UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Optimising Multi-Protein Interaction Screening through Machine Learning Algorithm Development

Enhancing Förster/Fluorescence Resonance Energy Transfer (FRET) Detection in Flow Cytometry

by Zhongran Edward Ni

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of

Distinguished Professor Dayong Jin,
Professor Ian Menz,
Doctor Leo Zhang

University of Technology Sydney
Faculty of Science
School of Life Sciences (SoLS)
School of Mathematical and Physical Sciences (MaPS)
Institute for Biomedical Materials & Devices (IBMD)

February 2023

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Zhongran Edward Ni*, declare that this thesis is submitted in fulfilment of the requirements for the award of *Doctor of Philosophy*, in the *School of Life Sciences and School of Mathematical and Physical Sciences*, *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: February 1st, 2023

DEDICATION

To my loving parents, who have always believed in me and supported me through every step of my journey. Your unwavering encouragement and love have made me the person I am today. This thesis is a small token of my gratitude and a reflection of all the lessons you have taught me. I dedicate this work to you with all my love and respect.

To my beloved Ms Kerui Zoe Xu, who has been my constant source of love, support, and inspiration. Your unwavering faith in me has pushed me to reach for the stars and strive for excellence. Your encouragement and understanding have sustained me through the toughest of times. This thesis is a testament to the love we share and a reflection of the journey we have been on together. I dedicate this work to you, my love, with all my heart and soul.

To the pursuit of scientific knowledge and the belief that advancements in technology can better the lives of humanity. This thesis represents my passion for understanding the world around us and using that knowledge to make a positive impact. I dedicate this work to my scientific career, in the hopes that it will contribute to the ongoing advancement of technology and the betterment of all people. Let us continue to strive for a brighter future through the power of science and technology.

ACKNOWLEDGEMENTS

I want to extend my sincerest gratitude to all those who have supported and encouraged me during these unprecedented times of the Covid-19 pandemic.

I offer my deepest appreciation to my esteemed supervisors, *Distinguished Professor Dayong Jin*, *Professor Ian Menz*, and *Doctor Le (Leo) Zhang*, who have been a source of unwavering guidance and support. I also want to express my gratitude to *Professor Yuen Yee Cheng* for her guidance and support throughout my academic journey. Your insights, wisdom, and patience have been invaluable during these challenging times. I am grateful for all you have done to help me succeed.

I offer my sincere thanks to my previous supervisors, *Doctor Lisa Sedger* and *Doctor Michael Johnson*, who have laid the foundation for my education and growth. Your encouragement and support have been a source of inspiration and profoundly impacted my academic career.

I offer my sincerest thanks to my friends, *Mr Steve Shuquan Su* and *Mr Daniel Yue Ma*, who have been a constant source of laughter and positivity. Your friendship and companionship have been a beacon of hope during the last three years, and I am honoured to have you in my life. Wishing you the very best as you embark on the final stretch of your PhD journey. May your hard work and determination pay off and may all your efforts be rewarded with success. I have no doubt that your scientific aptitude and passion for discovery will carry you far in your career. As we move forward and pursue new opportunities, I hope we will always look back on this moment with pride and satisfaction.

Finally, I am deeply grateful for the scholarships (International Research Training Program Scholarship and UTS President's Scholarship) from the University of Technology Sydney that have made this journey possible. Your support has allowed me to continue pursuing my dreams and achieving my goals.

To all of you, I offer my heartfelt thanks. This thesis acknowledges your unwavering support and guidance during these challenging times.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Appendix

List of Abbreviations

Algorithm related

Abbreviation	**Full name**
AIC	Akaike information criterion
ANN	Artificial neural network
ANOVA	Analysis of variance
BIC	Bayesian information criterion
CF	Collaborative filtering
CLR	Central logarithmic ratio
CoDA	Compositional data analysis
CV	Coefficient of variation
DBSCAN	Density-based spatial clustering of applications with noise
ESD	Extreme studentized deviation
FLAME	Flow analysis with automated multivariate estimation
FLOCK	Flow clustering without k
FMM	Finite mixture model
GLM	Generalized linear model
GMM	Gaussian mixture model
IJM	ImageJ Macro language
k-NN	k-nearest neighbour
lDDT	Local distance difference test
LR	Logarithmic ratio
MAE	Mean absolute error
MDL	Minimum description length
ML	Machine learning
MST	Minimum spanning tree
NNLS	Non-negative least squares
OLS	Ordinary least squares

PC	Principal component
PCA	Principal component analysis
RANSAC	Random sample consensus
RMSD	Root-mean-square deviation
SD	Standard deviation
SE	Spreading error
SEM	Standard errors around the mean
SOM	Self-organizing map
SPADE	Spanning-tree progression analysis of density-normalized events
WLS	Weighted least squares

Biology related

Abbreviation	**Full name**
BFP	Blue fluorescent protein
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
CD4	Cluster of differentiation 4
CD27	Cluster of differentiation 27
CFP	Cyan fluorescent protein
cIAP	Cellular inhibitor of apoptosis protein
CMV	Cytomegalovirus
CRDs	Cysteine-rich domains
CRM	Cytokine response modifiers
DD	Death domain
DMEM	Dulbecco's modified Eagle medium
eCFP	Enhanced cyan fluorescent protein
eGFP	Enhanced green fluorescent protein
eYFP	Enhanced yellow fluorescent protein
FADD	FAS-associated death domain

FBS	Fetal bovine serum
FP	Fluorescent protein
GFP	Green fluorescent protein
HEK-293T	Human embryonic kidney cells contains the SV40 large T antigen
huTNFRs	Human TNFRs
LB	Luria-Bertani (Broth)
LUBAC	Linear ubiquitin chain assembly complex
MCS	Multi-cloning site
miRFP703 nm)	Monomeric infrared fluorescent protein (maximum Em at 703
MLKL	Mixed lineage kinase domain-like
MPV	Monkeypox virus
MpvJ2R	Monkeypox virus encoded J2R protein
mRFP	Monomeric red fluorescent protein
MYXV	Myxoma virus
MyxMT2	Myxoma virus encoded MT2 protein
NCBI	National Centre for Biotechnology Information
NF-kB	Nuclear factor kappa-light-chain-enhancer of activated B cells
ORFs	Open reading frames
PBMCs	Peripheral blood mononuclear cells
PLAD	Pre-ligand-binding assembly domain
POIs	Proteins of interest
PPIs	Protein-protein interactions
RFP	Red fluorescent protein
RIPK	Receptor-interacting serine/threonine protein kinase
SDM	Site-directed mutagenesis
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
SODD	Silencer of death domain

TCR	T-cell receptor
Tm	Annealing temperature
TNF	Tumour necrosis factor
TNFR	Tumour necrosis factor receptor
TNFRSF	Tumour necrosis factor receptor superfamily
TNFSF	Tumour necrosis factor superfamily
TRADD	TNFR-associated death domain
TRAF	TNFR-associated factor
TRAPS	Tumour necrosis factor receptor associated periodic syndrome
VARV	Variola virus
VarG4R	Variola virus encoded G4R protein
vTNFR	Viral tumour necrosis factor receptor
WT	Wildtype
YFP	Yellow fluorescent protein

Chemistry related

Abbreviation	**Full name**
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
HBS	HEPES buffered saline
HEPES	N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid
IAA/Auxin	Indole-3-acetic acid
PBS	Phosphate-buffered saline
PFA	Paraformaldehyde
TB	Transformation buffer
TBE	Tris-borate-EDTA

Physics related

Abbreviation	**Full name**
E_{FRET}	FRET efficiency
Em	Emission
Ex	Excitation
FLIM	Fluorescence lifetime imaging microscopy
FRET	Förster/Fluorescence resonance energy transfer
FRET1	eCFP→eYFP FRET
FRET2	eYFP→mRFP FRET
FRET3	eCFP→mRFP FRET
FRET4	eCFP→eYFP→mRFP two-step FRET
FSC	Forward scattered light
HTS	High-throughput screening
LED	Light-emitting diode
MFI	Mean fluorescent intensity
PMT	Photomultiplier tube
QY	Quantum yield
RT	Room temperature
SSC	Side scattered light

ABSTRACT

Förster or fluorescence resonance energy transfer (FRET) is a widely utilised technique to analyse protein-protein interactions (PPIs) for exploring biological processes. While microscopy was traditionally used to measure FRET, flow cytometry-based FRET has become more prominent in the last decade. Flow cytometry allows the multichannel high-throughput examination of FRET with great sensitivity and statistically robust quantification in many samples. As a superior alternative to traditional manual analysis, machine learning (ML) has attracted increasing interest in modern cytometry data analysis. It can automatically perform objective data-oriented investigations from large datasets with minimal human interventions.

Moreover, current FRET analyses are limited to chemically linked molecules for FRET calibration and lack high-performance data pre-processing. To achieve absolute single-cell quantification of natural (chemically unlinked) PPIs, this study established ML-powered algorithms for flow cytometry-based FRET detection of multi-protein interactions. This study presents designs, validations, performance tests, and application demonstrations of the algorithms in a FRET analysis workflow. The UltraFast algorithm presented F1 score over 0.91 for singlet data identification. The collaborative filtering-based algorithms' performance demonstrated error rates below 0.01% for both the correction of baseline subtraction error and autofluorescence prediction. Moreover, the downstream spectral unmixing process accomplished near zero (as low as 0.412) residual spillover and near zero (as low as 0.620) spread error, demonstrating more than 1000 folds of improvement compared to spectral unmixing without the

abovementioned pre-processing steps. Together, the pre-processing pipeline with these developed algorithms achieved unbiased, accurate and robust flow cytometry data pre-processing, including singlet identification, fluorescence background-subtraction-error correction, autofluorescence prediction and removal, and FRET spectral unmixing.

The pre-processed flow cytometry data further allowed the complete quantification of two-protein three-colour FRET, three-protein six-colour FRET, and investigation of multiple simultaneous intercellular signalling activities. In particular, the FRET calibration and FRET efficiency are improved to the absolute single-cell level quantification compatible with chemically-unliked free-interacting molecules compared to the current method using linked calibration approaches. The new FRET analysis approaches have been tested on five different machine models, including conventional and full-spectrum flow cytometers, and validated using data generated from eleven FRET experiments. Further utilising the compositional data analysis (CoDA) techniques, this study also provided interpretations for the dynamic FRET energy competitions and the compositional activation levels of multiple cellular signalling pathways. In conclusion, this study provides powerful solutions for clinical diagnostics and therapeutic screenings, enabling the search for the next-generation PPI-specific and signalling pathway-specific cures for many human diseases.

INTRODUCTION

Over the past two decades, Förster or fluorescence resonance energy transfer (FRET) applications in the life sciences have substantially increased (Figure 1). According to PubMed, the number of peer-reviewed journals with "Resonance Energy Transfer" as the topic increased from 1,499 before the year 2002 to 18,525 from the year 2002 to 2022 (total of 20,024 since the year 1963) [1], showing both the rapid development and the fast expanding speed of FRET methods.

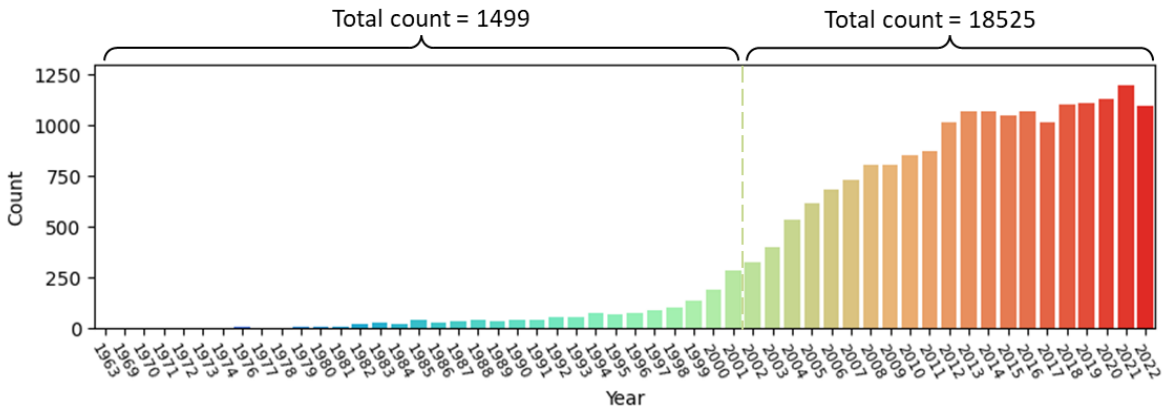


Figure 1 Publication metric on "FRET"
Publications per year found using "Resonance Energy Transfer" as the primary search string with advanced search field as [Title/Abstract] in PubMed https://pubmed.ncbi.nlm.nih.gov (Accessed on January 1st, 2023).

The fact that FRET is the most straightforward yet powerful technique for *in-situ* protein-protein interaction (PPI) super-resolved optical detection is driving this growth. FRET does not require the extraction of protein from samples, nor does it require the expression of the protein in the non-native host that may possess different post-translational modifications. FRET is a sensitive tool for investigating biological processes mediated by PPIs providing nanometre scale (<10 nm) resolution about intra- and intermolecular distances (Figure 2) [2, 3]. However, the fast-expanding popularity of

FRET has led to a significant variance in the methods of collecting, analysing, and interpreting the FRET data. Especially with the frequent advancement of fluorescence materials (fluorochrome dyes, fluorescent proteins, and fluorescent nanoparticles) and FRET quantification techniques (fluorescence emission, fluorescence-lifetime, fluorescence polarization anisotropy), investigations often rely heavily on special requirements of the fluorophores and detecting equipment. Moreover, the technical difficulties and the slow data acquisition speed are challenges for real-life clinical and pharmaceutical applications.

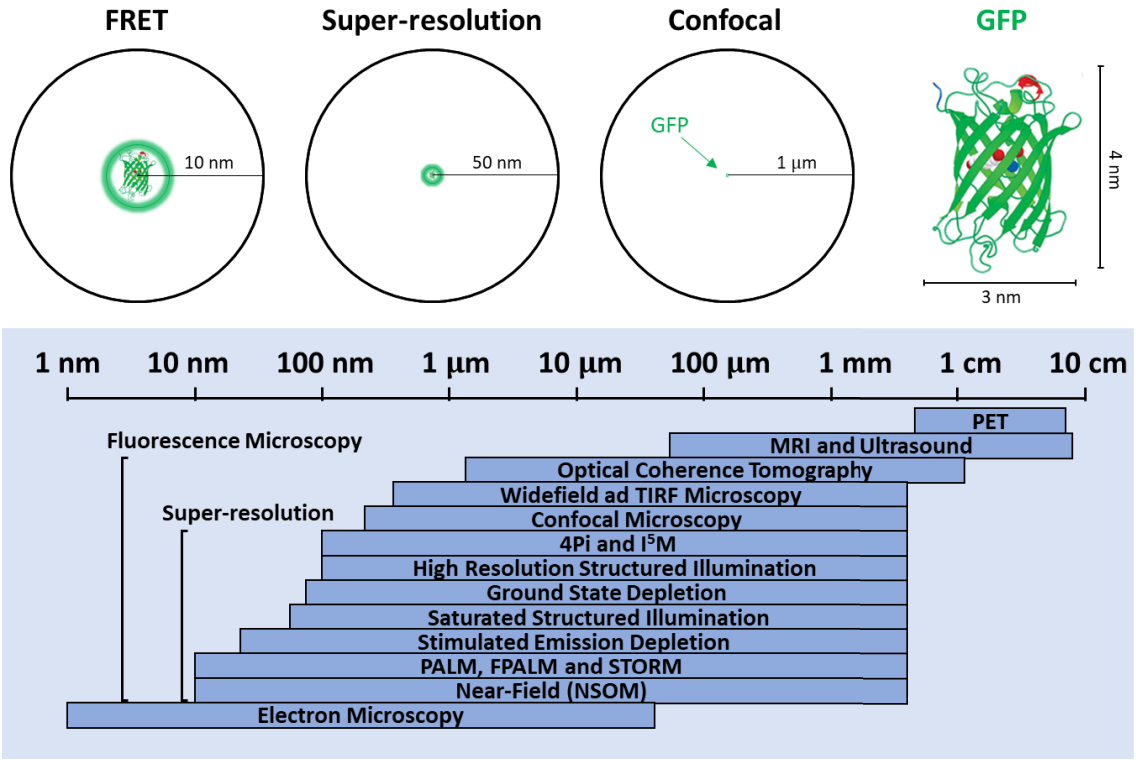


Figure 2 Super-resolved optical detection of FRET
FRET has a super-resolved optical detection range of < 10 nm. The relative size of GFP and the detection ranges of many imaging technologies are also displayed for comparison. (The bottom microscopy resolution section was modified from Yang et al. 2016 [4]).

Since the middle of the 1990s, FRET has been combined with flow cytometry technology to produce t interactions detect interactions with high-throughput rates, low equipment requirements, and high signal-to-noise sensitivities [5, 6]. Many studies have established standard data analytical workflows for flow cytometry and FRET detection [7, 8], but still, many limitations in the existing methods have impeded the true single-cell level full quantification of flow cytometry-based FRET detection. In particular, the current single-cell event identification method heavily depends on human-subjective, labour-intensive, and low-throughput manual gating strategies [9]. Moreover, as the sensitised emission, FRET has relatively low fluorescent intensity, especially when interacting molecules are further apart; therefore, FRET is extremely sensitive to background noise and autofluorescence [10]. Unfortunately, error-free background correction and autofluorescence removal methods are still yet to be developed. For FRET spectral unmixing, the fluorescent intensity-based ratiometric flow cytometry FRET method devised by Roszik et al. has provided an excellent formula for each detected fluorescent signal's spectral compositions [11]. However, the method requires a FRET calibration factor which can only be measured using the chemically linked FRET pairs and deducted using the population means, albeit the method has been widely used [12-16]. Using the linked calibration control and non-single-cell level calculation inhibits the accurate, true single-cell level detection of the PPI using flow cytometry-based FRET.

Moreover, with the upgrade on the two-protein FRET detection, many three-protein FRET systems have been developed for more complex biological PPI investigations [17-19]. As two different FRET acceptors interact with the same FRET donor in the three-way FRET, the FRET energy competition can happen among the three fluorophores.

However, none of the publications has successfully investigated the dynamic FRET energy compositions using three fluorophores on flow cytometry during three-way FRET. Furthermore, many flow cytometry-based techniques were developed as stand-alone PPI screening tools, although PPIs regulate many cellular signalling activities. Flow cytometry has the multi-channel detection capability (up to 64 fluorescence detection channels in full-spectrum flow cytometer [20]); therefore, flow cytometry-based FRET detection will be more appealing with the simultaneous detection of multiple cellular signalling reporters.

Lastly, traditional data analysis in cytometry utilises the inspection and manual classification of the 1-D histogram overlays or multiple 2-D scattering plots to perform data quality assessment, pre-processing, and analysis. However, the user-dependent bias makes it difficult to replicate among different operators. It becomes impractical when the numbers of samples and fluorescence channels vastly increase alongside the developments of modern multichannel and hyper-spectrum flow cytometry technologies. As a superior alternative, machine learning (ML) algorithms can provide automated, objective, and accurate data analysis when combined with the flow cytometry-based FRET technology.

The motivation of this project is to eliminate the obstacles which inhibit flow cytometry-based FRET from being the accurate single-cell level screening tool for PPI investigations with full quantitative power. This study aims to deliver an automated, unbiased, and fully quantitated, true single-cell level flow cytometry analytical pipeline powered by ML algorithms allowing the simultaneous detection of multi-protein FRET and cellular signalling pathway activities.

This dissertation starts with an overview of the previously published works on flow cytometry-based FRET. Then it presents several chapters oriented around the procedure from the construction of the biological samples for FRET investigation, the development and validation of the flow cytometry-based FRET algorithms, and the demonstration of utilising these algorithms to solve several real-life research problems.

Chapter one begins with the literature review to deliver background information on FRET principles and the FRET biological labelling methods. Followed by the flow cytometry-based workflow and currently existing algorithms, along with different flow cytometry-based FRET quantification methods.

Chapter two presents the molecular biology methods to prepare the plasmid vectors for fluorescent protein fused target protein expression and cellular signalling reporter expression, bacterial methods for plasmid vector maximal production and storage, and cell biology methods for mammalian expression of the fluorescently labelled PPI and cellular signalling detection. Followed by initial microscopy characteristics methods, flow cytometry configuration, and FRET detection methods. The rest and majority of this chapter is focused on the algorithm design, validation, performance testing and implementation, statistical analysis and data visualisation methods.

Chapters three to seven constitute the results of this study. Firstly, chapter three presents the examination and characterisation of the biological samples prepared for flow cytometry-based FRET detection and algorithm development. Then, chapter four describes the development of automated unsupervised solutions for data pre-processing, including (i) high-performance single-cell event identification, (ii) accurate baseline-subtraction correction, (iii) error-free nonlinear autofluorescence prediction

and removal, and (iv) FRET spectral unmixing for fluorescence spill-over compensation. Afterwards, chapter five demonstrates the development of single-cell level, fully quantitative two-protein three-colour FRET solutions with advanced FRET calibration and unlinked free-interacting FRET efficiency calculations. The algorithms were tested using both chemically linked enhanced green fluorescent protein (FP) and monomeric red fluorescent protein (eGFP-mRFP) FRET pairs and unlinked free interacting FP-labelled human tumour necrosis factor receptor (TNFR) proteins. Two applications were also demonstrated using the FRET algorithms for small molecule concentration detection in plant and human T cells antigen-induced surface protein interactions.

Moreover, chapter six further upgraded the FRET algorithms for the three-protein six-colour FRET investigations and validated using the trimeric interacting human TNFR1 meta-clusters. In addition, FRET algorithms were further combined with the compositional data analysis (CoDA) approach to investigate the energy dynamics during three-protein FRET. With the compositional perspective, the algorithms were tested on TNFR1 germline mutation autoinflammation disease and human-poxvirus interaction investigations. Furthermore, chapter seven takes one step beyond FRET and demonstrates the simultaneous detection of TNFR1 interactions, cell death reporting, and cell inflammation reporting.

In the final summary and perspective chapter, further attention was given to discussing the limitations and advantages of the flow cytometry FRET technique and the newly developed ML algorithms-powered workflow. Future study directions are further proposed to empower the FRET algorithms for more complex biological investigations.

Although no “one-size-fits-all” approach exists, this study provided a PPI-specific and cellular signalling pathway-specific screening platform with ML-powered analytical tools that should benefit many researchers to conduct next-generation customised therapeutic screening and disease diagnosis to provide precise solutions for human illness.

CHAPTER ONE

LITERATURE REVIEW

1.1 Principles and applications of FRET

Förster resonance energy transfer (FRET) is named after its inventor Theodor Förster [2]. The term “fluorescence resonance energy transfer” is often used when both the FRET-donor and FRET-acceptor molecules are fluorescent [21-23], despite FRET's energy not actually being transferred by fluorescence. In fact, FRET energy is transferred non-radiatively from an excited donor molecule to an acceptor molecule through a dipole-dipole resonance coupling mechanism [24, 25] (Figure 3).

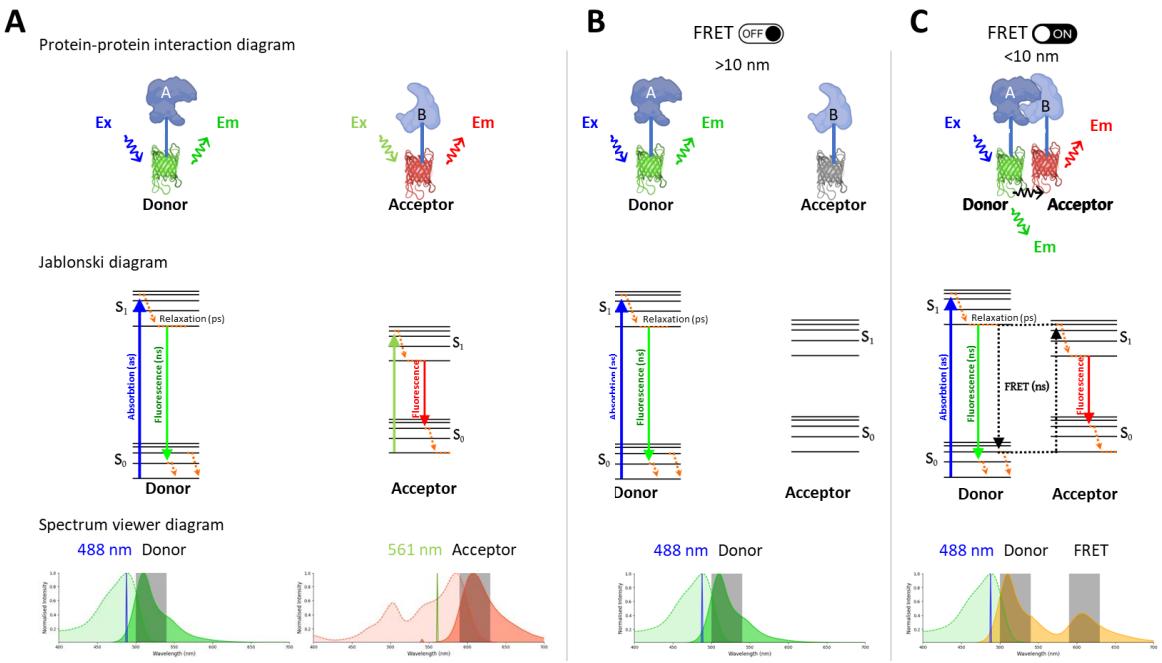


Figure 3 The energy transfers and fluorescence detection during FRET
FRET illustration using the protein-protein interaction diagrams of target proteins A tagged by green fluorescent protein (GFP) and B tagged by red fluorescent protein (RFP), the FRET dipole-dipole energy transfer Jablonski diagrams, and the spectrum viewer diagrams. FRET donor GFP and acceptor RFP are excited by the 488 nm and 561 nm laser (A). Ideally, only GFP will be excited by the 488 nm laser when there is no FRET event (B). FRET happens when proteins A-GFP and B-RFP are within proximity (< 10 nm), and a combination of the GFP donor and FRET emissions (yellow histogram) can be detected upon 488 nm laser excitation (C).
In general, three prerequisites must be met for the FRET procedure to take place:

(i) The emission spectrum of the FRET-donor molecule must sufficiently overlap with the excitation spectrum of the FRET-acceptor [26]

(ii) The FRET-donor must be in proximity to the FRET-acceptor molecule within the Förster radius, which is from 0.5 nm to 10 nm [27, 28]

(iii) The position between the FRET-donor and FRET-acceptor molecules must not be perpendicular to each other. FRET achieves maximal rate when donor and acceptor are in parallel and decreases as the angle between them increases [29]

However, FRET is not limited to two fluorophores, and some studies have shown the possibility of detecting a two-step, three-way FRET among three fluorophores [30-32]. FRET involving the usage of different donor and acceptor fluorophores is termed hetero-FRET. More interestingly, FRET can also happen using the same fluorophore as both a donor and an acceptor, with significantly overlapping excitation and emission from the same fluorophore, and it is termed homo-FRET [33]. FRET has many applications (Figure 4), with the most common usage of FRET being to detect protein-protein interactions and measure the relative proximity between two interacting protein entities. Moreover, intramolecular FRET can detect the concentration of small molecules with bioengineered FRET donor and acceptor on two ends of the same receptor protein, whilst the binding of the small molecules to the receptor triggers the conformational change, therefore, bringing the two ends closer along with the FRET-donor and FRET-acceptor molecules [34]. In addition, FRET can detect protease activity utilising chemically-linked FRET pairs already within the Förster radius in which the linker contains the target cleavage site of the tested protease. Upon enzymatic cleavage of the

linker, FRET-donor and FRET-acceptor are released from each other and no longer in proximity, and the FRET signal disappears. Beyond the abovementioned applications, FRET can also detect the changes in the test environment based on the altered properties of the fluorophores under different conditions, i.e., shifted FRET emission spectrum under different PH [35].

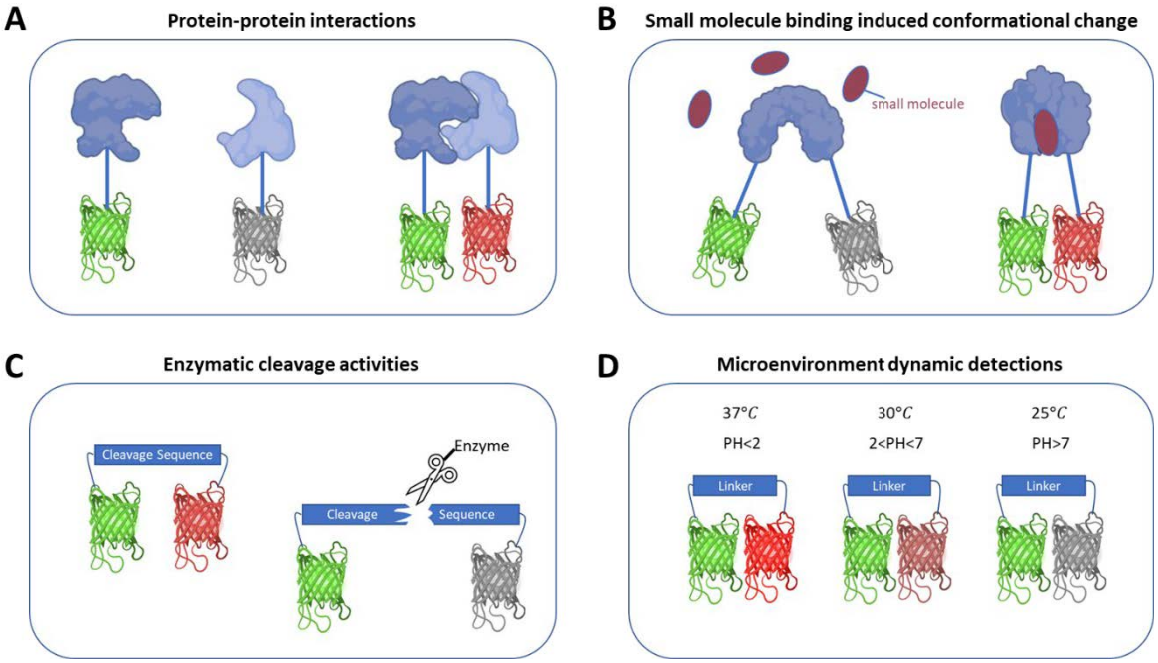


Figure 4 Applications of FRET
All diagrams show the fluorophore emissions upon donor laser excitation. Two proteins fused with GFP and RFP can only produce detectable RFP acceptor emission when interacting (A). Intracellular FRET happens on the small molecule binding protein when conformational change happens upon small molecule binding to the socket (B). The enzyme cleaves the linker between the FRETing donor-acceptor pair and frees the acceptor from the donor, and donor emission disappears (C). FRET efficiency changes and achieves different amounts of acceptor emissions under different testing environments (D).

1.2 FRET fluorophore labelling methods

Choosing the optimal donor and acceptor FRET pairs is crucial in achieving the highest possible FRET events in all cases. The choice should be based on the detecting hardware's configuration, such as laser and optical filters [36], as well as the availability of the fluorophores. In addition, the user must assess the spectral compatibility of the FRET fluorophores with other fluorescent materials being contemporarily assessed if conducting a multi-colour experiment. The biological sample's intrinsic autofluorescence should also be considered [37].

Fluorophore-conjugated antibodies are the easiest technique for studying interactions among endogenous biological proteins with FRET due to their high affinity and specificity for their binding domain [38]. Roy *et al.* (2008) have provided a practical evaluation of fluorochromes and conjugation strategies for FRET investigations [39]. The antibody labelling method also permits the introduction of numerous distinct fluorophores to a single target molecule, therefore providing epitope-specific interpretations of the FRET events if monoclonal antibodies are used [40]. If possible, a direct labelling approach with primary conjugated antibodies should always be used instead of secondary antibody labelling. Due to the increased fluorophore distance separation between the secondary antibodies, FRET events are often less efficient and prone to false-negative results. For labelling a large molecular complex assembly with secondary antibodies, it is also feasible that FRET will occur between otherwise more distant targets [41]. However, the polyclonal characteristics of the secondary antibodies may make it challenging to interpret the FRET results at the biological level. Even with the direct antibody labelling approach, caution is necessary when conducting the FRET experiment.

