*

◉ Yes
○ No

Collection, use or disclosure of sensitive information
Racial, ethnic information, political, religious and philosophical beliefs, sexual activity or identity, and trade union membership
*

◉ Yes
○ No

Activity that potentially infringes the privacy or professional reputation of participants, providers or organisations
e.g. observation in the workplace, collection of commercially confidential information, etc.
Commercially confidential information = Any information which is not in the public domain or publicly available, and where disclosure may undermine the economic interest or competitive position of the owner of the information (TGA adopted definition from European Medicines Agency (EMA)).

N.B. if canvassing opinion via consensus methods i.e. Delphi (?), answer "No" here
*

○ Yes
◉ No

Establishment of a register or databank of identifiable data for possible use in future research projects (Chapter 3.2, p.27) (?)

*

◉ Yes
○ No

Collection, transfer(?) and/or banking of human biospecimens.
e.g. tissue, blood, urine, sputum etc.(?)

*

◉ Yes
○ No

Any significant alteration to routine care or service provided to participants
e.g. deviation from standard care or usual practice
*

○ Yes
◉ No

Prospective assignment of human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes(?) (Chapter 3.14-3.17) *

○ Yes
◉ No

Potential for participants to experience harm (i.e. anything more than discomfort)(?)
e.g. physical, psychological, devaluation of personal worth, social, economic and/or legal (Chapter 2.1, p.12)
*

○ Yes
◉ No

High Risk

Section 2: Project information

Project title

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Application ID (automatically generated):

ETH21-5821

Application Title:*

Genetic identification of unidentified human remains

Please note that the HREC is now granting a standard approval period for the research proposals.
The approval period for your project will be specified in your approval letter.
Please also note that research should not commence until ethics approval has been granted. The Committee cannot grant retrospective approval for data that has already been collected.

Ethics category code (automatically selected):*

Human

Is this a resubmission of a previous application?*

◉ Yes
○ No

Is this a pilot study? *

○ Yes
◉ No

Has a pilot study been conducted as part of this project? *

○ Yes
◉ No

Please save and continue to the next page

Consultation

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Have you undertaken any consultation in preparing this application?*

◉ Yes
○ No

Please describe*

The project will fall under the National DNA Program for Unidentified and Missing Persons, situated within the Australian Federal Police. The project has been discussed extensively with key stakeholders within the AFP including the Chief Forensic Scientist (Dr Sarah Benson), the Biometrics Coordinator (Rachael Kennedy), the Lead Scientist of the Massively Parallel Sequencing Project (Dr Paul Roffey), the Research and Innovation Coordinator (Dr Nathan Scudder) and the National Missing Persons Coordination Centre Coordinator (Jodie McEwan).

Please save and continue to the next page

Section 3: Personnel

Investigators

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Are there external investigators or personnel listed on this protocol?*

○ Yes
◉ No

Is this application for a student project?*

○ Yes
◉ No

Positions in the personnel table

Position type:	In the personnel table use the following positions from the drop-down list:
Chief Investigator/Supervisor	1Chief Investigator (students must **not** be listed as Chief Investigator)
Co Investigator	3Assoc. Investigator
Co Supervisor	Co-Supervisor
Research Student	5Research Student
Project Administrator	7Project Administrator

Note: Further options are available in the drop down list.

Instructions on how to add a person to the personnel table:

1. Click on "Add"
2. Start typing the details (first name, last name or Staff ID) in the search bar.
3. Click on "Add selected"
4. The extra information panel will open, select their position from the drop-down list. If they are the primary contact (e.g. Chief Investigator/Supervisor), tick "Yes" under 'Primary contact' and then select "OK"

- **Student research:** Students must add their supervisors to their application and must mark their primary supervisor as a Chief Investigator and as a primary contact. Students must be listed as "5Research student" under the column 'Position' to ensure the application is properly submitted to their supervisor.
- **Ratifications/Transfers:** If this list differs from that of the original application, you must provide evidence that any additional investigators have been added via amendment to the lead/external HREC [attach relevant amendments and evidence of approval].

Internal personnel listed on this ethics protocol:*

1	Primary	Yes
	ID	
	Surname	McNevin
	Given Name	Dennis
	Full Name	Prof Dennis Blair McNevin
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Dennis.McNevin@uts.edu.au
	Work Number	
2	Primary	No
	ID	
	Surname	Ward
	Given Name	Jodie
	Full Name	A/Prof Jodie Ward
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Jodie.Ward@uts.edu.au
	Work Number	
3	Primary	
	ID	
	Surname	Watson
	Given Name	Jessica
	Full Name	Miss Jessica Lee Watson
	Position	5Research Student
	Type	Internal
	AOU	
	Managing Unit	
	Email Address	Jessica.Watson@uts.edu.au
	Work Number	

If any details are incorrect or missing please contact the Ethics Secretariat on (02) 9514 9772 or by email.

Please provide additional (or preferred) contact details of any of the people listed on the project if necessary (4000 character limit)

This question is not answered.

Please provide details of any formal qualifications (REF NS 1.1(e)) of each person listed on the project (4000 character limit)*

Dennis McNevin
PhD in Biochemical Engineering (University of Sydney) Graduate Diploma of Education (University of Western Sydney)
Bachelor of Engineering with Honours Class 1 (University of NSW)

Jodie Ward
PhD in Forensic Molecular Biology (Australian National University) Graduate Certificate in Management (Charles Sturt University) Graduate Certificate in Higher Education (University of Canberra)
Bachelor of Science with Honours Class 1 (Australian National University)

Jessica Watson
Bachelor of Forensic Science with Honours Class 1 and University Medal (University of Technology Sydney)

Please outline the experience of each person listed on this project relevant to this application (4000 character limit)*

Dennis McNevin is Professor of Forensic Genetics in the Centre for Forensic Science (CFS) at UTS. He has also been an academic at the University of Sydney, the Australian National University and the University of Canberra. Dennis has published widely in the forensic genetics literature and has established an international reputation for innovation in the use of massively parallel sequencing (MPS) for forensic genetics, the use of preservatives for room temperature storage of DNA and extraction of DNA from difficult substrates such as hair and bone. He supervises a range of higher degree by research students, graduated 11 PhD students (five as primary supervisor) and 33 Honours students (25 as primary supervisor). His graduates have found employment in the Australian Federal Police (AFP) graduate program as well as in AFP Forensics, NSW Forensic and Analytical Science Service (FASS) and Victoria Police Forensic Services Department (FSD). He has held two ARC Linkage Grants and currently manages the program wide ethics for the CFS.

Jodie Ward is the Program Lead of the Australian Federal Police (AFP) National DNA Program for Unidentified and Missing Persons. In addition, she is Director of the Australian Facility for Taphonomic Experimental Research (AFTER) and an Associate Professor in the Centre for Forensic Science at UTS. This unique Joint Appointment sees her lead the research, development and application of DNA based human identification techniques for missing persons casework in Australia. For the last decade, Jodie has focused on acquiring specialised skills in the recovery, analysis and interpretation of DNA from compromised and skeletonized human remains. Previously, she has held operational forensic DNA specialist roles with the AFP and NSW Police Force, and academic roles with the National Centre for Forensic Studies in Australia. Jodie manages the program wide ethics for AFTER, the National DNA Program's research strategy and a number of PhD and Honours students.

Jessica Watson is the PhD student associated with this project. She completed her Honours on "DNA identification of compromised human remains". She is co-author of a paper on "An in-field evaluation of rapid DNA technology for disaster victim identification", submitted to the International Journal of Legal Medicine.

Primary AOU*

SCI.School of Mathematical and Physical Sciences

Managing Unit

Science

Please save and continue to the next page

Section 4: Funding

Funding details

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Have you received funding in relation to this research?*

- (●) Yes
- ( ) No

Do you have a RM 'My Proposals' ID number?*

- ( ) Yes
- (●) No

List the source of funding (e.g. funding body / type)
(REF NS Page 8 "When is ethical review needed?", 2.2.6(h), 3.3.5(a), 3.3.18(b), 4.8.6, 5.2.7, and 5.7)
(2000 character limit)*

Australian Federal Police (AFP) National DNA Program for Unidentified and Missing Persons and an AFP Innovation Grant

Total amount of funding obtained, including in-kind contribution (please indicate which is applicable)
*

$50,000

What is your relationship to the funding source? (e.g. grant recipient, industry partner, contractor, employee, office-bearer, personal, other)*

The Australian Federal Police (AFP) is both an industry partner and employer for the chief investigators.

A/Prof Ward holds the following joint appointments:
* Director, Australian Facility for Taphonomic Experimental Research, School of Mathematical & Physical Sciences, Faculty of Science, UTS
* Program Lead, National DNA Program for Unidentified and Missing Persons, Specialist Protective Command, AFP

Professor McNevin is seconded at 0.5 FTE by the AFP for the National DNA Program for Unidentified and Missing Persons, Specialist Protective Command, AFP, to lead the research and development of new genetic technologies for Program use.

Please save and continue to the next page

Funding continued

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Is there any potential conflict of interest for you as a researcher because of the funding or commercial arrangements?*

○ Yes
◉ No

Are there any constraints on the research as a result of the funding arrangements, e.g. to intellectual property, publication, etc? (Section 4, The Code)*

◉ Yes
○ No

Please provide details of any constraints on the research as a result of the funding arrangements*

The AFP reserves the right to review and approve any research articles prior to publication.

Please save and continue to the next page

Section 5: Methodology

Description

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

The purpose of this section is to place your research in context for the HREC and demonstrate your ability to conduct the research. The HREC may only approve research which is methodologically sound. Remember to use simple language that can be understood by people from a variety of backgrounds. Avoid jargon and acronyms.

What are the hypotheses/goals/aims/objectives of your research? Please include a brief description using plain English explaining your research aims (approximately 100 words) (4000 character limit)*

The aim of the research is to evaluate and apply new genetic technologies to identify unidentified human remains as part of the Australian Federal Police (AFP) National DNA Program for Unidentified and Missing Persons.

Note: Clinical Trials, Recruitment of Participants and Data Collection are dealt with later in the application so you do not need to describe them in detail below

Please provide a brief description of the research design including research questions and proposed methods for conducting the research (approximately 250 words) (4000 character limit)*

We aim to understand the relationship between genotype and (a) biogeographical ancestry (BGA), (b) externally visible characteristics (EVCs) and (c) familial relationships in order to provide information for future missing persons investigations. The project will involve application of the following technologies to the identification of unidentified human remains:
* Advanced DNA extraction methods
* Whole mitochondrial genome, Y chromosome and X chromosome sequencing for lineagebased searching
* Autosomal DNA sequencing for short range familial searching, and prediction of BGA and EVCs
* Whole genome and/or medium-high density kinship informative SNP genotyping for long range familial searching (forensic genetic genealogy)

We propose to test the ability of genetic technologies to predict the (a) BGA, (b) EVCs and (c) familial relationships of fully informed and consenting participants. This information will be obtained from questionnaires and then predicted from genotypes derived from the DNA voluntarily donated by the participants. The predictions will be compared with the questionnaire data. We will also test the abilities of these genetic technologies to be applied to decomposed and skeletonised human remains from the Australian Facility for Taphonomic Experimental Research (AFTER) for which a separate program wide ethics approval already exists (ETH182999). Human remains present unique challenges for genotyping because the DNA is often of a low quality and quantity. It is important to evaluate, and subsequently optimise, these new technologies with authentic forensic samples such as bones to ensure that genotypes can be obtained from these types of samples for future missing persons investigations. Finally, the optimised and validated genetic technologies will be applied to unidentified human remains sourced (with permission) from police and coronial agencies of Australia's States and Territories as part of the National DNA Program for Unidentified and Missing Persons.

What do you hope the outcome(s) of this research will be? (4000 character limit)*

The outcomes of this research will be validated genetic methods for the identification of Australia's unidentified human remains.

Who do you think will benefit from this research? (4000 character limit)*

This research will benefit the relatives of deceased missing persons who will be given definitive knowledge of the fate of their missing relatives. It will also benefit the coronial systems of Australia's States and Territories which may have many unsolved missing persons cases resolved.

Please provide a brief description of the significance of your research (approximately 100 words) (4000 character limit)*

Many Australian families live in a perpetual trauma because they are unaware of the fate of a missing relative. The extent of the problem is such that the Australian Federal Police (AFP) have instigated a National DNA Program for Unidentified and Missing Persons. Some of these families will never see their relative again but they may be able to have the fate of their relative confirmed, if they are deceased. A very small but nevertheless important consolation for these families is to at least have certainty.

DNA is considered to be the gold standard for forensic identification. For many missing persons cases involving unidentified human remains, DNA is difficult to obtain because of the highly degraded nature of the remains. By using advanced DNA extraction methods (for DNA extraction from bones and teeth, for example), DNA may be recovered in cases where there has been no success using standard methods in the past.

Once DNA has been extracted, a DNA profile has traditionally been generated using polymerase chain reaction (PCR) followed by capillary electrophoresis for fragment length analysis of short tandem repeats (STRs) which are short, repetitive lengths of DNA. In this project, application of massively parallel sequencing (MPS) to produce DNA sequences should provide greater sensitivity for profiling low abundance, degraded DNA.

DNA sequences will then be used to identify unidentified human remains by:
* Comparing whole mitochondrial genome sequences with putative maternal relatives
* Comparing Y chromosome sequences with putative male paternal relatives
* Comparing autosomal DNA and X chromosome sequences with putative relatives
* Uploading medium and high density SNP genotypes to genealogy databases to identify putative relatives

It is estimated that only 2% of the population need to be represented in genealogy databases in order to identify at least one third cousin for any suitable unknown SNP genotype. This global reach offers the potential to identify the hundreds of missing persons that cannot be identified using the Australian law enforcement DNA database and that remain as cold cases in Australia's coronial system.

Please save and continue to the next page

Literature review & references

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Please give a brief literature review. The aim is to explain how your research fits into the context of other research in the area (REF NS 1.1(c)) (4000 character limit with spaces)
Please note that you cannot paste links into the online form
*

Capillary electrophoresis (CE) has been the method of choice for the detection of fluorescently labelled forensic DNA markers for over 15 years [1,2]. It has been applied to short tandem repeat (STR) genotyping and single nucleotide polymorphism (SNP) detection [3,4]. CE is limited by the number of markers that can be reliably multiplexed within a single PCR assay as well as available fluorophores with nonoverlapping emission spectra to distinguish multiple amplicons in a single sample. Although CE is reliable and has revolutionised DNA analysis of evidential samples, higher throughput platforms are now available and are capable of detecting hundreds to thousands of markers simultaneously without depleting large quantities of the DNA extract [57].

Forensic STR profiles are useful for identification if there is a reference profile available for comparison. If there is no such reference, we can obtain other useful information. DNA obtained from biological evidence is no longer limited to STR analysis for forensic identity and is being used to infer biogeographical ancestry (BGA) and externally visible characteristics (EVCs) using SNPs. Further, STR databases can be interrogated to find close genetic relatives of any unknown DNA donor, a process known as short range familial searching [8,9]. The relatively high mutation rates for STRs (relative to SNPs) means that this approach is only reliable for close genetic relatives (siblings, parents, children).

Very recently, the reach of familial searching has been greatly extended with the advent of forensic genetic genealogy (FGG) or long range familial searching. Two genetic relatives will share a distribution of uninterrupted haplotypes of various lengths, depending on their degree of relationship. The longer the lengths shared between two individuals and the more of them, the shorter the lineage paths between them (via most recent common ancestors: MRCAs) [10]. This approach requires either high density genotypes of the type produced by hybridisation arrays (microarrays) or whole genome sequences of the type produced by massively parallel sequencing (MPS).

MPS, widely used in the medical research and diagnostics fields, is a promising technology candidate for forensic DNA analysis [5,6,11,12]. MPS enables the detection of every nucleotide within a target region of DNA and therefore allows analysis of polymorphism variants with single base pair resolution. Thousands of genetic markers can be genotyped simultaneously, thereby reducing the amount of evidential sample required. A number of MPS platforms have been developed which differ in their degree of accuracy, coverage, read lengths, throughput and versatility [13-15]. These can all be applied to the identification of unidentified human remains by predicting BGA, EVCs and putative relatives.

A limitation of applying FGG to unidentified human remains is that the techniques have historically required a large amount of reasonably high quality DNA. The quantity and quality of DNA retrieved from decomposed and skeletonised human remains is much less than usually required for MPS. New FGG assays have been developed to improve the ability to recover the thousands of SNPs required for effective kinship analysis from challenging forensic samples. Such approaches can provide sufficient SNP density to apply FGG to unidentified human remains. As an example, Professor McNevin has employed targeted amplicon sequencing to a number of challenging samples [16,19] and this research aims to extend their use for FGG.

Please list the references only used in the literature review and cited in your application
NOTE: Do not include references you have not used in this application (4000 character limit)
*

[1] K. Lazaruk, P.S. Walsh, F. Oaks, et al. Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. Electrophoresis. 19(1) (1998) 8693.

[2] J.M. Butler, E. Buel, F. Crivellente, et al. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. ELECTROPHORESIS. 25(1011) (2004) 13971412.

[3] B. Quintáns, V. ÁlvarezIglesias, A. Salas, et al. Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. Forensic Science International. 140(2–3) (2004) 251257.

[4] B. Sobrino, M. Brión, A. Carracedo. SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Science International. 154(2–3) (2005) 181194.

