scientific reports

OPEN

Genomic alterations and possible druggable mutations in carcinoma of unknown primary (CUP)

Hamidreza Aboulkheyr Es[1,2], Hamid Mahdizadeh[1,3], Amir Abbas Hedayati Asl[3] & Mehdi Totonchi[1,3,4,5]✉
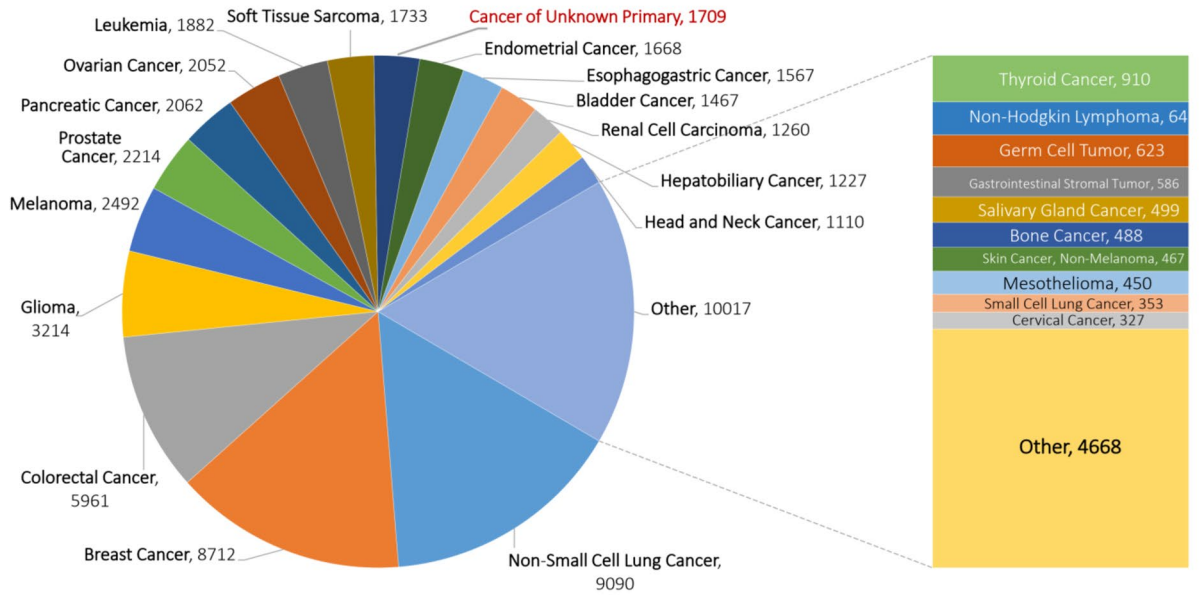
Carcinoma of Unknown Primary (CUP) is a heterogeneous and metastatic disease where the primary site of origin is undetectable. Currently, chemotherapy is the only state-of-art treatment option for CUP patients. The molecular profiling of the tumour, particularly mutation detection, offers a new treatment approach for CUP in a personalized fashion using targeted agents. We analyzed the mutation and copy number alterations profile of 1709 CUP samples deposited in the AACR Project Genomics Evidence Neoplasia Information Exchange (GENIE) cohort and explored potentially druggable mutations. We identified 52 significant mutated genes (SMGs) among CUP samples, in which 13 (25%) of SMGs were potentially targetable with either drugs are approved for the know primary tumour or undergoing clinical trials. The most variants detected were *TP53* (43%), *KRAS* (19.90%), *KMT2D* (12.60%), and *CDKN2A* (10.30%). Additionally, using pan-cancer analysis, we found similar variants of *TERT* promoter in CUP and NSCLC samples, suggesting that these mutations may serve as a diagnostic marker for identifying the primary tumour in CUP. Taken together, the mutation profiling analysis of the CUP tumours may open a new way of identifying druggable targets and consequently administrating appropriate treatment in a personalized manner.

Of all patients diagnosed with cancer, 2% present as metastatic carcinoma of unknown primary site (CUP)[1]. It is classified as any metastatic epithelial tumour where, following extensive clinical history, physical examination, radiological studies and histopathological investigations, failed to identify the primary site of tumours[2]. According to the European Society for Medical Oncology (ESMO) guidelines for the treatment of patients with favourable-risk CUP, the administration of various regimens of chemotherapy alone or in combination with radiotherapy or hormonal therapy has been proposed as only standard treatment guidelines[3]. Because of CUP tumour heterogeneity, the current clinical trials are challenging to perform, resulting in a poor prognosis with a median survival of less than 12 months and 5-year survival of 14%[4]. Thus, there is an urgent need to improve treatment modalities and prolong patients' survival with CUP[5].
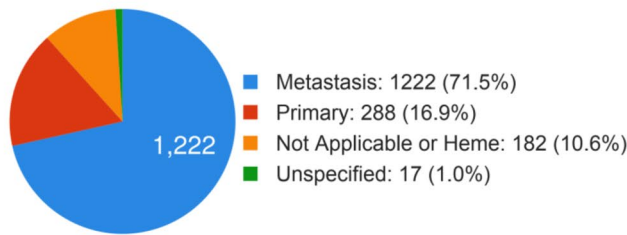
Personalized cancer medicine using genomics technologies opened new ways to treat various types of cancer using the identification of targetable mutations[6–10]. Recent studies have highlighted the crucial role of precision medicine in patient stratification and the selection of effective treatment in malignant types of cancer[11–17]. Moreover, several studies have reported improved overall survival in patients with advanced and metastatic cancers who have received genetically matched targeted therapies[18,19]. In CUP tumours, the implementation of this approach may improve treatment by targeting tumour-specific and druggable somatic variants in a personalized manner[4]. The AACR Project Genomics Evidence Neoplasia Information Exchange (GENIE) has recently collected the genomic information, including mutations and copy number variation of the wide range of solid tumours, including CUP from both primary and metastatic tumours[20–22]. Using these public data, we analyzed the genomic mutations and copy number alterations of 1709 CUP samples to provide insight into the genetic makeup of these tumours and determined potentially druggable targets.