Antibodies are essentially large Y-shaped proteins with a size of around 140 kDa. Antibodies have bivalent binding capabilities with two epitope recognition sites that can be problematic since antibodies have the propensity to form unnatural clusters and interfere with the FRET measurement. Additionally, antibody-binding can potentially block the protein-protein interaction domains and act as an antagonist to prevent PPI and lead to artificial labelling-induced false-negative FRET results [42]. It can also be troublesome if the bivalent binding antibody crosslinks two proteins and artificially produces FRET emission and false-positive PPI results [43]. Moreover, the employment of antibodies generally eliminates the ability to study living cells because cells must be fixed and permeabilised to stain intracellular proteins [44].

An alternative option is to use fluorescent proteins (FPs) to label the target proteins, and cells are transfected with plasmids to express FPs fused to specific proteins of interest (POIs) for exogenous expression. One of the earliest well-characterised FPs is the green fluorescent protein (GFP), which was firstly purified by Shimomura *et al.* (1962) from the jellyfish *Aequorea victoria*, and cloned by Prasher *et al.* (1992), then independently expressed in heterologous cells by Chalfie *et al.* and Inouye *et al.* in 1994 [45-48]. Since then, a wide variety of FPs has been discovered and engineered, offering a broad range of options for absorbance and emission spectra, oxygen or pH sensitivity, quantum yield, and excitation coefficients [49-51].

When conducting FRET investigations, monomeric FPs should always be used. Many FPs can undergo oligomerisation without needing the target proteins to interact, producing FRET-positive results caused by the interacting FPs instead of the target proteins [52]. Numerous monomeric FPs variations have been derived and are suitable for FRET

applications [41, 52, 53]. In addition, the folding, maturation rate, photostability, and PH sensitivity of the FPs, which ultimately influence the quantum yield (QY) of the FPs, may all impact the FRET events [54-60]. For FRET measurements, fluorophores with higher QY and photostability are preferable [61-63].

Modern fluorescent proteins, blue (BFP), cyan (CFP), green (GFP), yellow (YFP), red (RFP), and Infrared (iRFP) encompass the entire visible spectrum, with the majority of them originally deriving from GFP [64]. Based on the abovementioned variables, some frequently used FRET combinations had previously been examined. For instance, CFP and YFP have long been the most widely used FRET fluorophores for FRET [40], but they have limitations, such as their pH dependence and the comparatively low QY [65]. Using bioengineered eGFP, eCFP, eYFP, mClover3, and mRuby2 exhibited enhanced fluorescence properties and provided improved opportunities for detecting rapid molecular interactions, and thus more workable options for the detection of interactions [32, 66, 67]. Bajar *et al.* 2016 have provided a comprehensive guide to fluorescent protein FRET pair selections depending on their relative strengths and weaknesses [68].

FPs are commonly expressed as amino (N-) or carboxy (C-) terminus fusions with the POIs to explore their biological functions in living cells. Most FPs are relatively large molecules with a molecular mass of around 25 kDa [69]. Therefore, the fused fluorescent proteins may change the target protein's folding and interaction and compromise the inherent biological functions of these POIs [70]. It is necessary to evaluate FRET signals and cellular biological functions with FP-fused receptors as an additional layer of experimental control [12, 71, 72]. Additionally, to ensure adequate FP motility, the

fluorescent proteins might be linked to the POIs using a linker sequence with up to 2 to 30 amino acids [12, 32, 70]. However, one should be mindful that introducing a linker with a different sequence may change the FP's distance and orientation, which may therefore reduce the effectiveness of FRET.

Additionally, the fluorophores may be oriented in a turned-away position due to the unanticipated complex formation between the two proteins, leading to an interaction without FRET. Another option is to introduce the FP into the target protein in the space between functional domains, which might cause the FP to coexist undisturbed with the target protein [73, 74]. In addition, a study also showed that using a minimal number of amino acids in the linker at the C-terminus of cellular receptor proteins can produce a higher FRET detection while maintaining the biological function of the receptor proteins in activating intracellular signalling pathways [70].

1.3 Flow cytometry-based FRET analytical workflow and algorithms

Traditionally FRET is examined using microscopes. FRET can be measured using the fluorescence microscopy-based direct determination of quenched FRET-donor fluorescent intensity and increased sensitised FRET-acceptor fluorescent intensity [75]. Fluorescence lifetime imaging microscopy (FLIM) measures FRET by examining the time change for a FRET-donor fluorophore that stays in its excited state [76, 77]. In addition, polarised light microscopy analyses the anisotropy dynamics for homo-FRET [33, 78]. However, microscopy FRET has limitations due to its low throughput, high equipment dependency, low signal-to-noise ratio, and prone to false-positive results. Therefore, during the mid-90s, flow cytometry-based FRET was developed and proved tremendously successful in detecting protein-protein interaction in live cells. It

complements the microscopic FRET technique by allowing the collection of tens of thousands of events per second. Modern flow cytometers can also perform data acquisition using the robotic high-throughput screening (HTS) module with the capacity to collect hundreds of samples in under an hour. In addition, the most attractive feature of flow cytometry is the high-parameter detection, with the current full-spectrum flow cytometers able to perform the simultaneous detection of 64 fluorescence signals. Flow cytometry's capacity to analyse many samples with multiple parameters in a short time makes flow cytometry the most widely used screening tool for drug discovery and clinical diagnosis. More importantly, unlike microscopy-based FRET, flow cytometry-based FRET does not require special equipment, which means FRET can be easily detected on almost any conventional flow cytometer.

The flow cytometry-based FRET analytical workflow consists of logical stages before the user can perform fluorescent quantifications, data visualisation and biological interpretations, including FRETs and PPIs (Figure 5). The pre-processing stages consist of data import, followed by anomaly detection, singlet identification, autofluorescence removal, spectral unmixing, data transformation, and subpopulation clustering. There are many pieces of software and programming packages/libraries that provide solutions for each stage of the flow cytometry analytical pipeline. However, many unresolved challenges still hinder truly accurate FRET quantification, and this chapter will review them in the following sections.

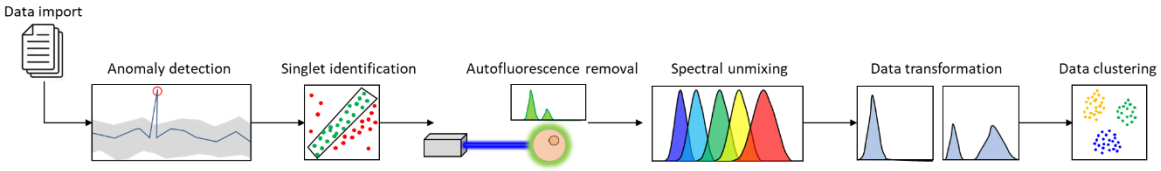


Figure 5 Flow cytometry data pre-processing
Flow cytometry data analysis has many stages after data import. It includes anomaly detection, singlet identification, autofluorescence removal, spectral unmixing, data transformation and data clustering before one can perform statistical analysis.

1.3.1 Anomaly detection and data cleaning

Anomaly detection and data cleaning are essential as the first step in the cytometry data analysis pipeline for achieving highly accurate, statistically robust, and biologically meaningful results. Data acquisition in cytometers should occur consistently with the sample flow rate. Flow cytometry relies significantly on the data collecting rate's consistency to guarantee accurate detection of the data. Numerous variables can cause the fluctuation of the collection rate. For example, sample clumps can clog the fluidic collection tube, rapidly decreasing the data collection rate. Likewise, air bubbles in the fluidic system can also affect the data acquisition quality. In addition, unstable laser powers and malfunctioning detectors contribute to erroneous fluorescent intensity readouts.

Moreover, flow cytometers often generate negative fluorescence intensity values during data acquisition. The negative values are due to the background signal's subtraction as part of the baseline removal procedure. In addition, negative numbers in fluorescence intensity may also result from compensation calculations and autofluorescence removal during post-acquisition data processing. Flow cytometry data, including any abovementioned anomalies, can compromise downstream analyses.

Currently, flowAI is one of the most popular data anomaly detection algorithms for flow cytometry [79]. The flowAI algorithm follows a straightforward logic with three data quality assessment stages: (i) the flow rate check – which detects abrupt changes in flow rate (ii) the fluorescence signal quality check – which detects instability of signal acquisition (iii) the dynamic range check – removes outliers below the lower detection limit and margin events beyond the upper detection limit. For fluidic flow rate consistency, flowAI reconstructs the flow rate from the number of cells per unit of time and defines the consistent flow stream as a flow rate with non-periodic oscillations but a constant variation. A generalised extreme studentized deviation (ESD) test is used to discover the flow rate irregularity. For the fluorescence signal quality check, flowAI plots the fluorescent intensity against the time of each recorded event to determine whether the data has a stable distribution over the acquisition time. At the last stage, data logged with the same value as the maximal detection range are removed to prevent saturated signals. However, for negative values, flowAI employs the Z-score test and only removes the ones beyond the default threshold; therefore, a portion of the data with negative fluorescent intensity is still presented in the collected signals, which is problematic for fluorescence intensity interpretation and many downstream data analyses.

The flowClean algorithm is another widely used data anomaly detection technique [80]. The development idea of flowClean is that a sample has several cell populations, each of which should have a uniform data acquisition rate and fluorescence intensity. FlowClean splits cells into populations based on the median fluorescence intensity of each detection channel. Then, it checks acquisition rate stability using the Poisson logarithmic likelihood [80]. Then, the fluorescence intensity stability examination uses

CoDA [81] and change point analysis [82] techniques. Nevertheless, flowClean replaces data with zero and negative values with small random positive values, which, despite making the data acceptable for logarithmic-based transformation, might artificially generate errors in data quantification.

1.3.2 Singlet identification

One of the advantages of flow cytometry over other population mean-based bulk measurement techniques is its ability to identify fluorescent signals with single-cell level resolution. Flow cytometry offers information at the single-cell level from numerous cellular subtypes in the sample. It displays the binominal distribution and heterogeneity of the data (Figure 6A). Identifying true single-cell events (singlets) from the doublets, multiplets, debris and background noise is critical in flow cytometry data analysis. Failure to distinguish singlet from cell aggregates might result in erroneous data interpretation (Figure 6B). For example, when measuring fluorescent intensities, flow cytometry measures the total fluorescent intensity (MFI) from doublets, records the MFI as a single particle event and produces inaccurate intensity readouts (Figure 6B first row). In addition, in cell phenotyping tasks, doublets containing different single markers will be recorded as multiple markers expressing cell types (Figure 6B second row). Moreover, in FRET measurement, an actual FRET-negative single-cell event requires the cell to be double-labelled with both FRET-donor and FRET-acceptor fluorophores, indicating the genuine non-interacting status of the POIs. However, false non-interacting results can be detected when cells are individually and separately labelled with only the FRET-donor fluorophore or only the FRET-acceptor fluorophore and form cell aggregates (Figure 6B third row).

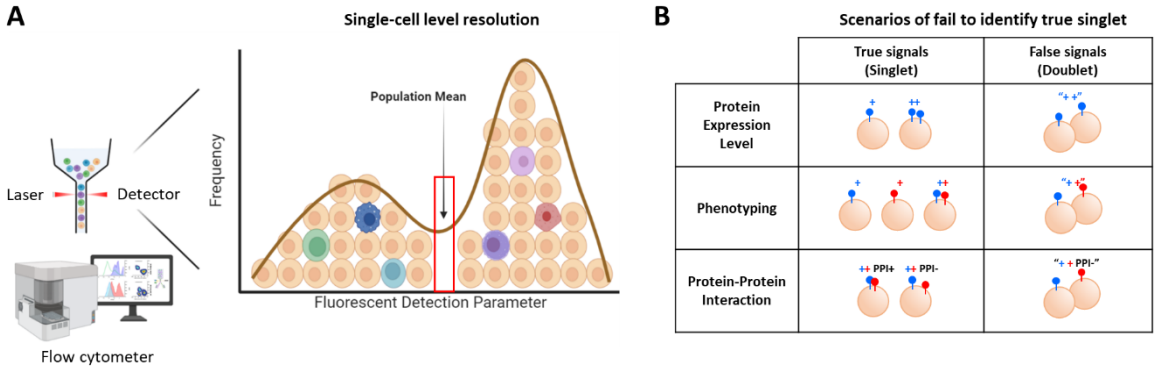


Figure 6 Detection of singlets is critical for flow cytometry data analysis.
Flow cytometry measurement at single-cell level revealing sample heterogeneity (A). Scenarios of unsuccessful singlet events identification for flow cytometry data analysis (B).

Flow cytometers measure scattered light signals on two perpendicular detectors, including forward and side scattered light signals (FSC and SSC), in addition to measuring fluorescence emissions. The FSC measure the relative cell sizes, and the SSC measure the relative cell granularity, indicating the cellular content's complexity. For every signal being measured by flow cytometers, the machine record three signal characteristics, including the width, height, and area of the signal peaks. All modern flow cytometers are calibrated such that these six characteristics can be used to determine the position of singlets within the entire sample by examining a series of 2D scatterplots (Figure 7), FSC-A vs SSC-A, FSC-H vs SSC-H, FSC-A vs FSC-H, SSC-A vs SSC-H, FSC-H vs FSC-W, SSH vs SSC-W. In detail, an actual singlet event produces only one signal peak, so the area and height of the peak should be in a perfect linear relationship. In addition, the relative time of a single cell passing the laser (the peak width) should be independent of the signal strength (the peak height) (Figure 7A).

Based on the calibration among all six scattered light signals, there is a well-established and commonly adopted method for identifying singlets using a six-step manual threshold drawing procedure known as gating (Figure 7B) [9]. In software like FlowJo [83], a user can automatically apply the six-step manual gating established from one sample to the rest of the samples with a few clicks. However, when cells are treated differently, the profiles on the six-step gating plots will vastly vary (Figure 7B). For example, if the expression of specific proteins is upregulated at different levels, the SSC-A/H will also increase or decrease according to the change in cellular content complexity. Additionally, if cells undergo apoptosis (programmed cell death) or proliferation (division), the cell size will also decrease or increase respectively. Such varied profiles in the six-step gating plots indicate that manual adjustments are often needed for each sample to achieve accurate singlet identification in practice.

The manual gating approach can be labour-intensive and often subjective when investigating a large sample batch. So far, very few studies have developed automated approaches for singlet identification. In contrast, most studies in the automated gating algorithm development have only focused on cell subtype identification using fluorescence detection channels (see section 1.3.4). However, most cell-type identification algorithms require the input data to be single-cell data cleaned of debris, artefacts, and cell aggregates. Among the very few singlet identification algorithms, FlowGateNist is the most recently developed Python algorithm that can perform automated processing on small cells, such as bacteria and yeast [84].

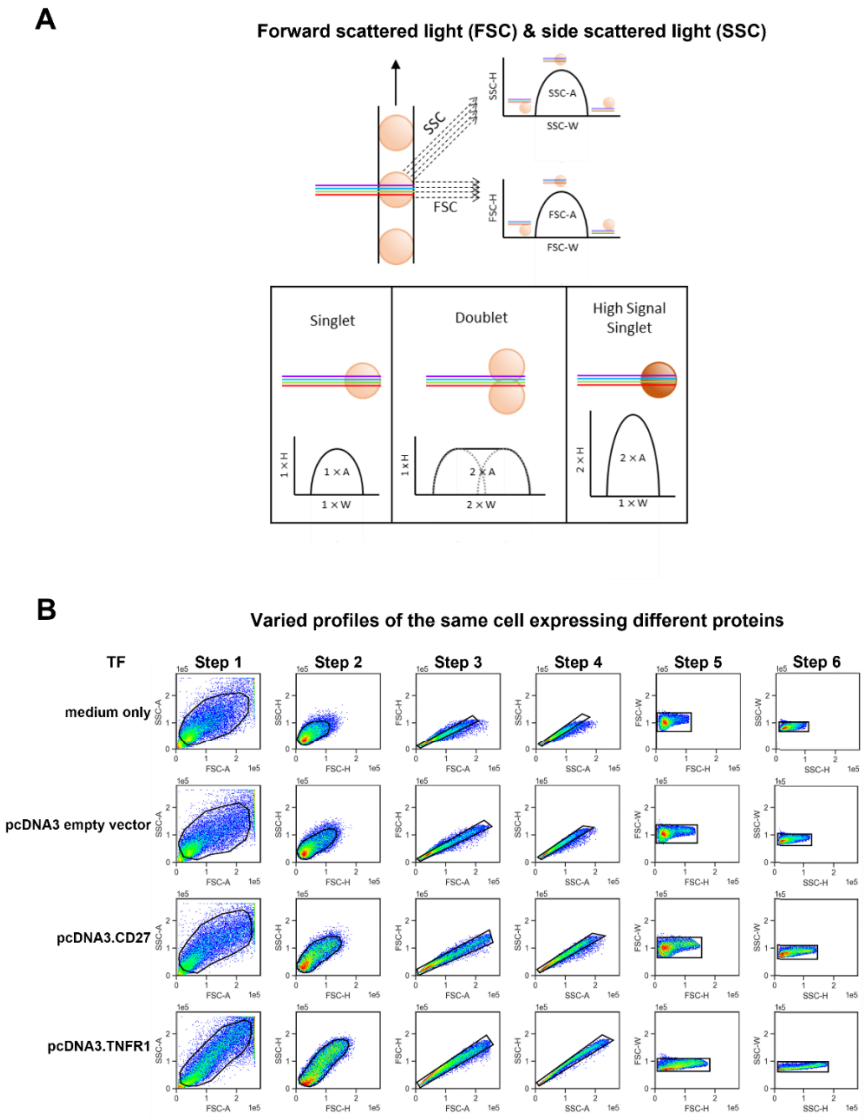


Figure 7 Flow cytometry FSC and SSC measurement, and singlet manual gating
Flow cytometry measures FSC and SSC, and each has width, height and area parameters (A). A well-established 6-step manual gating strategy for singlet identification requires further manual adjustments when cells are treated differently (B).

FlowGateNIST uses a GMM technique and a comparison between measured cell samples and buffer-only blank samples for automatic gating to distinguish between cell events and background events. Then it compares the height, area, and width parameters of the FSC/SSC signals to distinguish between singlet and multiplet events using a multi-dimensional GMM. However, since GMM requires the population to be convex, FlowGateNist requires iterative computations with different numbers of clusters and

merging them to identify the nonconvex singlet population using BIC, so the computation efficiency is slow. In addition, FlowGateNist is not suitable for large-size mammalian cell samples, which prevents it from being used in human biology studies, and this is also the limitation for FlowCal that was developed using example data for engineered bacteria (*E. coli*) and yeast (*S. cerevisiae*) [85]. Unlike FlowGateNist which uses GMM, FlowCal identifies singlets and uses a 2D histogram binning method to identify locations with the highest event density in a two-dimensional plot (typically the side-scatter vs. forward-scatter signals). The regions with the highest density, with a user-specified proportion of the total events, are gated as cell events, whereas the regions with lower densities are classified as non-cell background events. The TASBE flow analytics software package primarily focuses on flow cytometry data calibration and normalisation, but it also contains automated gating using GMM to discriminate cell events from background events [86]. Unlike FlowGateNist, TASBE uses a robust 2D Gaussian estimator to fit only the data collected in the FSC-A and SSC-A 2D scatterplots and identify the core region with the highest cell density. However, the TASBE algorithm also requires a user-specified percentage of the total to be gated as cells, whilst events beyond the central region and pre-defined percentage are eliminated as either background or multiplet events.

1.3.3 Spectral unmixing and autofluorescence removal

Fluorescence spectral unmixing (also known as compensation) is the adjustment process to remove spectra overlaps from multiple fluorophores detected across multiple detectors in flow cytometry data analysis. For flow cytometry-based FRET investigations, the overlap between the emission and excitation spectra of the donor and acceptor

molecules is one of the essential requirements for FRET to occur. However, cross excitation of the acceptor fluorophore with the donor excitation laser, the donor emission bleed-through into the acceptor emission spectrum range, and spillover between the acceptor and the sensitised FRET signals due to their similar emission spectra can all potentially impact the accuracy for FRET quantification. The spillover can be minimised by selecting FRET pairs and assigning detection channels that permit optimal donor excitation without acceptor cross-laser excitation. In addition, the acceptor detection filter must be chosen with no or minimum donor emission bleed-through. FRET pairings can also be selected from fluorophores whose spectra are maximally separated with large Stoke's shifts, but it should be emphasised that decreasing the spectral overlap will also decrease the rate for FRET events.

The key factor of almost all spectral unmixing algorithms is to minimise the spreading error (SE) after compensation since the SE impedes the pure fluorescent intensity measurement and reduces the detection sensitivity for the FRET detection. The SE for a given detector pair can be empirically determined by taking the square root of the squared differences in robust standard deviations of detected fluorescence intensity between the unstained sample and the single fluorescence control (see section 2.8.4) [87]. The most recently developed AutoSpill spectral unmixing algorithm and many other compensation techniques using linear regression models, such as ordinary least squares (OLS), non-negative least squares (NNLS), weighted least squares (WLS), or the generalized linear model (GLM) to predict the spillover coefficients between each pair of the cross-talked fluorescence detection [88]. The spillover coefficients are calculated using the single-colour controls and are defined as the linear ratio of the fluorescence

signals in the primary detection channel to the signal in the non-primary detection channel [89]. For example, in FRET detection, the spillover coefficient of the FRET-donor and FRET-acceptor channels is determined using the FRET-donor single colour control. It is calculated using the ratio of the donor-laser-excited donor fluorescence emission detected in the donor channel over the acceptor-laser-excited donor fluorescence emission detected in the acceptor channel.

Furthermore, many studies have treated autofluorescence as another single fluorescence emission and removed it from the labelling fluorescence using spectral unmixing algorithms [90-92], and this is also the approach used in flow cytometry-based FRET analysis [8]. However, autofluorescence is an intrinsic emission mixed with various types of biomaterials from the sample; therefore, treating autofluorescence as a single fluorescence emission can potentially induce bias and lead to unresolvable residual spreading errors during spectral unmixing [93]. Instead, the autofluorescence should be treated as the combination of many fluorescence emissions, and the prediction and removal of the autofluorescence should be handled using non-linear-based algorithms. Unfortunately, to date, the non-linear approach has still not yet been formally examined.

1.3.4 Cell subtype identification

Traditionally, cell subtype identification is performed with manual gating in flow cytometry data analysis to distinguish positively labelled cells from the unlabelled populations. However, with the advancement of the flow cytometer technology with more detection parameters, manual cell subtype identification becomes an implausible task for human eyes. Unsupervised clustering algorithms quickly became attractive for cell subtypes identification in flow cytometry data analysis, and this is because they can

produce unbiased results using multiple detection parameters simultaneously with minimal human input.

The K-Means clustering algorithm is one of the first automated data clustering techniques applied to cytometry data analysis in 1985 [94]. The K-Means algorithm clusters data by separating them into k groups with equal variance so that data within the same group share the same centroid [94]. Despite the fact that K-Means is very simple to implement, it has several limitations: (i) it requires a user-specified number of clusters k that often requires researchers to have pre-established knowledge about the cluster numbers of the sample, (ii) it assumes clusters are convex and isotropic, which means that samples with irregular shapes will not be clustered correctly, (iii) the accuracy also rely on the initially randomly assigned locations of the centroids, which means the algorithm needs to be repeated with various centroids initialization locations to achieve optimal results (Figure 8).

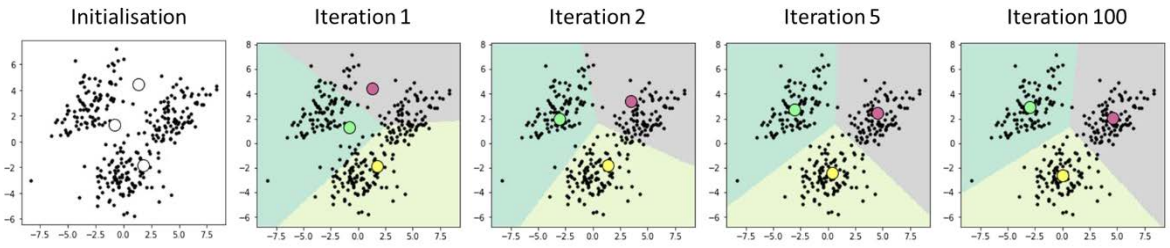


Figure 8 Illustration of K-Means clustering algorithm
The first 100 iterations of K-means with user-specified hyperparameter k = 3. The three large dots in each graph are the location of the centroids calculated using the data within each cluster.

Many recent algorithms, including X-Means, G-Means, flowPeaks, flow clustering without k (FLOCK) and flowMeans, have been developed for cell subtype clustering without needing to input a value for k in flow cytometry data analysis [95-99]. In a sense,

these algorithms iteratively cluster data with different k values and use evaluation techniques such as Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) or Minimum Description Length (MDL) to find the optimal k value automatically. In particular, the FLOCK algorithm uses a grid-based approach to find high-density regions that enable K-Means clustering to converge significantly faster than random initialization of cluster centroids and simultaneously determines an appropriate value for k based on the number of dense regions [99]. While the flowMeans algorithm, like the more recently improved K-Means algorithm, is a time-efficient yet accurate approach for automatically identifying cell subtypes in cytometry data analysis [95]. The flowMeans algorithm can identify cell populations with irregular shapes by using several centroids to represent the same population, i.e., spilling the data into many mini-clusters and merging them. In a sense, flowMeans enables high-throughput FACS data analysis pipelines that overcome the initialisation, population shape limitation, and repeated model selection issues of the conventional K-Means clustering.

Density-based spatial clustering of applications with noise (DBSCAN) is a non-parametric data clustering algorithm that does not require a user-specified k value for the number of clusters. DBSCAN can detect noise while performing data clustering (without assigning the noise data to any group [100, 101]. It begins by detecting core points using a user-defined data neighbourhood size (ε) and the minimal number of data required within each neighbourhood ($MinPts$). Once the core points have been identified, data within the overlapped core points neighbourhood and the adjacent non-core points are clustered together (Figure 9). Intuitively, DBSCAN identifies clusters as high-density regions separated by low-density regions. The FlowGrid algorithm, as the more recent

DBSCAN-based flow cytometry clustering algorithm, has combined the grid-based approach of the FLOCK algorithm with the power of DBSCAN, and it can handle the high-parameter data with irregular shapes with significantly reduced computation time [102].

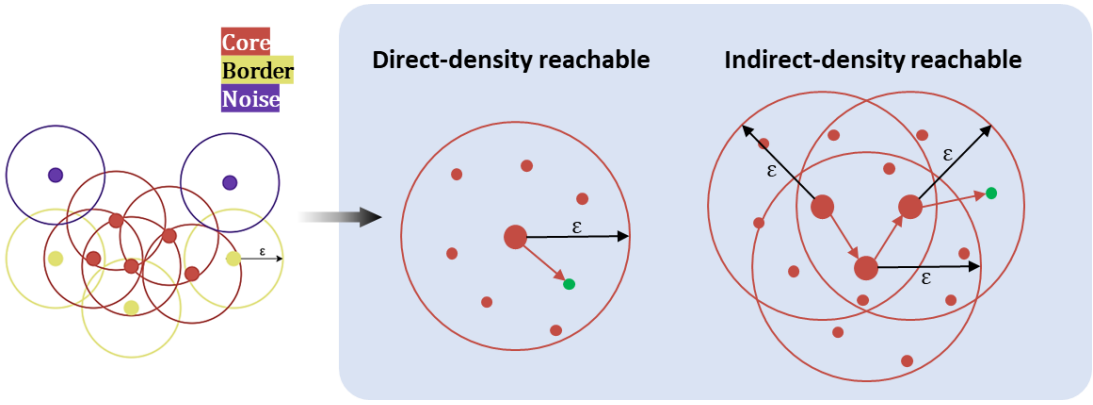


Figure 9 Illustration of DBSCAN clustering algorithm
The red dots are core points defined by $MinPts$ within the ε radius of the centre data, and the yellow dots are the border data being merged into the core clusters. The purple dots are noise data. The little green dots are the data being 'reached' by the centre data within each core cluster.

A finite mixture model (FMM) is a group of clustering algorithms that, as the name implies, performs unsupervised data clustering using a mixture of distribution models (Figure 10). Popular flow cytometry data clustering algorithms include Gaussian mixture model (GMM), flowClust, flowMerge and flow analysis with automated multivariate estimation (FLAME). GMM assumes the cell population distribution combines multiple Gaussian distributions from each cluster [103-107]. However, adopting Gaussian distributions might result in misleading cluster splits due to outliers and skewness in the data. The flowClust algorithm solves these issues using a Box-Cox transformation to decrease the data probability distribution skewness and then a t-distributions mixture model to represent the cell population [105]. The t-distribution enables a higher tolerance for outlier impacts in the cell populations because it has higher tails at both

ends than the Gaussian distribution data (Figure 10B) [105]. The flowMerge algorithm extends the flowClust framework by employing a cluster merging procedure to further support a concave cell population represented by multiple probability distribution components [104].

Additionally, the flowMerge algorithm provides an automatic cluster number selection method, making it compatible with the high-throughput cytometry data analysis pipelines with automation. Alternatively, FLAME does not require the Box-Cox transformation. It uses a mixture of skew t-distributions to increase the model's flexibility and adaptability to fit cytometry data with skewness [106]. This approach permits unsupervised learning with non-Gaussian populations, which are tolerant of outliers and data probability distribution with heavy tails.

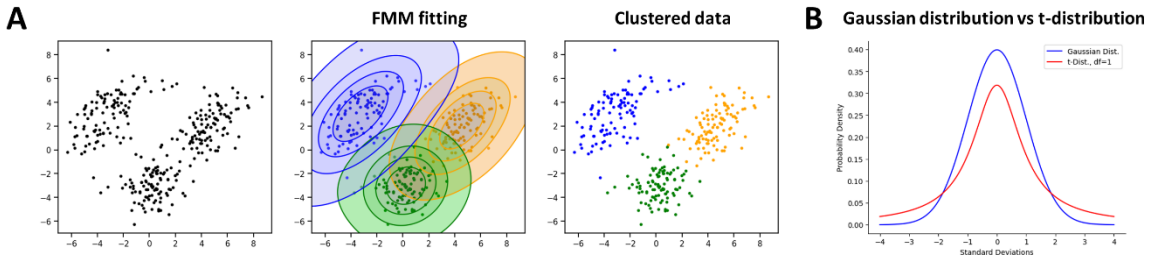


Figure 10 Illustration of FMM clustering algorithm
FMM clustering three groups of data with the 2D Gaussian models (A). Each coloured contour circle (blue, green, orange) indicates a 2D Gaussian model that best fits the data. Data within the circles are identified as the same cluster. A comparison of the Gaussian distribution and the t-distribution (B). The t-distribution has higher tails and is more tolerant of outliers.

The agglomerative approach of the hierarchical clustering algorithm is a bottom-up solution that starts with assigning each cell to its own cluster and sequentially merging them (Figure 11). The spanning-tree progression analysis of density-normalized events (SPADE) is a versatile machine-learning algorithm for high-dimensional single-cell flow

cytometry data clustering using the agglomerative hierarchical method [108]. The SPADE algorithm first minimises the data density variation using density-dependent down-sampling, which ultimately balances the large and small cell population sizes and boosts the relative frequency of rare populations. Then, density-normalised single-cell data are merged with their nearest neighbours based on the calculated distances, and the merging steps run iteratively until the number of remaining clusters reaches a user-defined threshold. A minimum spanning tree (MST) that links the density-normalised data is constructed to produce a tree-like graphic representation of the sample. Lastly, SPADE uses up-sampling to allocate the complete original data to clusters where their nearest neighbours belong.

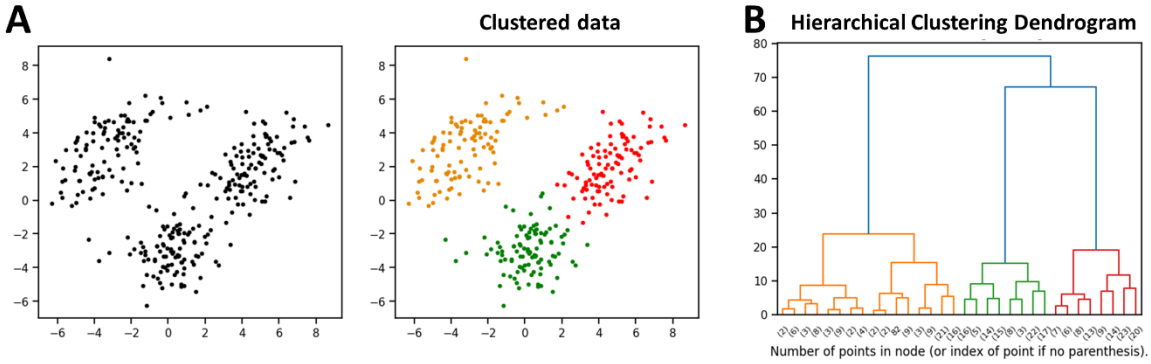


Figure 11 Illustration of hierarchical clustering
Data is being clustered using the agglomerative hierarchical clustering algorithm (A). The dendrogram shows the distance between data and how data are merged and clustered based on the distance (B).