[5] S.B. Seo, J.L. King, D.H. Warshauer, et al. Single nucleotide polymorphism typing with massively parallel sequencing for human identification. International Journal of Legal Medicine. 127(6) (2013) 10791086

[6] E. Berglund, A. Kiialainen, A.C. Syvanen. Next generation sequencing technologies and applications for human genetic history and forensics. Investigative Genetics. 2(1) (2011) 23. DOI:10.1186/20412223223.

[7] B. Merriman, Ion Torrent R&D Team, J.M. Rothberg. Progress in Ion Torrent semiconductor chip-based sequencing. ELECTROPHORESIS. 33(23) (2012) 33973417.

[8] Maguire, C.N., et al., Familial searching: A specialist forensic DNA profiling service utilising the National DNA Database® to identify unknown offenders via their relatives—The UK experience. Forensic Science International: Genetics (2014) 8(1):19.

[9] Bieber, F.R., C.H. Brenner, and D. Lazer, Finding criminals through DNA of their relatives. Science (2006) 312(5778):13151316.

[10] Speed, D. and D.J. Balding, Relatedness in the postgenomic era: Is it still useful? Nature Reviews Genetics (2015)16(1):3344.

[11] E. Giardina, A. Spinella, G. Novelli. Past, present and future of forensic DNA typing. Nanomedicine 6(2) (2011) 257270.

[12] R. Li, Y. Li, X. Fang, et al. SNP detection for massively parallel whole genome resequencing. Genome Research. 19(6) (2009) 11241132.

[13] L. Liu, Y. Li, S. Li, et al. Comparison of next generation sequencing systems. Journal of Biomedicine and Biotechnology (2012) 11.

[14] A. Ratan, W. Miller, J. Guillory, et al. Comparison of sequencing platforms for single nucleotide variant calls in a human sample. PLoS One. 8(2) (2013) e55089.

[15] S. Suzuki, N. Ono, C. Furusawa, et al. Comparison of sequence reads obtained from three next generation sequencing platforms. PLoS One. 6(5) (2011) e19534.

[16] R Daniel, C Santos, C Phillips, M Fondevila, RAH Van Oorschot, D McNevin. A SNaPshot of next generation sequencing for forensic SNP analysis. Forensic Science International: Genetics 14 (2015) 5060.

[17] M AlAsfi, D McNevin, B Mehta, D Power, ME Gahan, R Daniel. Assessment of the Precision ID Ancestry Panel. International Journal of Legal Medicine 132(6) (2018) 15811594.

[18] B Mehta, R Daniel, C Phillips, S Doyle, G Elvidge, D McNevin. Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq. Electrophoresis 37(21) (2016) 28322840.

[19] C Phillips, D McNevin, KK Kidd, R Lagacé, S Wootton, M De La Puente, R. Daniel. MAPlex A massively parallel sequencing ancestry analysis multiplex for AsiaPacific populations. Forensic Science International: Genetics 42 (2019) 213226.

Please save and continue to the next page

Methods and methodologies

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

In order to consider your research, the HREC will need to know what it will involve for your participants (REF NS 3.1)

What kinds of methods and methodologies will you use in your research? (More than one box may be checked)*

- ☑ Quantitative
- ☑ Qualitative

Does your research involve collection and/or use of secondary data? (e.g. existing / routinely collected data etc.)*

- ◉ Yes
- ○ No

Please provide a description of the secondary data source(s) below*

The research will involve generation of deidentified DNA profiles from:
* Fully informed and consenting DNA donors
* Cadavers from the Australian Facility for Taphonomic Experimental Research (AFTER) for which a separate program wide ethics approval already exists (ETH182999)
* Unidentified human remains sourced (with permission) from police and coronial agencies of Australia's States and Territories

Where permission has been granted, a limited subset of de-identified genetic data will be uploaded to forensic databases for comparison with database profiles:
* EMPOP mitochondrial DNA database (https://empop.online/), maintained by the University of Innsbruck. This database predicts maternal lineage relative to representative populations from around the world.
* YHRD Y chromosome marker database (https://yhrd.org/), maintained by Charité Universitätsmedizin Berlin. This database predicts paternal lineage relative to representative populations from around the world.
* FROGkb database (http://frog.med.yale.edu/FrogKB/), maintained by Yale University. This database predicts genetic ancestry relative to representative populations from around the world.
* Snipper app suite (http://mathgene.usc.es/snipper/), maintained by the University of Santiago de Compostela. This database predicts genetic ancestry relative to representative populations from around the world.
* HIrisplex-S webtool (https://hirisplex.erasmusmc.nl/), maintained by Erasmus MC, Rotterdam. This database predicts eye, hair and skin colour.

Where permission has been granted, a limited subset of de-identified genetic data will also be uploaded to DNA databases that will match parts of donors' DNA to the DNA of any genetic relatives also present on the databases (this genetic data is reconfigured, anonymised and/or encrypted when uploaded and will not be visible to any other database user):
* DNASolves (https://dnasolves.com/), maintained by Othram (https://othram.com/), a forensic genomics vendor
* GEDmatch (https://www.gedmatch.com), maintained by Verogen, Inc. (https://verogen.com), a forensic genomics vendor.
* FamilyTreeDNA (https://www.familytreedna.com/), maintained by Gene By Gene (https://genebygene.com), a genetic genealogy vendor.

Any search results in these databases (eg. ancestry and/or phenotype predictions, familial relations) or genetic relationships with other participants will not be disclosed to donors or their familial relations. DNA profiles generated as part of this research will not be uploaded to criminal law enforcement DNA databases (eg. National Criminal Investigation DNA Database). However, where permission has been granted, a limited subset of de-identified genetic data will be uploaded to internal AFP databases as reference purposes.

Please save and continue to the next page

Quantitative

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Section 1: Quantitative Methodologies*

- [x] Experimental
- [ ] Quasi-experimental
- [x] Correlational research
- [ ] Survey Design
- [ ] Meta analysis
- [ ] Other *(Please describe below)

Section 2: Quantitative methods*

- [ ] Written survey
- [ ] Online survey/research
- [ ] Other* (please describe below)
- [ ] Pre-post/testing
- [ ] Telephone survey
- [x] Questionnaires
- [ ] Access to records
- [ ] Clinical trial
- [x] Statistical analysis
- [ ] Content analysis
- [ ] Physiological testing/assessment

What **quantitative** methodology and methods will you be using in this research? More than one box may be checked.

Please save and continue to the next page

Qualitative

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

What **qualitative** methodology and methods will be using in this research?

Section 1: Qualitative methodology*

- ☐ Auto-ethnography
- ☐ Historical research
- ☑ Other *(Please describe below)
- ☐ Action research
- ☐ Narrative enquiry
- ☐ Biographical research
- ☐ Case study
- ☐ Phenomenology
- ☐ Indigenous research paradigm
- ☐ Discourse analysis
- ☐ Grounded theory

Please describe other methodologies*

There will be three types of samples employed in the research.

From consenting and fully informed genetic relatives, DNA will be used to verify familial relationships by:
* Comparing maternal genetic lineage from mitochondrial DNA whole genome sequences and upload to the EMPOP database
* Comparing paternal genetic lineage from Y chromosome sequences (if male) and upload to the YHRD database
* Comparing autosomal and X chromosome DNA sequences
* Upload to the GEDmatch, Othram, Family Tree DNA, EMPOP and YHRD databases as described previously.
The individuals in this group represent known familial relationships. Although there is a possibility that they may not know true genetic relationships, they will serve as controls against which to test the ability of genetic technologies to predict familial relationships.

From consenting and fully informed (unrelated) DNA donors, DNA will be used to predict:
* maternal genetic lineage from mitochondrial DNA whole genome sequences and upload to the EMPOP database
* paternal genetic lineage from Y chromosome sequences (if male) and upload to the YHRD database
* biogeographical ancestry (BGA), from autosomal DNA sequences and upload to the FROGkb database and Snipper app suite
* externally visible characteristics (EVCs), from autosomal DNA sequences and upload to the HIrisPlex-S database
The individuals in this group will be asked to complete a questionnaire relating to their BGA and EVCs. Although there is a possibility that they may not know their true BGA, they will serve as controls against which to test the ability of genetic technologies to predict maternal lineage, paternal lineage, BGA and EVCs.

From DNA extracted from human remains, DNA will be used to determine suitability (DNA quantity and quality) for prediction of:
* maternal genetic lineage, from mitochondrial DNA whole genome sequences and upload to the EMPOP database
* paternal genetic lineage, from Y chromosome sequences (if male) and upload to the YHRD database
* father-daughter, motherson or motherdaughter relationships, from X chromosome sequences
* BGA, from autosomal DNA sequences and upload to the FROGkb database and Snipper app suite
* EVCs, from autosomal DNA sequences and upload to the HIrisPlex-S database
* putative genetic relatives, by upload to GEDmatch, Othram, Family Tree DNA, EMPOP and YHRD databases as described previously. The sources of the remains will be:
* The Australian Facility for Taphonomic Experimental Research (AFTER) at UTS (for which a program wide ethics approval already exists: ETH18 - 2999). This group will be used to test the ability of genetic technologies to obtain genetic DNA from degraded, decomposed and skeletonised human remains and to subsequently predict maternal lineage, paternal lineage, BGA, EVCs and genetic relatives.
* Various Australian police and coronial agencies (with their permission) submitted to the National DNA Program for Unidentified and Missing Persons. This group represents operational casework for which new genetic technologies can aid the human identification process.

Section 2: Qualitative methods*

- ☐ Participants observation
- ☐ Covert observation
- ☐ Life story or oral history
- ☐ Focus groups
- ☐ Structured interviews
- ☐ Semi-structured interviews
- ☐ Unstructured interviews
- ☑ Other * (Please describe below)
- ☐ On-line research
- ☐ Psychological testing/assessment
- ☐ Verbal protocol
- ☐ Journaling
- ☐ Artifact analysis
- ☐ Document/Policy analysis
- ☐ Access to records
- ☐ Audio/video recording

Please describe other methods being used. and provide as much information as possible including how this method will be conducted, how many participants will be involved (from each participant group if there is more than one group/cohort), the amount of time required of participants for this, whether it will be recorded, and any other information applicable (4000 character limit)*

The methods are as described above. The numbers of participants are justified by a statistical power calculation, below.

Please save and continue to the next page

Section 6: Research participants/subjects part 1

Recruitment of participants

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

In line with the National Statement, the definition of participants includes not only those humans who are the primary focus of the research but also those who will be affected by the research. The HREC regards the principle of respect for persons as of paramount importance. (REF NS 1.1 (d), 1.6-1.9, 1.10, 2.1).

How will you initially select and contact your participants? More than one box may be checked, if appropriate*

- ☑ Advertisement/flyer
- ☑ E-mail
- ☐ Telephone
- ☑ Internet
- ☑ Organisation
- ☑ Personal contact
- ☐ Letter
- ☐ Other contact method to be used

Outline how you will obtain participants' contact details*

Participants will be recruited from the staff and students at UTS, staff at AFP, members of the forensic community, and families and friends of the researchers (or other participants). Some participants may be familiar to the researchers. Email lists and social media advertisements (eg. posts from UTS Centre for Forensic Science or Australian and New Zealand Forensic Science Society social media accounts) will be employed, inviting voluntary participation.

Email lists will be, for example staff and student lists at UTS and staff lists at AFP. These lists are held by management (eg. Science Faculty at UTS, Forensic Management Team at AFP Majura). They were created to disseminate work-related or study-related information to staff and students respectively. The lists are moderated and recruitment emails must be approved by management before dissemination.

Social media advertisements will not be targeted. All followers will be able to view recruitment notices.

Please describe your recruitment plan/strategy*

In the event someone wishes to participate after receiving a recruitment notice, they may (voluntarily) initiate contact with the researchers (using the contact details provided) and further information (the participant information sheet) will be provided so that they may make an informed decision, prior to scheduling their attendance at collection locations. Recipients of the recruitment notices may also (voluntarily) share it with colleagues, friends and relatives who, in turn, may (voluntarily) initiate contact with the researchers.

Participation in this research is voluntary and participants will need to actively contact the researchers in order to participate (no targeted recruitment of students or employees is to occur).

Participants will be recruited in this order:
* Firstly, recruit participants from the AFP via staff email distribution lists
* If insufficient participant numbers are obtained, participants will then be recruited from the UTS Faculty of Science (Dean of Science approved) via Science email distribution lists and social media (e.g. UTS Science Facebook and Twitter pages)
* If there are still insufficient participant numbers obtained, participants will be recruited from the general public using available social media platforms (e.g. Forensic Science Society Facebook and Twitter pages)
* If there are still insufficient participant numbers obtained, participants will be recruited from whole of UTS via relevant UTS email distribution lists and social media (e.g. UTS Facebook and Twitter pages) (NOTE: if this is necessary, the researchers will seek consent from the Deputy Vice-Chancellor, Education & Students, beforehand)

Samples and questionnaire responses will ideally be sourced from participants who are local to Sydney and Canberra and at this time there is no intention to provide compensation for travel costs between the participants home and collection locations. Ideally, collection will occur during normal attendance at UTS campus(es) and the AFP Forensics Complex. In some cases, the researchers may travel to participants (with their permission) in order to collect tissue samples and questionnaire responses. When there are no other options, participants in remote locations may have a participant information sheet, consent form, questionnaire and buccal swab mailed to them and they may return their consent form, tissue sample and questionnaire responses by mail.

How many participants do you intend to recruit? (If you are intending to recruit different groups of participants, please answer all relevant questions for each group, e.g. control group, test group, etc) (4000 character limit)*

We will require no more than 250 consenting and fully informed unrelated DNA donors (to be sourced from staff and students at UTS, staff at AFP, members of the forensic community and associated family/friends). Any genetic relatives of these participants who are themselves fully informed and consenting participants will form a second group of related DNA donors. The numbers of these relatives who volunteer to participate will then dictate the number of familial relationships we can test, including:
* Siblings
* Parent / child
* Grandparent / grandchild
* Great grandparent / great grandchild
* Uncle or aunt / niece or nephew
* Great uncle or aunt / great niece or nephew
* First cousins
* Second cousins
* First cousins once removed
* Second cousins once removed

Explain how and why you have chosen this number. If the research is quantitative, explain the power calculations; if the research is qualitative, explain why the proposed number is likely to result in adequate data). For guidance, check ourFact Sheets.*

To test for differences between any two categorical independent variables (eg. European versus non European biogeographical ancestry), a statistical power calculation assuming unequal variances between the two variables (the worst case scenario) reveals that the number of samples required is given by a function depending on the standard deviations of the continuous dependent variables (eg. percentage European BGA), the means of the continuous dependent variables for the two independent variables and test statistic thresholds for significance (α) and power (β), respectively. Any measured differences should be at least as great as the standard deviation. Because there will be multiple comparisons, we choose a conservative α (high significance). For a twotailed test with 99 % significance (α = 0.99) and 80 % power (β = 0.8), we require at least 25 samples for each comparison in order to achieve 80 % power. The number of comparisons will be most for biogeographical ancestry (BGA) including:
1. African versus nonAfrican BGA
2. European versus nonEuropean BGA
3. SouthwestAsian versus nonsouthwestAsian BGA
4. SouthAsian versus nonsouthAsian BGA
5. NorthAsian versus nonnorthAsian BGA
6. EastAsian versus noneastAsian BGA
7. SoutheastAsian versus nonsoutheastAsian BGA
8. Oceanian versus nonOceanian BGA
9. NorthAmerican versus nonnorthAmerican BGA
10. SouthAmerican versus nonsouthAmerican BGA
This means we will need a total of 10 × 25 = 250 samples from unrelated DNA donors.

We will also test for differences between familial relationships (eg. sibling versus not related). The relationships to be tested include:
* Siblings
* Parent / child
* Grandparent / grandchild
* Great grandparent / great grandchild
* Uncle or aunt / niece or nephew
* Great uncle or aunt / great niece or nephew
* First cousins
* Second cousins
* First cousins once removed
* Second cousins once removed
Again, we will need 25 comparisons for each relationship (if we assume that, in the worst-case scenario, there are no more than two participants in any one family). This means that we may need another 250 genetic relatives in addition to the ones above. However, this number will be reduced if there are more than two participants in any one family group. We can also reduce the number of relationships we test in the event that we don't recruit enough family members.

Describe your inclusion and exclusion criteria for participants*

DNA donors will be restricted to fully informed and consenting adults over the age of 18 years old.

Please save and continue to the next page

Participant involvement

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

What time commitment will the research involve for your participants?
NOTE: This information must be included in any information to participants
(4000 character limit)*

DNA donors will be asked to provide a noninvasive tissue sample (eg. buccal or inner cheek swab, hair, nail clippings, previously extracted wisdom tooth) and answer a questionnaire about their biogeographical ancestry (BGA) and externally visible characteristics (EVCs) in the presence of a researcher. The whole process will take approximately 15 minutes.

DNA donors from the same family group will additionally be asked to declare their familial relationship with other fully informed and consenting family members.

In what location will the research/data collection take place?
NOTE: This information must be included in any information to participants
(4000 character limit)*

DNA collection will be conducted in two locations:
1. Centre for Forensic Science, Building 4, Level 5, UTS, Ultimo, NSW
2. Australian Federal Police Forensic Complex, Majura, ACT

In addition and when required, collection of tissue samples and questionnaire responses may also be conducted in participants' homes or a location convenient to the participant.