[1]Department of Genetics, Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, P.O. Box: 16635-148, Tehran, Iran. [2]School of Biomedical Engineering, University of Technology Sydney, Sydney, Australia. [3]Department of Stem Cells and Developmental Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran. [4]School of Biological Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran. [5]Gasteroeterology and Liver Diseases Research Center, Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran. ✉email: m.totonchi@royaninstitute.org
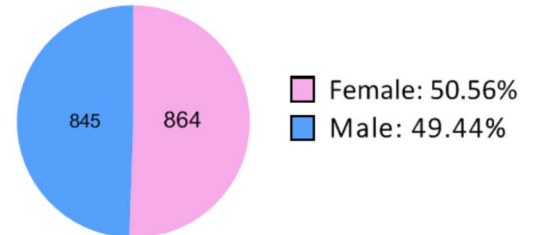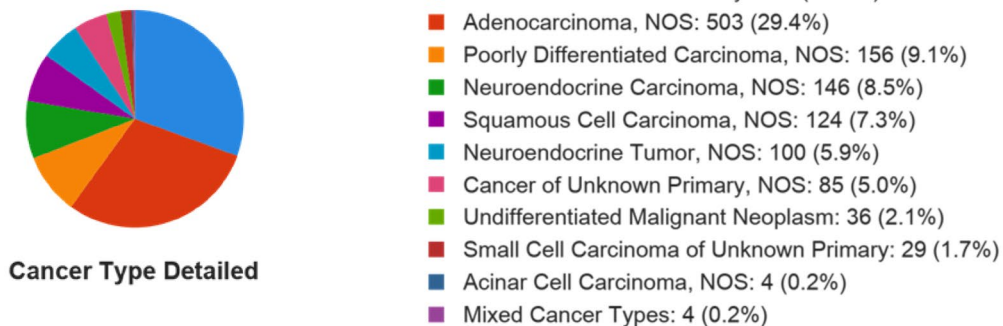
A

B

CUP samples

- Metastasis: 1222 (71.5%)
- Primary: 288 (16.9%)
- Not Applicable or Heme: 182 (10.6%)
- Unspecified: 17 (1.0%)

D

Gender

- Female: 50.56%
- Male: 49.44%

C

Cancer Type Detailed

- Cancer of Unknown Primary: 522 (30.5%)
- Adenocarcinoma, NOS: 503 (29.4%)
- Poorly Differentiated Carcinoma, NOS: 156 (9.1%)
- Neuroendocrine Carcinoma, NOS: 146 (8.5%)
- Squamous Cell Carcinoma, NOS: 124 (7.3%)
- Neuroendocrine Tumor, NOS: 100 (5.9%)
- Cancer of Unknown Primary, NOS: 85 (5.0%)
- Undifferentiated Malignant Neoplasm: 36 (2.1%)
- Small Cell Carcinoma of Unknown Primary: 29 (1.7%)
- Acinar Cell Carcinoma, NOS: 4 (0.2%)
- Mixed Cancer Types: 4 (0.2%)

Figure 1. Overview of the GENIE database. Distribution of tumour types among cases successfully sequenced and analyzed in this cohort.

Results

Clinical characteristic of samples. In total, 45,048 samples across 17 cancer types, including CUP, were included in this study. The sample type distribution was 24,567 primary and 15,484 metastasis tumours in GENIE cohorts. The hotspot regional mutations and copy number variations of these samples were available from GENIE and cBioportal. According to the information provided by GENIE, we divided samples into more than17 broader cancer types, including CUP samples (Fig. 1A). The cancer types containing the most samples were non-small cell lung cancer (9090 (15.3%)), breast invasive ductal carcinoma (8712 (14.7%)), colorectal can-

cer (5961 (10.0%)), Glioma (3214 (5.4%)), Melanoma (2492 (4.2%)), prostate cancer (2214 (3.7%)). The number of CUP samples registered in this cohort was 1709 (2.9%), dividing into 1222 metastatic (71.5%), 288 primaries (16.9%), 182 (10.6%) not applicable or heme and 17 (1.0%) unspecified (Fig. 1B) according to the ICD-O, ICD-O-3, and MSKCC OncoTree ontology classification. In addition, the CUP samples comprised various Not Otherwise Specified (NOS) cancer types, including but not limited to adenocarcinoma (503 (29.4%)), poorly differentiated carcinoma (156 (9.1%)), neuroendocrine carcinoma (146, (8.5%), and squamous cell carcinoma (124, (7.3%)) (Fig. 1C).For gender information among CUP patients, 864 (50.56%) of patients were female, and 845 (49.44%) were male (Fig. 1D).
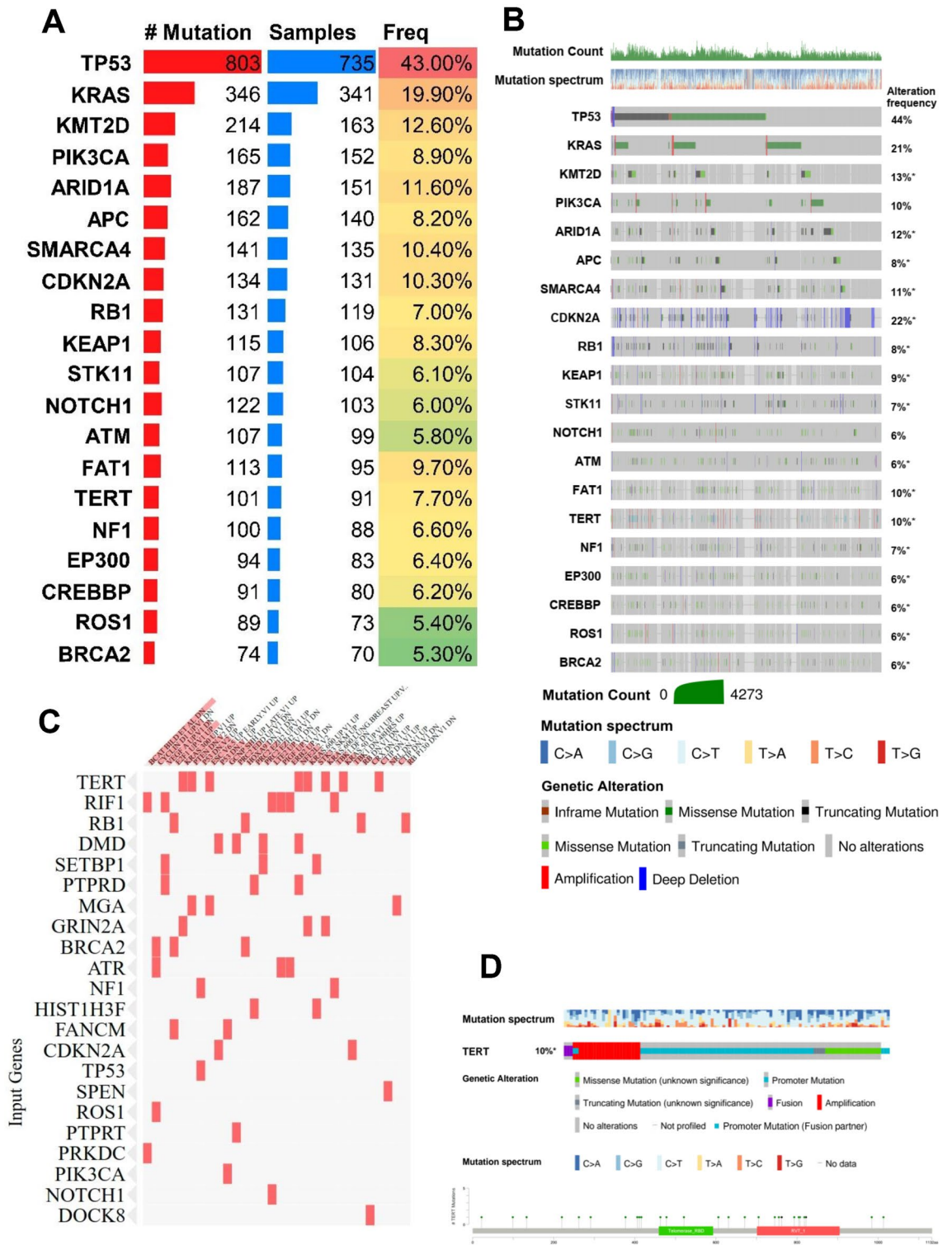
Significantly mutated genes (SMG) in CUP samples. We analyzed the most genomic mutations of hotspot regions at the gene level in CUP samples in GENIE according to the previously developed method[9,10]. In total, 52 SMG was identified (Fig. 2A; Supplementary Table 1). Among SMGs, the mutation rate of *TP53, KRAS, ARID1A, SMARCA4* and *KMT2D* were recorded significantly higher than other identified SMGs (Fig. 2B, Supplementary Table 1). The pathway enrichment analysis of identified SMGs resulted in SMGs' involvement in a wide range of cellular processes (Fig. 2C, Supplementary Table 2), including transcription factors/regulators, receptor tyrosine kinase signalling, cell cycle, IGF pathway-protein kinase B signalling, phosphatidylinositol-3-OH kinase (PI(3)K) signalling, Wnt/β-catenin signalling, *PDGF, FGF, EGF, TGF-β,* and Notch signalling pathways and integrin signalling pathway. The identification of MAPK, PI(3)K and Wnt/β-catenin signalling pathways is consistent with classical cancer studies. Notably, almost all samples had at least one non-synonymous mutation in one of the SMG. The average number of point mutations in these genes varies across samples, with the highest (512 mutations for *TP53* across 727 cases) and the lowest (15 mutations for *GLI3* across 15 cases) (Fig. 2B. Supplementary Table 1). This suggests that the numbers of both cancer-related genes (52 identified in this study) and cooperating driver mutations required during oncogenesis are few, although large-scale structural rearrangements were not included in this analysis. Interestingly, in line with the previous study performed by Zehir et al.[9] highlighting *TERT* promoter mutations across few primary tumours, we observed a similar mutation of *TERT* promoter among CUP samples (n = 91) (Fig. 2D). Although the clinical relevance of mutations in the *TERT* promoter remains incompletely understood, our results reaffirm the high prevalence of these alterations in patients with advanced solid tumours and suggest an association with disease progression and poor outcome. Additionally, the presence of similar mutations of *TERT* promoter in CUP and NSCLC samples suggests these mutations may serve as a diagnostic marker for identifying the primary tumour in CUP patients.