The PhenoGraph flow cytometry clustering algorithm uses an alternative approach based on the Louvain community detection method and the k-nearest neighbour (k-NN) approach [109]. PhenoGraph is a robust, data-driven clustering algorithm that requires only one hyperparameter input k to construct the k-NN graph. It is important to note

that this differs from the hyperparameter k required in K-Means for cluster numbers. Then PhenoGraph assigns single-cell flow cytometry data into different communities based on the neighbour density of each cell. It uses the modularity maximisation method to maintain the clustering connection between the intracommunity data points and removes the intercommunity connections for cluster segmentation. In theory, smaller k in PhenoGraph will result in more clusters by splitting the community network with higher resolution, and higher k will lead to a lower-resolution clustering with fewer cluster numbers. In practice, PhenoGraph is not very sensitive to k, i.e., the algorithm will return the same cluster number with a reasonable range of different k values [109].

A self-organizing map (SOM) is a type of artificial neural network (ANN) that is an unsupervised learning algorithm for data clustering as well as data visualization by producing a low-dimensional data representation [110-112]. The FlowSOM algorithm is developed particularly for flow cytometry data clustering, and the workflow of FlowSOM has four stages: (i) data normalization, (ii) constructing a SOM artificial neural network representation of the data clustering, (iii) building an MST visualization for the SOM clustering results, and (iv) perform the meta-clustering using a consensus hierarchical clustering algorithm [111]. The consensus hierarchical clustering algorithm works by random subsampling the observations multiple times and implements hierarchical clustering on each subsample. A final meta-clustering result is made with the initial SOM clustering results and based on how often the same clusters are further clustered to form a supercluster. The consensus hierarchical clustering approach gets better results than the basic hierarchical clustering approach in terms of clustering stability. However, FlowSOM requires a user-specified k value for the desired number of final clusters, the

size of the network grid, the learning rate, and the start and end neighbourhood radius for the neurons in SOM. The requirement of multiple hyperparameters has made FlowSOM a complicated yet powerful flow cytometry data clustering algorithm.

1.4 FRET quantification

FRET detection can be achieved by either measuring the quenching of FET-donor emissions or the increase of sensitized FRET-acceptor emissions. These can be achieved using chemically-linked single molecules, such as tandem dyes, the PE-Cy7 fluorophore [39], or chemically-linked fluorescent biosensor protein molecules using amino acid linkers [10, 68, 113]. Alternatively, FRET can be achieved in unlinked interacting molecular or protein multimers, such as two individual fluorescently conjugated monoclonal antibodies [114], two naturally fluorescent interacting proteins [115], or two synthetically expressed fluorescent fusion proteins [32], dependent upon the molecular distance between the fluorescent donor and acceptor elements within such interactions. Taking the ratio of sensitised FRET emission over the quenched donor emission can only provide semi-quantitative results [34] since the FRET fluorescent intensity reflects both the PPIs level, the expression level of those fluorescently labelled POIs, and the instrument configurations.

FRET efficiency (E) is an unbiased full quantification parameter for the FRET event, independent of the used flow cytometer configuration, the protein expression level, and FRET efficiency is inversely proportional to the sixth power of the distance between donor and acceptor fluorophores that come into proximity (Eq.1).

Equation 1 FRET efficiency and FRET donor-acceptor distance

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}$$

The parameter R is the actual distance between donor and acceptor fluorophores, and R_0 is the "Förster radius", which is the distance at which the FRET efficiency is 50% [3, 25]. The "Förster radius" R_0 is determined by the spectral properties of the FRET-donor and FRET-acceptor fluorophores, which can be calculated using the orientation of the fluorophores (k^2), the quantum yield of the donor (ϕ_D), the refractive index of the medium surrounding the fluorophores (n), and the spectral overlap integral of the donor and acceptor (J), as shown in (Eq2.) [116].

Equation 2 The Föster radius

$$R_0 = 0.211 * \sqrt[6]{k^2 n^{-4} \phi_D J(\lambda)} \ \& J(\lambda) = \frac{\int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda}{\int_0^\infty F_D(\lambda) d\lambda}$$

$F_D(\lambda)$ is the donor's normalised fluorescence emission (depending on wavelength), ε_A is the acceptor's extinction coefficient (in M^{-1} cm^{-1}) (dependent on wavelength), and λ is the wavelength. As the actual k^2 of FPs is unknown, the orientation between fluorophores is commonly considered to be 2/3, equivalent to a random orientation. A different k^2 value would result in a different R_0 value, but the trends between the various FRET pairs remain unchanged. Fluorescent proteins of the same spectral class frequently exhibit identical excitation and emission spectra. However, their extinction coefficients (ε_A) and quantum yields (ϕ_D) may differ. When the donors have identical emission spectra, Eq. 2 predicts that the FP with the greater quantum yield (ϕ_D) is the superior donor. Due to the bigger overlap integral, an acceptor with a more significant extinction coefficient than another acceptor with similar excitation spectra is anticipated to be a superior acceptor. The R_0 of several regularly used fluorophores in

FRET have already been calculated and discussed [51]. Moreover, FPBase (https://www.fpbase.org) is a highly recommended online R_0 calculation tool [117].

In order to achieve FRET efficiency calculations using the flow cytometry-based fluorescent intensity values, the detected fluorescence emission needs to be first spectrally unmixed to obtain clean donor, acceptor and sensitised FRET emission quantities. Spectrally unmixing requires single-colour controls for each fluorescent source, but sensitised FRET has no pure single-colour control since simultaneous donor and acceptor emissions always exist. FRET spectral unmixing can be accomplished with the mathematical formulation of the mixed fluorescent signals within each detection channel as the weighted combinations of the donor, acceptor, and FRET emissions with all potential spillover signals [10, 16]. This approach requires the FRET calibration factor α which is defined as the ratio between quenched donor emission and sensitised FRET emission (see section 2.7.6) [114, 118]. The α can be determined using linear regression to predict the intercept and coefficient (Eq.3) [10]. Chemically linked FRET pairs must be used to ensure the intercept and coefficient are constant numbers for α determination.

Equation 3 Linear regression approach to deduct α factor for FRET calibration

$$R_F - 1 = \frac{I_D E \alpha}{I_D(1-E)}$$

$$R_1 = \frac{I_D E \alpha}{I_A S_2}$$

$$\frac{1}{R_F - 1} = \frac{\varepsilon^D_{\lambda_D} C_D}{\varepsilon^A_{\lambda_D} C_A} + \frac{\varepsilon^D_{\lambda_D} C_D}{\varepsilon^A_{\lambda_D} C_A} \alpha \frac{1}{R_1}$$

Unlinked free-interacting FRET pair can also be used but can be extremely challenging and require the expression of the POIs driven by the same promoter and the assumption

of the known concentration ratio of the donor and acceptor molecules [32]. Once the α factor is determined, the FRET efficiency can be calculated (see section 2.7.6).

1.5 Prospective investigations

Many limitations in the flow cytometry data analytical workflow still need to be resolved before full FRET quantification can be achieved. Especially for the experiment using mammalian cells, an accurate, automated, rapid, and unbiased algorithm still needs to be developed. The usage of many unsupervised clustering algorithms is well evaluated for cell subtype identification using the fluorescence detection parameter from the flow cytometry data. Their performance in singlet identification using scattered light signals has not been determined. Advancements can be made based on the existing clustering algorithms, particularly utilising the FSC/SSC signals, which is a good starting point for promising results.

Currently, the gold-standard flowClean and flowAI algorithms both fail to resolve the negative-value issue caused by baseline-subtraction-errors. Flow cytometry produces multiparameter data, and for each individual cell data, negative values are often detected within some detection parameters and not all detectors. Therefore, the high-parameter measurements guarantee a great chance of at least one detector producing high-quality data for every cell. Furthermore, instead of removing the entire single-cell data, the specific poor-quality measurement can be removed and treated as missing data. Using collaborative learning algorithms, such as collaborative filtering (CF) recommendation algorithm, that predict the missing values using the existing high-quality data can provide accurate solutions for correcting the baseline subtraction-errors [119]. The CF recommendation algorithm is also a nonlinear approach for data

prediction; therefore, exploring the potential of using CF for autofluorescence prediction and removal can also shed light on achieving the clean flow cytometry fluorescence readout. Once the clean fluorescent measurement is achieved, the spread error of spectral unmixing based on FRET emission formulation and linear models can also be vastly eliminated. In addition, other robust linear fitting algorithms, such as random sample consensus (RANSAC), can also be employed as an iterative method for estimating the spillover coefficients. Spillover coefficients from the highly overlapped FRET emissions that may contain undetected outliers can further increase the accuracy of the linear spectral unmixing results to accomplish purely unmixed fluorescent signals.

The current FRET calibration using the α factor requires the population mean for constructing the linear regression model. Therefore, despite the single-cell resolution of the downstream FRET efficiency calculation, the accuracy can only be maintained at the population level. Therefore, there is an urgent need for a single-cell level FRET calibration algorithm, especially for the challenging chemically unlinked FRET system. Furthermore, a few studies have explored the capacity to conduct FRET measurements and FRET efficiency quantification using the three-molecule FRET system. However, the current approaches still treat each pair of the two molecules within the three-way FRET event as independent components. In other words, the energy competition from the single FRET donor to two different FRET acceptors has not been thoroughly investigated in flow cytometry-based FRET analysis. Algorithms that transform the data into compositional data for relative comparisons, such as the CoDA method, should be utilised for flow cytometry FRET data interpretation.

CHAPTER TWO

METHODS AND ALGORITHM DESIGNS

2.1 Molecular biology methods

2.1.1 Generation of human TNFR-fluorescent fusion protein expression plasmids

Human TNFR1 was chosen to examine and validate both the two-protein three-colour FRET and three-protein six-colour FRET algorithms. As previously described by Chan et al., human TNFR1 can form self-assembled homodimers through the pre-ligand-binding assembly domain (PLAD) [72], and Vanamee et al. also indicated that cytokine-recruited trimers of dimers in a hexagonal meta-cluster conformation [120]. Human CD27 and TNFR2 were used as FRET-negative controls for comparison with the TNFR1 FRET-positive detection. The major advantage of using TNFR1, TNFR2, and CD27 is that they are all from the same tumour necrosis factor receptor superfamily (TNFRSF) with similar structures, and TNFR1 does not form heterodimers with either TNFR2 or CD27 [32, 121, 122].

The human pcDNA3.TNFR1-eCFP, pcDNA3.TNFR1-eYFP, pcDNA3.TNFR2-eCFP and pcDNA3.TNFR2-eYFP plasmids were constructed as previously described and kindly provided by Dr Francis Chan [70]. All pcDNA3 plasmids are under the control of a cytomegalovirus (CMV) promoter. Human TNFR1 and TNFR2 open reading frames (ORFs) (without the stop codon) were cloned into the multi-cloning site (MCS) of pcDNA3 using the 5'-*Hind*III and 3'-*Xho*I sites, and the downstream eCFP and eYFP ORFs were cloned using the 5'-*Xho*I and 3'-*Xba*I sites. Therefore, the plasmids have the fused receptor-FP sequences with six nucleotides as the linker from the XhoI cloning site. All pcDNA3 vectors contain ampicillin and neomycin resistance genes for selection in both bacteria and mammalian host cells, respectively.

To generate the CD27 fusion protein pcDNA3 based -eCFP and -eYFP plasmids, the human CD27 cDNA sequence was first amplified from the mRNA that was isolated from human peripheral blood mononuclear cells by RT-PCR (First Strand cDNA synthesis kit, Life Technologies) using PCR forward primer 5'- AGCAGGTACCATGGCACGGCCACATCCCT GGTGG-3' and reverse primer 5'-TACTAACTCGAGGGGGGAGCAGGCAGGCTCCGGTT-3'. The PCR introduced the *Kpn*I and *Xho*I restriction enzyme sites (underlined) into the amplified products and enabled either TNFR1 or TNFR2 cDNA to be replaced with the CD27 cDNA.

To generate pcDNA3.TNFR1-mRFP, pcDNA3.TNFR2-mRFP and pcDNA3.CD27-mRFP plasmids, monomeric RFP ORF was PCR amplified from pVitro2-mRFP (a gift from Dr Rosetta Martiniello-Wilks) with the forward primer 5'-ATCCTCGAGATGGCCTCCTCCGA GGA-3' and a plasmid specific reverse primer 5'-AACCTGCTCCTAGGGTCGACAATCGAT-3' which contains an *Xho*I and *Avr*II (compatible with *Xba*I) restriction sites. Thus, the mRFP cDNA replaced eCFP or eYFP cDNAs in the pcDNA3 vectors.

2.1.2 Generation of TRAPS mutants as -eYFP fusion protein expression plasmids

To address the capacity of the FRET algorithms in analysing mutant proteins and WT proteins interactions, such as in the scenario of autoinflammatory diseases caused by germline mutations, the tumour necrosis factor receptor associated periodic syndrome (TRAPS) mutants TNFR1-eYFP were generated through site-directed mutagenesis (SDM). The SDM reactions were carried out using the pcDNA3-TNFR1-eYFP plasmid as the template and designed forward and reverse primer pairs as listed below (Table 1).

Table 1 Primers for site-direct mutagenesis of TRAPS plasmids

Domain	Location	Usual name	protein name	Sequence change	Forward Primer (Sequence)	Tm (F)	Reversed Primer (Sequence)	Tm (R)
PLAD/CRD1	Exon 2	D12E	p.(Asp41Glu)	c.123T>G	AGAAGAGAGAgAGTGTGTGTCCC	64°C	CCCTGTCCCCTAGGTGAG	67°C
PLAD/CRD1	Exon 2	C15Y	p.(Cys44Tyr)	c.131G>A	GATAGTGTGTaTCCCCAAGGAAAATATATC	64°C	TCTCTTCTCCCTGTCCCC	65°C
PLAD/CRD1	Exon 2	Y20H	p.(Tyr49His)	c.145T>C	TGTGTCCCCAAGGAAAAcATATCCACCCTC	63°C	TTTTCCTTGGGGACACACACTATCTCT	58°C
PLAD/CRD1	Exon 2	Y20D	p.(Tyr49Asp)	c.145T>G	TGTGTCCCCAAGGAAAAgATATCCACCCTC	63°C	TTTTCCTTGGGGACACACACTATCTCT	58°C
PLAD/CRD1	Exon 2	Y20C	p.(Tyr49Cys)	c.146A>G	CAAGGAAAATgTATCCACCCTC	58°C	GGGACACACACTATCTCTC	61°C
PLAD/CRD1	Exon 2	H22Y	p.(His51Tyr)	c.151C>T	TGTCCCCAAGGAAAATATATCtACCCTCAAAATA	60°C	TTTTCCTTGGGGACACACACTATCTCT	58°C
PLAD/CRD1	Exon 2	H22R	p.(His51Arg)	c.152A>G	TGTCCCCAAGGAAAATATATCCgCCCTCAAAATA	62°C	TTTTCCTTGGGGACACACACTATCTCT	58°C
PLAD/CRD1	Exon 2	H22Q	p.(His51Gln)	c.153C>G	AATATATCCAgCCTCAAAATAATTCGATTTG	62°C	TTCCTTGGGGACACACAC	64°C
PLAD/CRD1	Exon 2	N25D	p.(Asn54Asp)	c.160A>G	CCACCCTCAAgATAATTCGATTTG	59°C	ATATATTTTCCTTGGGGAC	56°C
PLAD/CRD1	Exon 2	I28S	p.(Ile57Ser)	c.170T>G	AATAATTCGAgTTGCTGTACC	57°C	TTGAGGGTGGATATATTTTC	56°C
PLAD/CRD1	Exon 2	C29R	p.(Cys58Arg)	c.172T>C	TAATTCGATTcGCTGTACCAAG	56°C	TTTTGAGGGTGGATATATTTTC	58°C
PLAD/CRD1	Exon 2	C29G	p.(Cys58Gly)	c.172T>G	TAATTCGATTgGCTGTACCAAG	58°C	TTTTGAGGGTGGATATATTTTC	58°C
PLAD/CRD1	Exon 2	C29F	p.(Cys58Phe)	c.173G>T	TCAAAATAATTCGATTTtCTGTACCAAGTG	55°C	AAATCGAATTATTTTGAGGGTGGATATATT	53°C
PLAD/CRD1	Exon 2	C29Y	p.(Cys58Tyr)	c.173G>A	TCAAAATAATTCGATTTaCTGTACCAAGTG	55°C	AAATCGAATTATTTTGAGGGTGGATATATT	53°C
PLAD/CRD1	Exon 2	C29S	p.(Cys58Ser)	c.173G>C	AATTCGATTTcCTGTACCAAGTG	57°C	ATTTTGAGGGTGGATATATTTTC	58°C
PLAD/CRD1	Exon 2	C29W	p.(Cys58Trp)	c.174C>G	ATTCGATTTGgTGTACCAAGTG	60°C	TATTTTGAGGGTGGATATATTTTC	58°C
PLAD/CRD1	Exon 2	C30R	p.(Cys59Arg)	c.175T>C	CAAAATAATTCGATTTGCcGTACCAAGTGCCA	60°C	AAATCGAATTATTTTGAGGGTGGATATATT	53°C
PLAD/CRD1	Exon 2	C30S	p.(Cys59Ser)	c.176G>C	CAAAATAATTCGATTTGCTcTACCAAGTGCCA	59°C	AAATCGAATTATTTTGAGGGTGGATATATT	53°C
PLAD/CRD1	Exon 2	C30Y	p.(Cys59Tyr)	c.176G>A	CAAAATAATTCGATTTGCTaTACCAAGTGCCA	58°C	AAATCGAATTATTTTGAGGGTGGATATATT	53°C
PLAD/CRD1	Exon 2	C30F	p.(Cys59Phe)	c.176G>T	CAAAATAATTCGATTTGCTtTACCAAGTGCCA	58°C	AAATCGAATTATTTTGAGGGTGGATATATT	53°C
PLAD/CRD1	Exon 2	C33G	p.(Cys62Gly)	c.184T>G	CGATTTGCTGTACCAAGgGCCACAAAGGAA	63°C	CTTGGTACAGCAAATCGAATTATTTTG	54°C
PLAD/CRD1	Exon 2	C33Y	p.(Cys62Tyr)	c.185G>A	CGATTTGCTGTACCAAGTaCCACAAAGGAA	60°C	CTTGGTACAGCAAATCGAATTATTTTG	54°C
PLAD/CRD1	Exon 3	G36E	p.(Gly65Glu)	c.194G>A	TGCCACAAAGaAACCTACTTG	59°C	CTTGGTACAGCAAATCGAATTATTTTG	63°C
PLAD/CRD1	Exon 3	T37I	p.(Thr66Ile)	c.197C>T	CAAGTGCCACAAAGGAATCTACTTGTACAA	59°C	TTCCTTTGTGGCACTTGGTACAGCAAATCGA	62°C
PLAD/CRD1	Exon 3	Y38C	p.(Tyr67Cys)	c.200A>G	AAAGGAACCTgCTTGTACAATGAC	61°C	GTGGCACTTGGTACAGCA	66°C
PLAD/CRD1	Exon 3	Y38S	p.(Tyr67Ser)	c.200A>C	CAAGTGCCACAAAGGAACCTcCTTGTACAA	62°C	AAGTAGGTTCCTTTGTGGCACTTGGTA	58°C
PLAD/CRD1	Exon 3	L39F	p.(Leu68Phe)	c.204G>C	CACAAAGGAACCTACTTcTACAATGACTGT	59°C	AAGTAGGTTCCTTTGTGGCACTTGGTA	58°C
PLAD/CRD1	Exon 3	D42DEL	p.(Asp71del)	c.211_213delGAC	TGTCCAGGCCCGGGGCAG	77°C	ATTGTACAAGTAGGTTCCTTTGTGGCACTTGG	72°C
PLAD/CRD1	Exon 3	D42E	p.(Asp71Glu)	c.213C>A	TGTACAATGAaTGTCCAGGCC	64°C	AGTAGGTTCCTTTGTGGC	62°C
PLAD/CRD1	Exon 3	C43R	p.(Cys72Arg)	c.214T>C	CCTACTTGTACAATGACcGTCCAGGCCCGG	67°C	GTCATTGTACAAGTAGGTTCCTTTGTG	57°C
PLAD/CRD1	Exon 3	C43G	p.(Cys72Gly)	c.214T>G	GTACAATGACgGTCCAGGCCC	65°C	AAGTAGGTTCCTTTGTGGC	62°C
PLAD/CRD1	Exon 3	C43Y	p.(Cys72Tyr)	c.215G>A	TACAATGACTaTCCAGGCCCG	65°C	CAAGTAGGTTCCTTTGTGG	60°C
PLAD/CRD1	Exon 3	C43S	p.(Cys72Ser)	c.215G>C	TACAATGACTcTCCAGGCCCG	63°C	CAAGTAGGTTCCTTTGTGG	60°C
PLAD/CRD1	Exon 3	C43F	p.(Cys72Phe)	c.215G>T	TACAATGACTtTCCAGGCCCG	63°C	CAAGTAGGTTCCTTTGTGG	60°C
PLAD/CRD1	Exon 3	P46L	p.(Pro75Leu)	c.224C>T	CAATGACTGTCCAGGCCtGGGGCAGGATA	66°C	GGCCTGGACAGTCATTGTACAAGTAGG	61°C
PLAD/CRD1	Exon 3	Q77H	p.(Gln77His)	c.231G>T	GCCCGGGGCAtGATACGGACT	70°C	CTGGACAGTCATTGTACAAGTAGGTTC	67°C
PLAD/CRD1	Exon 3	T50M	p.(Thr79Met)	c.236C>T	GCCCGGGGCAGGATAtGGACT	70°C	CTGGACAGTCATTGTACAAGTAGGTTC	67°C
PLAD/CRD1	Exon 3	T50K	p.(Thr79Lys)	c.236C>A	GCCCGGGGCAGGATAaGGACT	70°C	CTGGACAGTCATTGTACAAGTAGGTTC	67°C
PLAD/CRD1	Exon 3	C52R	p.(Cys81Arg)	c.241T>C	GGATACGGACcGCAGGGAGTGTGAGAGC	73°C	TGCCCCGGGCCTGGACAG	77°C
PLAD/CRD1	Exon 3	C52F	p.(Cys81Phe)	c.242G>T	GATACGGACTtCAGGGAGTGTGAGAGCGG	73°C	CTGCCCCGGGCCTGGACA	77°C
PLAD/CRD1	Exon 3	C52Y	p.(Cys81Tyr)	c.242G>A	GATACGGACTaCAGGGAGTGTGAGAGCGG	73°C	CTGCCCCGGGCCTGGACA	77°C
PLAD/CRD1	Exon 3	C52S	p.(Cys81Ser)	c.242G>C	GATACGGACTcCAGGGAGTGTGAGAGCG	72°C	CTGCCCCGGGCCTGGACA	77°C
PLAD/CRD1	Exon 3	C52W	p.(Cys81Trp)	c.243C>G	ATACGGACTGgAGGGAGTGTGAGAGCG	73°C	CCTGCCCCGGGCCTGGAC	78°C
PLAD/CRD1	Exon 3	R53G	p.(Arg82Gly)	c.244A>G	TACGGACTGCgGGGAGTGTGAGAGC	73°C	TCCTGCCCCGGGCCTGGA	78°C
CRD4	Exon 6	L167_G175del	p.(Leu196_Gly204del)	c.586_612del27	ACTGAGGACTCAGGCACC	66°C	GCACAACTTCGTGCACTC	64°C
CRD4	Exon 5	C139S	p.(Cys139Ser)	c.503G>C	GTGTGCACCTcCCATGCAGGT	68°C	GGTGTTCTGTTTCTCCTGGC	65°C
CRD4	Exon 5	C139A	p.(Cys139Ala)	c.502_503TG>GC	CGTGTGCACCgcCCATGCAGGT	63°C	GTGTTCTGTTTCTCCTGG	59°C
CRD4	Exon 5	C150S	p.(Cys150Ser)	c.536G>C	GAAAACGAGTcTGTCTCCTGTAG	58°C	TCTTAGAAAGAAACCTGC	56°C
CRD4	Exon 5	C150A	p.(Cys150Ala)	c.535_536TG>GC	AGAAAACGAGgcTGTCTCCTGTAGTAAC	56°C	CTTAGAAAGAAACCTGCATG	58°C
CRD4	Exon 5	C153S	p.(Cys153Ser)	c.545G>C	TGTGTCTCCTcTAGTAACTGTAAG	58°C	CTCGTTTTCTCTTAGAAAGAAAC	59°C
CRD4	Exon 5	C153A	p.(Cys153Ala)	c.544_545TG>GC	GTGTGTCTCCgcTAGTAACTGTAAGAAAAG	55°C	TCGTTTTCTCTTAGAAAGAAAC	57°C
CRD4	Exon 6	C156S	p.(Cys156Ser)	c.554G>C	TGTAGTAACTcTAAGAAAAGCCTG	56°C	GGAGACACACTCGTTTTC	59°C
CRD4	Exon 6	C156A	p.(Cys156Ala)	c.553_554TG>GC	CTGTAGTAACgcTAAGAAAAGCCTGGAGTG	57°C	GAGACACACTCGTTTTCTC	59°C
CRD4	Exon 6	C162S	p.(Cys162Ser)	c.572G>C	AGCCTGGAGTcCACGAAGTTG	64°C	TTTCTTACAGTTACTACAGGAG	59°C
CRD4	Exon 6	C162A	p.(Cys162Ala)	c.571_572TG>GC	AAGCCTGGAGgcCACGAAGTTG	58°C	TTCTTACAGTTACTACAGG	56°C
CRD4	Exon 6	C166S	p.(Cys166Ser)	c.584G>C	ACGAAGTTGTcCCTACCCCAG	63°C	GCACTCCAGGCTTTTCTTAC	64°C
CRD4	Exon 6	C166A	p.(Cys166Ala)	c.583_584TG>GC	CACGAAGTTGgcCCTACCCCAG	59°C	CACTCCAGGCTTTTCTTAC	60°C
After CRD4	Exon 6	I170N	p.(Ile199Asn)	c.596T>A	CTACCCCAGAaTGAGAATGTTAAG	60°C	GCACAACTTCGTGCACTC	64°C
After CRD4	Exon 6	I170T	p.(Ile199Thr)	c.596T>C	CTACCCCAGAcTGAGAATGTTAAG	64°C	GCACAACTTCGTGCACTC	64°C
After CRD4	Exon 6	V173D	p.(Val202Asp)	c.605T>A	ATTGAGAATGaTAAGGGCACTGAG	61°C	CTGGGGTAGGCACAACTT	65°C
After CRD4	Exon 6	G204C	p.(Gly204Cys)	c.610G>T	GAATGTTAAGtGCACTGAGGACTC	60°C	TCAATCTGGGGTAGGCAC	64°C
After CRD4	Exon 7	S197C	p.(Ser226Cys)	c.677C>G	TGCCTTTTATgCCTCCTCTTC	61°C	AAGACCAAAGAAAATGACC	56°C

To begin the SDM process, 12.5 µl of Q5 Hot Start High-Fidelity 2X Master Mix (New England BioLabs), 1.25 µl of each of the 10 µM forward and reverse primer, 2 µl of 10 ng/µl WT pcDNA3.TNFR1-eYFP template plasmid DNA, and 9 µl of the nuclease-free water, were assembled in a thin-walled PCR tube with a final volume of 25 µl per SDM reaction. The reagents are completely mixed before performing the routine PCR on the thermocycler (SimpliAmp^{TM}, Thermo Fisher Scientific). After 30 seconds of the initial denaturation at 98 °C, a total of 25 cycles of PCR amplification were carried out with denaturing at 98 °C for 10 seconds, re-annealing at 1-2 °C below the lowest Tm of the primer (Table 1) for 30 seconds and an extension of 30 seconds/Kb DNA at 72 °C. A final extension was performed at 72 °C for 2 minutes, and then the PCR products were held at 4 °C before collection and storage.

2.1.3 Generation of poxviruses vTNFR-mRFP fusion protein expression plasmids

With the aim of further testing the robustness of the FRET algorithms, the investigation of human TNFR1 interactions was expanded to include the poxviruses-encoded viral TNFRs (vTNFRs). The vTNFRs used in this study include the smallpox Variola virus-encoded G4R, Myxoma virus-encoded MT2 and Monkeypox virus-encoded J2R. All vTNFRs were previously generated as the pcDNA3.VarG4R-MycHis, pcDNA3.MyxMT2-MycHis and pcDNA3.MpxJ2R-MycHis plasmids and kindly provided by Dr Alexander Gale [123].

To ensure optimal expression of the Variola G4R and Monkeypox J2R ORFs in both *Homo sapien* and *Mus musculus* cell lines, the sequence for Variola G4R and Monkeypox J2R were codon optimised by Codon Devices Inc., USA, using the Variola India 1967 strain sequence (NCBI accession: NP_042240.1) and the Monkeypox 1996 Zaire strain

sequence (NCBI accession: AF380138.1), respectively. The codon optimisation altered the nucleotide sequences without altering the amino acid sequences to match the available tRNA pool of *Homo sapien* and *Mus musculus* cell lines. In addition, the cloning site sequences of *Kpn*I, *Xho*I, *Hind*III and *Xba*I were also altered to allow subcloning using these restriction enzymes at specific cloning sites in the plasmid vector. The plasmids containing codon-optimised vTNFRs-MycHis sequences were provided by Dr Sarah Sherwood [124]. All vTNFRs-MycHis plasmids were digested with the *Hind*III and *Xho*I restriction enzymes and cloned into the pcDNA3.TNFR1-mRFP plasmid by replacing the human TNFR1 cDNA using the same restriction enzymes. Due to biohazard restrictions on handling the Variola virus and Monkeypox virus, the ORFs of G4R and J2R were synthesised with WHO and smallpox committee approval.

2.1.4 Generation of pGL4.23.NF-κB-[miniP]-miRFP703 inflammation reporter plasmids

TNFR1-mediated activation of NF-κB regulated cellular reporters were generated using the pGL4.23.[miniP] vector (Promega) to evaluate the simultaneous detection of the TNFR1 receptor interaction FRET emissions and the intercellular inflammation signalling pathway activities. The first step of constructing the inflammation reporter plasmid is to generate the 2X NF-κB response element (in bold) using the designed oligonucleotides 5'-GATCCTA**GGGAATTCCCGGGAATTCCC**TAGGATC-3' with the same *Avr*II restriction enzyme digestion sites on each end. The oligonucleotides were diluted with the annealing buffer (10 mM Tris, pH 7.5, 50 mM NaCl and 1 mM EDTA) to a final concentration of 50 µM. A total of 200 µl oligonucleotides solution was then added to a thin layer PCR tube on the thermocycler. After an initial incubation at 95 °C for 2 minutes,

the temperature was cooled to 25 °C over 45 minutes for self-annealing. Following a quick centrifugation, the collected self-annealed 2X NF-κB response element dsDNA was then digested with the *Avr*II restriction enzyme in the digestion buffer. After 2 hours of digestion, the digested dsDNA was further purified using ethanol precipitation.

To construct 4X and 8X NF-κB response elements, the purified digested 2X dsDNA was ligated using the T4 ligase (New England BioLabs). All NF-κB response element fragments were cloned into the pGL4.23.[miniP] vector upstream of the minimal promoter sequence using the *Nhe*I (*Avr*II compatible) cloning site and this generated the pGL4.23.NF-κB-[miniP] plasmids.

To clone the miRFP703 fluorescent protein reporter sequence into the pGL4.23.NF-κB-[miniP] plasmids, the pmiRFP703 plasmid (Addgene) was used. The miRFP703 cDNA was first digested from the pmiRFP703 plasmid using *Nco*I and *Xba*I restriction enzymes. Then, the released miRFP703 cDNA was cloned into the pGL4.23.NF-κB-[miniP] plasmids downstream of the minimal promoter sequence using the same *Nco*I and *Xba*I cloning sites. The final process generated the final pGL4.23.NF-κB-[miniP]-miRFP703 inflammation reporter plasmids with 2X, 4X and 8X NF-κB response elements. Meanwhile, the miRFP703 cDNA was also cloned into the pGL4.23.[miniP] vector using the same method without NF-κB response element for use as the negative control.