What travel, if any, does the research involve for your participants?
NOTE: This information must be included in any information to participants
(4000 character limit)*

There is no travel required by participants. Participants may elect to post their selfadministered tissue sample, questionnaire responses and consent form if they are remote to the locations above and if a researcher is unable to visit them.

Please include any additional information relating to participants that you think relevant
NOTE: This information must be included in any information to participants
(4000 character limit)*

Participation in the research will be purely voluntary.

The collection of tissue samples and administration of the questionnaire will be performed according to the standard operating procedures (SOPs) developed for an existing program wide ethics approval (ETH182521 Collection of biological samples and associated phenotypes for forensic analysis).

DNA and questionnaire data will be deidentified. They will not be associated with the donor's name or any other personal information that may be used to identify the sample or data. The DNA and questionnaire data will only be associated with a unique sample code.

Participants may withdraw from the project at any time by quoting the unique sample code. In this case, their DNA and any information derived from it will be destroyed, unless the data has been incorporated into amalgamated data.

Describe and justify any benefit, payment or compensation the participants will receive. For research being conducted with Aboriginal and Torres Strait Islander People, the described benefits from research should have been discussed with and agreed to by the Aboriginal or Torres Strait Islander research stakeholders. (REF NS 2.1) and 4.7.8 & 4.7.9)
(4000 character limit)*

There will be no benefit to the participants beyond the experience of being involved in experimental genetic research.

There will be no payment or compensation for the participants.

Please save and continue to the next page

Consent

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Will you be obtaining written consent?*

- (●) Yes
- ( ) No

Please provide sample documents in attachments list at the end of the application form
Please use the following HREC templates when creating an information sheet and consent form: HREC templates

Do you believe there will be any special issues relating to consent in your research? (REF NS 1.13, 2.2, 2.3, Chapter 4)*

- (●) Yes
- ( ) No

Please describe what special issues may be related to consent in your research (4000 character limit)*

Participants in family groups may not have the genetic relationships to family members which they had presumed (eg. a parent may not be a biological parent). Genetic relationships or any other information derived from their DNA will not be disclosed to participants.

By obtaining whole genome sequences, it is possible that incidental healthrelated information may be obtained by someone who is trained to do so. This project does not aim to identify or analyse genetic markers for healthrelated information and the researchers are not qualified to do so. If any healthrelated information were to be (accidently or incidentally) made known to the researchers, this information would not be disclosed to participants.

Are the participants able to consent fully? (REF NS Chapter 2, 4.4, 4.5)*

- (●) Yes
- ( ) No

Please save and continue to the next page

Limited disclosure

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Does this research involve limited disclosure to participants? (REF NS 2.3)*

- ( ) Yes
- (●) No

Please save and continue to the next page

Ethical considerations specific to participants

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Indicate if your research will involve the following populations (as per the National Statement) other than as incidental participants (i.e. they are not included in the design of the project but may be participants) (REF NS Chapter 4) *

- ☐ Women who are pregnant and the human foetus
- ☐ Children and young people
- ☐ People in dependent or unequal relationships
- ☐ People highly dependent upon medical care who may be unable to give consent
- ☐ People with a cognitive impairment, an intellectual disability or a mental illness
- ☐ People who may be involved in illegal activities
- ☐ People who are incarcerated
- ☐ Aboriginal and Torres Strait Islander Peoples
- ☐ People in other countries
- ☑ None of the above

If your research is being conducted in Australia, does it involve Culturally and Linguistically Diverse (CALD) People (other than incidentally)?*

- ○ Yes
- ◉ No

Does your research involve Defence or the Department of Veteran Affairs in any way?*

- ○ Yes
- ◉ No

Please save and continue to the next page

Section 7: Research participants/subjects part 2

Risk/harm

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Risk or harm could be described as damage or hurt to the wellbeing, interests or welfare of an individual, institution or group. Harm could range from physical hurt or damage such as illness or injury, to psychological or emotional hurt or damage, such as embarrassment or distress. Please note that as a researcher, you are not necessarily immune from risk yourself and should give careful consideration to this question (REF NS 2.1).

NOTE:
It is **really** important that you carefully consider all ***potential*** risks that could occur, even if they seem negligible.
Please **do not** provide one-word answers to any of the questions below.
Describe, as best as you can, any possible risks to research participants, subjects and related groups
NOTE: This information must be included in any information to participants (4000 character limit)*

We will store deidentified DNA profiles in a secure, password protected environment at UTS and at the AFP but it is always possible that DNA profiles may be stolen by hacking. In such a case, it is possible that samples may be re-identified. This would require someone to compare a de-identified DNA profile with an overlapping profile known to be from the DNA donor. Because the field of genetics is constantly evolving, there is always a risk that the raw data of any genetic tests may reveal information from which a qualified person may extract health-related information.

If participants opt to have their deidentified DNA profile uploaded to one or more of the specified DNA databases then, again, it is always possible that their DNA profile may be stolen by hacking.

If participants opt to have their DNA profile compared with genetic relatives, there is a risk that they may find out about genetic relationships (or the absence of genetic relationships) of which they were not aware. This information will not be disclosed to the participants.

By obtaining whole genome sequences, it is possible that incidental healthrelated information may be obtained by someone who is trained to do so. This project does not aim to identify or analyse genetic markers for healthrelated information and the researchers are not qualified to do so. If any healthrelated information were to be (accidently or incidentally) made known to the researchers, this information would not be disclosed to participants.

How would you categorise the magnitude of potential risk? (e.g. inconvenience, discomfort, harmful, painful)
Explain why you believe this is so (4000 character limit)*

Having a DNA profile accessed without consent is a serious breach of privacy. DNA contains personal and sensitive information about the donor if it is able to be associated with them. Examples of information that could be revealed in this case include:
* Medical conditions and risks
* Ancestral origin
* Physical traits
* Genetic relationships with other individuals
* Unknown or unexpected children, parents, or siblings
* Biological sex

If participants opt to have their DNA profile compared with genetic relatives, discovering genetic relationships (or the absence of genetic relationships) of which they are not aware can be distressing. For example, they may find that a parent is not a biological parent.

If any healthrelated information were to be (accidently or incidentally) made known to the participants', this may breach their right "not to know" about health-related information derived from their DNA.

How would you categorise the likelihood of risk? (i.e. slight, possible, likely, probable, unavoidable)
Explain why you believe this is so (4000 characters)*

While posing a significant breach of privacy, the risk of having DNA stolen from AFP or UTS password protected data storage is very small. External databases (GEDmatch, Othram, FamilyTreeDNA, EMPOP, YHRD, FROG-kb, Snipper app suite and HIrisplex-S webtool) may be less secure and hacking represents more of a risk. However it is important to note that many law enforcement and forensic agencies use these platforms for forensic casework, after having undertaken appropriate privacy and risk assessments. Othram conducts all laboratory and analysis work in-house and has secure chain-of-custody and data handling procedures that are isolated from the internet. GEDmatch and FamilyTreeDNA reconfigure, anonymise and/or encrypt genetic data once uploaded and prevent visibility of 'research' data (once labelled as such on upload) to other users of the databases and to members of the general public. EMPOP, YHRD, FROG-kb, Snipper and HIrisPlex-S only accept limited subsets of de-identified genetic data. These data are not unique to a particular individual and could not be used to re-identify a DNA donor.

If participants opt to have their DNA profile compared with genetic relatives, discovering genetic relationships (or the absence of genetic relationships) of which they are not aware is possible. However, this information will not be disclosed to participants by the researchers so the risk is low.

This project does not aim to identify or analyse genetic markers for healthrelated information and the researchers are not qualified to do so. If any healthrelated information were to be (accidently or incidentally) made known to the researchers, this information would not be disclosed to participants. Therefore, the risk of healthrelated information being known to participants is low.

What strategies will you use to minimise and/or manage the risks? (4000 character limit)*

All DNA, questionnaire responses and any derived data will only be associated with a unique sample code for each donor. That is, it will be de-identified at the point of collection. It will not be possible for the researchers or anyone else to associate the donor with their DNA, questionnaire responses and any derived data. If the data were to be stolen, it could only be reidentified by comparison with another profile from a (known) donor.

Participants will be informed that if they opt to have their DNA profile compared with presumed genetic relatives, any genetic relationships (or the absence of genetic relationships) of which they were not previously aware will not be disclosed to them.

This project does not aim to identify or analyse genetic markers for healthrelated information and the researchers are not qualified to do so. If any healthrelated information were to be (accidently or incidentally) made known to the researchers, this information would not be disclosed to participants.

Discuss likely or possible risk to researchers (including yourself), and your strategies for minimising such risks (4000 character limit)*

There is some risk that the researchers may feel uncomfortable asking colleagues, friends and family members for DNA samples. This risk will be minimised by putting the onus on any potential participant to (voluntarily) initiate contact with the researchers after receiving a recruitment notice. They will then receive a participant information sheet to help them decide what (if anything) they would like to consent to.

Please save and continue to the next page

Pre-existing relationships

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Are there likely to be any pre-existing relationships with research participants? (e.g. employer/employee, colleague, friend, relation, student/teacher, etc) (REF NS 4.3)*

- (●) Yes
- ( ) No

Please describe (4000 character limit)*

Participants will be recruited from staff and students at UTS, staff at AFP, members of the forensic community and family and friends of the researchers (or other participants). Some may be familiar to the researchers. Email lists and social media advertisements will also be employed, inviting voluntary participation. Participants in family groups will be known to each other and will, in most cases, be genetically related to each other.

Email lists will be, for example staff and student lists at UTS and staff lists at AFP. These lists are held by management (eg. Science Faculty at UTS, Forensic Management Team at AFP Majura). They were created to disseminate work-related and study-related information to staff and students respectively.

Social media advertisements will not be targeted. All followers will be able to view recruitment notices.

How might these relationships influence their decision to participate, be affected by the proposed research or create potential ethical conflict? Please describe strategy for dealing with this (4000 character limit)*

It may be perceived by staff at AFP and staff and/or students at UTS that participation in the research is a requirement of their employment or that nonparticipation may adversely affect their employment or assessment. It will be made clear in the recruitment notice and Participant Information Sheet that this is not the case.

Family members may feel obligated to participate in the research if approached. It will be made clear that their participation is completely voluntary.

The decision to participate will require any potential participant to (voluntarily) initiate contact with the researchers after receiving a recruitment notice.

Describe how you will ensure that student assessment, employee security, etc., will not be adversely affected by participation in this research (4000 character limit)*

The research will be completely divorced from student assessment and employee security. Neither student assessment nor employee security are influenced be participation (or not) in this research. There is no policy or practice linking participation in this research with either student assessment nor employee security.

Will you be recruiting UTS staff and/or students as research participants?*

- (●) Yes
- ( ) No

You must obtain approval from the Dean or their nominee. Please attach a letter of support on the attachments page

Please save and continue to the next page

External organisations

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Indicate if your research will involve any of the following:*

- [ ] Institution
- [x] Organisation
- [ ] Community Group
- [ ] None of the above

Please describe what type(s) of institution / organisation / community group will be involved and how many will be involved (4000 character limit)*

The research is associated with the Australian Federal Police National DNA Program for Unidentified and Missing Persons. Associate Professor Ward is the Program Lead and Professor McNevin is seconded by the program (0.5 FTE). As such, they are AFP employees for the purposes of the research. There will also be a UTS PhD student (Jessica Watson) associated with the project.

Was the research generated from within the institution / organisation / community group?*

- (●) Yes
- ( ) No

Please provide details of how the research was generated from within the institution / organisation / community group (4000 character limit)*

Associate Professor Ward has been awarded an Australian Federal Police Innovation Grant of $50,000 in order to support a research and development program within the National DNA Program for Unidentified and Missing Persons, including sponsoring a PhD student to undertake a component of this research. The proposed research will contribute towards the aims of the program.

Please save and continue to the next page

External organisation consent

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Have you sought appropriate approval or support from the institution / organisation / community group involved?*

- (●) Yes
- ( ) No

Please attach a copy of any letter of approval/agreement at the end of this form

Do you intend to feed the research results back to the institution / organisation /community group?*

- (●) Yes
- ( ) No

Please describe how (4000 character limit)*

The results of the research will be the subject of internal AFP reports as well as publications in international, peer reviewed journals.

Does this research involve any contracts, including confidentiality agreements? (REF NS 3.2.12, 3.5.6) (Section 2.5 and 4, The Code)*

- (●) Yes
- ( ) No

Is your contract finalised?*

- (●) Yes
- ( ) No

Please detail any particular conditions that might have ethical implications for the research (e.g. access to data, publication, etc) **NOTE: You should attach an electronic copy of your contract to your ethics application** (4000 character limit)*

All work produced by the research, and copyright of that work, will be owned by the Commonwealth of Australia as represented by the AFP.

Please save and continue to the next page

Section 8: Data

Data collection & use

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Does your research involve access to student records at this University?*

- ( ) Yes
- (●) No

Provide an analysis plan outlining how the aims/objectives will be met, the statistical methods to be used, and who will be carrying out the analysis. *

The first stage of the project will involve testing the ability of genetic technologies to predict:
* maternal lineage, by comparison with questionnaire responses
* paternal lineage, by comparison with questionnaire responses
* biogeographical ancestry (BGA), by comparison with questionnaire responses
* externally visible characteristics (EVCs), by comparison with questionnaire responses
* genetic relatives, by comparison with declared familial relationships in families
This will involve DNA collected from fully informed and consenting participants, some of whom may be in the same family. DNA collected from these participants will be subject to the following analyses:
* Autosomal and Y chromosome short tandem repeat (STR) profiling using polymerase chain reaction (PCR) and both capillary electrophoresis and massively parallel sequencing (MPS)
* Whole mitochondrial genome sequencing using MPS
* Whole autosomal genome sequencing using MPS
* Targeted amplicon sequencing of the autosome and X and Y chromosomes for STRs and single nucleotide polymorphisms (SNPs) using MPS

Autosomal STR profiles will be subject to short range familial searching by comparison with fully informed and consenting genetic relatives.

Y chromosome STR profiles will be used to determine paternal lineage by:
* Comparison with fully informed and consenting genetic relatives
* Uploading to the YHRD Y STR database

Y chromosome SNP profiles will be used to determine paternal lineage by:
* Comparison with fully informed and consenting genetic relatives
* Comparison with published Y haplotypes
* Uploading to the YHRD Y SNP database

X chromosome STR profiles will be used to determine certain parental lineages by comparison with fully informed and consenting genetic relatives

Whole mitochondrial genome sequences will be used to determine maternal lineage by:
* Comparison with fully informed and consenting genetic relatives
* Comparison with published mtDNA haplotypes
* Uploading to the EMPOP mitochondrial DNA database

Autosomal SNP profiles will be subject to long range familial searching (genetic genealogy) by:
* Comparison with fully informed and consenting genetic relatives
* Uploading to the GEDmatch database, maintained by Verogen
* Uploading to FamilyTreeDNA, maintained by Gene By Gene
* Uploading to DNASolves, maintained by Othram

Autosomal SNP profiles will be subject to ancestry analysis by comparison with publicly available reference population databases using the following statistical tools:
* Multidimensional scaling (MDS) approaches including principle components analysis (PCA) and principle coordinates analysis (PCoA)
* Structure population genetic software
* Construction of phylogenetic trees
* Upload to the FROGkb database
* Upload to the Snipper app suite
Classification success rates will be used to produce confusion matrices for population assignment. Autosomal SNP profiles will be subject to phenotype analysis by using the HIrisplex-S webtool.

The second stage of the project will involve testing the ability of forensic genetic technologies to be applied to decomposed and skeletonised human remains from the Australian Facility for Taphonomic Experimental Research (AFTER) for which a separate programwide ethics approval already exists (ETH182999). DNA will be extracted from donated human bodies using boutique DNA extraction methods and used to optimise and validate the genetic analyses described above. These samples reflect the DNA quality and quantity of the types of forensic samples encountered in missing persons investigations.

The third and final stage of the project will involve application of optimised and validated methods to unidentified human remains cases (with permission) as part of the National DNA Program for Unidentified and Missing Persons.

Describe any foreseeable future use of this data; such as sharing with other researchers, secondary use for related research, publishing for unrelated research and non-research purposes and any other possible uses. Please note this information must be included in the participant information sheet.*

We propose to add the DNA extracted from samples collected in this research and questionnaire responses from fully informed and consenting participants to the bank of samples and responses collected under an existing ethics application (ETH182521: Collection of biological samples and associated phenotypes for forensic analysis).

We also propose to add a limited subset of deidentified genetic data from fully informed and consenting participants to population databases as follows:
* mitochondrial SNP genotypes to EMPOP
* Y STR and Y SNP profiles to YHRD
* Autosomal, X and Y STR and SNP profiles to internal reference databases maintained by the AFP for casework.
None of these are criminal DNA databases and they cannot be used to identify the participants nor reveal any information about them. Increasing the number of de-identified reference samples in these databases that are representative of the Australian population will contribute to a valuable resource for investigative forensic casework nationally and internationally.

Do you have a research data management plan?*

- (●) Yes
- ( ) No

Please save and continue to the next page

Section 9: Additional information

Other ethical issues

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

If there are any additional ethical issues which you do not believe have been covered by this form, please explain them for the HREC: (4000 character limit)*

We are not aware of any other ethical issues.