Mutual exclusivity and co-occurrence among SMGs. The 1035 pair-wise exclusivity and co-occurrence analysis of the 52 SMGs found 198 mutually exclusive (*P* value < 0.001) and 837 co-occurring (*P* value < 0.001) pairs (Fig. 3 and Supplementary Table 3) among cup samples. Pairs with significant exclusivity were include *KRAS* and *FAT1*, *KRAS* and *NOTCH3*, *KRAS* and *NF1*, *Kras* and *DMD* and *CDKN2A* and *RB1* in CUP samples. Additionally, the cohort analysis identified pairs with significant co-occurrence, including *KRAS* and *APC*, *TP53* and *APC*, *KRAS* and *CDKN2A*, *KRAS* and *STK11*, *KRAS, KEAP1*, and *SMARCA4* and *KEAP1,* highlighting the importance of these oncogenes in CUP tumours.

Copy number alteration among cup samples. The copy number variation differences within CUP samples resulted in the identification of 624 frequently amplified/deleted regions. Significant amplification of *MYC*, *FGF4* and *FGF19* observed in a small fraction of patients (Fig. 4A), while deletion of cell cycle-related genes *CDKN2B* and *CDKN2A* were detected in only 10% and 20% of patients, respectively (Fig. 4A). Further, we analyzed copy number alteration of the CUP-SMGs within CUP samples (Fig. 4B) and across primary tumours of 14 cancer types registered in GENIE (Fig. 4C, Supplementary Table 4). Among CUP samples, a deep deletion of *TP53*, *RB1*, *CDKN2A*, and *STK11* and amplification of *KRAS* and *PIK3CA* were observed. In a pan-cancer analysis, amplification of *KRAS* and *PIK3CA* in the breast (66 and 114 of cases) and non-small cell lung cancer (46 and 48 of cases), *TERT* in non-small cell lung cancer (114 of cases) and *ATR* in breast cancer (36 of cases), were the most amplified genes, while deletion in *CDKN2A* in glioma (676 of cases), *RB1* and *TP53* in small cell lung cancer (15 of cases) were observed in these 14 different cancer types (Fig. 4C). Among these genes with significantly altered copy numbers between CUP and primary tumours, a significant amplification of *TERT* promoter was observed in both CUP and non-small cell lung cancer samples compared to glioma and breast primary tumours suggesting that copy number variation of *TERT* may play a diagnostic role for the identification of the origin of CUP tumours (Fig. 4D).

Mutation frequency of CUP-SMGs across 17 known primary tumours. To identify similar and targetable mutation patterns in CUP, we analyzed and compared the genomic alteration frequency of identified CUP-SMGs in primary tumour types across 17 cancer types in GENIE (Fig. 5A). The majority of CUP-SMGs mutations were enriched in non-small cell lung cancer (4221 cases), colon cancer (4011 cases) and breast cancer (3376 cases) (Fig. 5A).