2.1.5 DNA quantification

Plasmid DNA was measured by 260nm absorbance (A260) using a Nanodrop-One Spectrophotometer (Thermo Fisher Scientific) according to the manufacturer's instructions. The quality of DNA was assessed by A260/A280 and A260/A230, where ratios of around 1.8 and 1.8-2.2, respectively, are defined as 'pure'.

2.1.6 Restriction enzyme digestion

Restriction enzyme digestions were performed with 1X optimal enzyme buffer, restriction enzymes (10 units/1 µg of DNA), 0.1 µg of plasmid DNA, and 0.1 mg/ml bovine serum albumin (BSA) (New England BioLabs). Reactions were incubated at the required temperatures for two hours, after which the enzyme was heat-inactivated for 20 min at 80 °C, except for XhoI and AvrII, which cannot be heat-inactivated.

2.1.7 DNA gel electrophoresis

Tris-borate-EDTA (TBE) agarose gels, usually 1% (w/v) unless indicated, were cast by dissolving 1 g of agarose (Astral Scientific) into 100 ml of TBE, using the microwave at power level-6 until completely dissolved. GelRed (Biotium) was added into the molten agarose gel solution at 1:10,000 (v/v) and mixed thoroughly before pouring into a casting tray (Bio-Rad) fitted with 8- or 20-well comb and allowed to solidify at RT. DNA samples and a 1Kb DNA ladder (Invitrogen) were mixed with 1-2 µL of 10X blue loading buffer (TaKaRa Bio) and loaded into wells. The DNA was electrophoresed at 80V (Bio-Rad PowerPac-300) until the loading buffer indicator dye had run approximately 3/4 of the length of the gel.

2.1.8 DNA gel extraction

Digested DNA fragments were visualised using InGenius3 UV (302nm) trans-illuminator (Syngene), and images were acquired using the GeneSys software (version 1.5.0.0). The DNA fragments of interest were excised with a sterile razor blade and collected into a 2 mL Eppendorf tube. The gel slices were weighted using a scale sensitive to 0.001g (A200S Analytical balance, Sartorius) and processed with the PureLink Quick Gel Extraction Kit (Invitrogen) following the manufacturer's protocol. Briefly, gel slices were dissolved with

3 gel-volumes of Solubilisation Buffer (L3) by incubation for 15 min at 50 °C in a heat block (AccuBlock). One gel-volume equivalent of 100% isopropanol (Sigma-Aldrich) was then added and centrifuged (5415D Centrifuge, Eppendorf) at 12,000 × g for 1 min at RT. The gel/DNA sample was subsequently purified using a Quick Gel extraction column: DNA/gel samples were centrifuged at 12,000 × g for 1 min at RT, washed with 500 µL Wash Buffer (W1, containing 80% ethanol) and centrifuged again. Residual ethanol was removed by a further centrifugation step at approximately 16,000 × g for 2 min. Finally, the DNA was eluted in 50 µL Elution Buffer (E5) by centrifugation.

2.1.9 DNA ligation

Plasmid DNA ligations were performed on ice in reaction mixtures containing 1 µL of 10X T4 DNA ligase buffer and 1 µL of T4 DNA ligase (New England Biolabs). Generally, approximately 100 ng of vector and cDNA insert, often in 3 reactions with vector to insert ratios or 1:3, 1:1, or 3:1, using dH2O to a final total volume of 10µL. Reaction mixtures were incubated on a single layer of ice at room temperature overnight, where the reaction temperature will cover a range of 4 to 20 °C or R/T, i.e. as the ice melts overnight. The ligase enzyme was heat-inactivated for 10 min at 65 °C before *E. coli* transformation of the newly ligated plasmid.

2.1.10 DNA sequencing

To validate the plasmid cloning results, extracted and purified plasmid DNAs were sent to Macrogen Korea for Sanger DNA sequencing using the 3730XL sequencers (Applied Biosystems). The plasmid DNA samples were diluted to a concentration of 100 ng/µl for delivery via post. The sequencing reactions were carried out by the facility staff, and the sequencing results were returned as ".ab1" files. Sequences were analysed as detailed

in the following section 2.6.2. The primers used for sequencing are listed below (Table 2).

Table 2 Primers for plasmid sequencing

Primer Names	Primer Sequence	Sense	Tm (°C)
pcDNA3 FWD (T7)	5'-TAATACGACTCACTATAGG-3'	Sense	53
pcDNA3 REV (SP6)	5'-ATTTAGGTGACACTATAG-3'	Antisense	51
eCFP/eYFP-N-REV	5'-CGTCGCCGTCCAGCTC-3'	Antisense	65
mRFP-N-REV	5'-GTACTGGAACTGAGGGGACAG-3'	Antisense	65
HuCD27 FWD	5'-TGTGGAGCCTGCAGAGCCTTGTC-3'	Sense	72
HuTNFR1 FWD	5'-ACTCAGGCACCACAGTGCTGTT-3'	Sense	70
HuTNFR2 FWD	5'-CCTTGTGCCTGCAGAGAGAAGC-3'	Sense	68
VarG4R FWD	5'-GACCAGCGAGCTGACTATCACA-3'	Sense	67
MpvJ2R FWD	5'-TCAATCTCTACTAGCGAGCTGACA-3'	Sense	66
MyxMT2 FWD	5'-ATGTCCTCGGTACACGTATTCCG-3'	Sense	67
pGL4.23 FWD (RV primer3)	5'-CTAGCAAAATAGGCTGTCCC-3'	Sense	61
pGL4.23 REV (EBV-rev)	5'-GTGGTTTGTCCAAACTCATC-3'	Antisense	60

Plasmid Sample	Primer Used for Sequencing
pcDNA3.HuCD27-eCFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), HuCD27 FWD, eCFP/eYFP-N-REV
pcDNA3.HuTNFR1-eCFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), HuTNFR1 FWD, eCFP/eYFP-N-REV
pcDNA3.HuTNFR2-eCFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), HuTNFR2 FWD, eCFP/eYFP-N-REV
pcDNA3.HuCD27-eYFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), HuCD27 FWD, eCFP/eYFP-N-REV
pcDNA3.HuTNFR1-eYFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), HuTNFR1 FWD, eCFP/eYFP-N-REV
pcDNA3.HuTNFR2-eYFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), HuTNFR2 FWD, eCFP/eYFP-N-REV
pcDNA3.HuCD27-mRFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), HuCD27 FWD, mRFP-N-REV
pcDNA3.HuTNFR1-mRFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), HuTNFR1 FWD, mRFP-N-REV
pcDNA3.HuTNFR2-mRFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), HuTNFR2 FWD, mRFP-N-REV
pcDNA3.HuTNFR1(TRAPS)-eYFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), HuTNFR1 FWD, eCFP/eYFP-N-REV
pcDNA3.VarG4R-mRFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), VarG4R FWD, mRFP-N-REV
pcDNA3.MpvJ2R-mRFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), MpvJ2R FWD, mRFP-N-REV
pcDNA3.MyxMT2-mRFP	pcDNA3 FWD (T7), pcDNA3 REV (SP6), MyxMT2 FWD, mRFP-N-REV
pGL4.23.NF-κB-[miniP]-miRFP703	pGL4.23 FWD (RV primer3), pGL4.23 REV (EBV-rev)

2.2 Bacterial Methods

2.2.1 Chemically competent bacteria cells preparation

For plasmid DNA storage, chemically competent bacteria cells were made for plasmid heat-shock transformation. *E.coli* DH5 α bacterial cells (genotype: *fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*) were provided by Dr Iain Duggin (UTS). Briefly, 500 mL Super Optimal Broth (SOB) (20 g/L

Tryptone, 5 g/L Yeast Extract, 0.584 g/L NaCl, 0.186 g/L KCl, 2.4 g/L $MgSO^4$, pH = 7.0) was inoculated by a single colony of DH5α *E. coli* picked from a freshly streaked Luria-Bertani (LB) Agar plate (10 g/L Tryptone, 5 g/L Yeast Extract, 10 g/L NaCl, 15 g/L Agar). The inoculated SOB was further incubated at 18 °C with shaking at 200-250 rpm until reaching an OD^{600} of 0.6. All incubations steps of bacteria were performed for 10 min at 4 °C, and centrifugations (CR-22GIII Centrifuge, Hitachi) at 3220×g for 10 min at 4 °C. Briefly, bacteria were incubated and then harvested by centrifugation, then resuspended with 80 mL of ice-cold Transformation Buffer (TB) (250 mM KCl, 15 mM $CaCl_2$, 55 mM $MnCl_2$, and 10 mM PIPES, pH = 6.7). After incubation, bacteria were harvested by centrifugation and resuspended in 20 mL of 93% TB and 7% Dimethyl sulfoxide (DMSO) (Sigma-Aldrich). Chemically competent bacteria were dispensed into pre-chilled 1.5 mL Eppendorf tubes and stored at -80 °C until use.

2.2.2 Heat-shock transformation

For each transformation, 100 µL of ice-cold (thawed) competent *E. coli* were gently mixed with plasmid DNA (5 µL, 100 ng/µL) in a pre-chilled sterile Eppendorf tube and incubated on ice for 20 mins. Cells were then incubated in a heat block for 1 min at 42 °C and immediately transferred back to on ice for 2 min. A 900 µL aliquot of SOB with catabolite repression (20 mM glucose) (SOC) was added, and the *E. coli* were incubated for 60 min at 37°C. Finally, the bacteria were centrifuged at 3000 × g for 5 min at RT (Eppendorf 5415D centrifuge), and the bacterial pellet was resuspended in 100 µL of fresh SOC, then plated onto LB Agar plates (with or without antibiotics) and cultured overnight at 37 °C. Transformed bacteria were stored in 15% (v/v) glycerol solution at -80 °C until use.

2.2.3 DNA mini-prep purification

Plasmid mini-preparations were performed using ISOLATE-II Plasmid Mini Kit (Bioline) following the manufacturer's instructions. Briefly, centrifugation steps were 12,000g for 1 min at RT. Transformed DH5α *E. coli* in 5 mL LB containing 80 µg/mL ampicillin was first harvested by centrifugation for 30 s. Cells were then resuspended in 250 µL Resuspension Buffer (P1) containing 0.4 mg/mL RNase-A, lysed by incubation with 250 µL alkaline Lysis Buffer (P2) for 5 min at RT, and neutralised with 300 µL acid Neutralisation Buffer (P3). The lysate was clarified by centrifugation for 5 min. A volume of 750 µL clarified supernatant was pipetted into a silica column and centrifuged. The column was then cleaned by a 2-step washing with 500 µL preheated (50 °C) Wash Buffer (PW1) and 600 µL Wash Buffer (PW2, containing 80% ethanol), where each wash step was followed by centrifugation. Residual ethanol was removed by centrifugation for 2 min before finally eluting the plasmid DNA in 50 µl of elution buffer (PE).

2.2.4 DNA maxi-prep purification

Plasmid maxipreparation was performed using the PureLink HiPure Plasmid Filter Purification Kit (Life Technologies) following a similar protocol, starting with a 200 mL culture volume. In addition, eluted plasmid DNA was precipitated with 10.5 mL isopropanol followed by centrifugation at 12,000 × g for 30 min at 4 °C. After centrifugation, and plasmid DNA was resuspended with 5 mL 70% ethanol followed by another centrifugation at 12,000 × g for 5 min at 4 °C and air-dried for 15 min at RT then finally resuspending in 500 µL TE buffer.

2.3 Cell Biology Methods – Mammalian Expression

2.3.1 Mammalian cell tissue culture

Human embryonic kidney cells contains the SV40 large T antigen (HEK-293T) were the gift of Dr Grant Logan (CMRI, Australia) [125]. HEK-293T cells were grown in Dulbecco's Modified Eagle medium (DMEM, pH = 7.4) (Gibco) supplemented with 5% heat-inactivated fetal bovine serum (FBS) (Invitrogen), 100 U/ml penicillin and 100 mg/ml streptomycin (Gibco), in a 37°C humidified 5% CO_2 incubator. HEK-293T cells were sub-cultured twice weekly at a ratio of ~1:8. In brief, media was discarded, and cells were washed with 0.9% normal saline (NaCl) (Baxter) incubated in 5 mL 0.25% (w/v) Trypsin-0.53 mM EDTA solution (Gibco) for 5-10min at 37 °C, to detach from the plastic culture vessel. Fresh DMEM containing 10% FBS was added; then, the cells were harvested by centrifugation at 350 × g for 5 min at 4 °C in 50 mL Falcon tubes. The cell pellet was resuspended with approximately 10 mL of fresh DMEM5 before transferring it into a new tissue culture flask.

2.3.2 Calcium phosphate DNA transfection

HEK-293T cells were seeded into 6- or 24-well tissue culture dishes (Corning) on the previous day and at confluency sufficient to reach ~70% confluency by the next day. Media was removed and replaced 2 hours before transfection. For 6-well dishes, a mixture containing a total of 2 µg of plasmid DNA for each plasmid, plus 10 µL of 2.5M $CaCl_2$ (Sigma-Aldrich) diluted to a total volume of 100µL in sterile dH2O water. The DNA mixture was then added dropwise with intermittent mixing (by vortexing) into 100 µL of N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid (HEPES) buffered saline (HBS) (280 mM NaCl, 1.5 mM Na_2HPO_4, 50 mM HEPES, 10 mM KCl, and 12 mM Dextrose, pH = 7.05)

and incubated for 20 min at RT. The precipitated DNA mixture was then evenly added to cells by pipetting. Culture media were replaced with fresh DMEM5 after culture for 4 hours, and the cells were incubated for a further 36 hours before harvesting, depending on the experimental needs. Transient transfection volume was scaled down proportionally for 48-well dishes.

2.3.3 Cell harvesting

Transfected cells were first detached from the bottom of the tissue culture vessel by gentle pipetting. Single-cell suspensions were prepared, where cells were gauze-filtered (100 µm gauze, Sefar) to remove cell aggregates that would otherwise cause blockage in the flow cytometer. Cells were then collected by centrifugation at 350 × g for 5 min at 4 °C. To eliminate potential biohazard, harvested cells were fixed with 1 ml 4% (v/v) paraformaldehyde (PFA) in 1 × Phosphate-Buffered Saline (PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.8 mM KH_2PO_4) prior to flow cytometry data acquisition.

2.3.4 Cell Live/Dead staining

The LIVE/DEAD™ Fixable Blue Dead Cell Stain Kit (Invitrogen^{TM}) was used to determine the viability of cells prior to PFA fixation. The stain is designed to be excited by a UV laser (350 nm) and detected at around 450 nm wavelength. For every 40 cell samples, 50 µl of anhydrous DMSO was added to one vial of the fluorescent reactive dye, which both were freshly thawed and brought to room temperature immediately before use. For each cell sample, collect around 1×10^6 cells in suspension and wash the cells once with 1 ml of PBS. After resuspending the cells with 1 ml of PBS, 1 µl of the reconstituted live/dead dye was added to each cell sample and mixed well with gentle vortexing. Following 30 minutes of incubation at RT in the dark, the cells were rewashed using 1

mL PBS and then resuspended with 1 ml 4% (v/v) PFA in PBS. PFA fixation was carried out at RT for 15 minutes, and then the cells were washed and resuspended for the last time using 1 ml PBS with 1 % (w/v) BSA.

2.4 Flow Cytometry

2.4.1 Single-cell sorting

To establish the true ground of the singlet and non-singlet populations, single-cell sorting was conducted using the non-transfected cells. The Symphony S6 cell sorter (Becton-Dickinson) performed cell sorting with samples' concentration of around 1×10^6 cells per ml PBS in suspension. The six-step singlet gating using FSC-A/SSC-A, FSC-H/SSC-H, FSC-A/FSC-H, SSC-A/SSC-H, FSC-H/FSC-W and SSC-H/SSC-W scatterplots were formed on site with the non-singlet identified as events outside those gates. Cells were sorted into tissue culture plates and directly examined using microscopy. Of note, there is a potential bias from the cell sorting samples due to the possibility that non-singlet events at the detection point might be disrupted into single cells by the high-speed fluidic sorting procedure.

2.4.2 The detection of two-protein three-colour FRET

To test the FRET algorithm for detecting TNFR1 homodimer interactions, cells were separately co-transfected using pcDNA3.TNFR1-eCFP + pcDNA3.TNFR1-eYFP for eCFP → eYFP FRET detection, pcDNA3.TNFR1-eYFP + pcDNA3.TNFR1-mRFP for eYFP → mRFP FRET detection, and pcDNA3.TNFR1-eCFP + pcDNA3.TNFR1-mRFP for eCFP → mRFP FRET detection. For FRET-negative controls, pcDNA3.TNFR1-eCFP + pcDNA3.CD27-eYFP, pcDNA3.TNFR1-eYFP + pcDNA3.CD27-mRFP, and pcDNA3.TNFR1-

eCFP + pcDNA3.CD27-mRFP transfected cells were used for each of the detection of above FRET signals, respectively. In order to test the reliability of the FRET algorithm, harvested cell samples were divided into duplicate sets of FACS tubes and data were acquired on an LSR-II flow cytometer (Becton-Dickinson) equipped with UV 355 nm (20 mW), violet 405 nm (25 mW), blue 488 nm (20 mW) and red 635 nm (20 mW) lasers and a Fortessa X20 flow cytometer (Becton-Dickinson) equipped with violet 405 nm (50 mW), blue 488 nm (100 mW) and red 640 nm (40 mW) lasers, on the same day. The cytometry acquisition settings were as follows: first, the FSC-A/SSC-A and SSC-A/SSC-H scatterplots were drawn to detect cells with the FSC threshold set at 5000, then the FSC-A/FSC-H, SSC-A/SSC-H, FSC-H/FSC-W and SSC-H/SSC-W scatterplots were used to define and gate on single cells. Next, the unstained pcDNA3 empty vector-transfected cells and single-colour controls (single transfections with CFP, YFP or RFP expression plasmids) were acquired and used to adjust the laser voltages and set the compensation values to minimize bleed-through into adjacent channels. This is necessary due to the broad emission spectra of these fluorophores.

For detector configuration, the eCFP donor, eCFP → eYFP and eCFP → mRFP FRET emissions were all detected on the 405 nm violet laser detector array, with 450/50 nm, 500 LP 546/10 nm and 595 LP 610/20 nm filters on the LSR-II and with 450/50 nm, 505 LP 540/10 nm and 600 LP 610/20 nm filters on the Fortessa X20. The YFP donor/acceptor, mRFP acceptor and eCFP→mRFP FRET emissions were all detected on the 488 nm blue laser detector array, and of note, due to the lack of 561 nm laser on both flow cytometers, the mRFP was excited with the 488 nm blue laser. In detail, the emissions were detected with 500 LP 530/30 nm, 635 LP 670/14 nm and 595 LP 610/20 nm filters

on the LSR-II and with 505 LP 542/27 nm, 635 LP 695/40 nm and 600 LP 610/20 nm filters on the Fortessa X20, respectively. The co-transfected cells were then examined using two-parameter dot plots to define single-colour FP and double-positive (eCFP and eYFP, or eYFP and mRFP, or eCFP and mRFP) co-transfected cells. Generally, data from 30,000 cells were collected using the FACS DIVA software (v8.0.2, Becton-Dickinson). FACS data were recorded and exported as 18-bit flow cytometry ".FCS3.0" files.

2.4.3 The detection of three-protein six-colour FRET

To test the FRET algorithm for detecting TNF-recruited TNFR1 homotrimer interactions, cells were co-transfected using the pcDNA3.TNFR1-eCFP + pcDNA3.TNFR1-eYFP + pcDNA3.TNFR1-mRFP plasmids. For comparison and validation, at each of the eCFP, eYFP and mRFP positions, the TNFR1 fusion protein plasmid was replaced by the TNFR2, CD27 fusion protein plasmid or simply replaced by the pcDNA3 empty vector. The LSR-II was used with the same hardware configuration described in the detection of two-protein three-colour FRET section for data acquisition. However, all eCFP, eYFP, mRFP, eCFP → eYFP, eYFP → mRFP and eCFP → mRFP detection channels were used simultaneously.

The FRET algorithm for the three-protein six-colour FRET analysis was also tested using the Aurora^{TM} full-spectrum flow cytometer (Cytek) data and the conventional multichannel filter-based LSR-II flow cytometer. The full-spectrum flow cytometer was equipped with a 405nm (100mW) laser, a 488nm (50mW) laser, a 561nm (50mW) laser and a 640nm (80mW) laser. The laser-detector modules each have 16, 14, 10, and 8 detector channels, spanning 420-829nm, 498-829nm, 567-829nm, and 652-829nm, respectively (Appendix Figure 1). FCS data were recorded using the SpectroFlo software

(version 2.2.0.3), with the FSC threshold set to 200,000. With 10,000 events in the single cell gate, data were recorded and saved as 22-bit ".FCS3.1" files.

2.4.4 The simultaneous detection of three-protein six-colour FRET with cell signalling

To evaluate the possibility of simultaneously detecting multiple cells signalling pathway fluorescence reporters and the three-protein six-colour FRET emissions, cells were co-transfected with the pcDNA3.TNFR1-eCFP + pcDNA3.TNFR1-eYFP + pcDNA3.TNFR1-mRFP plasmids and the pGL4.23.NF-κB-[miniP]-miRFP703 inflammation reporter plasmid. The pGL4.23.NF-κB-[miniP]-miRFP703 inflammation reporter plasmids with different copies of NF-κB response elements were used for each transfection to determine the optimal signal-to-background ratio. Also, live/dead staining was performed on transfected cells before sample harvesting.

For flow cytometry data acquisition, the LSR-II flow cytometer was used, and the configuration was upgraded from the three-protein six-colour FRET configuration mentioned in the last section. In addition to the eCFP, eYFP, mRFP, eCFP$\rightarrow$eYFP, eYFP$\rightarrow$mRFP and eCFP$\rightarrow$mRFP detection channels, a 410LP 450/50 nm detector was used on the 355nm UV laser array for live/dead detection, and a 710LP 730/45 nm detector was used for the miRFP703 inflammation reporter detection.

2.5 Microscopy

2.5.1 EVOS FL cell imaging system

The expression of TNFR-fluorescent fusion proteins in 293 cells was confirmed by the EVOS FL cell imaging system (Life Technologies) before harvesting for the flow cytometry

detection or live-cell imaging. Cells were inspected, and images were taken under a transmitted channel and three fluorescence channels with filter cubes: CFP (445/45 excitation; 510/42 emission), YFP (500/24 excitation; 524/27 emission) and RFP (531/40 excitation; 593/40 emission). The light-emitting diode (LED) intensities were fixed for illumination in each channel under 10x objective lens (LPlanFL PH2, NA 0.30, WD 1.2 mm), transmitted (50% LED-intensity), CFP (30% LED-intensity), YFP (30% LED-intensity) and RFP (40% LED-intensity). Images were captured with a built-in ICX285AQ colour charge-coupled camera (Sony).

2.5.2 Live-cell imaging

The live-cell A1R confocal microscopy (Nikon) was used to examine the optimal sample harvesting time for the detection of flow cytometry-based FRET. The A1R data characterised the expression profiles of the human CD27, TNFR1 and TNFR2 -eCFP, -eYFP and -mRFP fusion proteins. The HEK-293T cells were transfected, and after the initial 6 hours of incubation, a total of 48 hours of live-cell imaging was performed after replacing the transfected cell culture with the fresh medium.

The eCFP, eYFP, and mRFP emission was excited by the 30 mW 458 nm, 514 nm and 561 nm lasers and detected at 482/35 nm, 540/30 nm and 595/50 nm, respectively. Emissions were detected through a Plan Apo λ 20x lens (numerical aperture: 0.75, refractive index: 1.0). A cell without expressing TNFR1-RFP was also analysed to generate the background. In addition, the expression of miRFP703 inflammation reporters were also examined using the Nikon A1R confocal microscopy using the 50 mW 637 nm laser and detected at 650 nm LP.

2.5.3 Full-spectrum imaging

The maximum detection ranges of the eCFP, eYFP and mRFP fluorophores were examined using the Leica Stellaris 8 confocal microscopy platform. The excitation range was from 440 nm to 790 nm with a 1 nm step size, and the emission detection range was from 450 nm to 830 nm with a 5 nm step size.

2.6 Software and Online Server

2.6.1 Primer design

The oligonucleotide primer sequence design was conducted using the Primer Premier software (version 6.0). The appropriate oligonucleotide primer annealing temperature (Tm) for SDM using the Q5 high-fidelity DNA polymerase (New England BioLabs) was estimated using the online NEB Tm calculator (https://tmcalculator.neb.com). For PCR using non-NEB polymerase products, the appropriate annealing temperature was estimated using the SnapGene software (version 5.2.0).

2.6.2 Sequencing analysis

For sequencing analysis, the sequencing chromatogram ".ab1" files were analysed using Python with the SeqIO.read function in the Biopython package (version 1.80). Sequences were aligned using the Basic local alignment search tool (BLAST) from the National Centre for Biotechnology Information (NCBI, https://blast.ncbi.nlm.nih.gov/Blast.cgi). For sequence alignment of human and poxvirus vTNFRs, the Clustal Omega online server was used with the default parameters (https://www.ebi.ac.uk/Tools/msa/clustalo/).

2.6.3 TRAPS mutation sequencing design

The registry of hereditary autoinflammatory disorders mutations website was used (https://infevers.umai-montpellier.fr) to design SDM for TRAPS mutations within the PLAD/CRD1 and CRD4 homodimer interaction critical domains. The website provides TRAPS mutation sequence records of the TNFRSF1A gene. The TRAPS mutation sequence map was generated using the SnapGene software (version 5.2.0).

2.6.4 Flow cytometry singlet manual gating data generation

To generate the ground truth data for singlet identification using the manual gating strategy, the six-step manual gating was performed on collected flow cytometry data (as mentioned for single-cell sorting). FCS files were examined using FlowJo (version 10.8.1), and manual gating was conducted using the polygon gate drawing tool. Gated singlet and non-singlet data were exported as ".csv" files and used to evaluate the singlet identification algorithm performance.

2.6.5 Example FCS data for algorithm validation and application demonstrations

FlowRepository is a database of flow cytometry experiments where data gathered and annotated according to the MIFlowCyt standard can be queried and downloaded. It is widely utilised as a repository for experimental findings published in peer-reviewed flow cytometry journals. In addition to the samples created and relative data collected in this study, three collected flow cytometry datasets on FlowRepository (https://flowrepository.org/) were used to further validate and demonstrate the applications with the FRET algorithms developed in this study. For the chemically linked eGFP-mRFP FRET pairs (Repository ID: FR-FCM-ZZGR) [16], the small molecule IAA/Auxin detection in plant cells (Repository ID: FR-FCM-Z3FL) [34], and the human cluster of

differentiation 4 receptor to T-cell receptor (CD4-TCR) interaction detection (Repository ID: FR-FCM-ZZR6) [126], the flow cytometry datasets were downloaded and used in this study. These datasets were searched using the keyword “FRET” and selected as they have the proper single colour controls and are fully described in published peer-reviewed journals.

2.6.6 Microscopy fluorescent intensity quantification

To perform live-cell imaging quantification, the “.nd2” confocal imaging data file was analysed using the ImageJ software (version 1.53). Image data files were imported and analysed in batch mode using the ImageJ Macro language (IJM). In detail, the files were imported using the “Bio-Formats Macro Extensions” function, and then the multichannel data were split into several individual channels using the “Split Channels” function. Within each fluorescent channel, the MFI was calculated for the entire imaging field at each data collection time spot (for live-cell imaging) or at each excitation and emission step (for full-spectrum imaging). The measured MFI data were exported as “.csv” files.

2.6.7 Myxoma virus MT2 structure and MT2::TNFR1 interaction predictions

The Myxoma virus MT2 structure was predicted using the AlphaFold2 Colab online server (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) [127]. The Myxoma virus MT2 ORF (NCBI accession: AAA46632) was used as the query sequence. Sequence alignments/templates were generated through MMseqs2 and HHsearch. The best-predicted model was created as “.PDB” files and three prediction performance plots were also generated: (i) number of sequences per position,

(ii) predicted local distance difference test (lDDT) per position, which is the percentage of the model prediction confidence at each position, and (iii) the predicted alignment error. The MyxMT2 protein predicted structure was then examined using the ChimeraX (version 1.4) to align with the TNFR1 homodimer structure (PDB ID: 1NCF) using the Matchmaker function with the default parameters. After structure alignment, the alignment score and the root-mean-square deviation (RMSD) were generated.

2.7 Algorithm Designs and Implementations

2.7.1 Automatic optimal channel selections for full-spectrum flow cytometer

Unlike the conventional filter-based flow cytometers that record the fluorescent intensity using pre-configured filters, the full-spectrum flow cytometer measures the fluorescent data using all 48 channels (for 4-laser Cytek Aurora flow cytometer) or all 64 channels (for 5-laser Cytek Aurora flow cytometer). Because the fluorescence emission is broadly detected in multiple channels, an optimal channel was determined by principal component analysis (PCA). The channel with the highest variance contribution in the PC1 dimension was selected, indicating the most extensive separation between fluorescent-negative and fluorescent-positive populations. PCAs were performed using the “decomposition.PCA” function of the scikit-learn package (version 1.2.0) in Python.

2.7.2 Algorithm design for UltraFast singlet identification

The UltraFast algorithm was developed for automatic unsupervised flow cytometry data singlet identification. The algorithm has a total of six stages. The first stage of UltraFast was to remove the background noise data from the primary cell population using the 2D gaussian model fitting using the FSC-A and SSC-A channels. The method was previously described by Razo-Mejia et al. with the threshold set to 0.4 [128], and for UltraFast, the

threshold value was set to 0.95. The second stage was to normalise the data and shrink the massive data linear scale differences of the cells and small debris into a closer logarithmic scale range. The SSC-A, SSC-H, FSC-A and FSC-H data were transformed into the logarithmic scale, and the SSC-W and FSC-W parameters were not transformed (linear scale) to maintain the sensitivity for outliers. Bringing the logarithmic and linear scale to the same data range, the data were then standardised to have the same mean = 0 and variance = 1. The third stage of UltraFast further transforms the six-dimensional data (FSC-A/W/H + SSC-A/W/H) into two-dimensional space using the dimensionality reduction algorithm PCA. The two-dimensional data further allowed the process in stage 4 to split the data points into square grids for higher computational efficiency. Moreover, the grid number is the only hyperparameter that needs to be specified by users; the squared grid number defines the total number of grid boxes used for data transformation. Intuitively, the higher the grid number, the more stringent the UltraFast algorithm will be toward singlet identification. The data within each grid box was then used to calculate the population density, and the elbow point of density change was identified using the elbow method in stage 5. The elbow point was identified by connecting a line between the curve's start and end. Then, the elbow point method calculated the distance of each data point on the curve to the connected line. The data point with the largest distance to the line was identified as the elbow point representing the cut-off point between high-density and low-density data. In the last stage, the DBSCAN algorithm was used for unsupervised clustering. The grid size can be calculated from the grid number defined in stage 4, and then it was converted to the ε hyperparameter as the data neighbourhood radius for the DBSCAN algorithm. Moreover, the automatically identified population critical cut-off point using the elbow point

method was also converted to the $MiniPts$ hyperparameter for the DBSCAN algorithm. Therefore, the UltraFast algorithm can be automatically performed using only one user-specified hyperparameter, the grid number. Additionally, at the last stage, only the primary cluster identified by UltraFast was identified as singlets, and other tiny clusters that did not merge with the primary cluster were identified as non-singlet events.

Several well-established unsupervised clustering algorithms were also tested to compare the performance of the UltraFast singlet identification algorithms. The data was processed the same through stage 1 to stage 3 as UltraFast for a fair comparison. The algorithms were run with iterations with different values for the required hyperparameters, and the optimal singlet identification results were returned with the highest F1 score compared to the manual gating results. In particular, the FlowGrid [102], K-Means [94], G-Means [97], X-Means [98], flowMeans [95], GMM [107], flowClust [105], TASBE with 2D gaussian covariance chi-squared test [86] were used for comparison with the UltraFast algorithm developed in this study.