Please save and continue to the next page

Section 10: Attachments

Attachments

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

I have attached the following supporting documents

Budget page from funding application*
- (●) Yes
- ( ) N/A

Relevant contracts/agreements*
- (●) Yes
- ( ) N/A

Participant Information Sheet(s)*
- (●) Yes
- ( ) No

Informed consent form(s)*
- (●) Yes
- ( ) N/A

Evidence of approval from external institution, organisation or community group*
- (●) Yes
- ( ) N/A

Explanations of any technical terms used *
- ( ) Yes
- (●) N/A

Evidence of support from the Dean to use UTS staff/students as participants. If advertising university-wide, please provide evidence of support from the Deputy Vice-Chancellor (Education & Students)*
- (●) Yes
- ( ) N/A

Research data management plan (RDMP)*
- (●) Yes
- ( ) No

Standard Operating Procedures

N.B. May include a distress or disclosure protocol [see UTS HREC Disclosure Guidelines] under University policies and guidelines, Faculty of Health Low Risk protocol; procedures for participant screening, physiological, or biological sampling and/or laboratory or safety procedures where relevant.

*
- (●) Yes
- ( ) No

Documents attached to this application:

How to attach documents

1. Click on 'Add'
   Ensure the fields are as follows:
   - Document type- soft copy
   - Name: Include the document name and version number
   - Description: This field is optional
2. You can then either select the file you want to upload OR drag and drop it where it says 'Drop file here'
3. Click on 'OK'

Note: Please use the following HREC templates when creating an information sheet, consent form, verbal script, etc.: HREC templates. All submitted documents should be titled, and have have version control included in the footer.*

1	Document type	Soft copy
	Name	Evidence of support from the Dean of Science
	Reference (Document Title)	21001 Human Ethics Application - Dennis McNevin.pdf
	Description	
2	Document type	Soft copy
	Name	AFP Innovation Grant funding
	Reference (Document Title)	Successful funding bid_Email.pdf
	Description	
3	Document type	Soft copy
	Name	AFP funding approval
	Reference (Document Title)	2020 21 Innovation Fund Approved Projects.pdf
	Description	
4	Document type	Soft copy
	Name	AFP National DNA Program for Unidentified & Missing Persons
	Reference (Document Title)	National DNA Program_Summary for Media.pdf
	Description	
5	Document type	Soft copy
	Name	Research Data Management Plan
	Reference (Document Title)	RDMP.pdf
	Description	
6	Document type	Soft copy
	Name	SOPs for ETH18-2521
	Reference (Document Title)	ETH18-2521 SOPs.pdf
	Description	
7	Document type	Soft copy
	Name	Recruitment notice
	Reference (Document Title)	ETH21-5821 Recruitment.pdf
	Description	
8	Document type	Soft copy
	Name	Informed consent form
	Reference (Document Title)	ETH21-5821 Consent.pdf
	Description	
9	Document type	Soft copy
	Name	Questionnaire
	Reference (Document Title)	ETH21-5821 Questionnaire.pdf
	Description	
10	Document type	Soft copy
	Name	Participant information sheet
	Reference (Document Title)	ETH21-5821 Participant information.pdf
	Description	
11	Document type	Soft copy
	Name	Project budget
	Reference (Document Title)	Project budget.pdf
	Description	
12	Document type	Soft copy
	Name	ETH21-5821 - MCNEVIN - MREC outcome and comments 22 April 2021
	Reference (Document Title)	ETH21-5821 - MCNEVIN - MREC outcome and comments 22 April 2021.docx
	Description	
13	Document type	Soft copy
	Name	Responses to MREC

	Reference (Document Title)	ETH21-5821 - MCNEVIN - Responses to MREC.pdf
	Description	
14	Document type	Soft copy
	Name	Revised RDMP
	Reference (Document Title)	RDMP revised.pdf
	Description	
15	Document type	Soft copy
	Name	Revised recruitment notice
	Reference (Document Title)	ETH21-5821 Recruitment v5.docx
	Description	
16	Document type	Soft copy
	Name	Revised participant information sheet
	Reference (Document Title)	ETH21-5821 Participant information v6.docx
	Description	
17	Document type	Soft copy
	Name	Revised informed consent form
	Reference (Document Title)	ETH21-5821 Consent v6.docx
	Description	
18	Document type	Soft copy
	Name	Revised questionnaire
	Reference (Document Title)	ETH21-5821 Questionnaire v2.docx
	Description	

Reminder to student applicants:

1. Please note that once your application is submitted it will go directly to your supervisor and not to the Committee.
2. We **strongly** recommend notifying your supervisor that you have submitted your application in case of any technical issues, to avoid potential delays in the review process.
3. Once your supervisor endorses your application it will go to your Local Research Office for endorsement before coming to the Ethics Secretariat for review.
4. Your electronic application must be endorsed by your supervisor by the Local Research Office (LRO) submission deadline.
5. Please also ensure that the Primary AOU listed at the end of the Investigators page is updated to your supervisor's AOU. This will show in the table under 'Internal personnel listed below', once you add them. If you need any assistance with this please contact Research.Ethics@uts.edu.au or call 9514 9772. Please note that this is particularly important if you have a dual role as a staff/student as your application could go to the wrong faculty for review through the automated process.

Declaration

Declaration

I have answered all questions in the risk assessment truly and completely to the best of my knowledge

I will notify the UTS Human Research Ethics Committee of any variation to this research that may alter the level of risk associated with it

This research will be undertaken in compliance with the UTS Research Ethics and Integrity Policy or any replacement or amendment thereof

This research will be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research

Please click on the "Submit" button in the Actions menu.

Confirmation

Confirmation by Local Research Office High Risk

Application type*

Research (staff project)

Internal personnel listed on this ethics protocol*

1	Primary	Yes
	ID	[redacted]
	Surname	McNevin
	Given Name	Dennis
	Full Name	Prof Dennis Blair McNevin
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Dennis.McNevin@uts.edu.au
	Work Number	
2	Primary	No
	ID	[redacted]
	Surname	Ward
	Given Name	Jodie
	Full Name	A/Prof Jodie Ward
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Jodie.Ward@uts.edu.au
	Work Number	
3	Primary	
	ID	[redacted]
	Surname	Watson
	Given Name	Jessica
	Full Name	Miss Jessica Lee Watson
	Position	5Research Student
	Type	Internal
	AOU	
	Managing Unit	
	Email Address	Jessica.Watson@uts.edu.au
	Work Number	

Checked by:*

Lisa Merry

Date of review:*

22/03/2021

The Research Office has confirmed that: All information in this application and supporting documentation is correct and as complete as possible *

- (●) Yes
- ( ) No

Confirmation by ADR

Application type

Human

Internal personnel listed on this ethics protocol

1	Primary	Yes
	ID	[redacted]
	Surname	McNevin
	Given Name	Dennis
	Full Name	Prof Dennis Blair McNevin
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Dennis.McNevin@uts.edu.au
	Work Number	
2	Primary	No
	ID	[redacted]
	Surname	Ward
	Given Name	Jodie
	Full Name	A/Prof Jodie Ward
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Jodie.Ward@uts.edu.au
	Work Number	
3	Primary	
	ID	[redacted]
	Surname	Watson
	Given Name	Jessica
	Full Name	Miss Jessica Lee Watson
	Position	5Research Student
	Type	Internal
	AOU	
	Managing Unit	
	Email Address	Jessica.Watson@uts.edu.au
	Work Number	

Date of LRO review

23/03/2021

Declaration:

- I am aware that this research is being conducted within this Faculty/School/Centre.
- I am satisfied that the researchers have met all Faculty/School/Centre requirements in relation to this research
- This research will be undertaken in compliance with the UTS Research Ethics and Integrity Policy or any replacement or amendment thereof
- This research will be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research

*

(•) Yes
( ) No

Comments

This question is not answered.

Research Office use only

Research Office use only

Application Status

Approved

Approval Purpose

Research (staff project)

Current Committee

0Health and Medical Research Ethics Committee (Human)

TRIM number

RES21/412

Start date

09/06/2021

End date

09/06/2026

Date received

31/03/2021

Date Approved

09/06/2021

Date Reviewed

15/04/2021

Date Withdrawn

This question is not answered.

Special conditions*

NA

Appendix 2. Ethics approval from UTS HREC under project number ETH21-6606

ResearchMaster

Ethics Application

Application ID :	ETH21-6606
Application Title :	Genetic identification of unidentified human remains
Date of Submission :	21/10/2021
Primary Investigator :	Prof Dennis Blair McNevin (Chief Investigator)
Other Personnel :	Miss Jessica Lee Watson (5Research Student)
	A/Prof Jodie Ward (Chief Investigator)

Section 1: Ethics Portal

Select your application type

What type of application are you looking for?
Please **do not** change your application type without first consulting with the Ethics Secretariat (9514 9772).*

- ○ New application (including scope-checking for nil/negligible risk research)
- ○ Ratification of existing approval
- ○ Transfer of existing approval
- ○ Evaluation of teaching and learning activities
- ◉ Amendment to existing approval
- ○ Program approval

You have selected "amendment application". This option allows you to amend
an existing UTS HREC approved protocol **EXCEPT** for approved negligible risk declarations. If you want to update an approved negligible risk declaration, please submit a new application.
Please click "save" before continuing.

Please refer to the amendment table on our website regarding the requirements for amendments to existing approval before continuing.

Please indicate the risk classification of the original ethics approval?*

- ○ Nil/Neg risk
- ○ Low risk
- ◉ High risk

What should I know before I start?

Would you like more information on:

- ○ This system
- ○ The ethics process
- ◉ Purpose of the ethics review process

Purpose of the Human Research Ethics Review Process

The ethical review process is valuable as it:

- Provides the opportunity to reflect on the research methodology;
- Ensures the research integrity of the work and
- Increases the trust levels of the research participants and the publishers about the output of the research project.

The online ethics application simplifies the process:

- The form is intuitive so you won't need to answer every question just those that are applicable to your research
- Some parts of the form are auto-populated
- You can save the form and return to it later to complete and submit it
- Guidelines and instructions are incorporated into the form
- Reduced paper handling.

Importance of ethics procedures
Regardless of the level of risk, all staff and students are expected to abide by the standards outlined in the National Statement, the Australian Code, and guidelines established by the UTS Human Research Ethics Committee (HREC). It is the responsibility of researchers, both staff and students, to familiarise themselves with these.

Section 1A: Risk evaluation

Risk A

Determining the level of risk and review

- Please answer each question carefully **and consecutively**.
- For assistance with answering these questions please refer to the National Statement on Ethical Conduct in Human Research as per the chapters listed below.
- If you need to contact the Research Ethics Officer you can call (02) 9514 9772
- Click on the help buttons (?) for more information
- You can save your application at any time by clicking on the save button on the left hand side in the toolbar. For further information and help in completing your application go to our website.

Does your research involve:

Projects involving covert observation, active concealment, or planned deception of participants

e.g. covert observation of the hand-washing behaviour of hospital employees, undisclosed role-playing by a researcher, etc. Does NOT include observation in a public place WITHOUT the use of photographs, images, video or audio footage (Chapter 2.3, p.19)

*

○ Yes
◉ No

Targeted recruitment or analysis of data(?) from any of the groups listed below (or where any of these groups are likely to be significantly over-represented in the group being studied)

- Women who are pregnant and the human fetus (Chapter 4.1, p. 61)
- Children and young people (under 18 years) (Chapter 4.2, p. 65)
- People in dependent or unequal relationships (e.g. lecturer/student [except T&L], doctor/patient, employer/employee) (Chapter 4.3, p.68)
- People highly dependent on medical care who may be unable to give consent Chapter 4.4, p.68)
- People with a cognitive impairment, an intellectual disability, or a mental illness (may include the disadvantaged/homeless) (Chapter 4.5, p. 70)
- People who may be involved in illegal activities (including those affected e.g. victims of domestic violence) (Chapter 4.6, p.73)
- Aboriginal and Torres Strait Islander Peoples (Chapter 4.7, p.77)

*

○ Yes
◉ No

Targeted recruitment of people in / from countries that score <50 on the Corruption Perception Index (CPI) (check here)
This includes any cohorts from these countries, i.e. it is not restricted to marginalised groups within these countries*

○ Yes
◉ No

Collection, use or disclosure of personal information without consent of the participant(?)

- **a record which may include your name, address and other details about the participant (e.g. date of birth, financial information etc.)**
- **photographs, images, video or audio footage**
- **fingerprints, blood or DNA samples**

*

○ Yes
◉ No

Collection, use or disclosure of health information(?)

- personal information that is information or an opinion about
  - the physical or mental health or a disability (at any time) of an individual; or
  - an individual's expressed wishes about the future provision of health services to him or her, or
  - a health service provided, or to be provided, to an individual or
- other personal information collected to provide, or in providing, a health service, or
- other personal information about an individual collected in connection with the donation, or intended donation, of a individual's body parts, organs, body substances, or
- other personal information that is genetic information about an individual arising from a health service provided to the individual in a form that is or could be predictive of the health (at any time) of the individual or of a genetic relative of the individual, or
- healthcare identifiers

N.B Includes information collected through physiological testing or assessment. Examples include but are not limited to EEG, EMG, BMI, blood pressure, DEXA, etc.*

◉ Yes
○ No

Collection, use or disclosure of sensitive information
Racial, ethnic information, political, religious and philosophical beliefs, sexual activity or identity, and trade union membership
*

◉ Yes
○ No

Activity that potentially infringes the privacy or professional reputation of participants, providers or organisations
e.g. observation in the workplace, collection of commercially confidential information, etc.
Commercially confidential information = Any information which is not in the public domain or publicly available, and where disclosure may undermine the economic interest or competitive position of the owner of the information (TGA adopted definition from European Medicines Agency (EMA)).

N.B. if canvassing opinion via consensus methods i.e. Delphi (?), answer "No" here
*

○ Yes
◉ No

Establishment of a register or databank of identifiable data for possible use in future research projects (Chapter 3.2, p.27) (?)

*

◉ Yes
○ No

Collection, transfer(?) and/or banking of human biospecimens.
e.g. tissue, blood, urine, sputum etc.(?)

*

◉ Yes
○ No

Any significant alteration to routine care or service provided to participants
e.g. deviation from standard care or usual practice
*

○ Yes
◉ No

Prospective assignment of human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes(?) (Chapter 3.14-3.17) *

○ Yes
◉ No

Potential for participants to experience harm (i.e. anything more than discomfort)(?)
e.g. physical, psychological, devaluation of personal worth, social, economic and/or legal (Chapter 2.1, p.12)
*

○ Yes
◉ No

High Risk

Section 2: Project information

Project title

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Application ID (automatically generated):

ETH21-6606

Application Title:*

Genetic identification of unidentified human remains

Ethics category code (automatically selected):*

Human

Please search for your original ethics application by clicking on 'More criteria'.
Please note that you can only search for previously submitted applications where personnel listed on this application were also listed on the original one.
*

1	Ethics Category	Human
	Ethics Application Code	ETH21-5821
	Ethics Title	Genetic identification of unidentified human remains
	Start Date	09/06/2021
	End Date	09/06/2026
	Review Date	15/04/2021
	Application Status	Approved
	Other Comments	

Please save and continue to the next page

Section 3: Personnel

Investigators

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Please note that for amendment applications you only need to add the Chief Investigator/Supervisor, student(s) and any new personnel

Have new external investigators been added to this protocol?*

- ○ Yes
- ◉ No

Is this application for a student project?*

- ◉ Yes
- ○ No

Student applicants:

1. Please note that once your application is submitted it will go directly to your supervisor and not to the Committee.
2. We **strongly** recommend notifying your supervisor that you have submitted your application in case of any technical issues, to avoid potential delays in the review process.
3. Once your supervisor endorses your application it will go to your Local Research Office for endorsement before coming to the Ethics Secretariat for review.
4. Your electronic application must be endorsed by your supervisor by the Local Research Office (LRO) submission deadline.
5. Please also ensure that the Primary AOU at the end of this page is updated to your supervisor's AOU. This will show in the table under 'Internal personnel listed below', once you add them. If you need any assistance with this please contact Research.Ethics@uts.edu.au or call 9514 9772. Please note that this is particularly important if you have a dual role as a staff/student as your application could go to the wrong faculty for review through the automated process.

Are the student(s) listed on this protocol new to the protocol? (e.g. being added as part of the amendment application)*

- ○ Yes
- ◉ No

Positions in the personnel table

Position type:	In the personnel table use the following positions from the drop-down list:
Chief Investigator/Supervisor	1Chief Investigator (students must **not** be listed as Chief Investigator)
Co Investigator	3Assoc. Investigator
Co Supervisor	Co-Supervisor
Research Student	5Research Student
Project Administrator	7Project Administrator

Note: Further options are available in the drop down list.

Instructions on how to add a person to the personnel table:

1. Click on "Add"
2. Start typing the details (first name, last name or Staff ID) in the search bar.
3. Click on "Add selected"
4. The extra information panel will open, select their position from the drop-down list. If they are the primary contact (e.g. Chief Investigator/Supervisor), tick "Yes" under 'Primary contact' and then select "OK"

- **Student research:** Students must add their supervisors to their application and must mark their primary supervisor as a Chief Investigator and as a primary contact. Students must be listed as "5Research student" under the column 'Position' to ensure the application is properly submitted to their supervisor.
- **Ratifications/Transfers:** If this list differs from that of the original application, you must provide evidence that any additional investigators have been added via amendment to the lead/external HREC [attach relevant amendments and evidence of approval].