The most frequently mutated gene in this cohort was *TP53* (44% total samples) (Fig. 5B). Its mutations predominate in non-small cell lung cancer (46.36%; 2517 cases), colon cancer (65.55%; 2365 cases) and breast cancer (36.26%; 2060 cases) (Fig. 5B). *KRAS* is the second most commonly mutated gene, occurring frequently (> 10%) in most cancer types (pancreatic: 74.6%, colon cancer:44.24%, non-small cell lung cancer:30.93%) except hepatobiliary carcinoma, cervical cancer, bladder cancer, thyroid cancer, melanoma, small-cell lung cancer, head and neck carcinoma, prostate and breast cancer (Fig. 5B). *PIK3CA* mutations were frequented in breast cancer (36.7%) and cervical cancer (25.14%), being specifically enriched in luminal subtype tumours. Many cancer types carried mutations in chromatin re-modelling genes. In particular, histone-lysine N-methyltransferase genes

Figure 2. The most significant mutated genes in CUP samples. (**A**) Mutation frequency of SMGs. Genes with a cohort-level alteration frequency of > 5% or a tumour type–specific alteration frequency of > 30% are displayed. (**B**) Genomic alterations of 52 SMGs within CUP samples. (**C**) Pathway enrichment analysis of SMGs from MSigDB. (**D**) Genomic alterations identified in TERT promoter among CUP samples.(Mutation frequency were analyzed using the MuSiC suite version (0.4) and Mutatiom Mapper under cBioportal (https://www.cbioportal.org/mutation_mapper).
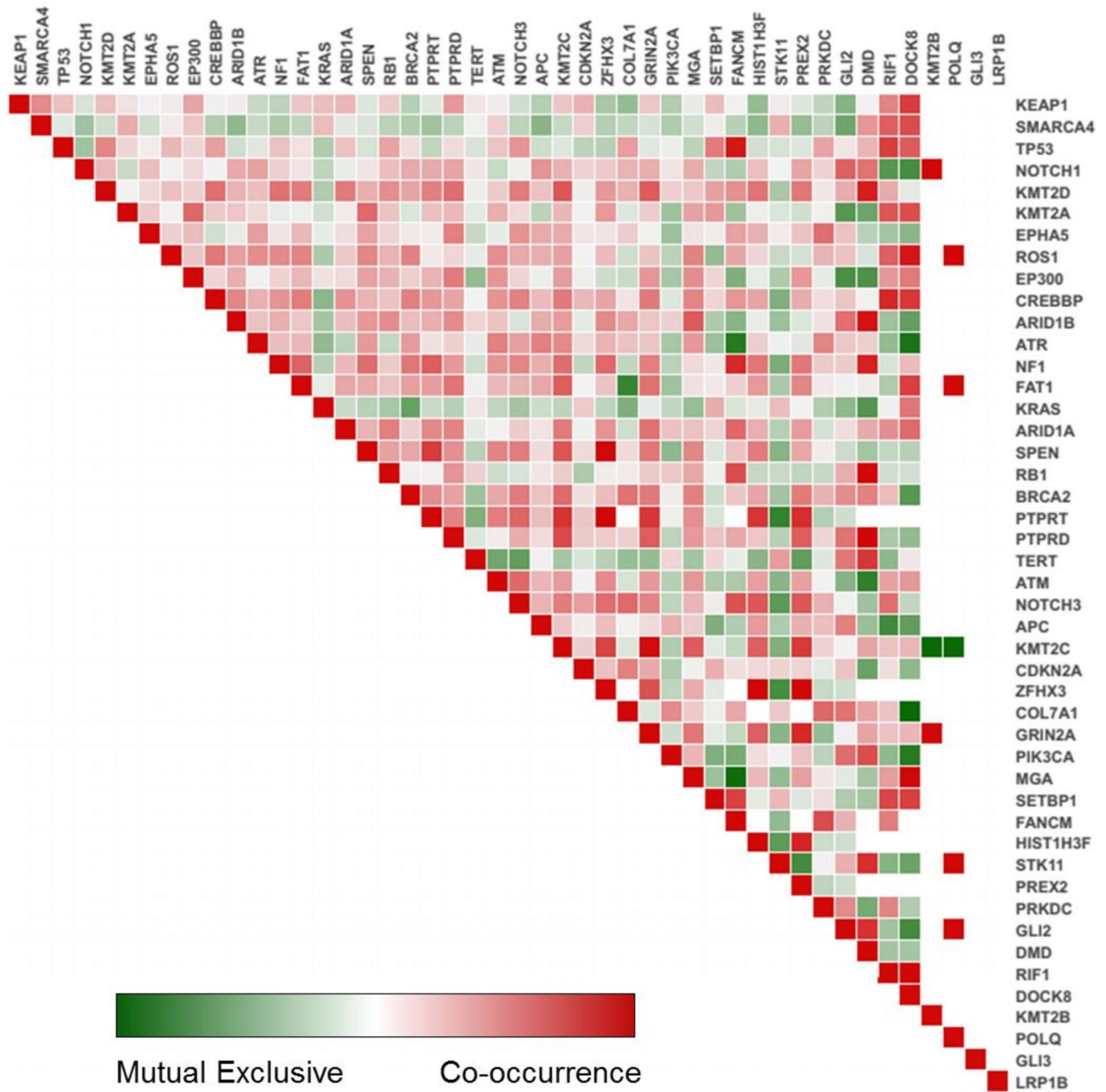
Figure 3. Mutual exclusivity and co-occurrence between identified SMG in CUP. The data retrieved from cbiopotal (https://genie.cbioportal.org) and anayized using Mutual exclusivity module under GiTools software (version 2.3.1).

KMT2D, *KMT2C* and *KMT2B* in bladder, lung and endometrial cancers, whereas the *KMT2A* is mostly mutated in non-small cell lung cancer and colon cancer. Mutations in *ARID1A* are frequent in non-small cell lung cancer, colon cancer, bladder cancer and breast cancer, whereas mutations in *KEAP1* and *STK11* was predominate in non-small cell lung cancer (8.62% and 11.75%, respectively) (Fig. 5B). *KRAS* mutations are typically mutually exclusive, with recurrent activating mutations (*KRAS* (Gly 12) and *KRAS* (Gly 13) common in colon cancer, non-small cell lung cancer and pancreatic cancer. We compared the most common hotspot mutations in *KRAS* between CUP, and other *KRAS* mutation enriched cancer types (Fig. 5C). The comparison of hotspot mutations resulted in the enrichment of G12D and G12R in pancreatic cancer, G12C, G12F and G13C in non-small cell lung cancer and CUP samples. These data highlight similarity of *KRAS* hotspot mutations between CUP and NSCLC.