2.7.3 Algorithm design for baseline subtraction-error correction

The unsupervised collaborative filtering (CF) recommendation algorithm was employed as the core algorithm with additional data processing and transformation to correct the negative fluorescent intensity values generated during baseline subtraction. Firstly, each single data point containing at least one negative error value was removed from the entire multichannel data, converting the negative data into missing data that needed to be predicted.

For predicting the correct fluorescent intensity of the missing data, the CF algorithm was implemented as below (Eq.4), where n_u is the number of cells, n_m is the number of

detection channels, $r(i,j) = 1$ if cell j has detected fluorescent intensity in channel i, and $y^{(i,j)}$ is the fluorescent intensity from cell j in detection channel i, and it only exists when $r(i,j) = 1$.

Equation 4 The collaborative filtering recommendation algorithm

Given $x^{(1)}, \ldots, x^{(n_m)}$ and their fluorescent intensities to estimate attributes $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(nu)}$:

$$Cost\ function = \min_{\theta^{(1)}, \ldots, \theta^{(n_u)}} \frac{1}{2} \sum_{j=1}^{n_u} \sum_{i:r(i,j)=1} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right)^2 + \frac{\lambda}{2} \sum_{j=1}^{n_u} \sum_{k=1}^{n} \left(\theta_k^{(j)} \right)^2$$

Given $\theta^{(1)}, \ldots, \theta^{(nu)}$, for $n_u =$ number of cells and $n_m =$ number of detection channels to predict $x^{(i)}$,..., $x^{(n_m)}$ for all data:

$$Cost\ function = \min_{x^{(1)}, \ldots, x^{(n_m)}} \frac{1}{2} \sum_{i=1}^{n_m} \sum_{j:r(i,j)=1} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right)^2 + \frac{\lambda}{2} \sum_{i=1}^{n_m} \sum_{k=1}^{n} \left(x_k^{(i)} \right)^2$$

Minimising $x^{(1)}, \ldots, x^{(n_m)}$ and $\theta^{(1)}, \ldots, \theta^{(nu)}$ simultaneously:

$$J\left(x^{(1)}, \ldots, x^{(n_m)}, \theta^{(1)}, \ldots, \theta^{(nu)}\right) = \min_{\substack{x^{(1)}, \ldots, x^{(n_m)} \\ \theta^{(1)}, \ldots, \theta^{(n_u)}}} \frac{1}{2} \sum_{(i,j):r(i,j)=1} \left((\theta^{(j)})^T x^{(i)} - y^{(i,j)} \right)^2$$

$$+ \frac{\lambda}{2} \sum_{i=1}^{n_m} \sum_{k=1}^{n} \left(x_k^{(i)} \right)^2 + \frac{\lambda}{2} \sum_{j=1}^{n_u} \sum_{k=1}^{n} \left(\theta_k^{(j)} \right)^2 , \ x \in \mathbb{R}^n, \ \theta \in \mathbb{R}^n$$

The number of attributes was set to 50, and the cost functions were iteratively and simultaneously updated for both θ and x, until the differences between the original good-quality data and the predicted good-quality data were minimised (< 1×10^8). Then the predicted values for original poor-quality negative data were used to replace those negative values in the original data. The formulation and the implementation of this background subtraction correction algorithm were performed in Python (version 3.9.0).

In particular, the cost function minimisation was designed using the conjugate gradient algorithm “optimize.fmin_cg” function in the SciPy package (version 1.9.3).

2.7.4 Algorithm design for autofluorescence prediction and removal

For autofluorescence prediction, the exact algorithm logic, design, and implementation were performed as the CF recommendation algorithm listed in the baseline subtraction correction section. The number of learned attributes was set to 30, and the data was structured differently before implementing the CF recommendation algorithm. In detail, the fluorescently unlabelled cell sample (e.g., mock-transfected cells) was concatenated with each of the fluorescently labelled cell samples, where the autofluorescence intensities of the fluorescently unlabelled data were the same as the detected fluorescent intensities. For autofluorescence prediction of the fluorescently labelled cell sample, the CF algorithm was implemented with the n_u is the total number of cells of the concatenated data, n_m is twice the number of the total detection channels, and $r(i,j) = 1$ for all data except for the upper left block of the autofluorescence values that needed to be predicted. Once the autofluorescence values were predicted, the autofluorescence was removed from the sample by subtracting the predicted single-cell level channel-wise autofluorescence value from the detected fluorescence for each individual cell in each detection channel.

2.7.5 FRET spectral unmixing

In order to perform FRET spectral unmixing, the original detected and pre-unmixed total fluorescent intensity of each detection channel needs to be formulated as the mixture of emissions from the involved fluorescence components (Eq.5-Eq.8). The fluorescent intensity of each laser-excited emission was notated as I, that I_1 to I_7 are the total

detected fluorescent intensities from the allocated detection channels for eCFP (cyan highlights), eCFP→eYFP FRET1 (green highlights), eCFP→mRFP FRET3 (purple highlights), eYFP (yellow highlights), eYFP→mRFP FRET2 (orange highlights), mRFP (red highlights) and autofluorescence (grey highlights). Since the autofluorescence was predicted and removed using the CF recommendation nonlinear unsupervised algorithm, thus the I_{AutoFl} in each equation can be removed and the I_1 to I_6 were detected fluorescent intensities after autofluorescence was removed. The $I_7(AutoFl)$ should no longer be needed and thus require one less detection channel assigned for autofluorescence detection.

Equation 5 The eCFP→eYFP FRET1 fluorescence components

$$I_1(eCFP) = 1 * I_{eCFP}(1 - E_{FRET1}) + \frac{S_7}{S_8} * \alpha_{FRET1} I_{eCFP} E_{FRET1} + S_7 * I_{eYFP} + S_{19} * I_{AutoFl}$$

$$I_2(eCFP \rightarrow eYFP) = S_1 * I_{eCFP}(1 - E_{FRET1}) + 1 * \alpha_{FRET1} I_{eCFP} E_{FRET1} + S_8 * I_{eYFP} + S_{20} * I_{AutoFl}$$

$$I_4(eYFP) = S_3 * I_{eCFP}(1 - E_{FRET1}) + \frac{1}{S_8}\varepsilon_1 * \alpha_{FRET1} I_{eCFP} E_{FRET1} + 1 * I_{eYFP} + S_{22} * I_{AutoFl}$$

$$I_7(AutoFl) = S_6 * I_{eCFP}(1 - E_{FRET1}) + \frac{S_{12}}{S_8} * \alpha_{FRET1} I_{eCFP} E_{FRET1} + S_{12} * I_{eYFP} + 1 * I_{AutoFl}$$

Equation 6 The eYFP→mRFP FRET2 fluorescence components

$$I_4(eYFP) = 1 * I_{eYFP}(1 - E_{FRET2}) + \frac{S_{16}}{S_{17}} * \alpha_{FRET2} I_{eYFP} E_{FRET2} + S_{16} * I_{mRFP} + S_{22} * I_{AutoFl}$$

$$I_5(eYFP \rightarrow mRFP) = S_{10} * I_{eYFP}(1 - E_{FRET2}) + 1 * \alpha_{FRET2} I_{eYFP} E_{FRET2} + S_{17} * I_{mRFP} + S_{23} * I_{AutoFl}$$

$$I_6(mRFP) = S_{11} * I_{eYFP}(1 - E_{FRET2}) + \frac{1}{S_{17}}\varepsilon_4 * \alpha_{FRET2} I_{eYFP} E_{FRET2} + 1 * I_{mRFP} + S_{24} * I_{AutoFl}$$

$$I_7(AutoFl) = S_{12} * I_{eYFP}(1 - E_{FRET2}) + \frac{S_{18}}{S_{17}}\varepsilon_3 * \alpha_{FRET2} I_{eYFP} E_{FRET2} + S_{18} * I_{mRFP} + 1 * I_{AutoFl}$$

Equation 7 The eCFP→mRFP FRET3 fluorescence components:

$$I_1(eCFP) = 1 * I_{eCFP}(1 - E_{FRET3}) + \frac{S_{13}}{S_{15}} * \alpha_{FRET3} I_{eCFP} E_{FRET3} + S_{13} * I_{mRFP} + S_{19} * I_{AutoFl}$$

$$I_3(eCFP \rightarrow mRFP) = S_2 * I_{eCFP}(1 - E_{FRET3}) + 1 * \alpha_{FRET3} I_{eCFP} E_{FRET3} + S_{15} * I_{mRFP} + S_{21} * I_{AutoFl}$$

$$I_6(mRFP) = S_5 * I_{eCFP}(1 - E_{FRET3}) + \frac{1}{S_{15}} \varepsilon_6 * \alpha_{FRET3} I_{eCFP} E_{FRET3} + 1 * I_{mRFP} + S_{24} * I_{AutoFl}$$

$$I_7(AutoFl) = S_6 * I_{eCFP}(1 - E_{FRET3}) + \frac{S_{18}}{S_{15}} * \alpha_{FRET3} I_{eCFP} E_{FRET3} + S_{18} * I_{mRFP} + 1 * I_{AutoFl}$$

Equation 8 The eCFP→eYFP→mRFP three-way FRET fluorescence components:

$$I_1(eCFP) = 1 * I_{eCFP}(1 - E'_{FRET1} - E'_{FRET3}) + \frac{S_7}{S_8} * \alpha_{FRET1} I_{eCFP} E'_{FRET1}(1 - E_{FRET2}) +$$

$$\frac{S_{13}}{S_{15}} * \alpha_{FRET3} I_{eCFP}(E_{FRET4} + E'_{FRET3}) + S_7 * I_{eYFP}(1 - E_{FRET2}) +$$

$$\frac{S_{13}}{S_{17}} \varepsilon_3 * \alpha_{FRET2} I_{eYFP} E_{FRET2} + S_{13} * I_{mRFP} + S_{19} * I_{AutoFl}$$

$$I_2(eCFP \rightarrow eYFP) = S_1 * I_{eCFP}(1 - E'_{FRET1} - E'_{FRET3}) + 1 * \alpha_{FRET1} I_{eCFP} E'_{FRET1}(1 - E_{FRET2}) +$$

$$\frac{S_{14}}{S_{15}} * \alpha_{FRET3} I_{eCFP}(E_{FRET4} + E'_{FRET3}) + S_8 * I_{eYFP}(1 - E_{FRET2}) +$$

$$\frac{S_{14}}{S_{17}} \varepsilon_3 * \alpha_{FRET2} I_{eYFP} E_{FRET2} + S_{14} * I_{mRFP} + S_{20} * I_{AutoFl}$$

$$I_3(eCFP \rightarrow mRFP) = S_2 * I_{eCFP}(1 - E'_{FRET1} - E'_{FRET3}) + \frac{S_9}{S_8} * \alpha_{FRET1} I_{eCFP} E'_{FRET1}(1 - E_{FRET2}) +$$

$$1 * \alpha_{FRET3} I_{eCFP}(E_{FRET4} + E'_{FRET3}) + S_9 * I_{eYFP}(1 - E_{FRET2}) +$$

$$\frac{S_{15}}{S_{17}} \varepsilon_3 * \alpha_{FRET2} I_{eYFP} E_{FRET2} + S_{15} * I_{mRFP} + S_{21} * I_{AutoFl}$$

$$I_4(eYFP) = S_3 * I_{eCFP}(1 - E'_{FRET1} - E'_{FRET3}) + \frac{1}{S_8} \varepsilon_1 * \alpha_{FRET1} I_{eCFP} E'_{FRET1}(1 - E_{FRET2}) +$$

$$\frac{S_{16}}{S_{15}} \varepsilon_5 * \alpha_{FRET3} I_{eCFP}(E_{FRET4} + E'_{FRET3}) + 1 * I_{eYFP}(1 - E_{FRET2}) +$$

$$\frac{S_{16}}{S_{17}} * \alpha_{FRET2} I_{eYFP} E_{FRET2} + S_{16} * I_{mRFP} + S_{22} * I_{AutoFl}$$

$$I_5(eYFP \rightarrow mRFP) = S_4 * I_{eCFP}(1 - E'_{FRET1} - E'_{FRET3}) + \frac{S_{10}}{S_8}\varepsilon_1 * \alpha_{FRET1} I_{eCFP} E'_{FRET1}(1 - E_{FRET2}) +$$
$$\frac{S_{17}}{S_{15}}\varepsilon_5 * \alpha_{FRET3} I_{eCFP}(E_{FRET4} + E'_{FRET3}) + S_{10} * I_{eYFP}(1 - E_{FRET2}) +$$
$$1 * \alpha_{FRET2} I_{eYFP} E_{FRET2} + S_{17} * I_{mRFP} + S_{23} * I_{AutoFl}$$

$$I_6(mRFP) = S_5 * I_{eCFP}(1 - E'_{FRET1} - E'_{FRET3}) + \frac{S_{11}}{S_8}\varepsilon_2 * \alpha_{FRET1} I_{eCFP} E'_{FRET1}(1 - E_{FRET2}) +$$
$$\frac{1}{S_{15}}\varepsilon_6 * \alpha_{FRET3} I_{eCFP}(E_{FRET4} + E'_{FRET3}) + S_{11} * I_{eYFP}(1 - E_{FRET2}) +$$
$$\frac{1}{S_{17}}\varepsilon_4 * \alpha_{FRET2} I_{eYFP} E_{FRET2} + 1 * I_{mRFP} + S_{24} * I_{AutoFl}$$

$$I_7(AutoFl) = S_6 * I_{eCFP}(1 - E'_{FRET1} - E'_{FRET3}) + \frac{S_{12}}{S_8} * \alpha_{FRET1} I_{eCFP} E'_{FRET1}(1 - E_{FRET2}) +$$
$$\frac{S_{18}}{S_{15}} * \alpha_{FRET3} I_{eCFP}(E_{FRET4} + E'_{FRET3}) + S_{12} * I_{eYFP}(1 - E_{FRET2}) +$$
$$\frac{S_{18}}{S_{17}}\varepsilon_3 * \alpha_{FRET2} I_{eYFP} E_{FRET2} + S_{18} * I_{mRFP} + 1 * I_{AutoFl}$$

In each equation above, the S parameters are the spillover coefficients calculated with the single colour control samples using the RANSAC robust linear regression (Eq.9) (see section 2.8.3 and Appendix Figure 2). The ε parameters are the excitation coefficients ratios of each FRET-donor and FRET-acceptor fluorophores measured at the specific laser excitation wavelength (Eq.10), which the data were obtained from the online database FPbase [117].

Equation 9 Spillover coefficients

$$S_1 = \frac{I_2^{eCFP}}{I_1^{eCFP}},\ S_2 = \frac{I_3^{eCFP}}{I_1^{eCFP}},\ S_3 = \frac{I_4^{eCFP}}{I_1^{eCFP}},\ S_4 = \frac{I_5^{eCFP}}{I_1^{eCFP}},\ S_5 = \frac{I_6^{eCFP}}{I_1^{eCFP}},\ S_6 = \frac{I_7^{eCFP}}{I_1^{eCFP}}$$

$$S_7 = \frac{I_1^{eYFP}}{I_4^{eYFP}},\ S_8 = \frac{I_2^{eYFP}}{I_4^{eYFP}},\ S_9 = \frac{I_3^{eYFP}}{I_4^{eYFP}},\ S_{10} = \frac{I_5^{eYFP}}{I_4^{eYFP}},\ S_{11} = \frac{I_6^{eYFP}}{I_4^{eYFP}},\ S_{12} = \frac{I_7^{eYFP}}{I_4^{eYFP}}$$

$$S_{13} = \frac{I_1^{mRFP}}{I_6^{mRFP}},\ S_{14} = \frac{I_2^{mRFP}}{I_6^{mRFP}},\ S_{15} = \frac{I_3^{mRFP}}{I_6^{mRFP}},\ S_{16} = \frac{I_4^{mRFP}}{I_6^{mRFP}},\ S_{17} = \frac{I_5^{mRFP}}{I_6^{mRFP}},\ S_{18} = \frac{I_7^{mRFP}}{I_6^{mRFP}}$$

$$S_{19} = \frac{I_1^{AutoFl}}{I_7^{AutoFl}},\ S_{20} = \frac{I_2^{AutoFl}}{I_7^{AutoFl}},\ S_{21} = \frac{I_3^{AutoFl}}{I_7^{AutoFl}},\ S_{22} = \frac{I_4^{AutoFl}}{I_7^{AutoFl}},\ S_{23} = \frac{I_5^{AutoFl}}{I_7^{AutoFl}},\ S_{24} = \frac{I_6^{AutoFl}}{I_7^{AutoFl}}$$

Equation 10 Excitation coefficient ratios

$$\varepsilon_1 = \frac{\varepsilon_{eYFP\ laser}^{eCFP}\varepsilon_{eCFP\ laser}^{eYFP}}{\varepsilon_{eCFP\ laser}^{eCFP}\varepsilon_{eYFP\ laser}^{eYFP}},\quad \varepsilon_2 = \frac{\varepsilon_{mRFP\ laser}^{eCFP}\varepsilon_{eCFP\ laser}^{eYFP}}{\varepsilon_{eCFP\ laser}^{eCFP}\varepsilon_{mRFP\ laser}^{eYFP}}$$

$$\varepsilon_3 = \frac{\varepsilon_{eCFP\ laser}^{eYFP}\varepsilon_{eYFP\ laser}^{mRFP}}{\varepsilon_{eYFP\ laser}^{eYFP}\varepsilon_{eCFP\ laser}^{mRFP}},\quad \varepsilon_4 = \frac{\varepsilon_{mRFP\ laser}^{eYFP}\varepsilon_{eYFP\ laser}^{mRFP}}{\varepsilon_{eYFP\ laser}^{eYFP}\varepsilon_{mRFP\ laser}^{mRFP}}$$

The pure FRET-donor, FRET-acceptor, FRET emission, and autofluorescence channel data were unmixed using matrix calculations: FRET1 (Eq.11), FRET2 (Eq.12), FRET3 (Eq.13), and three-way FRET (Eq.14). In three-way FRET the eCFP donor contributes to eCFP→eYFP FRET1, and eCFP→mRFP FRET3, so the FRET efficiency for FRET1 and FRET3 are annotated as E'_{FRET1} and E'_{FRET3} to distinguish them from the FRET efficiency notation for the two-molecule FRET. In addition, the three-molecule eCFP→eYFP→mRFP two-step relay FRET was termed FRET4 (Eq.15).

Equation 11 eCFP→eYFP FRET1 spectral unmixing

$$\begin{bmatrix} I_{unmixed\ 1} = I_{quenched\ eCFP} \\ I_{unmixed\ 2} = I_{sensitised\ FRET1} \\ I_{unmixed\ 4} = I_{eYFP} \\ I_{unmixed\ 7} = I_{AutoFl} \end{bmatrix}^T = \begin{bmatrix} I_{eCFP}(1 - E_{FRET1}) \\ \alpha_{FRET1} I_{eCFP} E_{FRET1} \\ I_{eYFP} \\ I_{AutoFl} \end{bmatrix}^T = \begin{bmatrix} I_1 \\ I_2 \\ I_4 \\ I_7 \end{bmatrix}^T \times \begin{bmatrix} 1 & S_1 & S_3 & S_6 \\ \frac{S_7}{S_8} & 1 & \frac{1}{S_8}\varepsilon_1 & \frac{S_{12}}{S_8} \\ S_7 & S_8 & 1 & S_{12} \\ S_{19} & S_{20} & S_{22} & 1 \end{bmatrix}^{-1}$$

Equation 12 eYFP→mRFP FRET2 spectral unmixing

$$\begin{bmatrix} I_{unmixed\ 4} = I_{quenched\ eYFP} \\ I_{unmixed\ 5} = I_{sensitised\ FRET2} \\ I_{unmixed\ 6} = I_{mRFP} \\ I_{unmixed\ 7} = I_{AutoFl} \end{bmatrix}^T = \begin{bmatrix} I_{eYFP}(1 - E_{FRET2}) \\ \alpha_{FRET2} I \ \ E_{FRET2} \\ I_{mRFP} \\ I_{AutoFl} \end{bmatrix}^T = \begin{bmatrix} I_4 \\ I_5 \\ I_6 \\ I_7 \end{bmatrix}^T \times \begin{bmatrix} 1 & S_{10} & S_{11} & S_{12} \\ \frac{S_{16}}{S_{17}} & 1 & \frac{1}{S_{17}}\varepsilon_4 & \frac{S_{18}}{S_{17}}\varepsilon_3 \\ S_{16} & S_{17} & 1 & S_{18} \\ S_{22} & S_{23} & S_{24} & 1 \end{bmatrix}^{-1}$$

Equation 13 eCFP→mRFP FRET3 spectral unmixing

$$\begin{bmatrix} I_{unmixed\ 1} = I_{quenched\ eCFP} \\ I_{unmixed\ 3} = I_{sensitised\ FRET3} \\ I_{unmixed\ 6} = I_{mRFP} \\ I_{unmixed\ 7} = I_{AutoFl} \end{bmatrix}^T = \begin{bmatrix} I_{eCFP}(1 - E_{FRET3}) \\ \alpha_{FRET3} I_{eCFP} E_{FRET3} \\ I_{mRFP} \\ I_{AutoFl} \end{bmatrix}^T = \begin{bmatrix} I_1 \\ I_3 \\ I_6 \\ I_7 \end{bmatrix}^T \times \begin{bmatrix} 1 & S_2 & S_5 & S_6 \\ \frac{S_{13}}{S_{15}} & 1 & \frac{1}{S_{15}}\varepsilon_2 & \frac{S_{18}}{S_{15}} \\ S_{13} & S_{15} & 1 & S_{18} \\ S_{19} & S_{21} & S_{24} & 1 \end{bmatrix}^{-1}$$

Equation 14 eCFP→eYFP→mRFP three-way FRET spectral unmixing

$$\begin{bmatrix} I_{unmixed\ 1} = I_{quenched\ eCFP} \\ I_{unmixed\ 2} = I_{sensitised\ FRET1} \\ I_{unmixed\ 3} = I_{sensitised\ FRET3} \\ I_{unmixed\ 4} = I_{quenched\ eYFP} \\ I_{unmixed\ 5} = I_{sensitised\ FRET2} \\ I_{unmixed\ 6} = I_{mRFP} \\ I_{unmixed\ 7} = I_{AutoFl} \end{bmatrix}^T = \begin{bmatrix} I_{eCFP}(1 - E'_{FRET1} - E'_{FRET3}) \\ \alpha_{FRET1} I_{eCFP} E'_{FRET1}(1 - E_{FRET2}) \\ \alpha_{FRET3} I_{eCFP}(E_{FRET4} + E'_{FRET3}) \\ I_{eYFP}(1 - E_{FRET2}) \\ \alpha_{FRET2} I_{eYFP} E_{FRET2} \\ I_{mRFP} \\ I_{AutoFl} \end{bmatrix}^T$$

$$= \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \\ I_7 \end{bmatrix}^T \times \begin{bmatrix} 1 & S_1 & S_2 & S_3 & S_4 & S_5 & S_6 \\ \frac{S_7}{S_8} & 1 & \frac{S_9}{S_8} & \frac{1}{S_8}\varepsilon_1 & \frac{S_{10}}{S_8}\varepsilon_1 & \frac{S_{11}}{S_8}\varepsilon_2 & \frac{S_{12}}{S_8} \\ \frac{S_{13}}{S_{15}} & \frac{S_{14}}{S_{15}} & 1 & \frac{S_{16}}{S_{15}}\varepsilon_1 & \frac{S_{17}}{S_{15}}\varepsilon_1 & \frac{1}{S_{15}}\varepsilon_2 & \frac{S_{18}}{S_{15}} \\ S_7 & S_8 & S_9 & 1 & S_{10} & S_{11} & S_{12} \\ \frac{S_{13}}{S_{17}}\varepsilon_3 & \frac{S_{14}}{S_{17}}\varepsilon_3 & \frac{S_{15}}{S_{17}}\varepsilon_3 & \frac{S_{16}}{S_{17}} & 1 & \frac{1}{S_{17}}\varepsilon_4 & \frac{S_{18}}{S_{17}}\varepsilon_3 \\ S_{13} & S_{14} & S_{15} & S_{16} & S_{17} & 1 & S_{18} \\ S_{19} & S_{20} & S_{21} & S_{22} & S_{23} & S_{24} & 1 \end{bmatrix}^{-1}$$

$$E_{FRET4} = E'_{FRET1} \times E_{FRET2}$$

2.7.6 FRET calibration and FRET quantification using FRET efficiency

The FRET efficiency depends on several factors, including (i) the overlap integral of the FRET-donor emission and FRET-acceptor excitation spectrum, (ii) the relative orientation of the FRET-donor and FRET-acceptor fluorophores in the dipole-dipole coupling, and (iii) the distance between the FRET-donor and FRET-acceptor molecules. However, in this FRET intensity-based study, FRET efficiency calculations require a previously defined α factor (Eq.15) [118, 129].

Equation 15 The α calibration factor

$$\alpha = \frac{I_A S \varepsilon^D_{\lambda_D} C_D}{I_D \varepsilon^A_{\lambda_D} C_A}$$

To conduct accurate single-cell level FRET efficiency calculations and FRET calibration for the unlinked naturally free-interacting FRET pairs, the ratio of R_2/R_1 was used (Eq.16), where R_1 is the ratio of sensitised FRET emission and the acceptor emission detected both in the FRET channel, and R_2 is the ratio of excitation coefficients-weighted quenched donor emission and the acceptor emission detected in the donor and acceptor channels, respectively. All fluorescent intensity values were the spectrally unmixed values from the previous steps. The S parameter is the sFRET-acceptor fluorophore's spillover coefficient calculated by the amount of emission detected in the FRET channel over the acceptor channel using the robust RANSAC linear regression. The C_D and C_A are the concentrations of FRET-donor and FRET-acceptor, which cannot be measured directly from the flow cytometry data due to donor quenching during FRET.

Equation 16 The ratios R_1 and R_2 for FRET calibration

$$R_1 = \frac{I_D E \alpha}{I_A S} \quad \& \quad R_2 = \frac{I_D (1-E) \varepsilon_{\lambda_A}^A}{I_A \varepsilon_{\lambda_D}^A}$$

$$\frac{R_2}{R_1} = \frac{I_D (1-E) \varepsilon_{\lambda_A}^A}{I_A \varepsilon_{\lambda_D}^A} * \frac{I_A S}{I_D E \alpha} = \frac{I_D (1-E) \varepsilon_{\lambda_A}^A}{I_A \varepsilon_{\lambda_D}^A} * \frac{I_A S}{I_D E} * \frac{I_D \varepsilon_{\lambda_D}^A C_A}{I_A S \varepsilon_{\lambda_D}^D C_D} = \frac{I_D (1-E)}{I_A} * \frac{\varepsilon_{\lambda_D}^A C_A}{\varepsilon_{\lambda_D}^D C_D} * \frac{1}{E}$$

The Beer-Lambert Law was used to eliminate the requirement of the FRET-donor and FRET-acceptor fluorophore concentrations (Eq.17). The flow cytometry collects cell fluorescence through the hyperdynamic focusing system, where the cells pass through a thin cuvette with the fixed diameter length L and are excited by multiple lasers. Therefore, the concentration terms were cancelled in the R_2/R_1 ratio (see Eq.16) that permitted the true single-cell level FRET efficiency and α factor calculations for the two-molecule FRET.

Equation 17 Single-level unlinked FRET calibration and FRET efficiency

$$I_D = \varepsilon_{\lambda_D}^D C_D \cdot L \quad I_A = \varepsilon_{\lambda_D}^A C_A \cdot L \quad (Beer\text{–}Lambert\ law)$$

$$\frac{R_2}{R_1} = \frac{1-E}{E} = \frac{1}{E} - 1 \quad \Rightarrow \quad E = \frac{1}{\frac{R_2}{R_1} + 1} \quad \Rightarrow \quad \alpha = \frac{I_D E \alpha}{I_D E} = \frac{I_D E \alpha}{\frac{I_D (1-E)}{(1-E)} E}$$

For three-molecule FRET as described in the eCFP→eYFP→mRFP relay FRET, the eCFP donor contributes to eCFP → eYFP FRET1, and eCFP → mRFP FRET3, so the FRET efficiencies ($E'_{FRET1}$and E'_{FRET3}) need to be adjusted (Eq.18). For eYFP→mRFP in the three-molecule FRET, because there is only one donor and one acceptor involved when using the eYFP laser, i.e., the eCFP-related FRETs are not being excited, so the E_{FRET2} can be calculated using the equation listed above (Eq.17).

Equation 18 FRET efficiency for three-way FRET

$$\frac{R2_{FRET1\prime}}{R1_{FRET1\prime}} = \frac{1 - E'_{FRET1} - E'_{FRET3}}{E'_{FRET1}} \;\&\; \frac{E'_{FRET1}}{1 - E'_{FRET3}} = \frac{1}{\frac{R2_{FRET1\prime}}{R1_{FRET1\prime}} + 1} = \theta_1 \quad \Rightarrow \quad E'_{FRET1} = \frac{\theta_1 - \theta_1\theta_3}{1 + \theta_1\theta_3}$$

$$\frac{R2_{FRET3\prime}}{R1_{FRET3\prime}} = \frac{1 - E'_{FRET1} - E'_{FRET3}}{E'_{FRET3}} \;\&\; \frac{E'_{FRET3}}{1 - E'_{FRET1}} = \frac{1}{\frac{R2_{FRET3\prime}}{R1_{FRET3\prime}} + 1} = \theta_3 \quad \Rightarrow \quad E'_{FRET3} = \frac{\theta_3 - \theta_1\theta_3}{1 + \theta_1\theta_3}$$

Additionally, the unquenched fluorescent intensity of each fluorophore before energy donation during FRET can also be deducted from the unmixed fluorescent intensities and calculated α factors (Eq.19). Of note, the accuracy of the FRET quantification using FRET efficiencies and FRET calibration factor α are heavily dependent on the accuracy of the pure fluorescent intensities calculated during spectral unmixing. Therefore, since spectral unmixing is primarily based on the baseline subtraction-error correction and autofluorescence removal, the accuracy level and the algorithm performance of all pre-processing stages are crucial for accurate FRET quantification.

Equation 19 Unquenched fluorescent intensities

$$\boxed{\boldsymbol{I_{Donor\ before\ donation} = I_{Donor\ after\ donation} + \frac{I_{Sensitized\ FRET\ emission}}{\alpha}}}$$

$$I_{eCFP} = I_{unmixed\ 1} + \frac{I_{unnixed\ 2}}{\alpha_{FRET1}}\ (for\ eCFP \rightarrow eYFP)$$

$$I_{eYFP} = I_{unmixed\ 4} + \frac{I_{unmixed\ 5}}{\alpha_{FRET2}}\ (for\ eYFP \rightarrow mRFP)$$

$$I_{eCFP} = I_{unmixed\ 1} + \frac{I_{unnixed\ 3}}{\alpha_{FRET3}}\ (for\ eCFP \rightarrow mRFP)$$

$$I_{eCFP} = I_{unmixed\ 1} + \frac{I_{unnixed\ 2}}{\alpha_{FRET1}} + \frac{I_{unmixed\ 3}}{\alpha_{FRET3}}\ (for\ eCFP \rightarrow eYFP \rightarrow mRFP)$$

$$I_{eYFP} = I_{unmixed\ 4} + \frac{I_{unmixed\ 5}}{\alpha_{FRET2}}\ (for\ eCFP \rightarrow eYFP \rightarrow mRFP)$$

2.7.7 FRET energy dynamic and cellular signalling pathway dynamic analyses

In the three-protein FRET system, the same eCFP donor simultaneously donates energy to both the eYFP acceptor and the mRFP acceptor. In order to resolve the energy dynamics, i.e., the competition between two acceptors, the relative FRET efficiencies can be used to reflect the relative energy flow using the CoDA algorithm. The compositional value transformation was conducted using data closure (Eq.20), taking the individual real space value and divided by the sum of all involved compositions.

Equation 20 Data closure for CoDA

$$C(n_1, \dots, n_j) = \{x_i, \ \dots, x_j\} \ where \ x_j = \frac{n_j}{\sum_{j'} n_{j'}}$$

The compositional ratio was further used by dividing each transformed value by the same reference value (Eq.21) to ensure sub-compositional coherence and eliminate the impact of the undiscovered hidden composition or other compositions not taken into account during data closure [130].