Internal personnel listed on this ethics protocol:*

1	Primary	No
	ID	
	Surname	Watson
	Given Name	Jessica
	Full Name	Miss Jessica Lee Watson
	Position	5Research Student
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Jessica.Watson@student.uts.edu.au
	Work Number	
2	Primary	No
	ID	
	Surname	Ward
	Given Name	Jodie
	Full Name	A/Prof Jodie Ward
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Jodie.Ward@uts.edu.au
	Work Number	
3	Primary	Yes
	ID	
	Surname	McNevin
	Given Name	Dennis
	Full Name	Prof Dennis Blair McNevin
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Dennis.McNevin@uts.edu.au
	Work Number	

If any details are incorrect or missing please contact the Ethics Secretariat on (02) 9514 9772 or by email.

Please provide additional (or preferred) contact details of any of the people listed on the project if necessary (4000 character limit)

This question is not answered.

Primary AOU*

SCI.School of Mathematical and Physical Sciences

Managing Unit

Science

Please save and continue to the next page

Student details

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Degree being undertaken (500 character limit)*

Doctor of Philosophy (Faculty of Science). My supervisors are Jodie Ward and Dennis McNevin.

Have you been successful in your doctoral/masters assessment? *

- ○ Yes
- ◉ No

Please indicate why you are applying for ethics approval at this stage, and when you will be seeking assessment or re-assessment? (4000 character limit)*

The original ethics application was made by my supervisor prior to my enrolment as a PhD student while I waited for the application to be processed. While in meetings with my supervisors and discussing the validation plans for the technology, we decided to submit an amendment.

I have not yet had my stage 1 assessment. It is due in May 2022.

Students, please read carefully: Once you have completed this application and followed the submission instructions, your application will go to your supervisor for review. Once your supervisor has reviewed and endorsed your application it will come to the Ethics Secretariat for a pre-review. This pre-review process helps ensure that your application is complete, has all necessary attachments, and that the quality of responses to the questions meets the Committee's expectations. Your application should therefore be submitted as early as possible. If you do not submit your application in time, it may be delayed and held off until the next closing date.

Section 5: Amendment form

Amendment details

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

Has your project title changed?*

- ○ Yes
- ◉ No

Please provide a brief summary of your research proposal based on your original ethics application and specify what stage the research is at (e.g. recruitment, data collection, etc). *

This research aims to understand the relationships between biogeographical ancestry (BGA), externally visible characteristics (EVCs), and familial relationships to provide information for future missing persons investigations. This will involve the application of the following technologies to the identification of unidentified human remains:
* Advanced DNA extraction methods
* Whole mitochondrial genome, Y chromosome and X chromosome sequencing for lineage-based searching
* Autosomal DNA sequencing for short-range familial searching, and prediction of BGA and EVCs
* Whole genome and/or medium-high density kinship informative SNP genotyping for long-range familial searching (forensic genetic genealogy)

We propose to test the ability of genetic technologies to predict the BGA, EVCs and familial relationships of fully informed and consenting participants. This information will be obtained from questionnaires and then compared to the information predicted from genotypes derived from the DNA voluntarily donated by the participants. We will also test the abilities of these genetic technologies to be applied to decomposed and skeletonised human remains from the Australian Facility for Taphonomic Experimental Research (AFTER) for which a separate program-wide ethics approval already exists (ETH18-2999). The optimised and validated genetic technologies will be applied to unidentified human remains sourced (with permission) from police and coronial agencies of Australia's States and Territories as part of the National DNA Program for Unidentified and Missing Persons.

The research is currently in the recruitment stage where we are collecting biological samples voluntarily donated by participants. The laboratory and data analysis protocols are being written for the first validation study.

Does your amendment involve any of the following changes:*

- ☐ Change to completion date
- ☐ Change to personnel
- ☑ Change to research instruments/participant material
- ☑ Change to research methodology
- ☑ Change to recruitment of participants
- ☐ Other

Does your amendment involve the addition of children as participants? (not as incidental)*

- ○ Yes
- ◉ No

Will changes include research be conducted using UTS staff and/or students? (not previously approved for)*

- ○ Yes
- ◉ No

What changes to your original ethics application are you proposing? (1500 character limit)*

We are proposing updating the methodology to include a comparison of our laboratory-generated genetic data against existing direct-to-consumer genetic data uploaded voluntarily by the participants to GEDmatch (a public genealogy database).

As a consequence, we have amended the participant's information forms and included an appendix with instructions for how to upload their genetic data, as well as the consent form and questionnaire. Our research methodology has been amended to include a comparative analysis of these data types for the accuracy of familial relationship detection.

We will be asking participants to voluntarily upload any of their existing direct-to-consumer genetic profiles to GEDmatch for the purposes of comparing with data produced with our new forensic tests if they have not previously done so. We have already notified volunteers that we would be uploading a limited subset of their genetic data to GEDmatch in the original ethics application forms. If they have a file already on GEDmatch, they are able to leave the data on the database or, should they not wish us to use the data, change their privacy settings to public or remove the data completely. If they have one or more direct-to-consumer genetic profiles, we have provided instructions on how to download it and upload it to GEDmatch, including account creation instructions. We do not ask the participants to pursue direct-to-consumer DNA profiling - we only ask them to upload pre-existing profiles.

If they are willing to have their direct-to-consumer genetic profile included as a part of this research, they can indicate this on the consent form and provide the GEDmatch ID profile (alias) on the questionnaire. GEDmatch provides the option for an alias for an individual's account which takes the place of their real name; we have recommended they use the unique sample code provided for their donated sample to remain de-identified and maintain their privacy.

Why do you wish to make these changes?*

In a missing persons investigation, family members of missing persons can elect to submit their DNA for direct-to-consumer profiling and upload the data to public genetic databases such as GEDmatch. This would only be useful in cases where there are no close relatives of the missing person in question as routine short tandem repeat (STR) genotyping can be used for short-range familial searching (parents, children and siblings).

Samples from unidentified human remains will be analysed using a forensic-specific kit called Kintelligence that contains a subset of the single nucleotide polymorphisms (SNPs) included in direct-to-consumer profile (10,000 SNPs versus up to 1 million SNPs). We wish to determine if this reduced number of SNPs provides genetic linkages to relatives with the same accuracy as the much larger direct-to-consumer profiles. If there are participants in this study with pre-existing direct-to-consumer profiles who are willing to make them available for comparison with Kintelligence, then we will have a point of comparison.

The original ethics application only allowed for comparison between relatives with genetic profiles generated from Kintelligence. If the research participants were to upload their direct-to-consumer DNA profiles to GEDmatch, we would be able to conduct mock searches of the profiles produced by Kintelligence against direct-to-consumer profiles, which would be the case in a real investigation.

Please save and continue to the next page

Impact of amendment on research participants part 1

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

This section requires you to consider the ways in which your proposed amendments may impact upon the ethical issues raised on your original application. Specifically, we ask you to outline the effects (if any) of your amendments on the following areas, and how you intend to deal with them. Does your amendment affect any of the following:

Outcome of your research?
Note: If you are making changes to data collection instruments, please specify what your intention is for the data which has already been collected. If the data which has already been included will still be used, how will this be analysed with new data, and how might this impact the validity of results / impact originally stated outcomes?*

- (●) Yes
- ( ) No

Please provide further information on how your amendment effects the outcome of your research*

The proposed changes to the ethics application do not affect data already collected. They will augment this data with a DNA profile that more closely resembles scenarios that would occur in reality. Upload of DNA profiles to GEDmatch allows estimation of the degree of relationship between two individuals through kinship coefficients and shared segments of DNA that are identical by descent. The original ethics application included assessing the accuracy of these relationships by comparing the DNA profiles generated inside the forensic laboratory by the researchers with a limited number of genetic markers (10,000).

Accuracy can also be assessed by comparing the forensic laboratory profiles against existing direct-to-consumer data uploaded to GEDmatch, consisting of many more genetic markers (in the order of one million). This is the scenario most likely in practice. The comparison of these two different kinds of profiles will allow us to determine the efficacy of comparing forensic-generated data against direct-to-consumer data to identify previously unidentified human remains.

Current or future applications for funding?*

- ( ) Yes
- (●) No

Recruitment of participants (quantity, methods)*

- (●) Yes
- ( ) No

Please provide further information on how your amendment effects the recruitment of participants (quality, methods).
Note: If changes are made to your recruitment strategy (e.g. inclusion criteria) will they be applied retrospectively?*

Changes to our recruitment strategy will be made retrospectively. Participants will be asked if they would consider uploading any existing genetic data from direct-to-consumer testing (e.g. Family Tree DNA, AncestryDNA, MyHeritage, 23andMe) to GEDmatch and make it available for law enforcement searching. We will not be requesting that participants provide their DNA for direct-to-consumer testing specifically for this project. Instructions will be provided for downloading their existing genetic data and uploading it to GEDmatch.

Anticipated risk or harm to participants and/or researchers?*

- (•) Yes
- ( ) No

Please explain how do you propose to minimise these risks*

The risk for participants who upload their direct-to-consumer genetic data to GEDmatch and make it available for law enforcement searching is that law enforcement officers from anywhere in the world will see their profile name (alias) and email address. This would occur if they are genetically related to a suspected offender or unidentified human remains. These officers will not be able to access the actual DNA profile. Other users of GEDmatch who could conduct searches where the participant is matched include genealogists and distant relatives.

Participants will be informed of this fact and are under no obligation to upload their genetic data. In the participant's information form and appendix, we have recommended that they use their participant ID for the GEDmatch profile ID. While GEDmatch requires a name on the account, they provide an option for an alias (GEDmatch profile ID), which would be visible to any user who finds a link to their profile as a result of a genetic relationship. In the creation of a GEDmatch account, users are required to provide an email address. We have advised that this email address could be used by law enforcement, genealogists or distant relatives to contact the participant.

At the conclusion of the study, they can opt to be notified so they can change their privacy settings or remove their data from GEDmatch.

Relationships (if any) between researchers and participants?*

- ( ) Yes
- (•) No

Please continue to the next page

Impact of amendment on research participants part 2

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

This section requires you to consider the ways in which your proposed amendments may impact upon the ethical issues raised on your original application. Specifically, we ask you to outline the effects (if any) of your amendments on the following areas, and how you intend to deal with them. Does your amendment affect any of the following:

Consent from Participants? *

- (•) Yes
- ( ) No

Please provide further information on how your amendment effects the consent from participants. Please attach revised consent form and information sheet if applicable (using tracked changes).*

The consent form has been updated to include a check box for the participant to mark whether they are willing to upload existing direct-to-consumer genetic data onto GEDmatch and make it available for law enforcement searching. The participant information form has been amended to advise of the consequences of doing so. Participants have the option to include contact information (email address) if they would like to be notified of the conclusion of the study so that they can remove their profile from GEDmatch or make it unavailable for law enforcement searching. The revised participant information and consent forms are attached.

Data collection, interpretation, storage and/or disposal? *

- ( ) Yes
- (•) No

Privacy and confidentiality of participants?*

- (•) Yes
- ( ) No

Please provide further information on how your amendment effects the privacy and confidentiality of participants?*

Genetic data uploaded to GEDmatch and made available to law enforcement searching will mean the individual's profile name will be available to law enforcement searching internationally. The participation information document has been updated to include that only their alias would be made available to law enforcement, not their genetic data, and would only be seen if they are genetically related to the individual being searched in the database. It has been recommended that participants who upload existing genetic data to GEDmatch use their participant ID for the GEDmatch profile ID. This will maintain their privacy in this study.

If law enforcement does obtain a match against a volunteer's online genetic data, they might be emailed in order to verify their identity. Law enforcement might require such verification for their genealogy research to build family trees and identify who could potentially be the person of interest. Without an Australian police warrant, participants are under no obligation to verify their identity to law enforcement or anyone else.

Are you required to submit requests for amendment to any external bodies to UTS? (e.g. an Area Health Service, other university)*

- ( ) Yes
- (•) No

Are there any other relevant ethical issues in relation to the proposed amendment? *

○ Yes
◉ No

Please continue to the next page

Amendment attachments

We recommend you save your application regularly while editing. You can save your application at any time by clicking on the save button.
For further information and help in completing your application go to our website

I have attached the following supporting (track changed) documents that require amendment from the approval of my original application:*

☑ consent form/information letter(s)
☑ surveys/questionnaires/outline of questions
☐ instruments for data collection
☐ approval for amendment from other institution
☑ other relevant attachments

Please list any other relevant attachments*

* Appendix to Participant Information Sheet (How to Upload to GEDmatch)

Documents attached to this application:

How to attach documents

1. Click on 'Add'
   Ensure the fields are as follows:
   - Document type- soft copy
   - Name: Include the document name and version number
   - Description: This field is optional
2. You can then either select the file you want to upload OR drag and drop it where it says 'Drop file here'
3. Click on 'OK'

*

1	Document type	Soft copy
	Name	Consent Form
	Reference (Document Title)	ETH21-6606 Consent.docx
	Description	Consent form updated with GEDmatch profile details.
2	Document type	Soft copy
	Name	Participant Information Form
	Reference (Document Title)	ETH21-6606 Participant Information.docx
	Description	Participant information form outlining the project and what will happen with samples volunteered. Updated to include uploading to GEDmatch and the risks associated.
3	Document type	Soft copy
	Name	Participant Information Appendix A
	Reference (Document Title)	ETH21-6606 Participant Information Appendix A.docx
	Description	Appendix to the participant information form with how to upload genetic data to GEDmatch.
4	Document type	Soft copy
	Name	Questionnaire
	Reference (Document Title)	ETH21-6606 Questionnaire.docx
	Description	Questionnaire updated with GEDmatch profile ID.

Please continue to the next page

Declaration

Declaration

I have answered all questions in the risk assessment truly and completely to the best of my knowledge

I will notify the UTS Human Research Ethics Committee of any variation to this research that may alter the level of risk associated with it

This research will be undertaken in compliance with the UTS Research Policy or any replacement or amendment thereof

This research will be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research

Please click on the "Submit" button in the Actions menu.

Confirmation

Confirmation by Local Research Office High Risk

Application type*

Amendment to existing approval

Internal personnel listed on this ethics protocol*

1	Primary	No
	ID	
	Surname	Watson
	Given Name	Jessica
	Full Name	Miss Jessica Lee Watson
	Position	5Research Student
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Jessica.Watson@student.uts.edu.au
	Work Number	
2	Primary	No
	ID	
	Surname	Ward
	Given Name	Jodie
	Full Name	A/Prof Jodie Ward
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Jodie.Ward@uts.edu.au
	Work Number	
3	Primary	Yes
	ID	
	Surname	McNevin
	Given Name	Dennis
	Full Name	Prof Dennis Blair McNevin
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Dennis.McNevin@uts.edu.au
	Work Number	

Please indicate the risk classification of the original ethics approval?*

- ○ Nil/Neg risk
- ○ Low risk
- ◉ High risk

Please contact the Ethics Secretariat.

Checked by:*

Lisa Merry

Date of review:*

14/10/2021

The Local Research Office has confirmed that: All information in this application and supporting documentation is correct and as complete as possible *

- ◉ Yes
- ○ No

Confirmation by ADR

Application type

Human

Internal personnel listed on this ethics protocol

1	Primary	No
	ID	[redacted]
	Surname	Watson
	Given Name	Jessica
	Full Name	Miss Jessica Lee Watson
	Position	5Research Student
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Jessica.Watson@student.uts.edu.au
	Work Number	
2	Primary	No
	ID	[redacted]
	Surname	Ward
	Given Name	Jodie
	Full Name	A/Prof Jodie Ward
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Jodie.Ward@uts.edu.au
	Work Number	
3	Primary	Yes
	ID	[redacted]
	Surname	McNevin
	Given Name	Dennis
	Full Name	Prof Dennis Blair McNevin
	Position	Chief Investigator
	Type	Internal
	AOU	SCI.School of Mathematical and Physical Sciences
	Managing Unit	Science
	Email Address	Dennis.McNevin@uts.edu.au
	Work Number	

Date of LRO review

14/10/2021

Declaration:

- I am aware that this research is being conducted within this Faculty/School/Centre.
- I am satisfied that the researchers have met all Faculty/School/Centre requirements in relation to this research
- This research will be undertaken in compliance with the UTS Research Ethics and Integrity Policy or any replacement or amendment thereof
- This research will be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research

*

(•) Yes
( ) No

Comments

This question is not answered.

Research Office use only

Research Office use only

Application Status

Approved

Approval Purpose

Amendment to existing approval

Current Committee

3-Expedited Review Committee (Human)

TRIM number

RES21/412/1

Date received

21/10/2021

Date Reviewed

02/11/2021

Date Approved

08/11/2021

Start date

08/11/2021

End date

This question is not answered.

Date Withdrawn

This question is not answered.

Special conditions

NA

Appendix 3. Participant Information Sheet

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

PARTICIPANT INFORMATION SHEET
ETH21-5821 GENETIC IDENTIFICATION OF UNIDENTIFIED HUMAN REMAINS

WHO IS DOING THE RESEARCH?

My name is Jessica Watson and I am a PhD student at the University of Technology Sydney (UTS). My project is part of the Australian Federal Police (AFP) National DNA Program for Unidentified and Missing Persons. My supervisors are Associate Professor Jodie Ward and Professor Dennis McNevin.

WHAT IS THIS RESEARCH ABOUT?

This research aims to find out about new and better methods for the genetic identification of unidentified human remains. This will include prediction of ancestral origin, physical traits and genetic relationships with other people. It will aid future missing persons investigations by providing investigative leads when all traditional methods and avenues of enquiry have been exhausted.