Targetable mutations and drug candidates. To identify or predict possible therapeutics based on genomic alterations identified in SMG in CUP samples, we performed a gene-drug association analysis using PanDrugs platforms[23]. The gene-drug associations classified into two groups called "Drug targets" in which drugs can directly target genes that contribute to disease phenotype, and "Biomarkers" where genes are representing a drug-response associated status while its protein products are not targetable[23]. From 262 identified
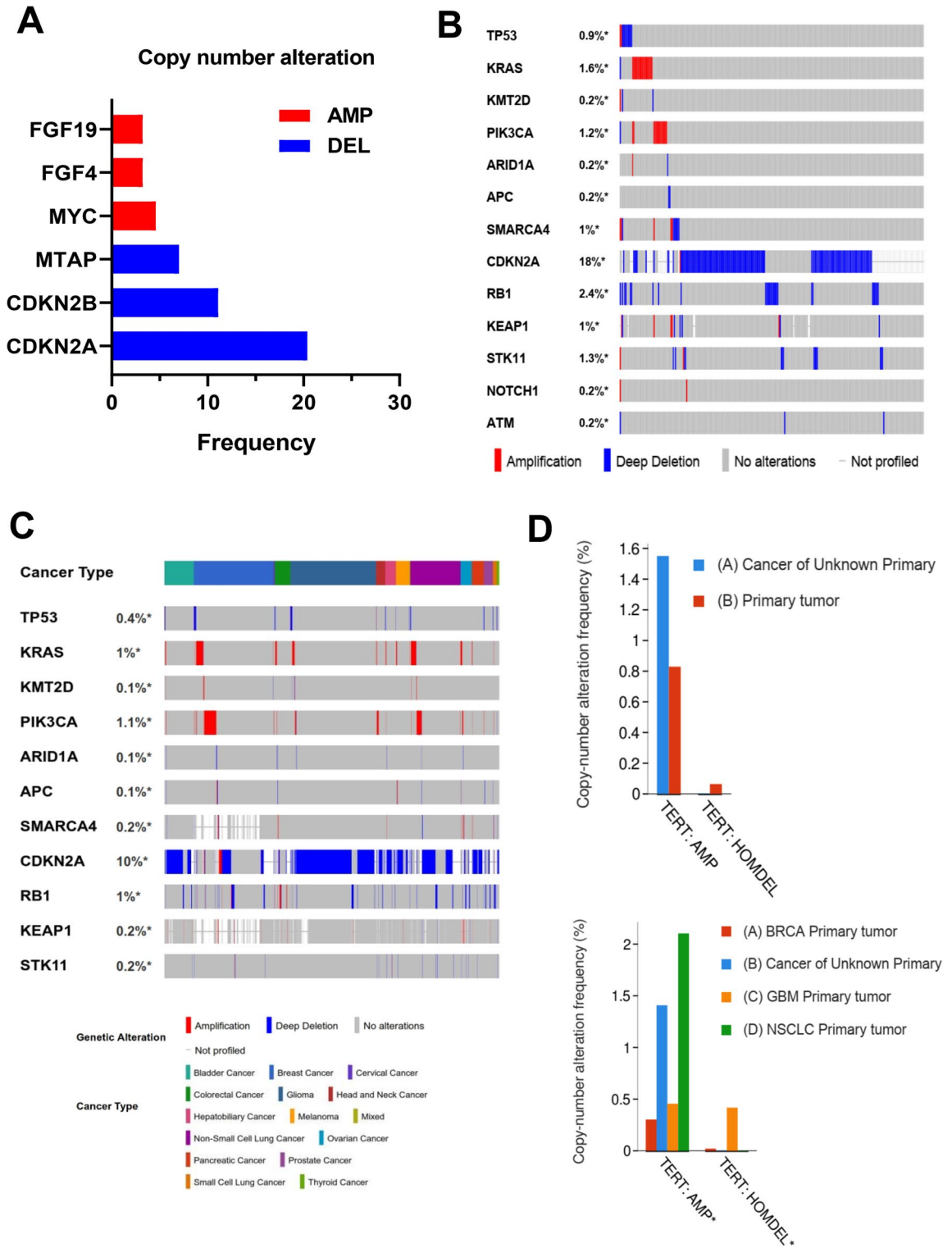
Figure 4. Copy number variation among CUP and other cancer types. (**A**) major copy number alteration detected in CUP samples. (**B**) Amplification and deletion status of identified CUP-SMGs within CUP samples and (**C**) other known primary tumours registered in GENIE database. (**D**) copy number variation analysis of *TERT* between CUP and all primary tumours (up panel) and NSCLC, GBM and BRCA (bottom panel). (data analyzed using cBioportal (http://www.cbioportal.org) version v3.2.11.
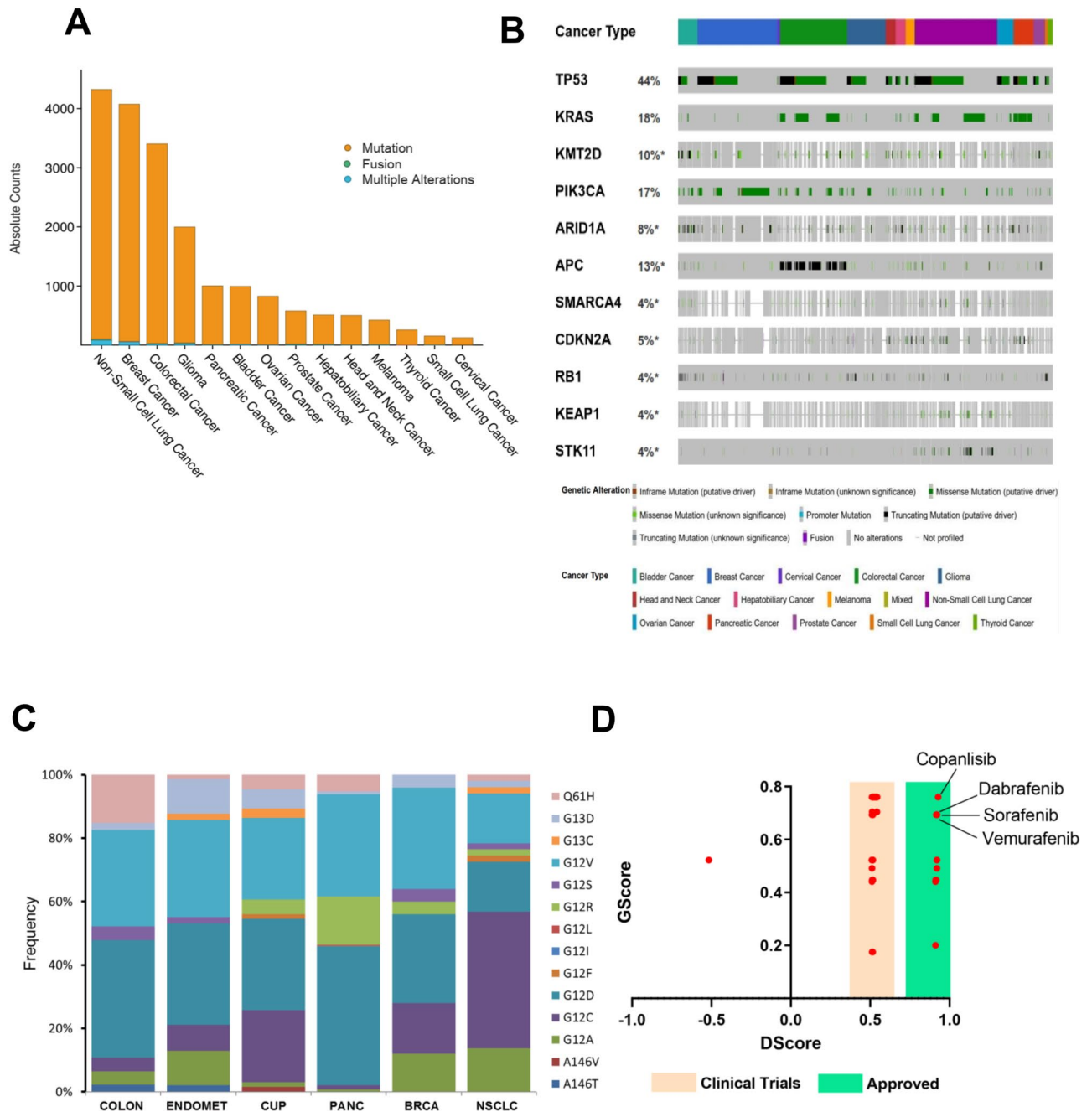
Figure 5. Mutation frequency of CUP-SMGs across 17 cancer types. (**A**, **B**) Distribution of genomic alterations in 52 CUP-SMGs across primary tumours of 17 different cancer types in GENIE cohort. (The oncoplot was generated using (https://genie.cbioportal.org). (**C**) Distribution of hotspot mutations identified in *KRAS* among six different cancer types including CUP. (**D**) The results of gene-drug association analysis using PanDrug platforms. The best candidate drugs with highest GScore and Dscore are labeled.