Equation 21 Ratio using the reference data to achieve sub-compositional coherence

$$R(x_i, x_{reference}) = \left(\frac{x_i}{x_{reference}}\right), \ i = 1, \dots, n$$

Logarithmic transformation was further conducted on the ratio values since the logarithmic ratio (LR) will bring the multiplicative ratio values into the additive vector scale, allowing the LR values to be directly used for downstream statistical analysis (Eq.22). These approaches were also applied to the detection of multiple cellular signalling pathway reporters. When the cells were spending energy to activate multiple signalling, the activation levels detected as the reporter fluorescent intensities were also compositional data; therefore, they needed to be interpreted using their LRs after data closure.

Equation 22 Logarithmic ratios for CoDA

$$LR(x_i, x_{reference}) = log\left(\frac{x_i}{x_{reference}}\right), \qquad i = 1, \dots, n$$

In this study, the geometric mean of the data was used as the reference value, and by dividing the logarithmic transformation on geometric mean balanced ratios, the data was centred with zero means. The final transformed values are called the central logarithmic ratios (CLR) (Eq.23). Having the zero mean will allow the direct application of PCA, which requires mean zero-centred input values, for data dimensionality reduction and data visualisation in the 2D scatterplot.

Equation 23 Central logarithmic ratios for CoDA

$$\boldsymbol{CLR(j) = log\left(\frac{x_j}{\left(\prod_j x_j\right)^{\frac{1}{J=3}}}\right) = log(x_j) - \frac{1}{J=3}\sum_j log(x_j)},$$

$$j = [I_{eCFP}\ to\ (eCFP \rightarrow eYFP),\ I_{eCFP}\ to\ (eCFP \rightarrow mRFP)]\ \&\ [I_{eYFP}\ to\ (eYFP \rightarrow mRFP)]$$

Multivariate linear models were constructed to explain how FRET efficiencies ($E'_{eCFP \rightarrow eYFP}$ and $E'_{eCFP \rightarrow mRFP}$) were impacted by the competing dynamic energy transfers from eCFP to eYFP and mRFP acceptors (Eq.24). In addition, the model also included the impact of the energy transfer from the eYFP donor to the mRFP acceptor. The independent variables are the CLR of the sensitised FRET emission values after CoDA transformation, and the dependent variables are the logarithmically transformed FRET efficiencies of eCFP→eYFP or eCFP→mRFP, respectively. An error term is also added to capture the residue errors. The parameters θ_1, θ_2 and θ_3 represent the magnitudes of the impacts.

Equation 24 Multivariate linear models for dynamic energy transfers during FRET

$$log(E'_{eCFP\rightarrow eYFP}) = \theta_1 \, log\left(\frac{I_{eCFP} \; to \; (eCFP \rightarrow eYFP)}{gMean(eCFP_{components})}\right) +$$

$$\theta_2 \, log\left(\frac{I_{eCFP} \; to \; (eCFP \rightarrow mRFP)}{gMean(eCFP_{components})}\right) + \theta_3 \, log\left(\frac{I_{eYFP} \; to \; (eYFP \rightarrow mRFP)}{gMean(eYFP_{components})}\right) + e$$

$$log(E'_{eCFP\rightarrow mRFP}) = \theta_1 \, log\left(\frac{I_{eCFP} \; to \; (eCFP \rightarrow eYFP)}{gMean(eCFP_{components})}\right) +$$

$$\theta_2 \, log\left(\frac{I_{eCFP} \; to \; (eCFP \rightarrow mRFP)}{gMean(eCFP_{components})}\right) + \theta_3 \, log\left(\frac{I_{eYFP} \; to \; (eYFP \rightarrow mRFP)}{gMean(eYFP_{components})}\right) + e$$

2.8 Statistical analysis

2.8.1 Hypothesis tests

Permutation t-test for the means of two independent samples that does not assume equal variance was used for the hypothesis test (Python "stats.ttest_ind" function from SciPy package version 1.9.3) with the number of random permutations set to 100. The null hypothesis of the permutation t-test is that two independent samples have identical averages, and the null hypothesis rejection significance levels were p < 0.05 *, p < 0.01 **, and p < 0.001 ***.

The two-sided independent t-test (Welch's t-test) was also used for the hypothesis test between two groups of samples with the calculated means ($\bar{X}$), variances (s), and sample size (N_i).

Equation 25 Welch's t-test

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{s^2_{\overline{X_1}} - s^2_{\overline{X_2}}}}, \quad s^2_{\overline{X_i}} = \frac{s_i}{\sqrt{N_i}}$$

The Kruskal-Wallis H-test was used for the median of more than two independent samples. The H-test is a non-parametric version of the analysis of variance (ANOVA), which does not assume equal data size or variance (Eq.26). The null hypothesis of the Kruskal-Wallis H-test is that the population median of all of the groups is equal. The null hypothesis rejection significance levels were $p < 0.05$ *, $p < 0.01$ **, and $p < 0.001$ ***.

Equation 26 Kruskal–Wallis one-way analysis of variance

$$H = (N-1)\frac{\sum_{i=1}^{g} n_i(\bar{r}_{i\cdot} - \bar{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(r_{ij} - \bar{r})^2}$$

For notation details, see Kruskal and Wallis [131].

2.8.2 Singlet identification accuracy

The $F1$ score was used as the harmonic mean of the precision and recall to determine the accuracy of singlet identification algorithms (Eq.27). The $F1$ score has the most excellent value of 1 and the worst value of 0. Precision and recall both contribute equally to the $F1$ score. The manual gated samples were used as the ground truth values. The $F1$ score was calculated as follows, where TP is the true-positive count, FP is the false-positive count, and the FN is the false-negative count:

Equation 27 The *F1* score for singlet identification accuracy evaluation

$$precision = \frac{TP}{(TP + FP)}$$

$$recall = \frac{TP}{(TP + FN)}$$

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

2.8.3 Linear model statistics

The robust linear regression was used to determine the spillover coefficients for spectral unmixing and the small molecule concentration detection linearity. The random sample consensus (RANSAC) algorithm was used as the iterative robust linear regression approach for accurate estimation of the parameters using random inliers from the datasets because it is more robust towards datasets with potential outliers. The Python function "linear_model.RANSACRegressor" was used from the scikit-learn package (version 1.2.0), and the R^2 scores and mean absolute errors were calculated as the spillover linearity scores and the spillover coefficients fitting errors.

The Pearson correlation coefficient was used to measure the linear relationship between two fluorescent intensity values from two different detection channels using the Python SciPy (version 1.9.3) package's "stats.pearsonr" function.

For fitting the multivariate linear regression model for the three-protein FRET energy dynamic analysis, the ordinary least squares linear regression was used with Python scikit-learn (version 1.2.0) package's "linear_model.LinearRegression" function. The R^2 score was used to evaluate the amount of data explained by the linear model.

2.8.4 Spectral unmixing spread error

The spectral unmixing spread error was previously formulated by Nguyen et al. as the compensation spread error [87]. The spillover spreading for a given detector/spillover channel pair can be estimated by taking the square root of the difference in squared robust standard deviations of recorded fluorescence intensity from unstained negative and single-colour control populations (Eq.28):

Equation 28 The spread error

$$^{R}\delta_{FL1} = {}^{R}C^{*}_{FL1}(95) - {}^{R}C^{*}_{FL1}(50)$$

$$^{S}\delta_{FL1} = {}^{S}C^{*}_{FL1}(95) - {}^{S}C^{*}_{FL1}(50)$$

$$\Delta C^{*}_{FL2} = {}^{S}C^{*}_{FL2}(50) - {}^{R}C^{*}_{FL2}(50)$$

$$\Delta\delta_{FL1} = \sqrt{{}^{S}\delta^{2}_{FL1} - {}^{R}\delta^{2}_{FL1}} \qquad SE^{FL2}_{FL1} = \frac{\Delta\delta_{FL1}}{\sqrt{\Delta F_{FL2}}}$$

In the equation above, the optimal and secondary detection channels are notated as $FL1$ and $FL2$, and the single-colour and unstained reference controls are S and R. For detected emissions, the parameters (C^{*}) are the fluorescence intensity values of the spectrally unmixed single-colour control samples, and the δ are the standard deviations adjusted to use the 50th and 95th percentiles of the fluorescent intensity values. The calculated spread errors were the final SE values. Of note, the SE was originally calculated using the 50th and 84th percentiles [87] but was set to use the 50th and 95th percentiles to increase robustness when evaluating the FRET algorithm performance.

2.9 Data visualisaton

2.9.1 PCA for dimensionality reduction

To visualise the CoDA transformed data with the five-dimensional quenched eCFP, quenched eYFP, sensitised eCFP → eYFP FRET, sensitised eCFP → mRFP FRET and sensitised eYFP → mRFP FRET data, PCA was used with the CLR transformed five-dimensional data to reduce them into a two-dimensional scatter plot for easy data visualisation. Again, PCAs were performed using the “decomposition.PCA” function of the scikit-learn package (version 1.2.0) in Python.

2.9.2 Data plots and graphs

All histograms, scatter plots, violin plots, bar graphs, radar plots, spectrum viewer graphs and flow cytometer configuration graphs were generated using the python matplotlib (version 3.6.2) and seaborn (version 0.12.1) packages.

2.10 Data and code availability

All data and codes used in this study are available on the GitHub host server (https://github.com/Edward-Z-Ni/CUBE) and the University of Technology Sydney eResearch GitLab server (https://code.research.uts.edu.au/12831196/cube). Access can be granted upon request. The GitHub and GitLab repositories are called “CUBE”, which stands for the “Cytometry Utilities Box Expansion”.

CHAPTER THREE

BIOLOGICAL SAMPLES PREPARATION FOR FRET AND CELL SIGNALLING

3.1 Introduction

This thesis presents the development of an *in-situ* flow cytometry-based PPI detection platform at the true single-cell level. The platform provides unsupervised ML-powered FRET algorithms to analyse the simultaneous detection of multiprotein FRET and numerous intracellular signalling pathway activities. For the accurate quantification of unlinked and naturally free interacting PPIs involving two and three protein molecules, the algorithms achieve error-free baseline corrected, autofluorescence removed, and spectrally unmixed FRET readouts. Human receptor proteins from the tumour necrosis factor receptor superfamily (TNFRSF) and *poxviridae*-encoded human TNFR homologous – viral TNFRs (vTNFRs) biological samples were prepared in this study to evaluate the capability of applying this flow cytometry-based PPI detection platform in conducting biologically meaningful and clinically significant investigations. The human TNFRs (huTNFRs) and vTNFRs were constructed as the fluorescent protein (FP) fused proteins to permit the mammalian cell expression and fluorescence detection by this flow cytometry-based FRET platform.

The human TNFRSF consists of cytokine receptors which are identified by their highly conserved, cysteine-rich domains (CRDs) in the extracellular region [132]. TNFRSF proteins are categorised by their ability to bind cytokine ligands from the tumour necrosis factor superfamily (TNFSF) via those conserved CRDs (Figure 12). TNFR1 receptors contain four extracellular CRDs (CRD1, CDR2, CRD3, and CRD4) [133]. Previous studies have suggested that the TNF ligand binds to the CRD2 and CRD3 of the TNFR1 receptors, and the ligand-induced TNFR receptor trimerization causes the cell surface TNFR1 aggregation and intracellular signalling activations [134].

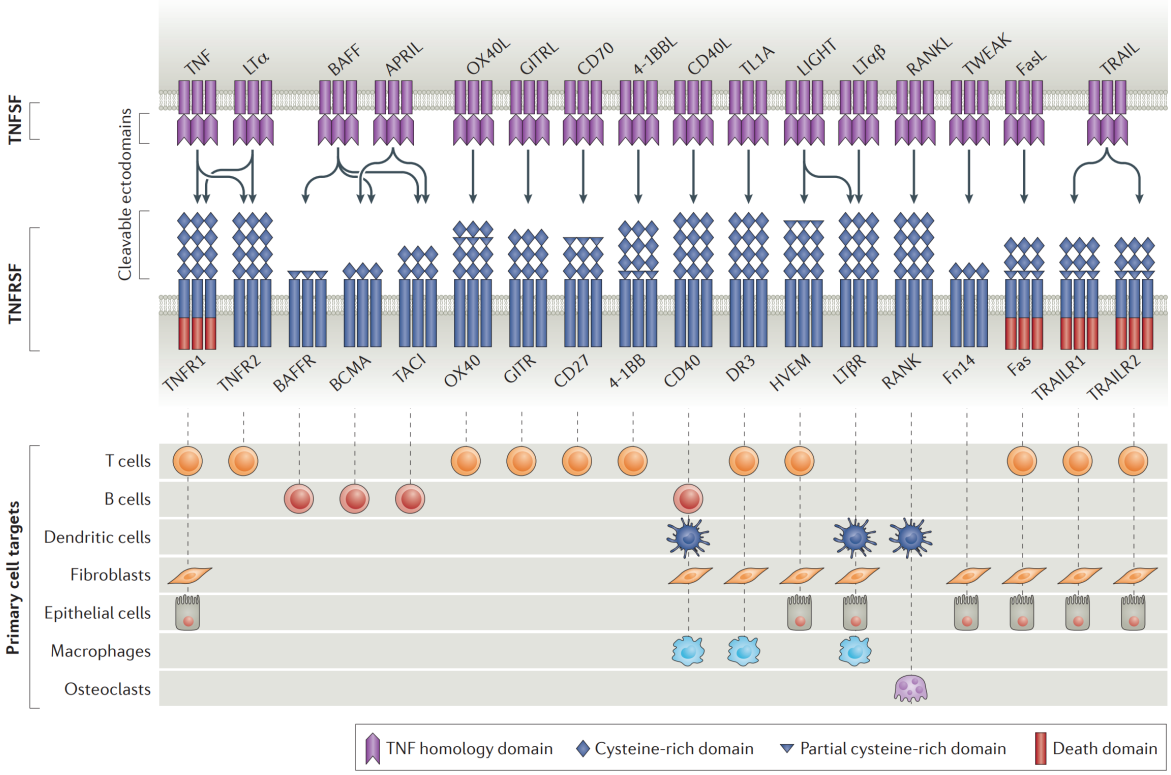


Figure 12 Select members of the TNF and TNFR superfamily
TNF superfamily ligands (TNFSP, purple; top) are trimeric proteins that bind to TNF superfamily receptors (TNFRSF, blue; bottom) and contain variable copies of cysteine-rich domains (CRDs). It also displays the primary cell types that respond to TNF-TNFR signalling. Receptors containing death domains (DD, red) can promote apoptotic cell death (TNFR1, Fas, TRAIL1 and TRAIL2). *(Reprint license number: 5461390797269, license data: Jan 03, 2023)* [132]

More recent studies showed that there is also a “pre-ligand assembly domain” (PLAD) located within the N-terminus of CRD1 on TNFR1 [71, 135]. Thus, it has been suggested that TNFRs exist as pre-assembled dimeric complexes prior to TNF binding and can further form, a more complex trimeric conformation as trimers of dimers upon TNF binding [136]. Therefore, TNFR1 was utilised for this study due to its inherent capacity to form transmembrane dimers and symmetrical ligand-bound trimers [120, 137-139] that allow the validation of both the 2-protein 3-colour FRET and the 3-protein 6-colour FRET algorithms (see section 5.2.2 and section 6.2.1). For FRET-negative controls, CD27

and TNFR2 were used as non-TNFR1-interacting structurally similar proteins from the same protein superfamily [134].

Fluorescent proteins served as versatile probes for all biological protein samples prepared in this study. Green fluorescent protein (GFP) is a naturally arising fluorescent protein (FP), first described in the jellyfish *Aequorea victoria* [22] but also found in many other organisms [64, 140, 141], and red fluorescent protein (DsRed) was characterised from the *Discosoma* coral. The molecular characterisation of GFP has led to the careful strategic mutation of specific amino acids that alters GFP's absorption, excitation and emission spectra [14, 50]. Cyan FP (CFP) and yellow FP (YFP), or their enhanced versions eCFP and eYFP, are molecular derivatives of GFP [142], whereas monomeric RFP (mRFP) is derived from DsRed [51, 53]. The detection of naturally occurring fluorescence in these organisms can be utilised to monitor their existence (i.e. relative abundance) and, thereby, the health of ecosystems. Moreover, these FPs are now typical reagents used for molecular biological investigations – both as recombinant fluorescent fusion proteins and as FRET reporters [32]. Furthermore, their biophysical properties have been well-characterised, including photon absorption efficiencies, quantum yields, fluorescence excitation and emission coefficients [51, 62, 63].

All the above mentioned FPs have been previously reported as FRET reporters for TNFR1 dimeric PPI investigations. In this study, human TNFR1, TNFR2 and CD27 were constructed as the -eCFP, -eYFP and -mRFP fusion proteins for the detection of TNFR1 dimeric PPI using three FRET pairs, i.e., eCFP→eYFP, eYFP→mRFP, and eCFP→mRFP. These FRETs are possible because the eCFP and eYFP FRET donors' emission spectrums are largely overlapped with the excitation spectrums of their respective eYFP and mRFP

FRET acceptors. However, given the role of TNFR1 as a trimer, there is an urgent need to develop the FRET assay with trichrome FRET capabilities. Since the eCFP donor emission spectrum overlaps with both the eYFP and mRFP excitation spectrums, the three-way FRETs (concurrent eCFP→eYFP, eYFP→mRFP and eCFP→mRFP) and a two-step FRET (eCFP→eYFP→mRFP) can all be detected for the trimeric TNFR1 examination (see section 5.2.2 and section 6.2.1).

The cell surface TNFR1 interactions permit the regulation of various human intracellular signalling pathways. In particular, the cell death pathway and the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) activated proinflammatory pathway are critical for human anti-tumour and antiviral immune responses [134, 143, 144]. Many pharmaceutical therapeutics screens and clinical diagnostics target the TNFR1-regulated cell death and proinflammatory signalling pathways as a tool to search for anti-inflammatory, anti-viral, and anti-tumour drugs and detecting autoinflammation diseases, respectively [145-149]. Therefore, it is crucial to deliver the simultaneous detection of the TNFR1 PPI FRET reporters and the cellular signalling reporters to understand how TNFR1 receptor interactions link to the intracellular signalling pathway activities. This study uses the UV-violet fixable live/dead dye (Thermo Scientific) for cell death detection. Since there is no commercial reporter for detecting the NF-κB activated proinflammatory signalling pathway that can be easily adapted to the FRET system, I generated the pGL4.23.NF-κB-[miniP]-miRFP703 reporters for the detection of NF-κB activation. The proinflammatory reporter uses the pGL4.23 minimal promoting expression vector cloned with the NF-κB response elements and the monomeric infrared fluorescent protein (miRFP703) for the detection of fluorescence signals.

In addition to the normal TNFR1 receptor biology, the germline mutation of TNFR1 can cause a rare autoinflammatory disease termed TNFR-associated periodic syndrome (TRAPS). TRAPS is a condition characterised by recurrent episodes of fevers that typically last three weeks to a few months, with frequencies varying between 6 weeks to a few years [150-152]. The mechanism of TRAPS pathology has not been fully understood, and how TRAPS-related mutations in TNFR1 protein alter the TNFR1 PPIs has not been thoroughly investigated. Therefore, to demonstrate the possibility of utilising the FRET platform developed in this study to detect the aberrant PPIs caused by germline mutations, TNFR1 TRAPS mutants have also been generated using site-direct mutagenesis. All TRAPS mutations located in the TNFR1 CRD1 and CRD4 (PPI critical domains) ORFs were used to generate TRAPS TNFR1-eYFP fusion constructs and compared with the wildtype (WT) TNFR1-eYFP in the abovementioned three-way FRET system.

The virulence of a virus infection is strongly influenced by the capacity of the virus to subvert the hosts' anti-viral defences. Therefore, it is unsurprising that many viruses have adopted multiple strategies to counteract the TNFR1 biology [153-157]. For poxviruses (*poxviridae*), a family of large dsDNA enveloped viruses, their large genome size (180-230 kb dsDNA45) confers the capacity to include multiple ORFs that encode vTNFRs whose function appears to subvert the hosts' TNFR1-mediated antiviral immune responses. These vTNFRs can be detected as soluble glycoproteins in the supernatant of poxvirus-infected cells [158], which bind and potently neutralise TNFR1-mediated cell death [159]. The functional importance of vTNFR is exemplified by the fact that vTNFR ORFs are found within the genomes of virtually all poxviruses [160, 161], and vTNFRs

sometimes are known by the more general term “cytokine response modifiers” (CRM) [162]. Previous studies primarily focused on the TNF cytokine “decoy” mechanisms of the vTNFR, and few studies were concentrated on the interactions between vTNFR and huTNFR1 receptor proteins. In this study, to demonstrate the power of the FRET platform in detecting the PPIs between vTNFR and huTNFR1, the vTNFR cDNAs from selected poxviruses (variola virus, myxoma virus, and monkeypox virus) were used to generate the -mRFP FRET component and replace the huTNFR1-mRFP in the FRET assay. By combining with the huTNFR1-eCFP and TRAPS-eYFP, it further permits the detection of protein interactions among human WT TNFR1, human TRAPS TNFR1, and vTNFR. In this study, I hypothesise that TRAPS mutations of TNFR1 are potential evolutionary trade-offs, and TRAPS TNFR1 mutants can disarm the poxvirus’s immune evasion mechanism, which targets the human WT TNFR1 proteins. This further demonstrates the power of the FRET platform developed in this study for more complex viral-host interaction and evolutionary competition studies. In summary, this chapter reports the generation of all the biological components (all summarised in Figure 13) that were required to establish the FRET assays and to demonstrate the capabilities and robustness of the ML-powered flow cytometry-based FRET platform.

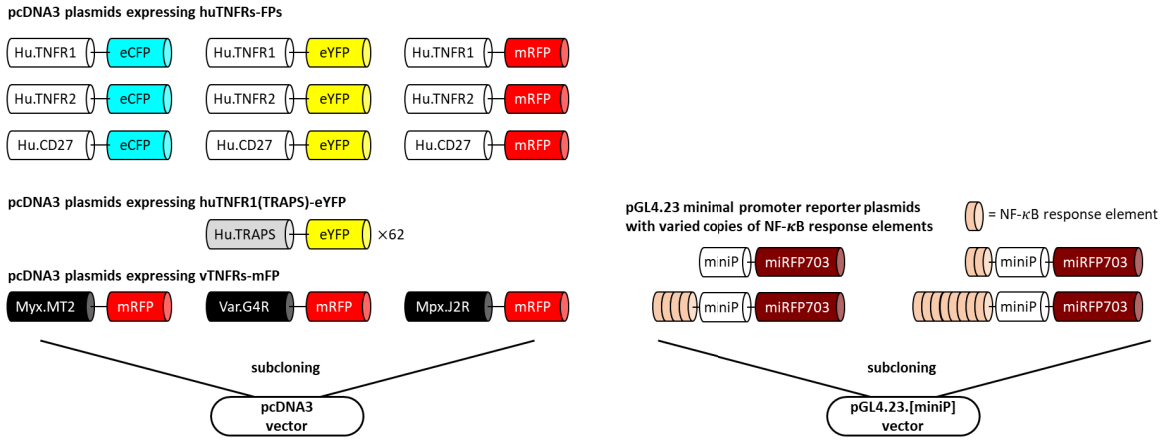


Figure 13 Biological samples prepared in this study
Human TNFR1, TNFR2, and CD27 were generated as the -eCFP, -eYFP, and -mRFP pcDNA3 fusion constructs. A total of 62 human TNFR1 TRAPS mutants were also generated as the -eYFP pcDNA3 fusion constructs. In addition, the myxoma virus MT2, variola virus G4R and monkeypox J2R viral TNFRs were generated as the -mRFP pcDNA3 fusion constructs. For proinflammatory signalling detection, the pGL4.23 minimal promoter plasmid was used to generate miRFP fluorescence reporters with 0, 2, 4, and 8 copies of NF-κB response elements.

3.2 Results:

3.2.1 Generation of TNFR1, TNFR2, CD27 as -eCFP, -eYFP, -mRFP fusion constructs

A total of nine TNFRs-FP fusion plasmid vectors, pcDNA3.TNFR1-eCFP, pcDNA3.TNFR1-eYFP, pcDNA3.TNFR1-mRFP, pcDNA3.TNFR2-eCFP, pcDNA3.TNFR2-eYFP, pcDNA3.TNFR2-mRFP, pcDNA3.CD27-eCFP, pcDNA3.CD27-eYFP and pcDNA3.CD27-mRFP were sequenced to confirm no mutation or miss-paired DNAs from the subcloning process. The sequences were compared to the NCBI nucleotide database using blast search, in which the TNFR1, TNFR2, and CD27 sequences were aligned with the *Homo sapiens* references, NCBI accession: AH003016.2, NCBI accession: AH006638.2, and NCBI accession: BC012160.1 respectively. The eCFP, eYFP, and mRFP sequences were aligned with the cloning vector and synthetic construct references, NCBI accession: OL452018.1,

NCBI accession: LT727258.1, and NCBI accession: AF506027.1 respectively. In detail, for human TNFRs' sequencing results, the TNFR1 sequence showed 99.82% identity to the reference data with a different nucleotide at the 36th position (CCG → CCA, proline→proline, see Figure 14 and Figure 15). The TNFR2 and CD27 sequences both showed 100% identity to the reference data (Appendix Figure 3 to Appendix Figure 8). For FP sequencing results, the eCFP and eYFP both showed 100% identity to the online reference sequence (see Figure 14, Figure 15, Appendix Figure 3, Appendix Figure 4, Appendix Figure 6 and Appendix Figure 7), while the mRFP showed 99.85% identity to the online reference sequence with a varied nucleotide at the 525th position (GCC→GCT, alanine → alanine, see Figure 16, Appendix Figure 5 and Appendix Figure 8). The sequencing variations can result from either the PCR error or single nucleotide polymorphisms in the original sequence template. In both cases, the variations do not impact the translated protein amino acid sequences or the protein structures. Thus, all pcDNA3.TNFR1-eCFP, pcDNA3.TNFR1-eYFP, pcDNA3.TNFR1-mRFP, pcDNA3.TNFR2-eCFP, pcDNA3.TNFR2-eYFP, pcDNA3.TNFR2-mRFP, pcDNA3.CD27-eCFP, pcDNA3-CD27.eYFP, and pcDNA3.CD27-mRFP have been successfully constructed and verified by DNA sequencing and can be utilised in FRET assays.

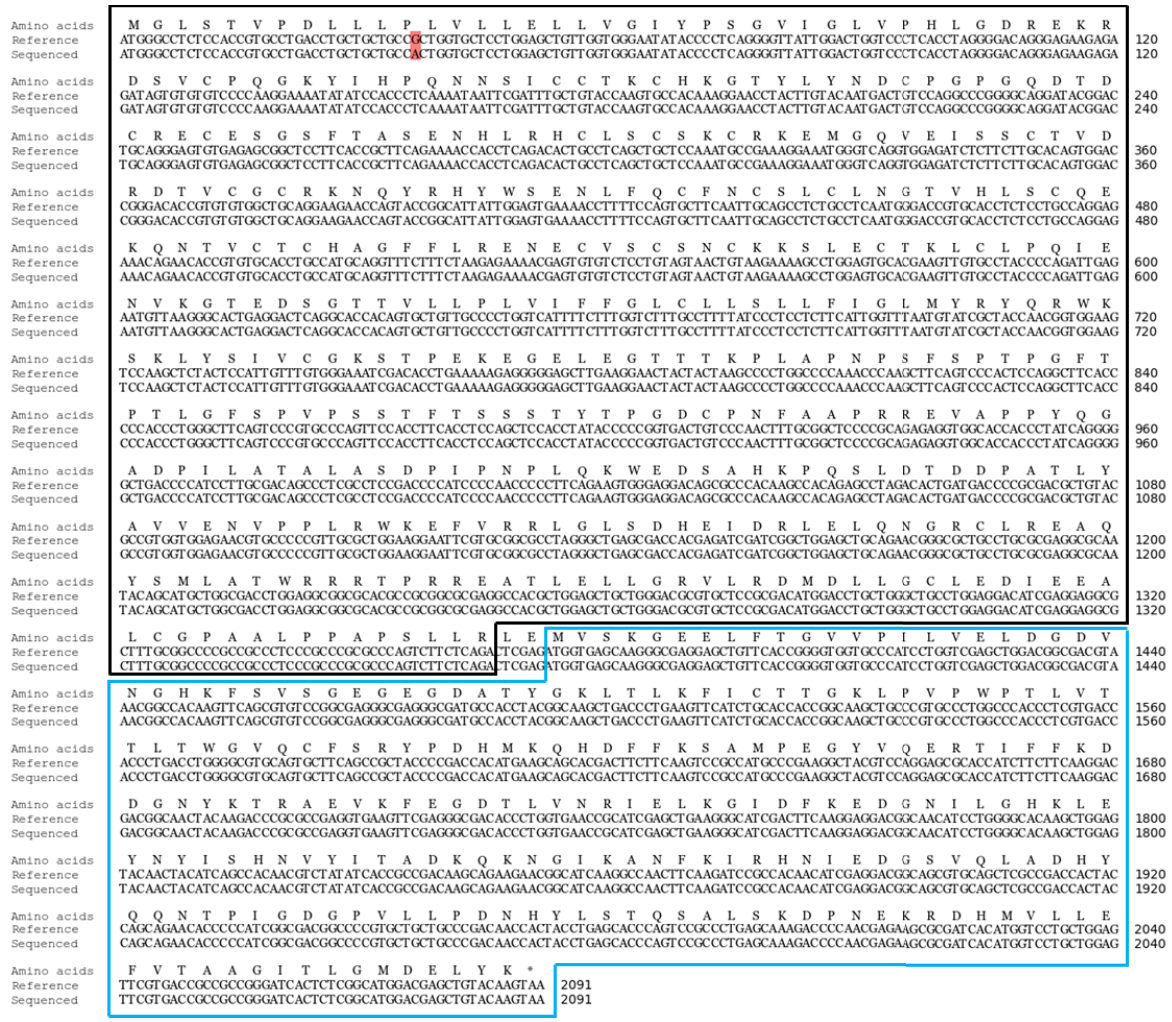


Figure 14 TNFR1-eCFP sequencing result

Human TNFR1 (black box) subcloned with the eCFP (cyan box) to form a fusion fluorescence receptor TNFR1-eCFP. The sequencing result of human TNFR1 shows a silent mutation: CCG→CCA, proline→proline (red) at the 36th nucleotide. No mutation in the eCFP sequence.

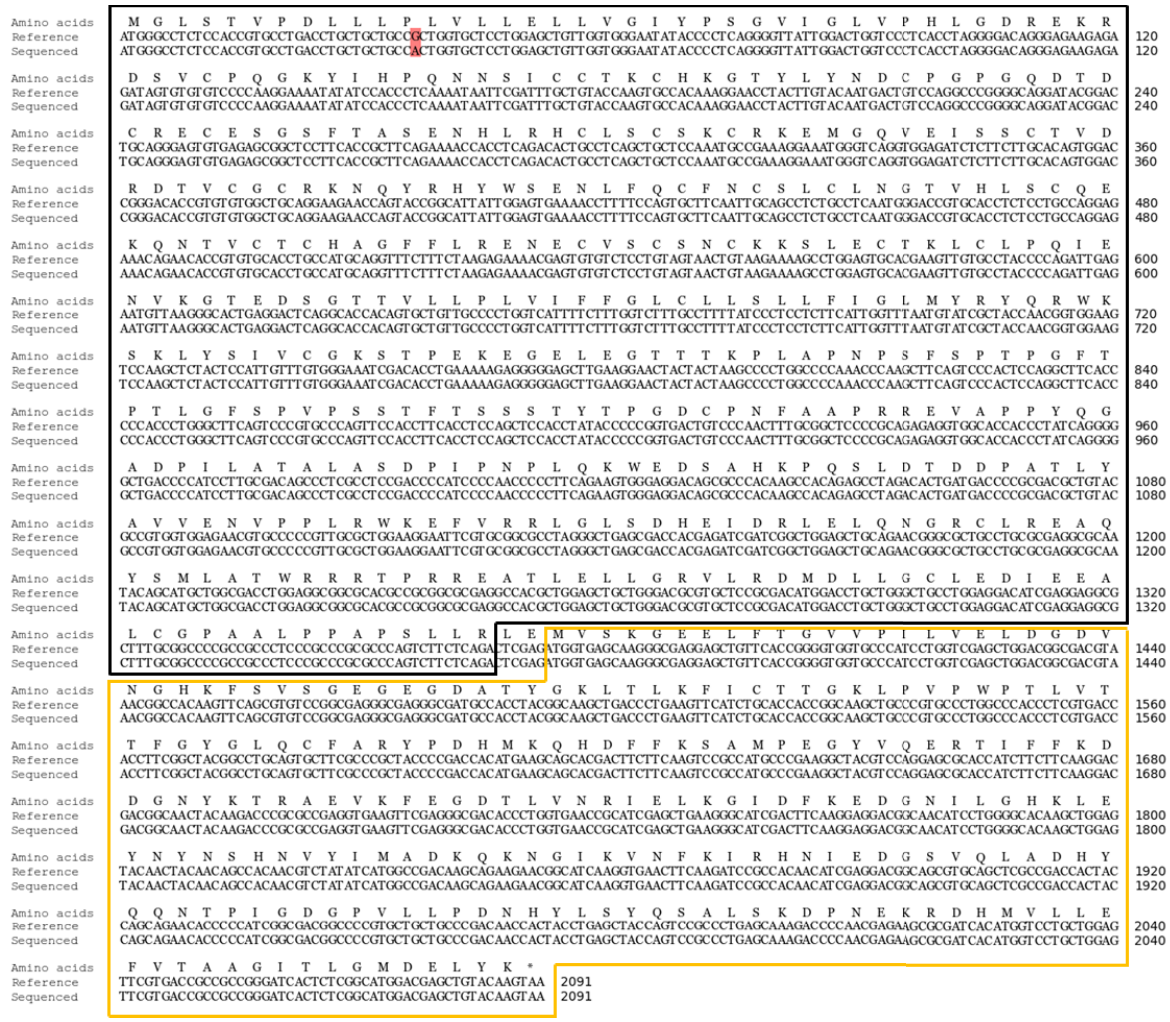


Figure 15 TNFR1-eYFP sequencing result

Human TNFR1 (black box) subcloned with the eYFP (yellow box) to form a fusion fluorescence receptor TNFR1-eYFP. The sequencing result of human TNFR1 shows a silent mutation: CCG→CCA, proline→proline (red) at the 36th nucleotide. No mutation in the eYFP sequence.