FUNDING

Funding for this project has been received from the National DNA Program for Unidentified and Missing Persons and an AFP Innovation Grant.

WHY HAVE I BEEN ASKED?

You have been invited to participate in this study because we need to understand the relationship between genotype (the DNA code) and (a) ancestry, (b) physical appearance and (c) familial relationships. You have been contacted because you are on a student/staff email list, you have subscribed to receive certain social media notifications, you are a colleague/friend/family member of another participant or you are known to us in some other capacity.

IF I SAY YES, WHAT WILL IT INVOLVE?

- If you decide to participate, you will be invited to provide a tissue sample (e.g. self administered buccal or inner cheek swab, hair, nail clippings, previously extracted tooth) for a range of genetic (DNA) tests and answer a questionnaire about your ancestry and physical appearance. These will all be de-identified, that is, your tissue sample and questionnaire responses will only be associated with a unique participant ID known only to you and the researchers. It will not be associated with any identifying information (i.e. your name). This will take approximately 15 minutes of your time.
- You may also opt to have a limited subset of your de-identified genetic data uploaded to the following databases (this genetic data is not unique to a particular individual and could not be used to identify you):
  - EMPOP mitochondrial DNA database (https://empop.online/), maintained by the University of Innsbruck. This database will predict your maternal lineage relative to representative populations from around the world.
  - YHRD Y chromosome marker database (https://yhrd.org/) maintained by Charité Universitätsmedizin Berlin. This database will predict your paternal lineage relative to representative populations from around the world.
  - FROGkb database (http://frog.med.yale.edu/FrogKB/), maintained by Yale University. This database will predict your genetic ancestry relative to representative populations from around the world.
  - Snipper app suite (http://mathgene.usc.es/snipper/), maintained by the University of Santiago de Compostela. This database will predict your genetic ancestry relative to representative populations from around the world.
  - HIrisplex-S webtool (https://hirisplex.erasmusmc.nl/), maintained by Erasmus Medical Centre, Rotterdam. This database will predict your eye, hair and skin colour.
- You may also opt to have a limited subset of your de-identified genetic data permanently submitted to forensic reference population databases (EMPOP, YHRD and/or internal AFP reference databases maintained by the researchers) for future forensic casework purposes.
- You may also opt to have your DNA and/or questionnaire responses used for future research purposes that are an extension of this project.
- If you have a relative who is also participating, you may also opt to:

  - have your DNA profile compared with known genetic relatives, if they also participate and consent, in order to correlate DNA with familial relationships
  - have your de-identified DNA sequenced for kinship informative markers by Othram (https://othram.com/), a private, US based forensic genomics vendor (the entire laboratory and analysis process is performed in-house with secure chain-of-custody and data handling procedures that are isolated from the internet)
- If you have a relative who is also participating, you may also opt to have a limited subset of your de-identified genetic data uploaded to the following DNA databases which will match parts of your DNA to the DNA of any genetic relatives also present on the databases (this genetic data is reconfigured, anonymised and/or encrypted when uploaded and will not be visible to any other database user):
  - DNASolves (https://dnasolves.com/), maintained by Othram.
  - GEDmatch (https://www.gedmatch.com), maintained by Verogen, Inc. (https://verogen.com), a private, US based forensic genomics vendor.
  - FamilyTreeDNA (https://www.familytreedna.com/), maintained by Gene by Gene (https://genebygene.com), a private, US based genetic genealogy vendor.
- If you have a relative who is also participating and you have a DNA profile from a direct-to-consumer company (e.g. Ancestry.com, 23andMe, MyHeritage, FamilyTreeDNA), would you consider migrating your data to GEDmatch (https://gedmatch.com) and making it available to law enforcement searching? Instructions on how to do this are in Appendix A. In order to identify the profile on GEDmatch and maintain your privacy, we recommend using the provided unique sample code as your GEDmatch profile ID.
- Your DNA profile will not be used in any active missing persons investigations or uploaded to any law enforcement criminal DNA database.

The participation process is summarised in a separate flowchart (see below).

ARE THERE ANY RISKS/INCONVENIENCE?

Yes, there are some risks/inconvenience.

We will store your de-identified DNA profile in a secure, password protected environment at UTS and the AFP for the purposes of this research only. It is always possible, however, that your DNA profile may be stolen by hacking. If you opt to have your de-identified DNA profile uploaded to a DNA database then, again, it is always possible that your DNA profile may be stolen by hacking. In such a case, it is also possible that your DNA profile could be compared with another DNA profile of yours in order to re-identify it. This could result in personal, sensitive or health information about you to become known to a third party, but only if they could re-identify your DNA.

If you opt to have your DNA profile compared with known genetic relatives, it is possible that that the presumed relationship is not genetic (e.g. a parent may not be a biological parent). Any genetic relationship derived from your DNA will not be disclosed to you. Some genetic tests may produce genetic data that contains health-related information about you. This project does not aim to identify or analyse genetic markers for health-related information and the researchers are not qualified to interpret this information. If any health-related information were to be (accidently or incidentally) made known to the researchers, this information would not be disclosed to you.

If you upload your direct-to-consumer profile to GEDmatch and it is available for law enforcement searching, law enforcement officers from anywhere in the world will see your profile name but will not be able to access your actual DNA profile. Your profile name could be linked to that of a suspected offender, if you are genetically related. If your unique sample code is your GEDmatch profile ID, law enforcement will have more trouble identifying you, while allowing the researchers to easily find your de-identified profile.

In summary, because the field of genetics is constantly evolving, there is always a risk that the raw genetic data of any genetic tests may reveal information from which a qualified person may extract personal, sensitive or health data, including:

- Medical conditions and risks
- Ancestral origin
- Physical traits
- Genetic relationships with other individuals
- Unknown or unexpected children, parents, or siblings
- Biological sex

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

DO I HAVE TO SAY YES?

Participation in this study is voluntary. It is completely up to you whether or not you decide to take part.

WHAT WILL HAPPEN IF I SAY NO?

If you decide not to participate, it will not affect your relationship with the researchers, with other participants or with the UTS or AFP. If you wish to withdraw from the study once it has started, you can do so at any time without having to give a reason, by contacting me (jessica.watson@uts.edu.au) and quoting your unique sample code.

If you withdraw from the study, the following will be destroyed:

- Your tissue sample and any DNA extracted from it
- Any data derived from your DNA (unless they have been aggregated in a de-identified form)
- Your questionnaire responses
- Any data derived from questionnaire responses (unless they have been aggregated in a de-identified form)

It may not be possible to withdraw your data from the study results if these have already been aggregated and had your unique sample code removed.

CONFIDENTIALITY

By signing the consent form you consent to the research team collecting and using personal information about you for the research project. All this information will be treated confidentially. Your DNA, questionnaire responses and any derived data will only be associated with a unique sample code. That is, it will be de-identified. By default, it will not be possible for the researchers or anyone else to associate you with your DNA, questionnaire responses and any derived data so it is important to retain this code if you wish to withdraw from the project.

If you upload your direct-to-consumer genetic data, you may be contacted by police or genealogists to ask for your identity via the email address linked to your GEDmatch profile.

With your permission, we would like to store your information for future use in research projects that are an extension of this research project. They will be maintained in accordance with an existing ethics approval (ETH18-2521: Collection of biological samples and associated phenotypes for forensic analysis). In all instances, your information will be treated confidentially.

We plan to publish the results in internal AFP reports and in international, peer reviewed academic journals.

WHAT IF I HAVE CONCERNS OR A COMPLAINT?

If you have concerns about the research that you think the researchers can help you with, please feel free to contact me or my supervisors (jodie.ward@uts.edu.au, dennis.mcnevin@uts.edu.au).

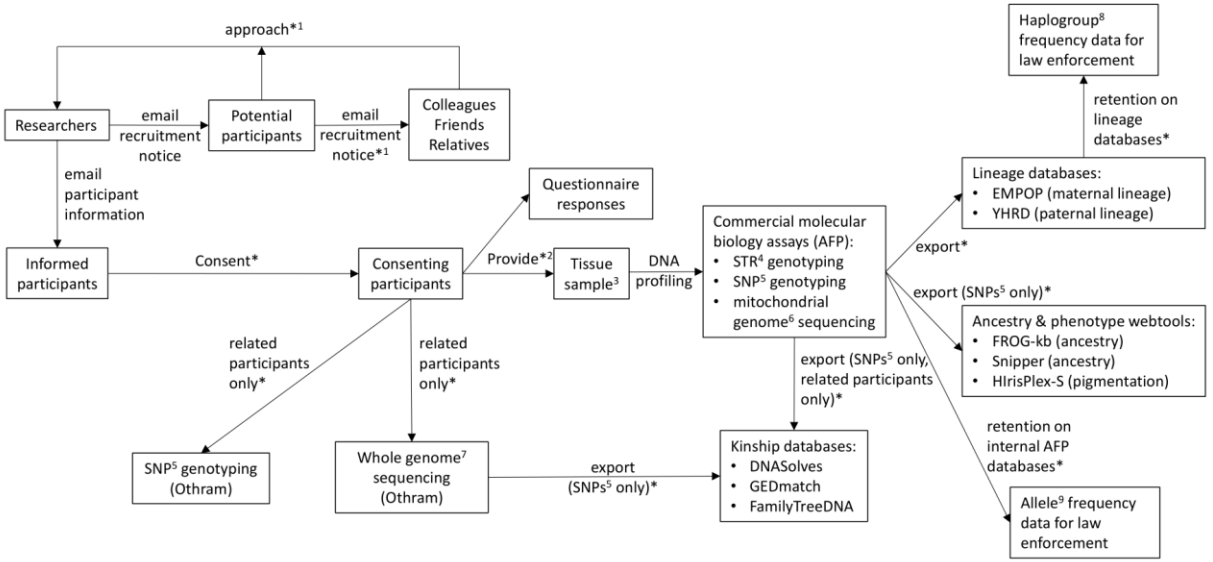
You will be given a copy of this form to keep.

YOUR UNIQUE SAMPLE CODE

Please retain your unique sample code, below, in case you wish to withdraw from the project at any time:

UNIVERSITY OF TECHNOLOGY SYDNEY

NOTE:
This study has been approved in line with the University of Technology Sydney Human Research Ethics Committee [UTS HREC] guidelines. If you have any concerns or complaints about any aspect of the conduct of this research, please contact the Ethics Secretariat on ph.: +61 2 9514 2478 or email: Research.Ethics@uts.edu.au], and quote the UTS HREC reference number. Any matter raised will be treated confidentially, investigated and you will be informed of the outcome.


approach*1
Researchers
email recruitment notice
Potential participants
email recruitment notice*1
Colleagues Friends Relatives
email participant information
Informed participants
Consent*
Consenting participants
Questionnaire responses
Provide*2
Tissue sample3
DNA profiling
Commercial molecular biology assays (AFP):
• STR4 genotyping
• SNP5 genotyping
• mitochondrial genome6 sequencing
Haplogroup8 frequency data for law enforcement
retention on lineage databases*
Lineage databases:
• EMPOP (maternal lineage)
• YHRD (paternal lineage)
export*
export (SNPs5 only)*
Ancestry & phenotype webtools:
• FROG-kb (ancestry)
• Snipper (ancestry)
• HIrisPlex-S (pigmentation)
related participants only*
related participants only*
export (SNPs5 only, related participants only)*
retention on internal AFP databases*
SNP5 genotyping (Othram)
Whole genome7 sequencing (Othram)
export (SNPs5 only)*
Kinship databases:
• DNASolves
• GEDmatch
• FamilyTreeDNA
Allele9 frequency data for law enforcement

UNIVERSITY OF TECHNOLOGY SYDNEY

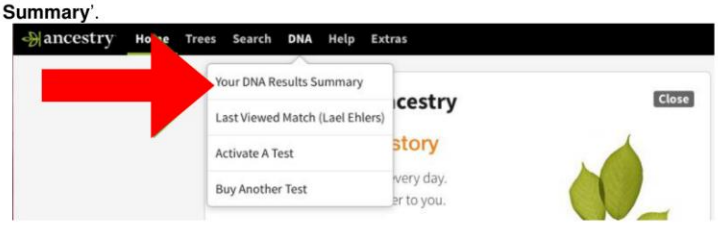
ETH21-5821 GENETIC IDENTIFICATION OF UNIDENTIFIED HUMAN REMAINS - FOOTNOTES

*You may opt in or out of any of these steps
[1]You may share the recruitment email with other colleagues, friends and relatives who can independently initiate contact with the researchers
[2]You will be asked to provide a tissue sample and questionnaire responses in the presence of one of the researchers
[3]Tissue samples may include buccal (inside cheek) swabs, plucked hairs, nail clippings, previously extracted tooth, etc.
[4]STR = short tandem repeats: repetitive parts of your DNA that are used to generate forensic DNA profiles
[5]SNP = single nucleotide polymorphisms: single base pair changes in your DNA that are used to generate forensic DNA profiles
[6]Mitochondrial genomes are complete mitochondrial DNA sequences located in the mitochondria of your cells that are inherited directly from your biological mother and used to generate forensic DNA profiles
[7]Whole genomes are complete DNA sequences that make up your entire set of chromosomes located in the nucleus of your cells that are inherited from your biological mother and father and used to generate forensic DNA profiles
[8]A haplogroup is a group of people who have similar genetic origins
[9]An allele is a unit of DNA inherited directly from one of your biological parents

Appendix A: How to Upload Your Genetic Data to GEDmatch

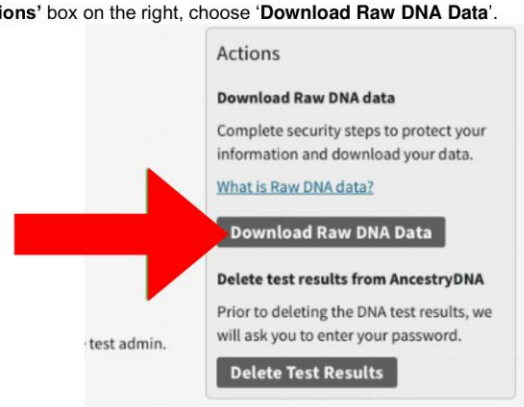
DOWNLOAD RAW DATA FROM ANCESTRYDNA

1. Login to your Ancestry.com account (https://www.ancestry.com.au/account/signin).
2. On the menu located at the top of the screen, under '**DNA'**, click on '**Your DNA Results Summary**'.



3. Click on the settings button ⚙ on the top right of the screen.

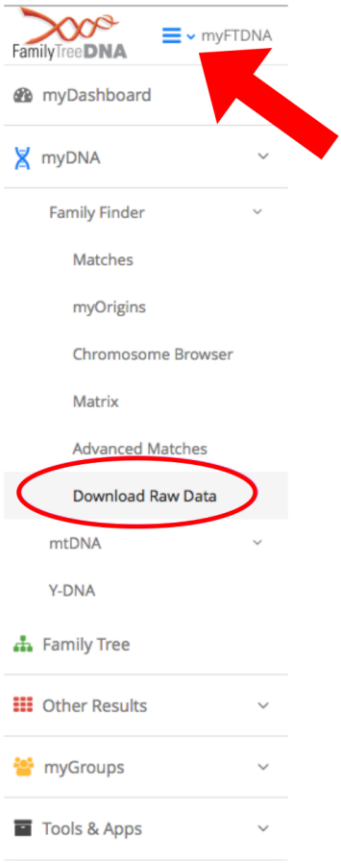
4. In the '**Actions'** box on the right, choose '**Download Raw DNA Data**'.



5. Input your password and email address and click the box acknowledging that you are responsible for your download; your raw DNA will be sent to this email address.
6. In the email, click on '**Confirm Data Download**'; this will take you back to AncestryDNA and download a zip file.

DOWNLOAD RAW DATA FROM FAMILYTREEDNA

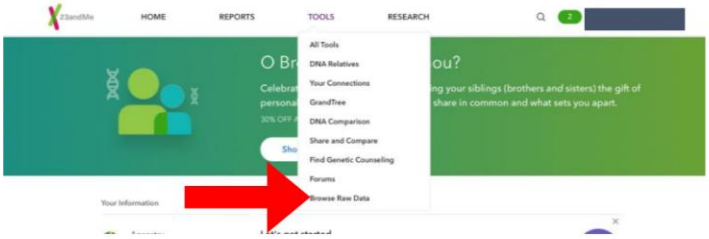
1. Login to your FamilyTreeDNA account (https://www.familytreedna.com/sign-in).
2. From the FamilyTreeDNA dropdown menu, under '**Family Finder**', click on '**Download Raw Data**'.



3. In the download menu, select both '**Build 37 Autosomal Raw Data**' and '**Build 37 X Chromosome Raw Data**' files to download.

DOWNLOAD RAW DATA FROM 23ANDME

1. Login to your 23andMe account (https://www.23andme.com/en-int/).
2. On the menu located at the top of the screen, under '**Tools**', click on '**Browse Raw Data**'.