interactions, 8.7% (23/262) was classified as a direct drug target, while 91.3% (239/262) of gene-drug interactions identified as Biomarker (Fig. 5D). Interestingly, we found five FDA approved drugs, Crizotinib (GScore: 0.76. Dscore: 0.95) and Copanlisib (GScore: 0.76. Dscore: 0.92), Debrafenib, Sorafenib, Vemurafenib, and Regorafenib as best candidates for targeting ALK/MET, PIK3CA, and BRAF inhibitors, respectively (Fig. 5D. Supplementary Table 5). Moreover, various off-label and clinically investigating compounds for targeting mutated *KRAS* were identified, although the GScore and DScore of these compounds did not reach a high score (Supplementary Table-5). Everolimus (mTOR inhibitor) and Bortezomib (26S proteasome inhibitor) were identified with the highest GScore and DScore compared to the other drugs candidates in this group (Supplementary Table-5). Taken together, these data highlight presence of at least one druggable variant and the potential of using genomic alteration guided targeted therapy in CUP patients.

Discussion

Currently, combination chemotherapy regimens have been considered as the first-line of therapy for CUP patients[24]. Personalized cancer therapy using the identification of druggable mutations has encouraged mutational profiling of various types of tumours, including metastasis tumours, for instance CUP[25-27]. This study analyzed the most significant mutated genes and identified the most prevalent variants in 1709 CUP samples. The gene-drug association studies suggested that at least one of the identified variants is linked to the known ,and approved targeted therapy agents or therapeutics are currently in clinical trial studies highlighting the potential of genomic alteration-based treatment approach for a patient with CUP. In line with this concept, numerous clinical studies have been reported durable treatment responses using mutation-matched targeted therapies drugs, including EGFR, BRAF, KIT, and MET[18,28-30].

Currently, targeted therapy agents Crizotinib and Copanlisib approved for the treatment of tumours that harbour mutations in *ROS1/MET/ALK* and *PIK3CA*, while therapeutic agents for the other identified variants, including *FGFR* family, *MYC*, *MET*, and *KRAS* are currently under investigation in active and ongoing clinical trials. A large proportion of the mutations detected in this study are associated with various signal transduction pathways, apoptotic regulation, cell cycle progression, and receptor tyrosine kinase signalling regulations. These results can be promising because the majority of available targeted drugs act through targeting one of these pathways, which are commonly altered in various types of cancer with known primary tumours[31-35]. The most mutated gene identified in this study was *TP53* (43%, 743/1709), with numerous non-synonymous coding region variants. Similar to these data, previous studies demonstrated the association of *TP53* mutations in metastatic progression in multiple cancer types, supporting the presence of high mutation load on *TP53* reported in CUP[36,37].

Other common variants detected in this cohort were observed in genes involved in activating and regulating key signal transduction pathways, including *BRAF* and *KRAS*. This is the first study to report various codon 12 variants of *KRAS* in CUP samples. The detection of codon 12 mutations in this cohort is consistent with the highly aggressive behaviours of CUP tumours[25,29]. Furthermore, characterizing the mutational status of *KRAS* has become clinically relevant in some malignancies because the presence of a *KRAS* mutation is known to stimulate resistance to some tyrosine kinase inhibitors[38-41]. Although currently no approved therapeutic agent to target and inhibit mutant *KRAS* activity available; however, recent clinical studies reported a partial response in CUP patients with a *KRAS*(G12D) mutation treated with Trametinib (MEK inhibitor)[30,42]. In this study, we also observed *KRAS*(G12C) variant in 25% of CUP samples. Recent promising results from Sotorasib (AMG-510); a specific covalent inhibitor of *KRAS*(G12C) in NSCLC suggest detecting this variant of *KRAS* as a possible druggable target in CUP patients[43]. Moreover, targeting *KRAS*(G12C) using Sotorasib in advanced solid tumours showed an encouraging anticancer activity which might be useful in CUP[44].

Similar to other studies, we also identified activating *BRAF* (V600E) mutations in 4.3% (74/1709) cases[24-26]. This offers the potential of using BRAF inhibitors such as Vemurafenib and Dabrafenib for CUP with *BRAF* (V600E) mutations. In line with these, through the gene-drug association analysis, we also observed a high GScore and DScore of BRAF inhibitors Dabrafenib and Vemurafenib for targeting V600E variant identified in CUP samples. Moreover, a clinical study showed a complete clinical response of CUP patients treated with *BRAF*(V600E) targeted therapy Vemurafenib in combination with immunotherapy agent Ipilimumab[45].

Mutations in *MET* and *ERBB2* (HER2) amplification were detected in 30 and 27 of cases,respectively, suggesting the possibility of targeting these receptor tyrosine kinase[28]. Targeting MET using Crizotinib for patients without exon-14 skipping combined with HER2 inhibitor Trastuzumab has been shown with success in CUP patients. The current success of HER2 and MET targeted therapies using Trastuzumab (for cases with a HER2 amplification status) and Crizotinib in a combination manner in advanced and metastatic tumours including *HER2* amplified and *MET*-mutant CUP tumours, suggest the further evaluation of these genes as druggable targets in patients with CUP[46]. Our results support those of other CUP studies highlighting the value of sequencing techniques, particularly gene mutation detection, to identify actionable targets[11,24-27].

Taken together, these data highlight the molecular heterogeneity of CUP tumours. The mutations detected across the majority of CUP cases included in this study highlight not only the genomic instability present in these tumours but also the potential application of targeted therapies for a significant proportion of patients with CUP, which might improve the prognosis and therapeutic decisions for these patients[12].