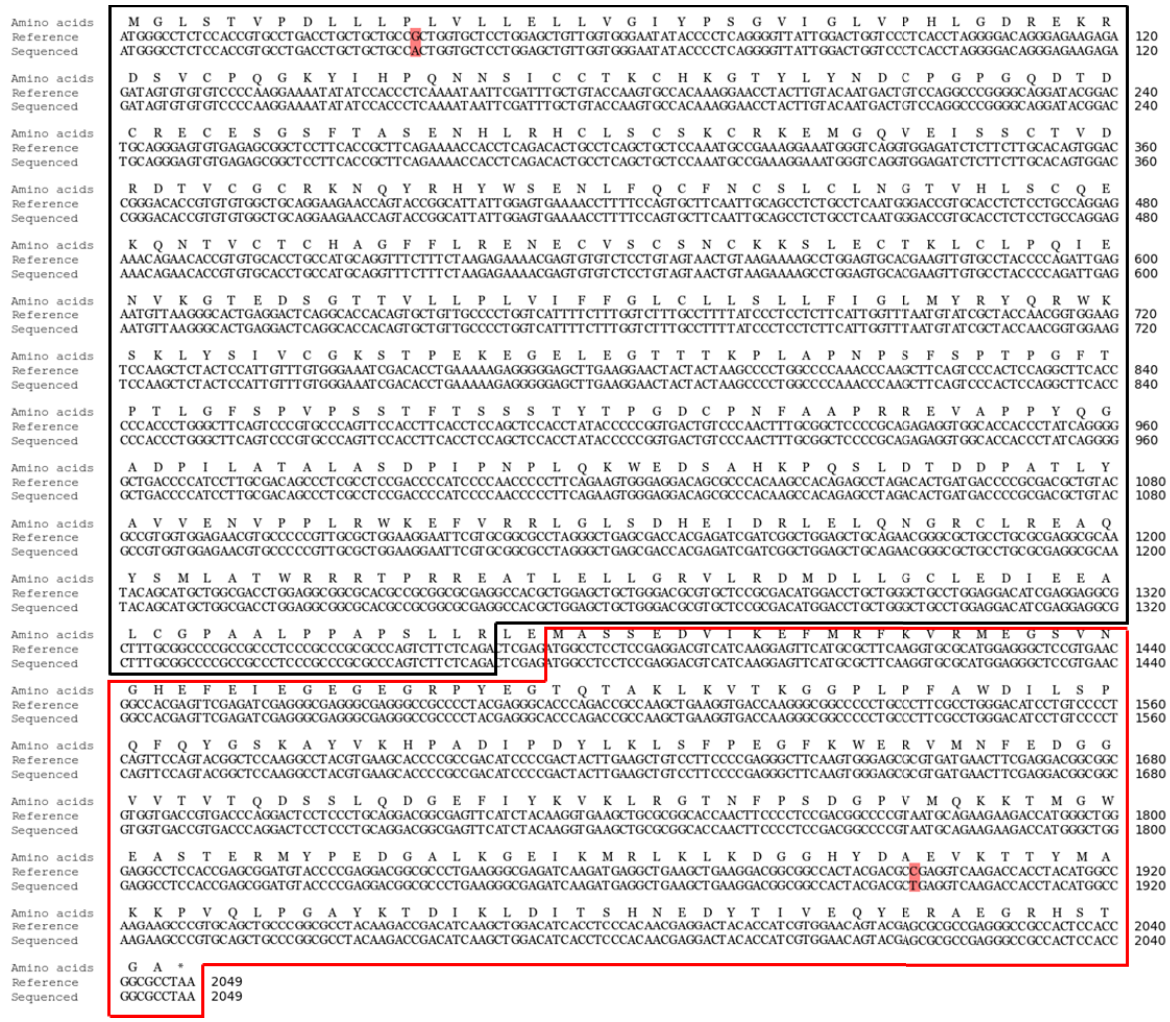


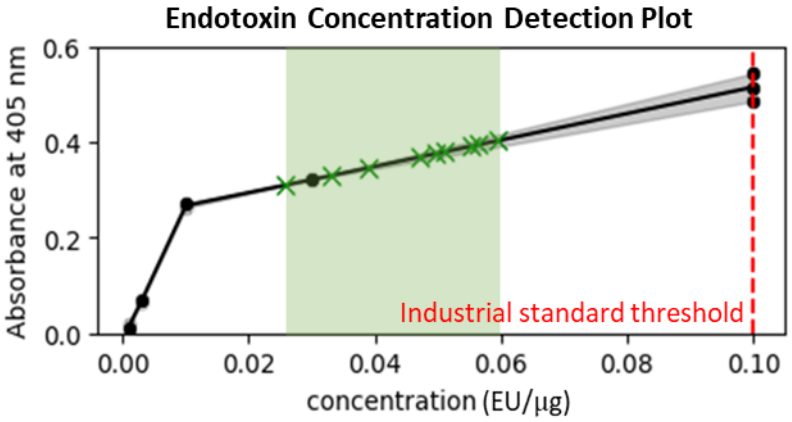
Figure 16 TNFR1-mRFP sequencing result

Human TNFR1 (black box) subcloned with the mRFP (red box) to form a fusion fluorescence receptor TNFR1-mRFP. Two silent mutations were identified, one in the TNFR1 ORF: CCG→CCA, proline→proline (red) at the 36th nucleotide, and the other one in the mRFP ORF: GCC→GCT, alanine→alanine (red) at the 525th nucleotide.

3.2.2 Establish the optimal sample harvesting period using live-cell imaging

Human cells recycle excess or unusable proteins into amino acids, which balances the need for other protein syntheses. The energy preserved during protein recycling also provides the capacity to rapidly mobilise proteins to their sites of action when the proper signal is received [163]. Transiently transfected HEK-293T cells express the fusion proteins from plasmid vectors without host genome integration. The plasmid DNAs will not be replicated during cell replication, and the expressed fusion protein level will decrease after maximal expression due to the transient transfection and protein recycling process. It is necessary to determine the optimal cell culture times post transfection to ensure the HEK-293T cells express each of the TNFR1, TNFR2, CD27-eCFP, -eYFP and -mRFP fusion constructs at high levels. The culturing time cannot be so long that the expressed fusion proteins or the transiently transfected plasmid DNAs undergo degradation and recycling.

Another factor that can also impact the plasmid transfection and expression efficiency is the potential endotoxin contamination from the *E.coli* bacterial host during the plasmid DNA maxiprep extraction procedure. Therefore, endotoxin levels of the maxiprep plasmid DNAs were examined using the chromogenic LAL endotoxin assay to rule out any influence on the final fusion protein level from endotoxin contamination. The concentrations of endotoxins from the nine TNFRSF receptor-FPs samples were examined with values below the contamination threshold based on the industrial standard (Figure 17). This result is often described as “endotoxin-free” in commercial-level endotoxin-free plasmid extraction kits.



concentration	Linear regression model	R^2	p-value	S.E.
0.001-0.01 EU/μg	$A = 26.583 \times C - 0.011$	0.995	3.075e-05	1.270
0.01-0.1 EU/μg	$A = 2.756 \times C + 0.239$	0.991	1.857e-07	0.137

Figure 17 Endotoxin detection of plasmids
The concentrations of residual endotoxin were examined within two linear concentration ranges (black dots and lines). All samples have low endotoxin levels (green cross) that meet the industrial standard (red dashed line).

The expression dynamics of each TNFRs-FP fusion construct were examined utilising the Nikon A1R confocal live-cell time course imaging and the HEK-293T cells transfected with the endotoxin-free plasmids. After the initial 6 hours of incubation, the expression levels were recorded and calculated as the mean fluorescent intensity of each sample for a 48-hour (6 hours to 54 hours) tissue culture period post-transfection. The fluorescent intensity values have been normalised to the same scale. All TNFRs-eCFP expressions reached the maximum between 42 to 48 hours post-transfection and started to decline after 48 hours, with the TNFR1-eCFP starting to decrease slightly earlier after 47 hours (Figure 18 cyan plots). All TNFRs receptor-eYFP expressions reached the maximum between 38-46 hours post-transfection, except for TNFR1-eYFP reached the maximum

at 29 hours and started to decline afterwards (Figure 18 yellow plots). All TNFRs-mRFP expressions are similar to the -eCFP expressions, but no expression decrease was observed within the 48 hours recording period (Figure 18 red plots). The varied expression dynamics are potentially caused by the codon bias of the different receptor and fluorescent protein sequence combinations which depend on the HEK-293T host cell tRNA pool that is used for making those fusion proteins. The early declines of the TNFR1 constructs are also the results of the cell death mediated by the TNFR1 signalling. Therefore, the optimal harvesting time between 30-36 hours post-transfection has been chosen to ensure the maximal fluorescence detection levels for all fluorescence-fused receptors. In addition, TNFR1-expressed cells were observed with shrunk cell sizes compared to TNFR2 and CD27-expressed cells (Figure 18B), which is an early indicator for cell apoptosis.

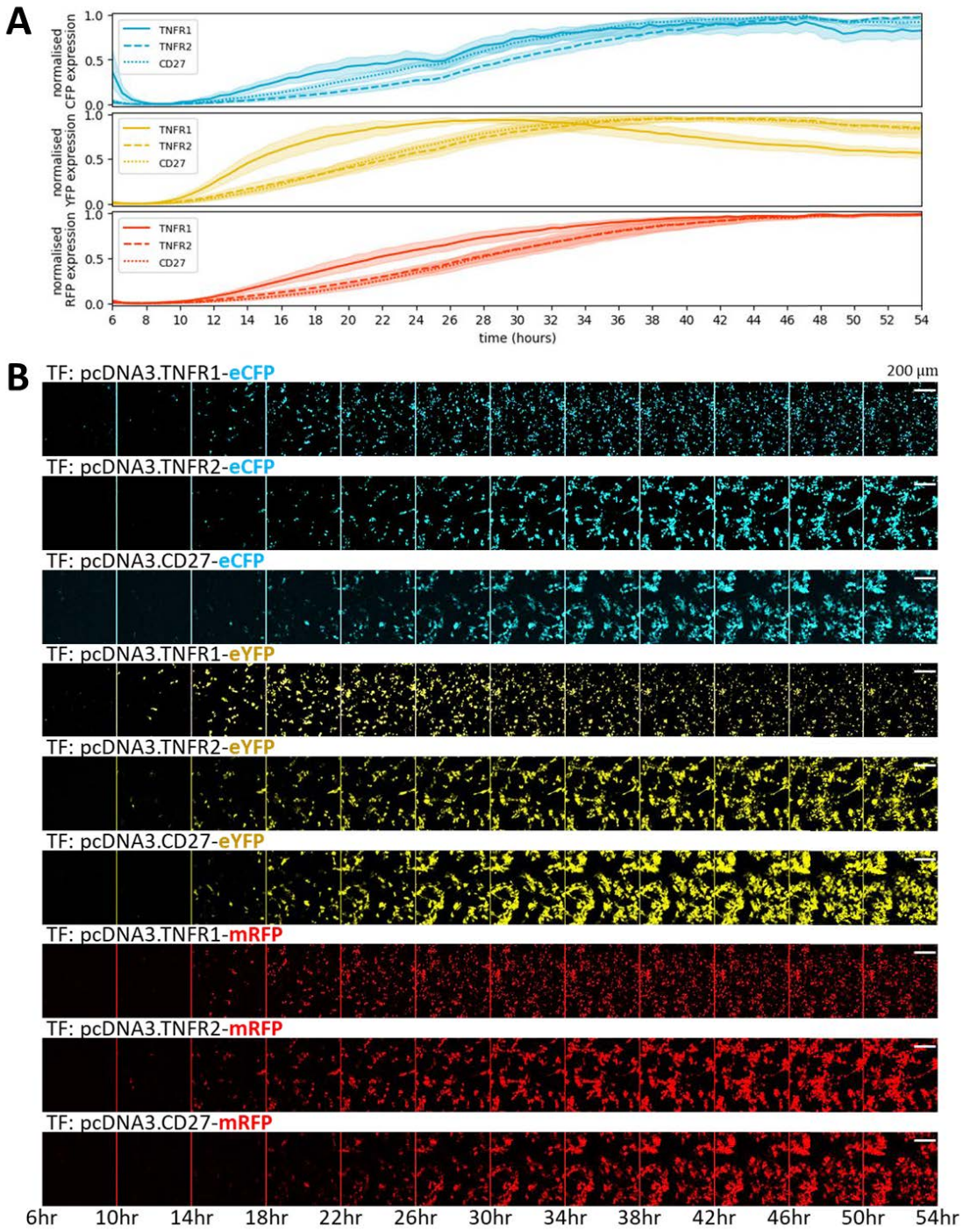


Figure 18 Live-cell imaging of TNFRs-eCFP/eYFP/mRFP expressions

The dynamic expression profiles of TNFR1 (solid lines), TNFR1 (dashed lines), and CD27 (dotted lines) as the -FP fusion proteins (cyan, yellow, and red) (A). Microscopy images were captured for TNFRs-eCFP/eYFP/mRFP 48 hours post-transfection (B).

3.2.3 Generation of TRAPS mutation-related TNFR1 as -eYFP fusion constructs

A total of 62 TRAPS TNFR1-eYFP constructs were generated using site-direct mutagenesis with the primers designed as listed in the method section (see section 2.1.2). TRAPS mutation constructs were confirmed by plasmid DNA sequencing and further compared with the WT TNFR1 sequence, and the results indicated that all 62 TRAPS TNFR1-eYFP have been successfully constructed with the designated mutations (Figure 19). Moreover, the expression of each TRAPS-eYFP fusion protein has been examined in HEK-293T cells with the EVOS microscope 36 hours after transient transfections. All generated TRAPS TNFR1-eYFP fusion proteins were successfully expressed and detected using the YFP (500/24 nm excitation, 542/27 nm emission) EVOS light cube (Figure 20).

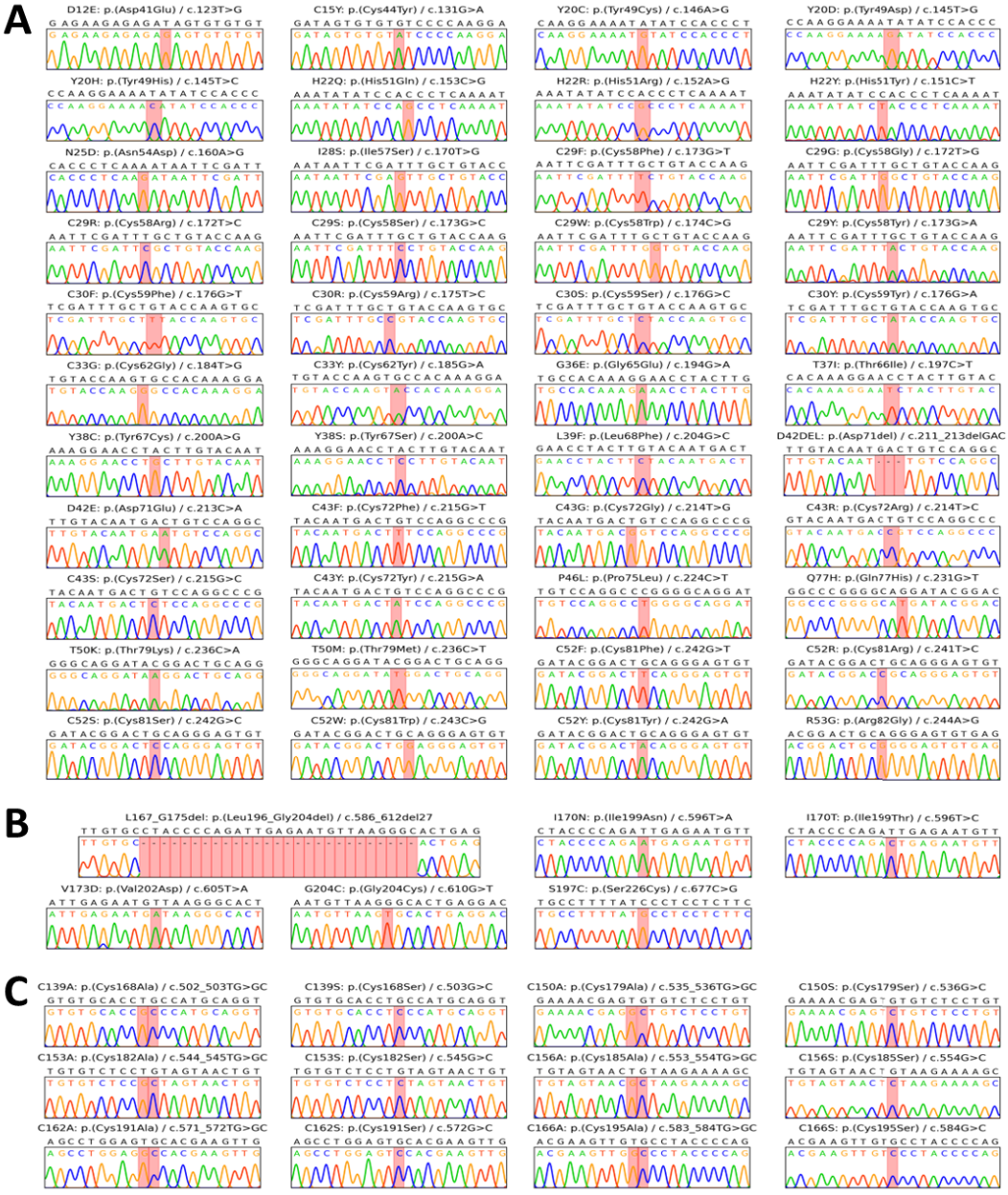


Figure 19 TRAPS TNFR1 sequencing results

A total of 44 naturally-existing TRAPS mutations in the CRD1/PLAD of TNFR1 (A), six naturally-existing TRAPS mutations around the CRD4 region (B), and 12 artificially generated cysteine→alanine and cysteine→serine generated mutations (C) were sequenced against the WT TNFR1 and indicated all desired mutations were successfully generated (red boxes).

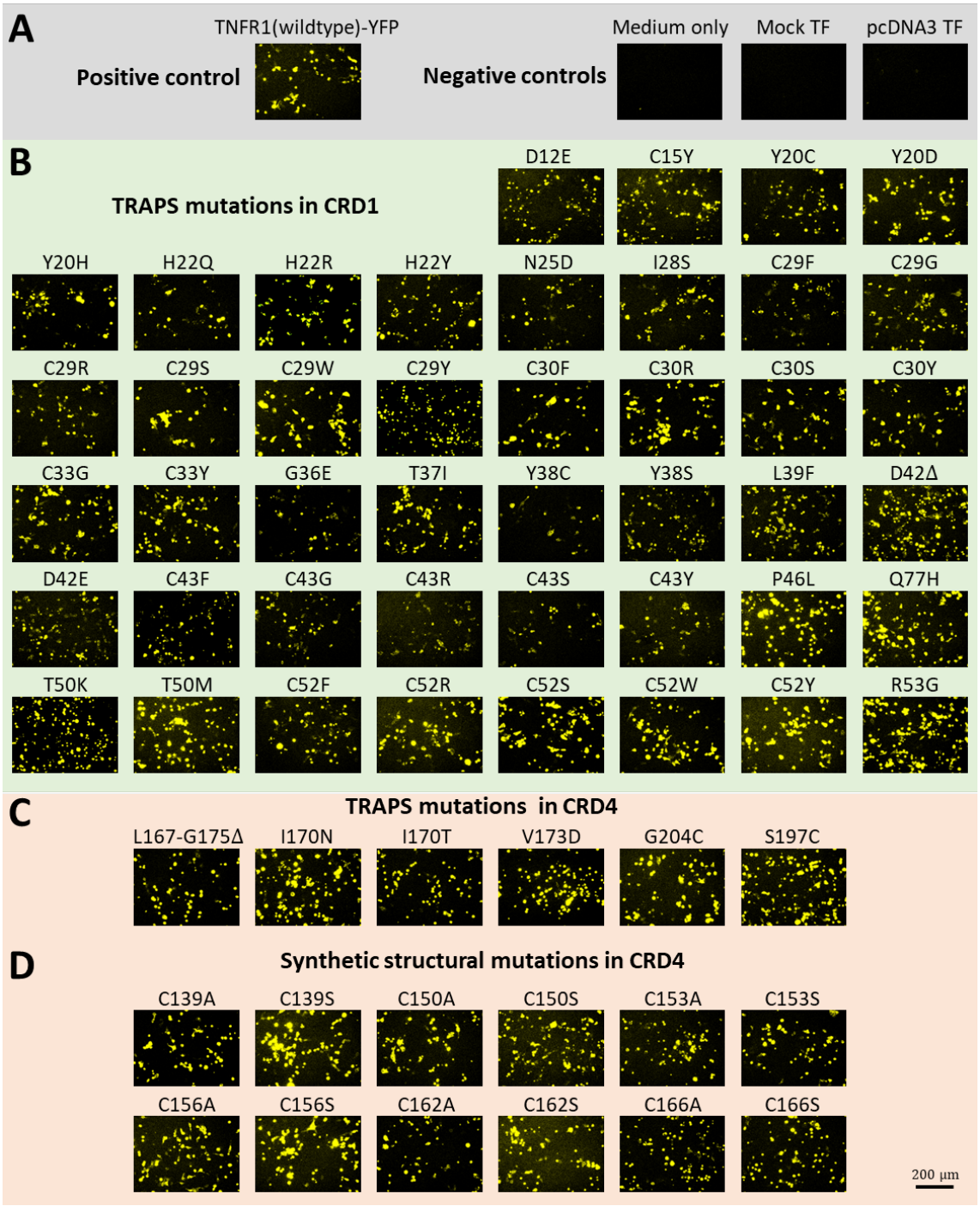


Figure 20 Fluorescence microscopy imaging of TRAPS TNFR1-eYFP expressions
A total of 62 TRAPS TNFR1-eYFP pcDNA3 plasmids were generated and transfected with the HEK-293T cells. All TRAPS TNFR1-eYFP mutant -eYFP fusion proteins were successfully generated and expressed in the mammalian cell line. Within all generated TRAPS mutants, 44 are naturally existing TRAPS mutations in the CRD1/PLAD of TNFR1 (A), six are naturally existing TRAPS mutations around the CRD4 region (B), and 12 artificially generated cysteine→alanine and cysteine→serine mutations in CRD4 (C).

3.2.4 Generation of vTNFR as -mRFP fusion constructs

A group of 3 poxviruses-encoded vTNFRs (MT2 encoded by the myxoma virus, G4R encoded by the variola virus and J2R encoded by the monkeypox virus) were selected to make the -mRFP fusion constructs. These were combined with the human TNFR1-eCFP and TNFR1-eYFP to demonstrate the FRET platform's capacity to investigate virus-host protein interactions. The vTNFR-mRFP fusion constructs were validated with the plasmid DNA sequencing and compared with the reference sequences from the NCBI database (NCBI accession: M95181.1 for MT2, NCBI accession: NC_001611.1 for G4R, and NCBI accession: JX878429.1 for J2R). In particular, the G4R and J2R were subcloned using the codon-optimised DNA plasmid template to guarantee the expression in human cells (Figure 22 yellow highlights and Figure 23 purple highlights). Therefore, the sequencing results showed that MyxMT2 has 100% sequence identity as the reference sequence (Figure 21), while VarG4R and MpxJ2R have 73.48% and 74.76% sequence identity as the reference sequence due to codon optimisation. When those subcloned sequences translated into protein, all vTNFR-mRFP have 100% amino acids sequence identity as their reference sequences in the NCBI database (Figure 21, Figure 22 and Figure 23 amino acid sequences).

Figure 21 MyxMT2-mRFP sequencing result

Myxoma virus-encoded MT2 vTNFR (black box) subcloned with the mRFP (red box) to form a fusion fluorescence receptor MyxMT2-mRFP. One silent mutation was identified in the mRFP ORF: GCC→GCT, alanine→alanine (red). No mutation in the MyxT2 ORF region was detected.

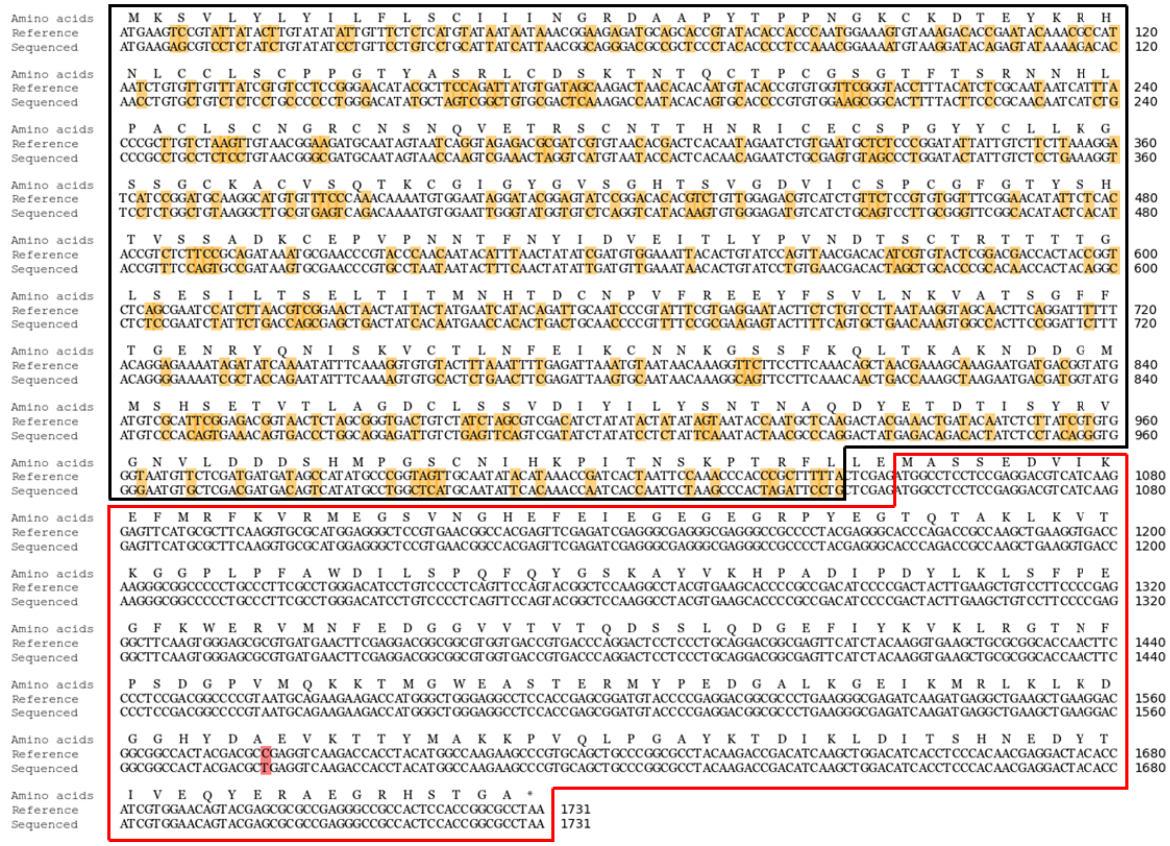


Figure 22 Codon optimised VarG4R-mRFP sequencing result

Variola virus-encoded G4R vTNFR (black box) subcloned with the mRFP (red box) to form a fusion fluorescence receptor VarG4R-mRFP. One silent mutation was identified in the mRFP ORF: GCC→GCT, alanine→alanine (red). All desired codon optimisations (yellow highlights) were identified, and no undesired mutation in the G4R ORF was detected.

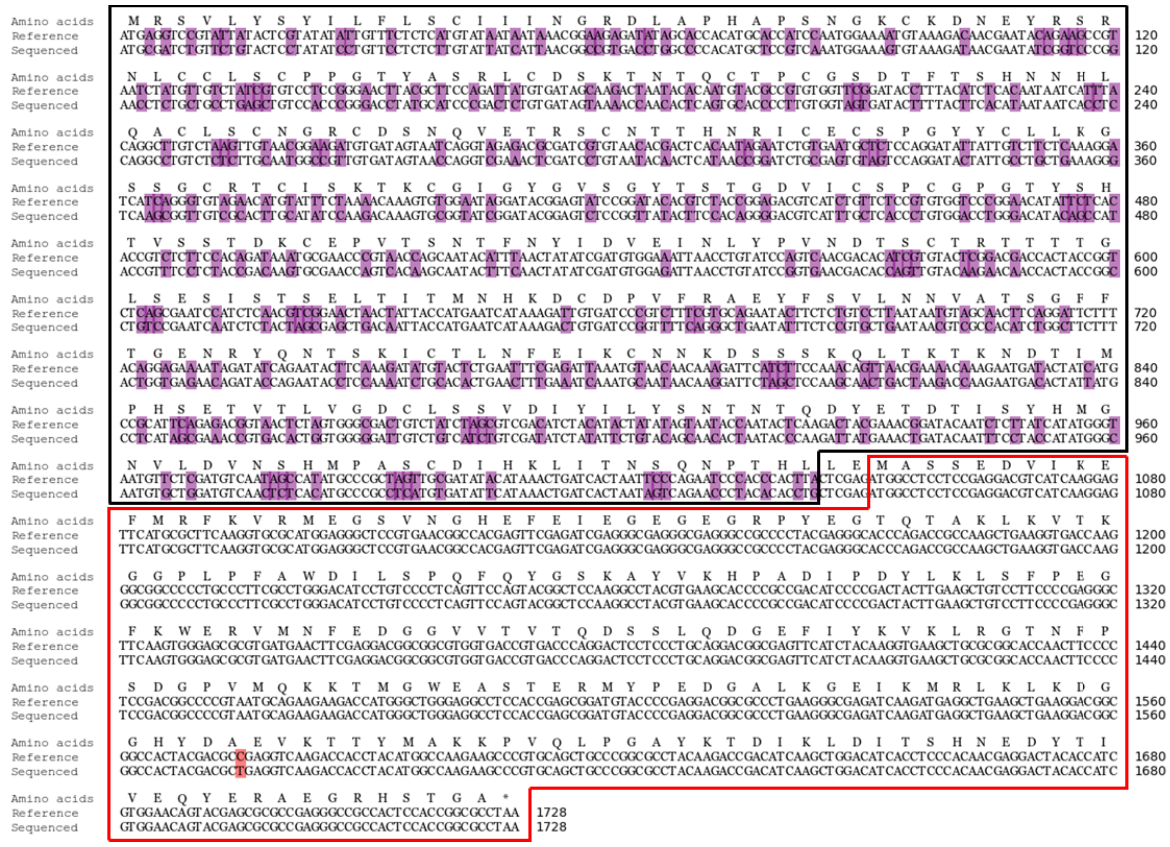


Figure 23 Codon optimised MpxJ2R-mRFP sequencing result

Monkeypox virus-encoded J2R vTNFR (black box) subcloned with the mRFP (red box) to form a fusion fluorescence receptor MpxJ2R-mRFP. One silent mutation was identified in the mRFP ORF: GCC→GCT, alanine→alanine (red). All desired codon optimisations (purple highlights) were identified, and no undesired mutation in the J2R ORF was detected.