3. Click on the '**Download**' tab at the top of the screen.

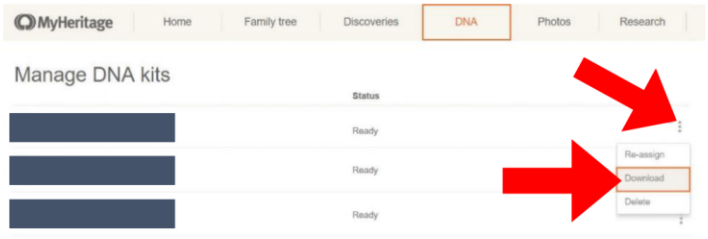
4. Scroll down to the '**Request Your Raw Data Download**' and click on '**Submit Request**'; your raw DNA will be sent to your registered email address.
5. In the email, click on '**Download Raw Data**' and input your 23andMe username and password.
6. Click on '**Download Raw Data**'.

UNIVERSITY OF TECHNOLOGY SYDNEY

DOWNLOAD RAW DATA FROM MYHERITAGE DNA

1. Login to your MyHeritage DNA account (https://www.myheritage.com/login).
2. On the menu located at the top of the screen, under '**DNA**', click on '**Manage DNA Kits**'.

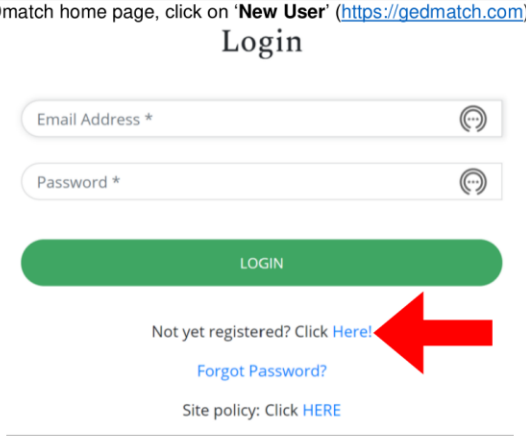
3. On the kit that you want to download, click the three dots ⋮ on the right side of the screen, click on '**Download**'.



4. A window will pop up with information on what you are going to download, click on '**Continue**'.
5. A new window will pop up asking you to accept the MyHeritage Terms of Service and Privacy Policy; if you elect to accept, tick the box and click on '**Continue**'.
6. An email will be sent to your registered email address.
7. In the email, click on the download link '**Click Here to Continue with the Download**'; this will redirect you to the MyHeritage website.
8. Enter your MyHeritage password and click '**Download**'.

UPLOAD YOUR RAW DATA TO GEDMATCH

1. On the GEDmatch home page, click on '**New User**' (https://gedmatch.com).



2. Fill out the required form: your name, alias, email address, and password. The alias takes the place of your name real name; we recommend using the unique sample code provided. When providing an email address, understand this email address will be visible to law enforcement and could be used to contact you by police, genealogists and distant relatives from around the world.
3. Check your email for a confirmation code.
4. Add the code from your email to the GEDmatch site to complete registration.
5. In the '**File Uploads**' box, select your DNA testing company.
6. You will be taken to a data upload page for your appropriate DNA testing company. Enter the required information.
7. Under the privacy options, select the bubble for '**Opt-in**'; this will allow your genetic data to be searched against all genetic data on the GEDmatch database.
8. At the bottom of the page, select '**Choose File**' and click on the downloaded zip file. Click on '**Upload**'.
9. [*FamilyTreeDNA Only*] Follow the same procedure to upload your XDNA data by selecting '**FTDNA X-DNA**'.

For ease of identification and your privacy, we recommend using your provided unique sample code as your GEDmatch profile ID (alias). This will enable us to link your genetic data with the DNA profile produced in this research project.

Appendix 4. Consent Form

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

CONSENT FORM
ETH21-5821 GENETIC IDENTIFICATION OF UNIDENTIFIED HUMAN REMAINS

I _____ agree to participate in the research project *Genetic identification of unidentified human remains* (ETH21-5821) being conducted by Jessica Watson from the University of Technology Sydney (UTS) at the Australian Federal Police (AFP) (jessica.watson@uts.edu.au). I understand that funding for this research has been provided by the AFP National DNA Program for Unidentified and Missing Persons and an AFP Innovation Grant.

I have read the Participant Information Sheet or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research as described in the Participant Information Sheet.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time without affecting my relationship with the researchers from UTS or the AFP.

I understand that I will be given a signed copy of this document to keep.

I agree to provide the following sample types for genetic analyses (please ✓ what you consent to and × what you do not consent to):
☐ Buccal (inner cheek) swab
☐ Hair
☐ Nail clippings
☐ Other (please specify) _____

I agree to (please ✓ what you consent to and × what you do not consent to):
☐ Answer a questionnaire on my ancestry and physical appearance
☐ Have my DNA compared with known genetic relatives (who also consent)
☐ Have my DNA sequenced by Othram (https://othram.com/) for kinship informative markers
☐ Have my DNA used for future research purposes
☐ Have my questionnaire responses used for future research purposes
☐ I agree to migrate my existing direct-to-consumer genetic data onto GEDmatch (https://gedmatch.com) and make it available for law enforcement searching
☐ I would like to be notified when the study has concluded to remove my profile from GEDmatch or change my settings on GEDmatch

Email: _____

I agree to have a limited subset of my de-identified genetic data uploaded to (please ✓ what you consent to and × what you do not consent to):
☐ EMPOP mitochondrial DNA database (https://empop.online/), maintained by the University of Innsbruck
☐ YHRD Y chromosome marker database (https://yhrd.org/), maintained by Charité Universitätsmedizin Berlin
☐ FROG-kb database (http://frog.med.yale.edu/FrogKB/), maintained by Yale University
☐ Snipper app suite (http://mathgene.usc.es/snipper/), maintained by the University of Santiago de Compostela
☐ HIrisplex-S webtool (https://hirisplex.erasmusmc.nl/), maintained by Erasmus Medical Centre, Rotterdam
☐ DNASolves (https://dnasolves.com/), maintained by Othram (https://othram.com/)
☐ GEDmatch (https://www.gedmatch.com), maintained by Verogen, Inc. (https://verogen.com)

☐ Family Tree DNA (https://www.familytreedna.com/), maintained by Gene by Gene (https://genebygene.com)

To aid future forensic casework, I agree to have a limited subset of my de-identified genetic data permanently included on (please ✓ what you consent to and × what you do not consent to):
☐ EMPOP mitochondrial DNA database (https://empop.online/)
☐ YHRD Y chromosome marker database (https://yhrd.org/)
☐ Internal AFP databases

I am aware that I can contact Jessica Watson (jessica.watson@uts.edu.au) or her supervisors (jodie.ward@uts.edu.au, dennis.mcnevin@uts.edu.au) if I have any concerns about the research.

_____ ____/____/____
Name and Signature [participant] Date

_____ ____/____/____
Name and Signature [researcher or delegate] Date

Appendix 5. Questionnaire

UTS UNIVERSITY OF TECHNOLOGY SYDNEY

QUESTIONNAIRE

ETH21-5821 GENETIC IDENTIFICATION OF UNIDENTIFIED HUMAN REMAINS

YOUR UNIQUE SAMPLE CODE

Below is your unique sample code. This is the only identifying information associated with your questionnaire responses and DNA. This number is not linked to your personal identity. If at any time you wish to withdraw them, please quote this code.

GEDMATCH PROFILE

If you have opted to migrate existing genetic data to GEDmatch, then please provide your GEDmatch profile ID. For your privacy and data security, it is recommended you use your unique sample code.

YOUR RELATIVES

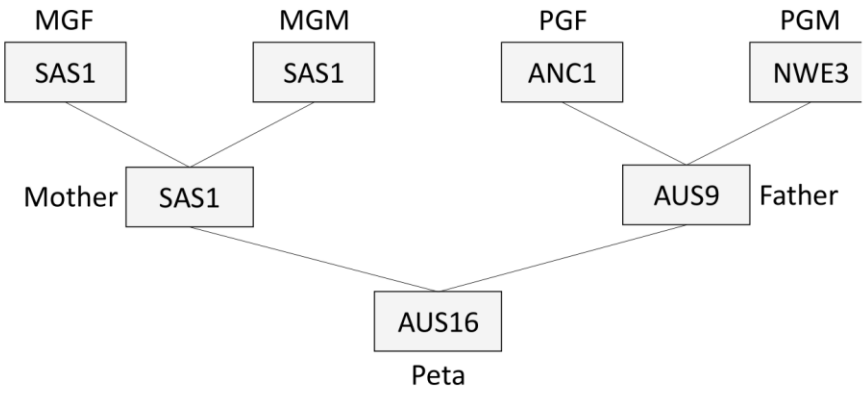
If you have opted to declare your familial relationship with other consenting participants, then please complete the table below. Please be as specific as possible about the relationship, eg:

- Brother
- Half-sister
- Mother
- Son
- Uncle on mothers' side
- First cousin on father's side

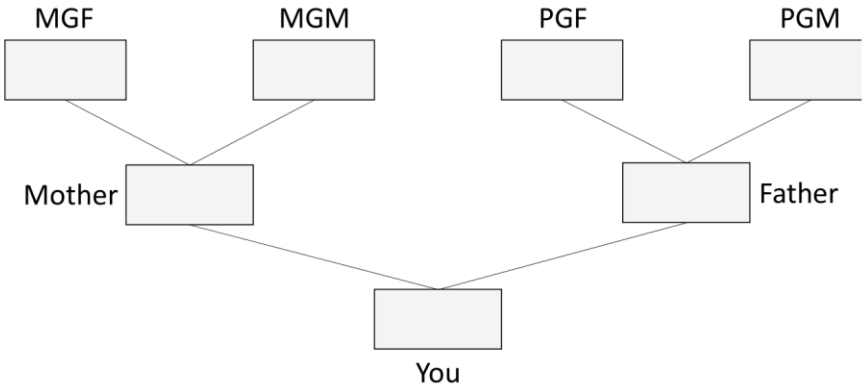
Sample code	**Relation to you**

BIOGEOGRAPHICAL ANCESTRY

Please indicate the ancestry of your grandparents, your parents and yourself by filling in the boxes below using the codes on the next page. For example, both Peta's parents were born in Australia. Her maternal grandparents were Indian and her mother considers herself to be of Indian decent. Her paternal grandfather was of English descent and her grandmother was French. Peta and her father both consider themselves to be Australian. Her ancestral chart could be completed as follows:



Please complete your ancestral chart…



Sub-population codes:

Sub-Saharan African
SSA1 South Africa; **SSA2** Zimbabwe; **SSA3** Mauritius; **SSA4** Kenya; **SSA5** Ethiopia; **SSA6** Nigeria; **SSA7** Somalia; **SSA8** Gambia; **SSA9** Ghana; **SSA10** Other (specify)
North African
NAF1 Egypt; **NAF2** Sudan; **NAF3** Libya; **NAF4** Morocco; **NAF5** Algeria; **NAF6** Tunisia; **NAF7** Other (specify)
Middle Eastern
MEA1 Lebanon; **MEA2** Iraq; **MEA3** Iran; **MEA4** Turkey; **MEA5** Syria; **MEA6** Saudi; **MEA7** Israel; **MEA8** UAE; **MEA9** Libya; **MEA10** Other (Specify)
Southern European
SEU1 Italy; **SEU2** Greece; **SEU3** Malta; **SEU4** Cyprus; **SEU5** Portugal; **SEU6** Spain; **SEU7** Gibraltar; **SEU8** Other (specify)
Eastern European
EUE1 Poland; **EUE2** Croatia; **EUE3** Macedonia; **EUE4** Bosnia/Herzegovina; **EUE5** Russia; **EUE6** Serbia; **EUE7** Hungary; **EUE8** Romania; **EUE9** Ukraine; **EUE10** Czech; **EUE11** Slovenia; **EUE12** Slovakia; **EUE13** Latvia; **EUE14** Bulgaria; **EUE15** Albania; **EUE16** Lithuania; **EUE17** Estonia; **EUE 18** Belarus; **EUE19** Montenegro; **EUE20** Other (specify)
Anglo/Celtic
ANC1 England; **ANC2** Scotland; **ANC3** Ireland; **ANC4** Wales; **ANC5** Channel Islands; **ANC6** Jersey; **ANC7** Isle of Man; **ANC8** Other (specify)
North/West European
NWE1 Germany; **NWE2** Netherlands; **NWE3** France; **NWE4** Austria; **NWE5** Switzerland; **NWE6** Sweden; **NWE7** Denmark; **NWE8** Finland; **NWE9** Belgium; **NWE10** Norway; **NWE11** Other (specify)
South Asian
SAS1 India; **SAS2** Sri Lanka; **SAS3** Bangladesh; **SAS4** Maldives; **SAS5** Other (specify)
Central Asian
CAS1 Pakistan; **CAS2** Nepal; **CAS3** Afghanistan; **CAS4** Bhutan; **CAS5** Kazakhstan; **CAS6** Uzbekistan; **CAS7** Armenia; **CAS8** Azerbaijan; **CAS9** Georgia; **CAS10** Other (specify)
North/East Asian
NEA1 China/Taiwan; **NEA2** Korea; **NEA3** Japan; **NEA4** Mongolia; **NEA5** Other (specify)
South East Asian
SEA1 Philippines; **SEA2** Vietnam; **SEA3** Malaysia; **SEA4** Indonesia; **SEA5** Thailand; **SEA6** Singapore; **SEA7** Cambodia; **SEA8** Myanmar; **SEA9** Laos; **SEA10** Other (specify)
Oceanian
OCE1 Fiji; **OCE2** Papua New Guinea; **OCE3** Samoa; **OCE4** Tonga; **OCE5** Cook Islands; **OCE6** Solomon Islands; **OCE7** New Caledonia; **OCE 8** Vanuatu; **OCE 9** Maori; **OCE 10** Other (specify)
American
AME1 American European; **AME2** American African; **AME3** American Hispanic; **AME4** Brazil; **AME5** Chile; **AME6** Columbia; **AME7** Argentina; **AME8** El Salvador; **AME9** Peru; **AME10** Uruguay; **AME11** Venezuela; **AME12** Mexico; **AME13** Ecuador; **AME14** West Indies; **AME 15** Other (specify)
Australasian
AUS1 Australian Aboriginal; **AUS2** Torres Strait Island; **AUS3** Australian African; **AUS4** Australian North African; **AUS5** Australian Middle Eastern; **AUS6** Australian Southern European; **AUS7** Australian Eastern European; **AUS8** Australian Anglo/Celtic; **AUS9** Australian North/West European; **AUS10** Australian South Asian; **AUS11** Australian Central Asian; **AUS12** Australian North/East Asian; **AUS13** Australian South East Asian; **AUS14** Australian Oceanian; **AUS15** Australian American; **AUS16** Other (specify)

Please answer as many of these questions as possible. You can choose not to answer any question.

1. What is your height in centimetres (cm)? _____

2. What is your weight in kilograms (kg)? _____

3. What is your age in years? _____

4. What is your shoe size? _____

5. Your biological gender is:
   ☐ Male
   ☐ Female

6. Which of the following best describes your natural eye colour?
   ☐ Blue
   ☐ Grey
   ☐ Green
   ☐ Hazel
   ☐ Brown
   ☐ Confirmation by collector

7. Which of the following best describes your skin colour on areas never exposed to the sun at age 20 (or now if you are not yet 20 years old)?
   ☐ Very pale
   ☐ Pale
   ☐ Intermediate
   ☐ Dark
   ☐ Dark to black
   ☐ Confirmation by collector

8. Which of the following best describes your hair texture at age 20 (or now if you are not yet 20 years old)?
   ☐ Straight
   ☐ Wavy
   ☐ Curly
   ☐ Extremely curly (“frizzy”)
   ☐ Confirmation by collector

9. What percentage of grey hair do you now have (choose the closest)?
   ☐ 0 %
   ☐ 10 %
   ☐ 25 %
   ☐ 50 %
   ☐ 75 %
   ☐ 100 %
   ☐ Confirmation by collector

10. In which of the following age brackets did you first notice that your hair was starting to go grey?
    - ☐ never
    - ☐ 20 to 30
    - ☐ 30 to 40
    - ☐ 40 to 50
    - ☐ 50 to 60
    - ☐ 60 to 70
    - ☐ after 70

11. Which of the following best describes your natural hair colour at age 18?
    - ☐ Black
    - ☐ Dark brown
    - ☐ Brown
    - ☐ Dark blond
    - ☐ Blond
    - ☐ Red
    - ☐ Grey

    ☐ Confirmation by collector

12. Which of the following best describes your natural hair colour now?
    - ☐ Black
    - ☐ Dark brown
    - ☐ Brown
    - ☐ Dark blond
    - ☐ Blond
    - ☐ Red
    - ☐ Grey

    ☐ Confirmation by collector

13. Your chin is best described as:
    ☐ Cleft ☐ Not cleft ☐ Not sure

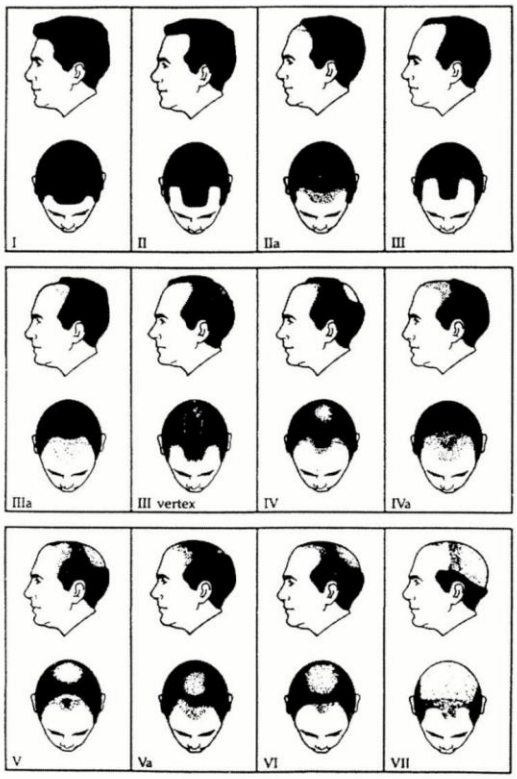
    ☐ Confirmation by collector

14. Your ear lobes are best described as:
    ☐ Attached ☐ Not attached ☐ Not sure

    ☐ Confirmation by collector

15. Which hand do you usually use to write a letter legibly?
    - ☐ Left
    - ☐ Either
    - ☐ Right

16. If you are a male, tick or cross the box which best describes your hair (loss) at the present time.



☐ Confirmation by collector

17. Which image below best describes your hairline:

☐ Widow's peak ☐ No widow's peak ☐ Not sure

☐ Confirmation by collector

Thank you for your time and co-operation!