Material and methods

Data collection.   GENIE v5.0 provided the mutation, copy number variation, gene fusion and clinical information of 59,442 tumour samples[21]. Most onco-types were classified into 17 categories according to Oncotree (http://oncotree.mskcc.org/oncotree/). The onco-types not included in these 17 categories were excluded from our analysis. Raw data were downloaded from Synapse (syn17112456, https://www.synapse.org/) and provided by the GENIE project using either R commands or cbioportal (https://genie.cbioportal.org/)[47,48]. The preprocessing protocols for these data are described in the GENIE-provided data guide.

Significantly mutated genes (SMG) analysis.   The SMG analysis performed according to the previously developed criteria and protocols[20,21]. We used the MuSiC suite[49] to identify significant genes for CUP samples and also for Pan-Cancer tumours according. This test assigns mutations to seven categories: AT transition, AT transversion, CG transition, CG transversion, CpG transition, CpG transversion and indel, and then uses statistical methods based on convolution, the hypergeometric distribution (Fisher's test P value < 0.05), and likelihood to combine the category-specific binomials to obtain overall P values. Notably, the genes with a cohort level alteration frequency of ≥ 5% or a tumour type-specific alteration frequency of ≥ 30% were included in our analysis, while tumours having no mutation, or more than 500 mutations were excluded in this study. Dif-

ferentially mutated sites were plotted using Mutation-Mapper module in cBioportal. (https://www.cbioportal.org/mutation_mapper).

Copy number variation analysis. Copy number alteration data were available at AACR Project GENIE, in cbioportal. In the present study, we selected the 17 most common cancer types for comparing their copy number variation frequencies with CUP samples. We calculated the changes in the average frequency of copy number variation (amplification and deletion) of CUP and Pan-cancer samples using provided R code in cbioportal.

Mutual exclusivity and co-occurrence analysis. We used Fisher's exact test to identify pairs of SMGs with significant (*P* value < 0.001 by Benjamini–Hochberg) exclusivity and co-occurrence. We identified significant pairs by analyzing all CUP samples together. Then we used a de novo driver exclusivity algorithm known as Dendrix[50] to identify sets of approximately mutually exclusive mutations on all samples together. The plotting for mutual exclusivity and co-occurrence was performed using Gitools software (version 2.3.1)[51].

Data availability

The genomic data from the GENIE dataset used in this study are openly available for download in https://www.synapse.org, reference number [syn17112456][21]. All data generated and described in this article are available from the corresponding web servers and portal and are freely available to download for noncommercial purposes, without breaching participant confidentiality.

References

1. Rassy, E. & Pavlidis, N. The currently declining incidence of cancer of unknown primary. *Cancer Epidemiol.* **61**, 139–141. https://doi.org/10.1016/j.canep.2019.06.006 (2019).
2. Oien, K. A. & Dennis, J. L. Diagnostic work-up of carcinoma of unknown primary: From immunohistochemistry to molecular profiling. *Ann. Oncol.* **23**(Suppl 10), x271–x277. https://doi.org/10.1093/annonc/mds357 (2012).
3. Fizazi, K. *et al.* Cancers of unknown primary site: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **26**(Suppl 5), v133–v138. https://doi.org/10.1093/annonc/mdv305 (2015).
4. Moran, S., Martinez-Cardus, A., Boussios, S. & Esteller, M. Precision medicine based on epigenomics: The paradigm of carcinoma of unknown primary. *Nat. Rev. Clin. Oncol.* **14**, 682–694. https://doi.org/10.1038/nrclinonc.2017.97 (2017).
5. Hainsworth, J. D. & Greco, F. A. Cancer of unknown primary site: New treatment paradigms in the era of precision medicine. *Am. Soc. Clin. Oncol. Educ. Book* **38**, 20–25. https://doi.org/10.1200/EDBK_100014 (2018).
6. Frampton, G. M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* **31**, 1023–1031. https://doi.org/10.1038/nbt.2696 (2013).
7. Campbell, J. D. *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616. https://doi.org/10.1038/ng.3564 (2016).
8. Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291–304 e296. https://doi.org/10.1016/j.cell.2018.03.022 (2018).
9. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713. https://doi.org/10.1038/nm.4333 (2017).
10. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339. https://doi.org/10.1038/nature12634 (2013).
11. Varghese, A. M. *et al.* Clinical and molecular characterization of patients with cancer of unknown primary in the modern era. *Ann. Oncol.* **28**, 3015–3021. https://doi.org/10.1093/annonc/mdx545 (2017).
12. Rassy, E. & Pavlidis, N. Progress in refining the clinical management of cancer of unknown primary in the molecular era. *Nat. Rev. Clin. Oncol.* **17**, 541–554. https://doi.org/10.1038/s41571-020-0359-1 (2020).
13. El Rassy, E. & Pavlidis, N. The current evidence for a biomarker-based approach in cancer of unknown primary. *Cancer Treat. Rev.* **67**, 21–28. https://doi.org/10.1016/j.ctrv.2018.04.011 (2018).
14. Hainsworth, J. D. *et al.* Molecular gene expression profiling to predict the tissue of origin and direct site-specific therapy in patients with carcinoma of unknown primary site: A prospective trial of the Sarah Cannon research institute. *J. Clin. Oncol.* **31**, 217–223. https://doi.org/10.1200/JCO.2012.43.3755 (2013).
15. Hayashi, H. *et al.* Randomized phase II trial comparing site-specific treatment based on gene expression profiling with carboplatin and paclitaxel for patients with cancer of unknown primary site. *J. Clin. Oncol.* **37**, 570–579. https://doi.org/10.1200/JCO.18.00771 (2019).
16. Moran, S. *et al.* Epigenetic profiling to classify cancer of unknown primary: A multicentre, retrospective analysis. *Lancet Oncol.* **17**, 1386–1395. https://doi.org/10.1016/S1470-2045(16)30297-2 (2016).
17. Rassy, E. *et al.* The role of site-specific therapy for cancers of unknown of primary: A meta-analysis. *Eur. J. Cancer.* **127**, 118–122. https://doi.org/10.1016/j.ejca.2019.12.016 (2020).
18. Tan, D. S. *et al.* Molecular profiling for druggable genetic abnormalities in carcinoma of unknown primary. *J. Clin. Oncol.* **31**, e237-239. https://doi.org/10.1200/JCO.2012.44.3937 (2013).
19. Robinson, D. R. *et al.* Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297–303. https://doi.org/10.1038/nature23306 (2017).
20. Micheel, C. M. *et al.* American Association for Cancer Research Project Genomics Evidence Neoplasia Information Exchange: From inception to first data release and beyond-lessons learned and member institutions' perspectives. *JCO Clin. Cancer Inform.* **2**, 1–14. https://doi.org/10.1200/CCI.17.00083 (2018).
21. Consortium, A. P. G. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov.* **7**, 818–831. https://doi.org/10.1158/2159-8290.CD-17-0151 (2017).
22. Litchfield, K., Turajlic, S. & Swanton, C. The GENIE is out of the bottle: Landmark cancer genomics dataset released. *Cancer Discov.* **7**, 796–798. https://doi.org/10.1158/2159-8290.CD-17-0604 (2017).
23. Pineiro-Yanez, E. *et al.* PanDrugs: A novel method to prioritize anticancer drug treatments according to individual genomic data. *Genome Med.* **10**, 41. https://doi.org/10.1186/s13073-018-0546-1 (2018).
24. Gatalica, Z. *et al.* Comprehensive tumor profiling identifies numerous biomarkers of drug response in cancers of unknown primary site: Analysis of 1806 cases. *Oncotarget* **5**, 12440–12447. https://doi.org/10.18632/oncotarget.2574 (2014).