Appendix 6. Sample Information

Supplementary Table 1. All samples used in this thesis and their sample type, DNA quantities (small autosomal (SA) target, large autosomal (LA) target and Y chromosome (male) DNA target) and their degradation index (DI). The chapters that include the samples are indicated.

Sample ID	Sample Type	SA Target (ng/μL)	LA Target (ng/μL)	Male DNA Target (ng/μL)	DI	Chapter					
						3	4	5	6	7	8
NA24385	Positive control	N/A	N/A	N/A	N/A	✓	✓				
2800M	Positive control	N/A	N/A	N/A	N/A			✓			
Family 1	Reference-type, buccal swab	0.291	0.359	0.000	0.81		✓		✓	✓	✓
Family 2	Reference-type, buccal swab	0.020	0.006	0.000	3.33	✓	✓		✓	✓	✓
Family 3	Reference-type, buccal swab	0.043	0.037	0.043	1.15	✓	✓		✓	✓	✓
Family 4	Reference-type, buccal swab	0.726	0.443	0.461	1.64		✓		✓	✓	✓
Family 5	Reference-type, buccal swab	0.297	0.203	0.293	1.46	✓	✓		✓	✓	✓
Family 6	Reference-type, buccal swab	0.221	0.095	0.000	2.32	✓	✓		✓	✓	✓
Family 7	Reference-type, buccal swab	0.066	0.058	0.072	1.14	✓	✓	✓	✓	✓	✓
Family 8	Reference-type, buccal swab	6.262	5.388	0.000	1.16		✓		✓	✓	✓
Family 9	Reference-type, buccal swab	3.907	3.996	0.000	0.98		✓		✓	✓	✓
Family 10	Reference-type, buccal swab	3.137	4.758	0.000	0.66		✓	✓	✓	✓	✓
Family 11	Reference-type, buccal swab	20.449	15.095	16.548	1.35		✓		✓	✓	✓
ETH21-5821-5	Reference-type, buccal swab	0.099	0.053	0.000	1.86			✓	✓		
ETH21-5821-21	Reference-type, buccal swab	4.903	2.962	0.000	1.66			✓	✓		
ETH21-5821-40	Reference-type, buccal swab	14.410	16.208	0.000	0.89		✓		✓		
ETH21-5821-41	Reference-type, buccal swab	6.774	7.607	0.000	0.89			✓	✓		
ETH21-5821-42	Reference-type, buccal swab	7.617	6.902	7.938	1.10			✓	✓		
ETH21-5821-43	Reference-type, buccal swab	21.095	31.369	0.000	0.67			✓	✓		
ETH21-5821-44	Reference-type, buccal swab	2.322	3.564	0.000	0.65		✓		✓		
ETH21-5821-45	Reference-type, buccal swab	1.859	1.938	0.000	0.96		✓		✓		
ETH21-5821-46	Reference-type, buccal swab	3.022	3.807	3.041	0.79			✓	✓		
ETH21-5821-47	Reference-type, buccal swab	4.013	6.046	5.722	0.66			✓	✓		
ETH21-5821-48	Reference-type, buccal swab	8.295	10.944	0.000	0.76			✓	✓		
ETH21-5821-49	Reference-type, buccal swab	4.412	3.920	0.000	1.13			✓	✓		
ETH21-5821-50	Reference-type, buccal swab	3.978	3.053	0.000	1.30		✓		✓		
ETH21-5821-51	Reference-type, buccal swab	3.508	3.367	3.287	1.04			✓	✓		
ETH21-5821-52	Reference-type, buccal swab	14.156	14.547	0.000	0.97			✓	✓		
ETH21-5821-53	Reference-type, buccal swab	9.525	9.913	0.000	0.96		✓		✓		

ETH21-5821-55	Reference-type, buccal swab	20.087	17.082	0.000	1.18				✓		
ETH21-5821-56	Reference-type, buccal swab	1.183	2.747	0.000	0.43		✓		✓		
ETH21-5821-60	Reference-type, buccal swab	1.165	1.332	0.000	0.87				✓		
ETH21-5821-61	Reference-type, buccal swab	1.339	0.811	0.000	1.65			✓	✓		
ETH21-5821-62	Reference-type, buccal swab	1.887	2.289	0.000	0.82			✓	✓		
ETH21-5821-63	Reference-type, buccal swab	5.520	7.049	0.000	0.78			✓	✓		
ETH21-5821-64	Reference-type, buccal swab	1.633	1.426	1.406	1.15			✓	✓		
ETH21-5821-65	Reference-type, buccal swab	1.052	1.260	0.000	0.84			✓	✓		
ETH21-5821-66	Reference-type, buccal swab	1.925	0.657	0.000	2.93			✓	✓		
ETH21-5821-67	Reference-type, buccal swab	0.351	0.329	0.000	1.07			✓	✓		
ETH21-5821-68	Reference-type, buccal swab	1.102	1.070	0.000	1.03			✓	✓		
ETH21-5821-70	Reference-type, buccal swab	1.336	0.494	0.000	2.71			✓	✓	✓	✓
ETH21-5821-71	Reference-type, buccal swab	6.399	7.852	0.000	0.81			✓	✓		
ETH21-5821-72	Reference-type, buccal swab	5.838	6.511	0.001	0.90				✓		
ETH21-5821-73	Reference-type, buccal swab	4.973	7.137	0.000	0.70				✓		
ETH21-5821-74	Reference-type, buccal swab	8.651	11.737	0.000	0.74			✓	✓		
ETH21-5821-75	Reference-type, buccal swab	6.840	11.933	0.000	0.57			✓	✓		
ETH21-5821-76	Reference-type, buccal swab	1.914	2.111	0.000	0.91			✓	✓		
ETH21-5821-77	Reference-type, buccal swab	2.976	3.470	0.000	0.86			✓	✓		
ETH21-5821-78	Reference-type, buccal swab	4.691	5.048	0.000	0.93			✓	✓		
ETH21-5821-79	Reference-type, buccal swab	3.474	5.331	3.262	0.65				✓		
ETH21-5821-80	Reference-type, buccal swab	6.975	8.368	0.000	0.83			✓	✓		
ETH21-5821-81	Reference-type, buccal swab	7.894	12.112	7.813	0.65			✓	✓		
ETH21-5821-82	Reference-type, buccal swab	2.519	2.823	2.569	0.89			✓	✓		
ETH21-5821-83	Reference-type, buccal swab	5.844	9.555	0.000	0.61			✓	✓		
ETH21-5821-84	Reference-type, buccal swab	7.726	9.007	0.000	0.86			✓	✓		
ETH21-5821-85	Reference-type, buccal swab	5.510	6.886	0.000	0.80			✓	✓		
ETH21-5821-86	Reference-type, buccal swab	5.380	4.258	7.499	1.26			✓	✓		
ETH21-5821-87	Reference-type, buccal swab	2.457	2.593	0.000	0.95			✓	✓		
ETH21-5821-88	Reference-type, buccal swab	8.221	13.237	0.000	0.62			✓	✓		
ETH21-5821-89	Reference-type, buccal swab	8.348	12.979	0.000	0.64			✓	✓		
ETH21-5821-90	Reference-type, buccal swab	4.619	6.629	0.000	0.70			✓	✓		
ETH21-5821-91	Reference-type, buccal swab	9.841	14.308	0.000	0.69			✓	✓		
ETH21-5821-92	Reference-type, buccal swab	8.980	8.298	0.000	1.08			✓	✓		
ETH21-5821-93	Reference-type, buccal swab	10.384	10.998	10.833	0.94				✓		

ETH21-5821-94	Reference-type, buccal swab	0.028	0.002	0.028	11.39			✓	✓		
ETH21-5821-95	Reference-type, buccal swab	2.029	1.460	0.000	1.39			✓	✓		
ETH21-5821-96	Reference-type, buccal swab	2.529	3.164	0.000	0.80			✓	✓		
ETH21-5821-97	Reference-type, buccal swab	9.256	10.533	9.958	0.88				✓		
ETH21-5821-98	Reference-type, buccal swab	10.112	10.572	0.000	0.96			✓	✓		
ETH21-5821-99	Reference-type, buccal swab	7.128	9.977	0.000	0.71			✓	✓		
ETH21-5821-100	Reference-type, buccal swab	4.360	5.532	0.000	0.79				✓		
ETH21-5821-101	Reference-type, buccal swab	9.606	12.608	0.000	0.76				✓		
ETH21-5821-102	Reference-type, buccal swab	5.874	6.086	5.341	0.97			✓	✓		
ETH21-5821-103	Reference-type, buccal swab	4.440	7.161	0.000	0.62				✓		
ETH21-5821-104	Reference-type, buccal swab	4.211	5.022	4.508	0.84			✓	✓		
Hair 1	Casework-type, hair	0.000	0.000	0.000	N/A		✓				
Blood 1	Casework-type, blood	0.311	0.293	0.000	1.06		✓		✓		
Blood 2	Casework-type, blood	8.500	14.400	0.000	0.59		✓		✓		
Nail 1	Casework-type, nail	1.366	0.956	0.982	1.43		✓		✓		
Nail 2	Casework-type, nail	0.939	0.747	1.072	1.26	✓	✓		✓		
Tooth 1	Casework-type, tooth	0.038	0.007	0.000	5.79	✓	✓	✓	✓		
Tooth 2	Casework-type, tooth	0.408	0.379	0.000	1.08	✓	✓	✓	✓		
Bone 1	Casework-type, bone	0.327	0.322	0.286	1.01	✓	✓	✓	✓		
Bone 2	Casework-type, bone	0.037	0.016	0.038	2.28		✓	✓	✓		
Bone 3	Casework-type, bone	0.016	0.006	N/A	2.57	✓	✓				
Bone 4	Casework-type, bone	5.372	1.082	5.674	4.96	✓	✓				
Bone 5	Casework-type, bone	0.705	0.009	0.650	82.94	✓	✓				
Bone 6	Casework-type, bone	0.548	0.277	0.000	1.98		✓	✓	✓		
Bone 7	Casework-type, bone	0.703	0.131	0.000	5.36		✓	✓	✓		
Bone 8	Casework-type, bone	0.305	0.133	0.000	2.30		✓	✓	✓		
Bone 9	Casework-type, bone	0.314	0.065	0.393	4.85		✓	✓	✓		
Bone 10	Casework-type, bone	0.151	0.085	0.169	1.78		✓	✓	✓		
Bone 11	Casework-type, bone	0.972	0.553	0.941	1.76		✓	✓	✓		
IDSNP-P399	Reference-type, buccal swab	0.247	0.232	0.289	1.06			✓			
IDSNP-P400	Reference-type, buccal swab	0.231	0.226	0.221	1.02			✓			
IDSNP-P402	Reference-type, buccal swab	0.416	0.410	0.000	1.01			✓			
IDSNP-P403	Reference-type, buccal swab	0.242	0.220	0.000	1.10			✓			
IDSNP-P404	Reference-type, buccal swab	0.527	0.627	0.000	0.84			✓			
IDSNP-P405	Reference-type, buccal swab	0.321	0.323	0.372	0.99			✓			
IDSNP-P406	Reference-type, buccal swab	0.367	0.371	0.381	0.99			✓			
IDSNP-P407	Reference-type, buccal swab	0.312	0.368	0.000	0.85			✓			

IDSNP-P408	Reference-type, buccal swab	0.688	0.669	0.000	1.03			✓			
IDSNP-P409	Reference-type, buccal swab	0.173	0.185	0.000	0.94			✓			
IDSNP-P410	Reference-type, buccal swab	0.413	0.397	0.566	1.04			✓			
IDSNP-P412	Reference-type, buccal swab	0.255	0.321	0.381	0.79			✓			
IDSNP-P413	Reference-type, buccal swab	0.111	0.131	0.175	0.85			✓			
IDSNP-P414	Reference-type, buccal swab	0.652	0.717	0.808	0.91			✓			
IDSNP-P415	Reference-type, buccal swab	0.384	0.445	0.000	0.86			✓			
IDSNP-P416	Reference-type, buccal swab	0.296	0.172	0.000	1.72			✓			
IDSNP-P417	Reference-type, buccal swab	0.205	0.314	0.381	0.65			✓			
IDSNP-P419	Reference-type, buccal swab	0.180	0.169	0.199	1.07			✓			
IDSNP-P420	Reference-type, buccal swab	0.249	0.286	0.310	0.87			✓			
IDSNP-P421	Reference-type, buccal swab	0.252	0.148	0.000	1.70			✓			
IDSNP-P422	Reference-type, buccal swab	0.287	0.241	0.275	1.19			✓			
IDSNP-P423	Reference-type, buccal swab	0.376	0.328	0.430	1.14			✓			
IDSNP-P424	Reference-type, buccal swab	0.513	0.600	0.711	0.86			✓			
IDSNP-P425	Reference-type, buccal swab	0.314	0.375	0.000	0.84			✓			
IDSNP-P426	Reference-type, buccal swab	0.219	0.254	0.284	0.86			✓			
IDSNP-P427	Reference-type, buccal swab	0.101	0.146	0.000	0.69			✓			
IDSNP-P439	Reference-type, buccal swab	0.392	0.487	0.492	0.81			✓			
IDSNP-P440	Reference-type, buccal swab	1.996	0.891	0.000	2.24			✓			
IDSNP-P441	Reference-type, buccal swab	2.654	1.037	0.000	2.56			✓			
IDSNP-P442	Reference-type, buccal swab	0.166	0.207	0.272	0.80			✓			
IDSNP-P443	Reference-type, buccal swab	0.307	0.416	0.478	0.74			✓			
IDSNP-P444	Reference-type, buccal swab	0.075	0.069	0.148	1.07			✓			
IDSNP-P445	Reference-type, buccal swab	0.083	0.084	0.000	0.99			✓			
IDSNP-P446	Reference-type, buccal swab	0.074	0.099	0.148	0.75			✓			
IDSNP-P447	Reference-type, buccal swab	0.109	0.166	0.211	0.65			✓			
IDSNP-P448	Reference-type, buccal swab	0.069	0.120	0.178	0.58			✓			
IDSNP-P449	Reference-type, buccal swab	0.406	0.566	0.000	0.72			✓			
IDSNP-P450	Reference-type, buccal swab	0.244	0.306	0.000	0.80			✓			
IDSNP-P451	Reference-type, buccal swab	0.100	0.166	0.000	0.60			✓			
IDSNP-P452	Reference-type, buccal swab	0.137	0.221	0.000	0.62			✓			
IDSNP-P453	Reference-type, buccal swab	0.136	0.167	0.000	0.81			✓			
IDSNP-P456	Reference-type, buccal swab	0.173	0.245	0.295	0.71			✓			
IDSNP-P458	Reference-type, buccal swab	0.394	0.567	0.621	0.70			✓			

IDSNP-P572	Reference-type, buccal swab	0.058	0.095	0.103	0.61			✓			
IDSNP-P573	Reference-type, buccal swab	0.198	0.329	0.000	0.60			✓			
IDSNP-P576	Reference-type, buccal swab	0.066	0.113	0.123	0.59			✓			
IDSNP-P578	Reference-type, buccal swab	0.429	0.773	0.683	0.55			✓			
IDSNP-P579	Reference-type, buccal swab	0.217	0.295	0.380	0.74			✓			
IDSNP-P580	Reference-type, buccal swab	0.169	0.248	0.222	0.68			✓			
IDSNP-P598	Reference-type, buccal swab	0.186	0.297	0.318	0.63			✓			
IDSNP-P605	Reference-type, buccal swab	0.092	0.170	0.208	0.54			✓			
IDSNP-P607	Reference-type, buccal swab	0.142	0.195	0.264	0.73			✓			
IDSNP-P608	Reference-type, buccal swab	0.113	0.158	0.186	0.72			✓			
IDSNP-P610	Reference-type, buccal swab	0.101	0.130	0.163	0.78			✓			
IDSNP-P611	Reference-type, buccal swab	0.243	0.334	0.433	0.73			✓			
IDSNP-P613	Reference-type, buccal swab	0.093	0.145	0.000	0.64			✓			
IDSNP-P615	Reference-type, buccal swab	0.082	0.119	0.134	0.70			✓			
Individual B	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓
Individual C	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓
Individual D	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓
Individual E	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓
Individual F	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓
Individual G	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓
Individual H	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓
Individual I	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓
Individual J	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓
Individual K	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓
Individual L	Microarray data uploaded to GEDmatch™	N/A	N/A	N/A	N/A						✓