25. Ross, J. S. *et al.* Comprehensive genomic profiling of carcinoma of unknown primary site: New routes to targeted therapies. *JAMA Oncol.* **1**, 40–49. https://doi.org/10.1001/jamaoncol.2014.216 (2015).
26. Loffler, H. *et al.* Molecular driver alterations and their clinical relevance in cancer of unknown primary site. *Oncotarget* **7**, 44322–44329. https://doi.org/10.18632/oncotarget.10035 (2016).
27. Tothill, R. W. *et al.* Massively-parallel sequencing assists the diagnosis and guided treatment of cancers of unknown primary. *J. Pathol.* **231**, 413–423. https://doi.org/10.1002/path.4251 (2013).
28. Stella, G. M. *et al.* MET mutations in cancers of unknown primary origin (CUPs). *Hum. Mutat.* **32**, 44–50. https://doi.org/10.1002/humu.21374 (2011).
29. Palma, N. A. *et al.* Durable response to crizotinib in a MET-amplified, KRAS-mutated carcinoma of unknown primary. *Case Rep. Oncol.* **7**, 503–508. https://doi.org/10.1159/000365326 (2014).
30. Kato, S. *et al.* Utility of genomic analysis in circulating tumor DNA from patients with carcinoma of unknown primary. *Cancer Res.* **77**, 4238–4246. https://doi.org/10.1158/0008-5472.CAN-17-0628 (2017).
31. Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: An evolving paradigm. *Nat. Rev. Cancer* **13**, 714–726. https://doi.org/10.1038/nrc3599 (2013).
32. Ramirez, M. *et al.* Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat. Commun.* **7**, 10690. https://doi.org/10.1038/ncomms10690 (2016).
33. Gottesman, M. M. Mechanisms of cancer drug resistance. *Annu Rev. Med.* **53**, 615–627. https://doi.org/10.1146/annurev.med.53.082901.103929 (2002).
34. Yuan, T. L. & Cantley, L. C. PI3K pathway alterations in cancer: Variations on a theme. *Oncogene* **27**, 5497–5510. https://doi.org/10.1038/onc.2008.245 (2008).
35. Roberts, P. J. & Der, C. J. Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* **26**, 3291–3310. https://doi.org/10.1038/sj.onc.1210422 (2007).
36. Powell, E., Piwnica-Worms, D. & Piwnica-Worms, H. Contribution of p53 to metastasis. *Cancer Discov.* **4**, 405–414. https://doi.org/10.1158/2159-8290.CD-13-0136 (2014).
37. Tang, Q., Su, Z., Gu, W. & Rustgi, A. K. Mutant p53 on the Path to Metastasis. *Trends Cancer* **6**, 62–73. https://doi.org/10.1016/j.trecan.2019.11.004 (2020).
38. Pao, W. *et al.* KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med.* **2**, e17. https://doi.org/10.1371/journal.pmed.0020017 (2005).
39. Del Re, M. *et al.* Contribution of KRAS mutations and c.2369C > T (p. T790M) EGFR to acquired resistance to EGFR-TKIs in EGFR mutant NSCLC: A study on circulating tumor DNA. *Oncotarget* **8**, 13611–13619. https://doi.org/10.18632/oncotarget.6957 (2017).
40. Ohashi, K. *et al.* Lung cancers with acquired resistance to EGFR inhibitors occasionally harbor BRAF gene mutations but lack mutations in KRAS, NRAS, or MEK1. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2127-2133. https://doi.org/10.1073/pnas.1203530109 (2012).
41. Misale, S. *et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532–536. https://doi.org/10.1038/nature11156 (2012).
42. Ross, S. J. *et al.* Targeting KRAS-dependent tumors with AZD4785, a high-affinity therapeutic antisense oligonucleotide inhibitor of KRAS. *Sci. Transl. Med.* https://doi.org/10.1126/scitranslmed.aal5253 (2017).
43. Canon, J. *et al.* The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **575**, 217–223. https://doi.org/10.1038/s41586-019-1694-1 (2019).
44. Hong, D. S. *et al.* KRAS(G12C) Inhibition with sotorasib in advanced solid tumors. *N. Engl. J. Med.* **383**, 1207–1217. https://doi.org/10.1056/NEJMoa1917239 (2020).
45. Roe, O. D. & Wahl, S. G. The undifferentiated carcinoma that became a melanoma: Re-biopsy of a cancer of an unknown primary site: A case report. *J. Med. Case Rep.* **11**, 82. https://doi.org/10.1186/s13256-017-1238-y (2017).
46. Clynick, B. *et al.* Genetic characterisation of molecular targets in carcinoma of unknown primary. *J. Transl. Med.* **16**, 185. https://doi.org/10.1186/s12967-018-1564-x (2018).
47. Cerami, E. *et al.* The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095 (2012).
48. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal* **6**, pl1. https://doi.org/10.1126/scisignal.2004088 (2013).
49. Dees, N. D. *et al.* MuSiC: Identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–1598. https://doi.org/10.1101/gr.134635.111 (2012).
50. Vandin, F., Upfal, E. & De Raphael, B. J. Novo discovery of mutated driver pathways in cancer. *Genome Res.* **22**, 375–385. https://doi.org/10.1101/gr.120477.111 (2012).
51. Perez-Llamas, C. & Lopez-Bigas, N. Gitools: Analysis and visualisation of genomic data using interactive heat-maps. *PLoS ONE* **6**, e19541. https://doi.org/10.1371/journal.pone.0019541 (2011).

Acknowledgements

Author contributions

H.A.E. participated in the study design and performed data extraction and analysis. He wrote the first draft of the manuscript additionally. H.M. participated in the study design and critically edited the manuscript. A.H.A. participated as clinical oncologists and scientific advisor in study design, M.T. also participated in the study design and editing of the manuscript and he has supported the project financially.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-94678-4.

Correspondence and requests for materials should be addressed to M.T.

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