The expression of each vTNFR-mRFP construct was also examined using the EVOS microscope, and HEK-293T cells transfected with the pcDNA3.MyxMT2-mRFP, pcDNA3.VarG4R-mRFP and pcDNA3.MpvJ2R-mRFP plasmids. All plasmids encoded vTNFR-mRFP fusion proteins were detectable using the RFP (531/40 nm excitation, 593/40 nm emission) EVOS light cube, substantiating that the newly constructed fusion proteins (MyxMT2, VarG4R and MpvJ2R) were each successfully expressed as mRFP labelled fusion proteins (i.e., red in colour) in HEK-293T cells (Figure 24).

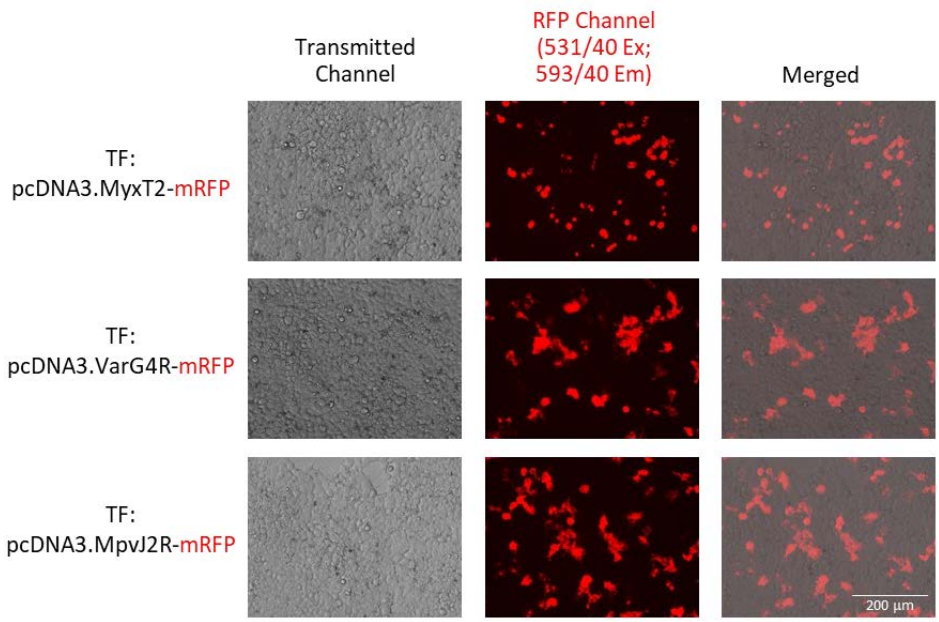


Figure 24 Fluorescence microscopy imaging of vTNFRs-mRFP expressions
Microscopy images captured in the transmitted and RFP fluorescence channels demonstrate the successful subcloning and mammalian expression of newly constructed pcDNA3.MyxT2-mRFP (top), pcDNA3.VarG4R-mRFP (middle), and pcDNA3.MpvJ2R-mRFP (bottom).

3.2.5 Generation of pGL4.23.NF-κB-[miniP]-miRFP703 proinflammatory reporter

Firstly, the *Avr*II digested 2× NF-κB response element oligonucleotides were incubated in T4 ligation buffer for self-ligation to achieve higher copies of insert NF-κB response element fragments (i.e., 4× and 8×). Then, the *Nhe*I digested pGL4.23.[miniP] vector was added to the subcloning reaction to form the pGL4.23.NF-κB-[miniP] plasmid with different numbers of insert copies. The ligation can be achieved because *Avr*II and *Nhe*I restriction enzymes generate compatible sticky ends during digestion (Figure 25A).

The newly constructed potential pGL4.23.NF-κB-[miniP] plasmids were then double-digested with *Sac*I and *Xho*I to release the entire insert fragment. Plasmids with different lengths of released insert fragments were selected for sequencing (Figure 25B). The sequencing results indicated pGL4.23.NF-κB-[miniP] with 2×, 4×, and 8× copies of inserts were successfully generated with the response element sequences located before the minimal promoter sequence (Figure 25C). Furthermore, the sequencing data also indicated that *Nco*I and *Xba*I double-digested miRFP703 reporter sequence was successfully cloned after the minimal promoter using the pGL4.23.NF-κB-[miniP] plasmids as the backbone (one example is shown in Figure 25D).

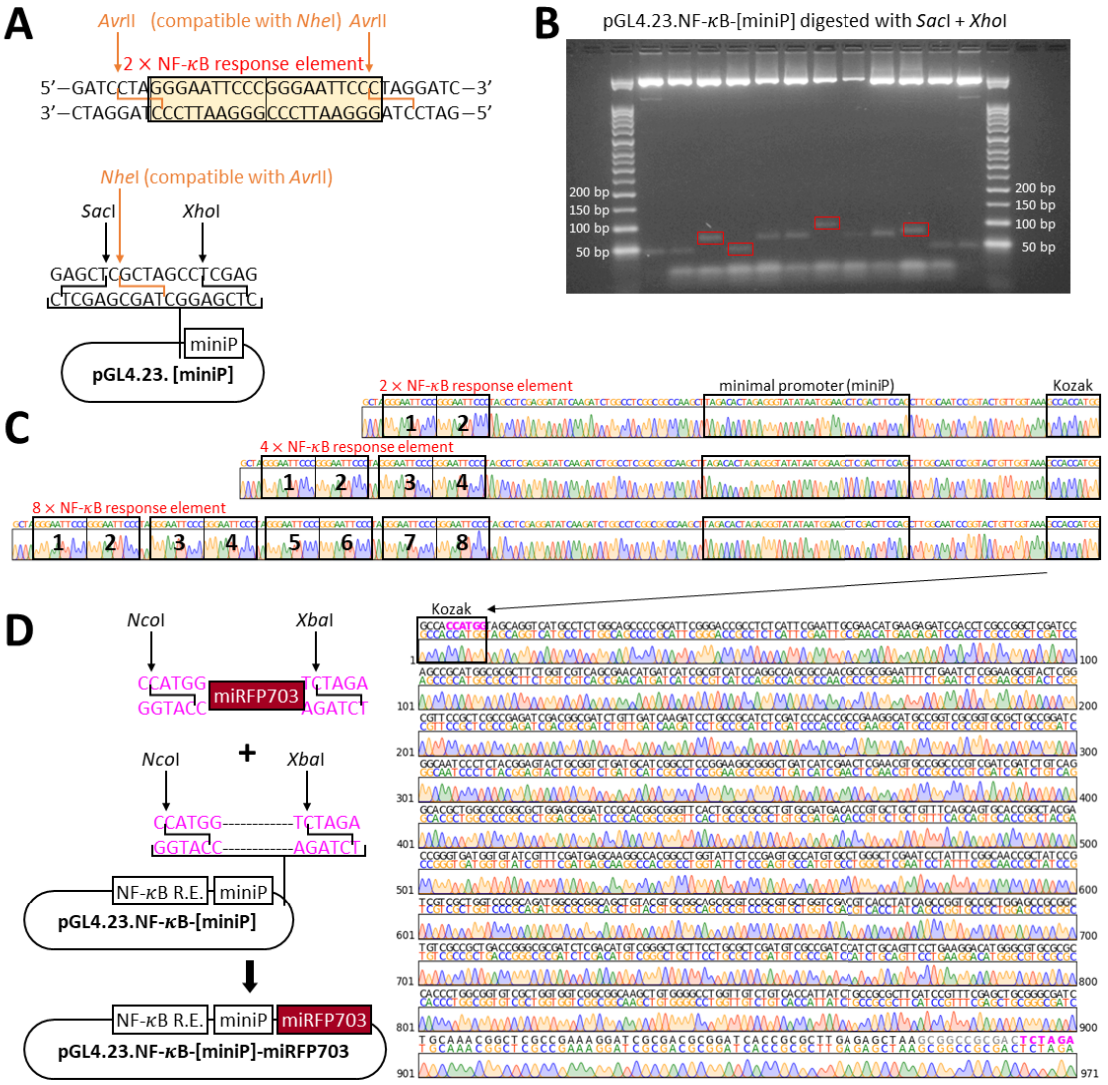


Figure 25 Sequencing results of the NF-κB-miRFP703 proinflammatory reporters
The construction of 2× NF-κB response elements insert fragments, which were subcloned into the pGL4.23 minimal promoter plasmid vector using the compatible *Avr*II and *Nhe*I sites (A). Insert fragments released by *Sac*I and *Xho*I double digestion (B). Sequencing results of the NF-κB response elements were successfully subcloned into the pGL4.23 vector before the minimal promoter sequence (C). The sequencing result indicated that the miRFP703 fluorescent protein reporter was successfully subcloned into the pGL4.23 vector after the minimal promoter sequence (D).

The NF-κB can be selectively activated when co-transfected with the non-fluorescence TNFR1 plasmids (Figure 26). In comparison, cells co-transfected with the non-fluorescence TNFR2 or CD27 plasmid and NF-κB reporter plasmids showed no NF-κB activation (Appendix Figure 9). The medium-only cells, pcDNA3 empty vector-transfected cells, and pGL4.23.[miniP]-miRFP703 plasmid (without NF-κB response element) transfected cells also showed no detectable miRFP fluorescence (Appendix Figure 10). Moreover, different number of copies of the NF-κB response elements induced different levels of miRFP703 expression. When co-transfected with TNFR1 plasmids, the 4× and 8× NF-κB response element samples both demonstrated higher miRFP703 expression levels higher than the sample without NF-κB response element, while the 2× NF-κB response element sample showed the same expression level as the no NF-κB response element sample, this means more than two copies of the NF-κB response element sequences are required to activate the miRFP703 expression through the minimal promoter. When this NF-κB proinflammatory reporter was used with the FRET reporters, the cells were harvested at 36 hours post-transfection. Thus, the miRFP703 detection levels were also calculated at 36 hours post-transfection to compare the activation using different number of copies of NF-κB response elements. The 2× NF-κB response element sample showed no significant difference (t-statistic = 0.088, p-value = 0.936) when compared to the NF-κB response element negative control, while both 4× and 8× NF-κB response element samples showed significantly higher miRFP703 expression than the negative control (t-statistic = -19.239, p-value = 0.002 and t-statistic = -15.498, p-value = 0.0004). The miRFP703 expression levels between the 4× and 8× NF-κB response element samples were not significantly different from each other (t-statistic = 2.949, p-value = 0.084).

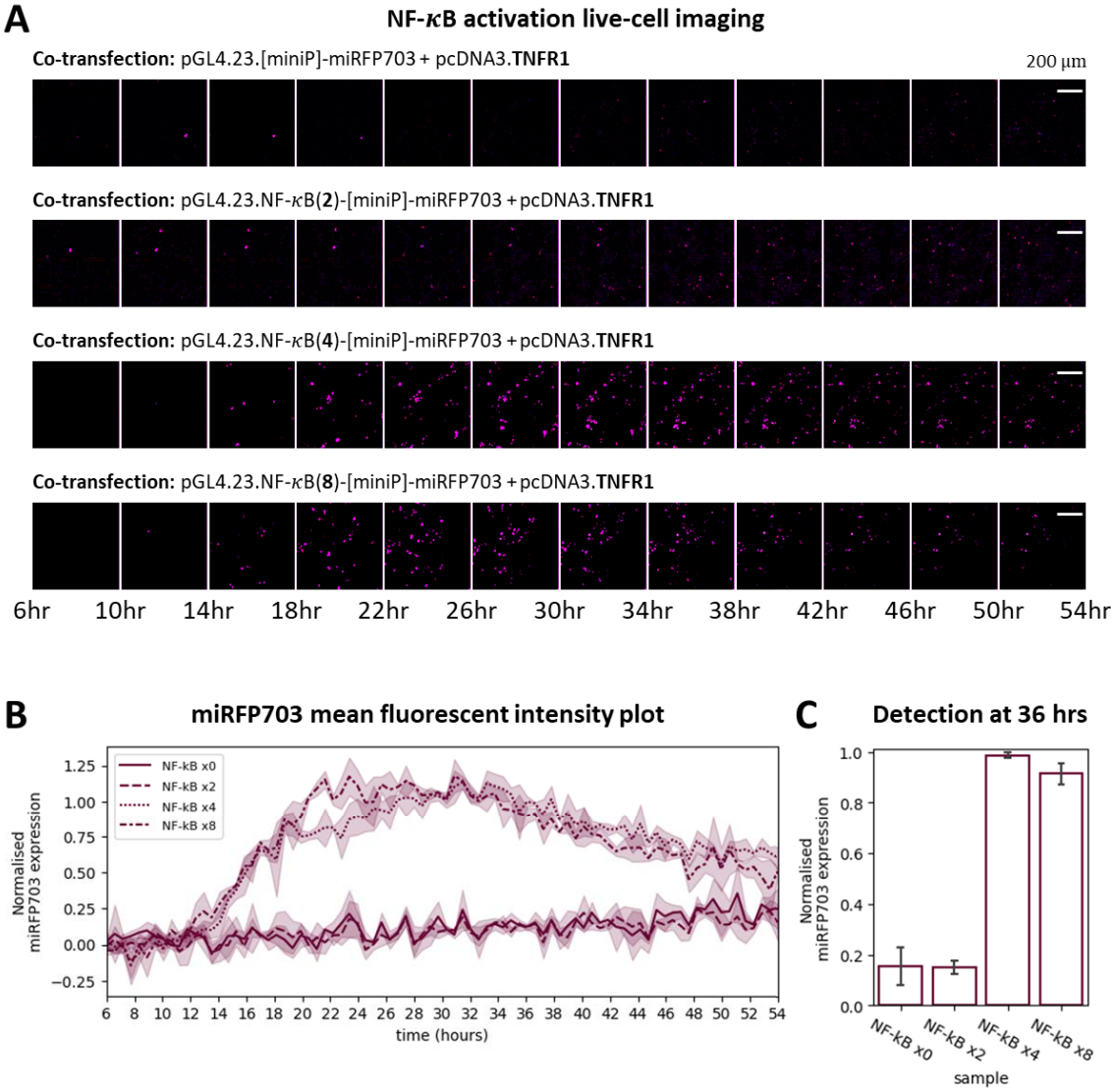


Figure 26 Live-cell imaging of NF-κB reporter and TNFR1 co-transfected HEK-293T cells
The NF-κB reporter with zero or 2× response elements had minimal miRFP expression, and the NF-κB reporter with 4× or 8× response elements started to express miRFP from 14 hours post-transfection (A). The expression dynamics of the NF-κB reporter with different number of copies of response elements (B). Normalised miRFP703 detection level calculated at the 36 hours post-transfection (C).

CHAPTER FOUR

FLOW CYTOMETRY

DATA PRE-PROCESSING

4.1 Introduction

Flow cytometry data pre-processing involves a series of logical stages before one can perform accurate downstream data analyses like FRET quantification. While the detection of FRET has been used to characterise many intracellular interactions, quantifying and comparing the FRET results are particularly difficult due to the complex nature of these processes within living cells; often, many variances can be produced during data pre-processing. Many software and programmed packages/libraries in python and R offer solutions for each pre-processing stage (see literature review section 1.3). Still, challenges in singlet identification, anomaly detection, autofluorescence removal, and spectral unmixing impede investigators from achieving error-free flow cytometry measurements. This lack of high-quality pre-proceed data has largely decreased the sensitivity, accuracy, and reliability of flow cytometry-based assays (including FRET assays) in academic research, clinical diagnosis, and pharmaceutical therapeutics screening.

Firstly, there are two most essential gating steps in a typical flow cytometry data analysis: (i) singlet identification to separate single cell events from cell aggregates (doublets or multiplets), cell debris and non-cell background events, and (ii) subtype identification to separate different cell types with distinct phenotypes into groups. The majority of automated flow cytometry gating algorithms have been focused on the second one in detecting various cell subpopulations utilising multidimensional fluorescence signals (see literature review section 1.3.4). On the one hand, flow cytometers detect well-calibrated scattered light signals that permit a series of gold-standard manual gating procedures to separate singlets from other data events. On the other hand, cells can

display varied scattered light profiles upon different treatments and require manual adjustment for individual samples. Therefore, the manual gating approach can be subjective, labour-intensive, and time-consuming for big-batch data analysis. To date, only a few methods have presented unsupervised algorithms for cell event and singlet identification to provide automated and objective solutions. The TASBE software package incorporates automated cell gating using a GMM to fit the scattered light data but can only identify cells from background events, not singlets from doublets or cell aggregates [86]. Razo-Mejia *et al.* established an unsupervised two-dimensional Gaussian function-based gating algorithm that fits the $\log_{10}$(FSC) and $\log_{10}$(SSC) data using covariance and chi-squared test. However, only events within the centre, high-density region (usually 40% of the total) are gated as cells using this approach [128]. The most recent algorithm is FlowGateNIST which uses iterative GMM with different cluster numbers to discriminate between cells and background events and, subsequently, between singlet and multiplet events. However, an optimised GMM model based on the BIC needs repetitive GMM fitting, which increases processing time. In addition, FlowGateNIST is limited to analysing small particles such as bacterial cells and is unsuitable for large mammalian cells [84]. Therefore, the desired singlets identification algorithm must be automatic, objective, rapid, and suitable for mammalian cell applications.

Secondly, flow cytometers constantly measure baseline fluorescence signals and subtract them from the detected fluorescent signals during data acquisition. Abnormal baseline subtractions can sometimes result in negative fluorescent intensity readout without actual physical or biological meaning. The negative values can also lead to

significant data loss since downstream statistical analyses require the removal of those negative values. The current widely used algorithms FlowClean and FlowAI can resolve aberrant detection errors such as saturated margin data, data with irregular fluidic flow rate and data with abnormal fluorescent intensity measurement. However, when handling the baseline subtraction-error, FlowClean simply replaces negative values with small random numbers, and flowAI only removes a portion of the negative data based on the calculated threshold, which both leave unresolved fluorescence quantification errors [79, 80]. A single-cell level baseline subtraction correction algorithm is urgently needed to prevent flow cytometry data loss and to achieve the detection of error-free fluorescent intensity values.

Thirdly, autofluorescence is a mixed fluorescent signal of various biomaterials in the biological sample. Autofluorescence can be problematic because it reduces the signal-to-noise ratio, resulting in decreased sensitivity and false-positives. Furthermore, autofluorescence spill-over can also obscure low-emitting signals, such as fluorophore-tagged proteins with low expression levels, weak sensitised FRET emissions and cellular signalling pathway reporters with low activation levels. The current methods are inaccurate for autofluorescence removal. They either subtract the sample-level mean autofluorescence intensity calculated from the non-fluorescent control samples or treat autofluorescence as a single colour and remove it with the linear spectral unmixing approach. The former conducts subtraction at the non-single-cell level, and the latter requires an extra detection channel with increased hardware requirements. Both approaches ignore that autofluorescence varies from cell to cell and reflects the biomaterials' heterogeneity in each sample. Therefore, a non-linear single-cell algorithm

is needed for precise autofluorescence prediction and removal to increase the signal-to-noise ratio, detection sensitivity and fluorescent intensity quantification accuracy.

Lastly, spectral unmixing is a fluorescent signal compensation technique that removes the spillovers from the spectrally overlapped photon emissions and cross-laser excited photon emissions from samples labelled with multiple fluorophores. All current spectral unmixing methods essentially minimise the spread error, indicating the level of residue error spreading after compensation. The major contribution to the spread errors are unresolved anomaly detection signals and autofluorescence, so the performance of the spectral unmixing is significantly dependent on the baseline subtraction correction and autofluorescence removal mentioned earlier. Moreover, compensation is especially challenging for FRET assays since it requires single-colour controls for all detected fluorescence signals since FRET signals have no true single-colour control because the FRET-donor and FRET-acceptor fluorescence emissions are simultaneously detected with the sensitised FRET fluorescence. Therefore, the spectral unmixing algorithm that can handle the FRET emissions is needed to achieve accurate flow cytometry-based FRET quantification.

In this chapter, I have developed several unsupervised algorithms to overcome the abovementioned unresolved challenges in flow cytometry data pre-processing. The performance of each pre-processing step and the overall pre-processing pipeline have been evaluated using a total of 11 flow cytometry-based FRET experiments consisting of three cell types and five different FRET pairs detected by five distinct flow cytometers (including conventional and spectral flow cytometers). This chapter provides ML-powered solutions that achieve automatic, objective, ultra-fast singlet identification,

followed by an error-free baseline-subtraction correction, autofluorescence removal and spectral unmixing. Together, they remove the obstacle in achieving accurate and unbiased flow cytometry data quantification and significantly increase the sensitivity and robustness of all flow cytometry assays, including the detection of FRET.

4.2 Results

4.2.1 UltraFast – an unsupervised algorithm for automatic singlet identification

Flow cytometers measure the relative cell size and granularity with forward-scattered (FSC) light and side-scattered (SSC) light detectors (Figure 27A). Each of the FSC and SSC signals has three measurable parameters, the heights (FSC-H, SSC-H) indicate the signal strength, the widths (FSC-W, SSC-W) indicate the relative time cost by a particle to pass through the laser, and the areas (FSC-A, SSC-A) represent the integrated area of the overall detected signal (Figure 27B). Despite the well-calibrated nature of flow cytometers allowing for a standardized six-step approach to separate singlets from other events, cells under different treatments or expressing different proteins can exhibit diverse profiles in these gating panels (Figure 27C). For instance, cells transfected with plasmid DNAs, especially those overexpressing human TNFR1 proteins, show noticeable enlargement and increased granularity (Figure 27C last row).

To elucidate, the use of manual gating in this study is primarily for baseline comparison. The ground truths of singlet identities were initially established using carefully adjusted manual gates for each sample, this approach was also assessed by cell sorting and microscopy imaging (see Appendix Figure 11). The performance of conventional manual gating baseline was performed on identifying singlets in a semi-automated way. In detail, the six-step gates were established on one sample, and the same set of gates was

automatically applied across the rest of the samples to test the accuracy of the batch manual gating approach.

This semi-automated approach separately used each of the untransfected sample, pcDNA3 empty vector transfected sample, pcDNA3.TNFR1 transfected sample and pcDNA3.CD27 transfected sample as the establishment control sample to generate an averaged performance. On average, the semi-automatic batch manual approach achieved an average F1 score of 0.7767 with an average precision of 0.9057 and an average recall of 0.7327 (Figure 27D). The standard deviation of the performance also has an extensive range from 0.0561 to 0.3802, indicating that the manual gating strategy's performance can be varied when utilised in the semi-automated batch mode.

Contrastingly, the UltraFast algorithm in the below sections operates on an entirely automated, unsupervised basis, eliminating the necessity for manual gating. It's essential to clarify that the semi-automated approach mentioned is utilised to provide a comparative baseline, not as a component of the UltraFast technique. UltraFast's innovation lies in its capability to automatically identify singlets, offering an improved, more consistent performance over manual or semi-automated strategies.

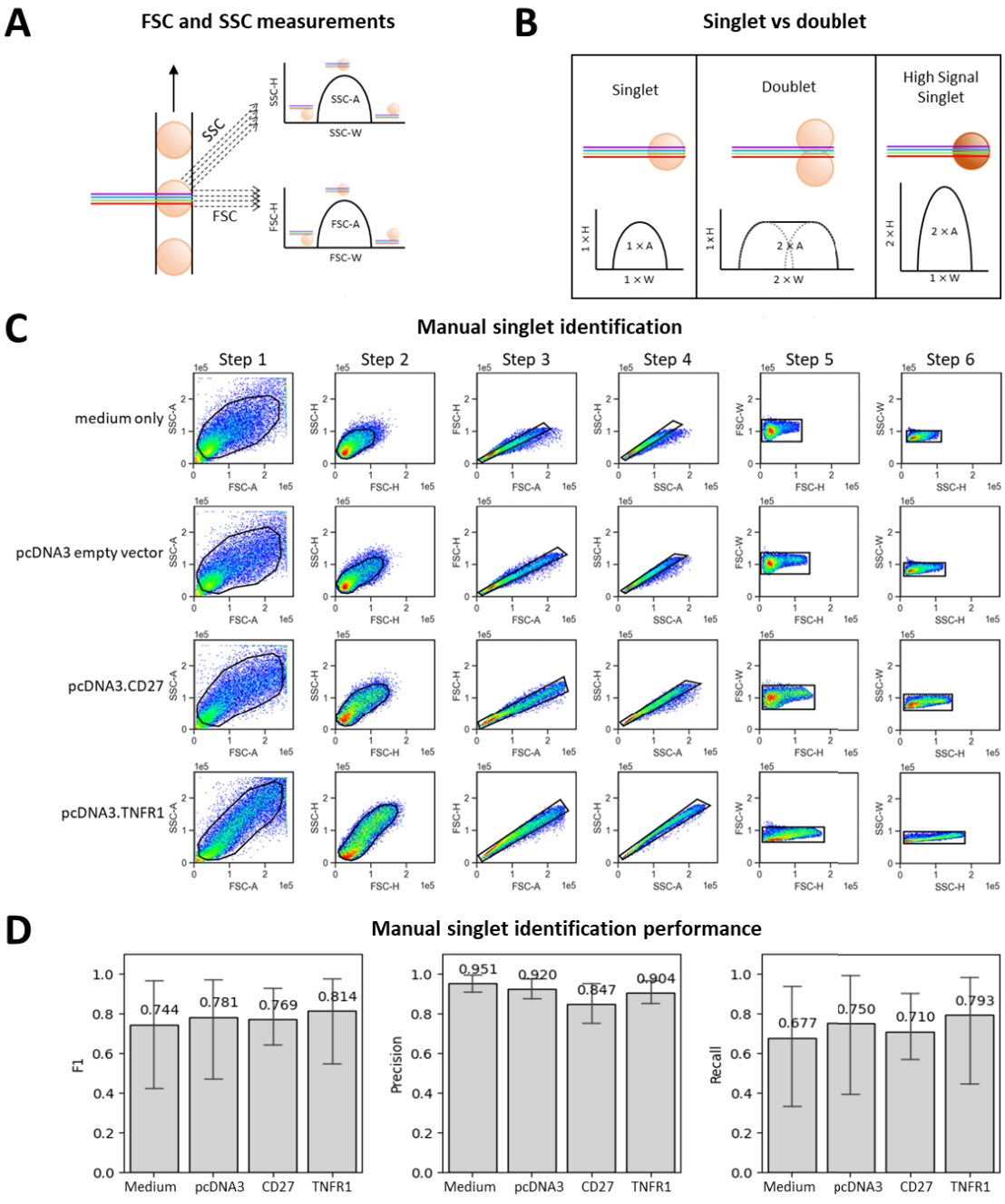


Figure 27 Semi-automated batch mode manual singlet identification

Flow cytometers measure FSC and SSC signals on two perpendicular detectors (A). Each FSC and SSC signal has width, height and area parameters (B). The six-step manual gatings were applied on HEK-293T cells with different transfections (C). The performance of singlet identification when utilising the semi-automated batch mode manual gating approach. The F1 scores, precision, and recall were presented as the bar graphs using different sample as the establishment control sample that generated the gates (D).

The unsupervised clustering algorithm, density-based spatial clustering of applications with noise (DBSCAN), has superior performance when clustering flow cytometry data with irregular shapes and much background noise (see Figure 29A). Several commonly used unsupervised clustering algorithms, K-Means, Ward, agglomerative clustering and GMM, were used to compare with the DBSCAN algorithm on four sets of synthetic data (Figure 28). Although all algorithms can handle the data with regular shapes and equal-size clusters well (Figure 28, first row), only DBSCAN perfectly clusters the irregularly shaped data (Figure 28, second row). DBSCAN also has a unique feature of noise identification (as those purple dots) that prevents the forced assignment of the noise into clusters (Figure 28, third row). Lastly, DBSCAN does not require user-supplied cluster numbers, preventing the incorrect separation of one group into numerous clusters (Figure 28, last row).

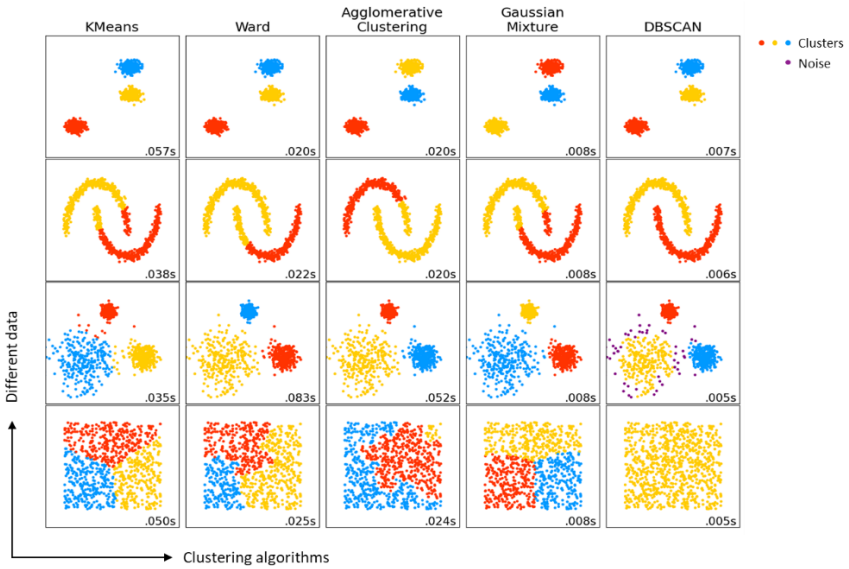


Figure 28 Unsupervised clustering algorithms comparison
K-Means, Ward, Agglomerative clustering, and GMM compared with the DBSCAN algorithm. Each data has 1000 data points. Each of the blue, red and yellow colour represents one cluster after unsupervised clustering, and the numbers at the lower right corners are indicating the computing time of the clustering task. The clustering comparison was performed on data with regular shapes and equal size (1st row), irregular shapes (2nd row), and different sizes with noise (purple colour dots) (3rd row). In addition, DBSCAN does not require user-supplied hyperparameter for cluster number, so it did not perform the incorrect separation of one group into numerous clusters (4th row).

For this project, I have developed a six-stage unsupervised algorithm called UltraFast that further raised the power of DBSCAN. The algorithm flowchart is summarised with graph illustrations (Figure 29A), and the detailed methods are described in section 2.7.2. The UltraFast algorithm can achieve unbiased, automated, singlet identification with less than half a second processing time per sample with 30,000 cells. The UltraFast algorithm achieves an average F1 score of 0.9193 with SD = 0.0090 (precision = 0.9305, recall = 0.9084), indicating a high true-positive singlet identification with low false-negative and low false-positive detection rates. This result outperforms many other clustering algorithms, including FlowGrid (mean F1 = 0.7484, SD = 0.0579). K-Means (mean F1 = 0.4300, SD = 0.0676), G-Means (mean F1 = 0.4257, SD = 0.0666), X-Means (mean F1 = 0.4247, SD = 0.0662), flowMeans (mean F1 = 0.7607, SD = 0.0586), GMM (mean F1 = 0.6053, SD = 0.0577), flowClust (mean F1 =0 .7171, SD = 0.0064), 2D gaussian fitting with covariance and chi-squared test (mean F1 = 0.6821, SD = 0.0624) (Figure 29B), for each algorithm the same data transformation from UltraFast stage-1 to stage-3 were utilised to achieve fair comparison (Figure 29B). The singlet identification results have been validated with several cell lines, including U2OS, HL60, HEK293T, THP1 and human T cells from human peripheral blood mononuclear cells (PBMCs), and the gating profiles were compared with the six-step manual gating using scatterplots (see Appendix Figure 12 with an example using HEK293T cell line). In addition, the UltraFast provides a “grid number” hyperparameter in step four which is approximate to the strictness of singlet identification that can be further fine-tuned by human experts. The optimisation of this hyperparameter can also be automatically tuned through iterations which shows a relatively stable performance from a broad range of grid numbers (50×50 to 150×150),

in which the F1 scores>0.9 was maintained using 82×82 to 122×122 number of grids (Figure 29C).

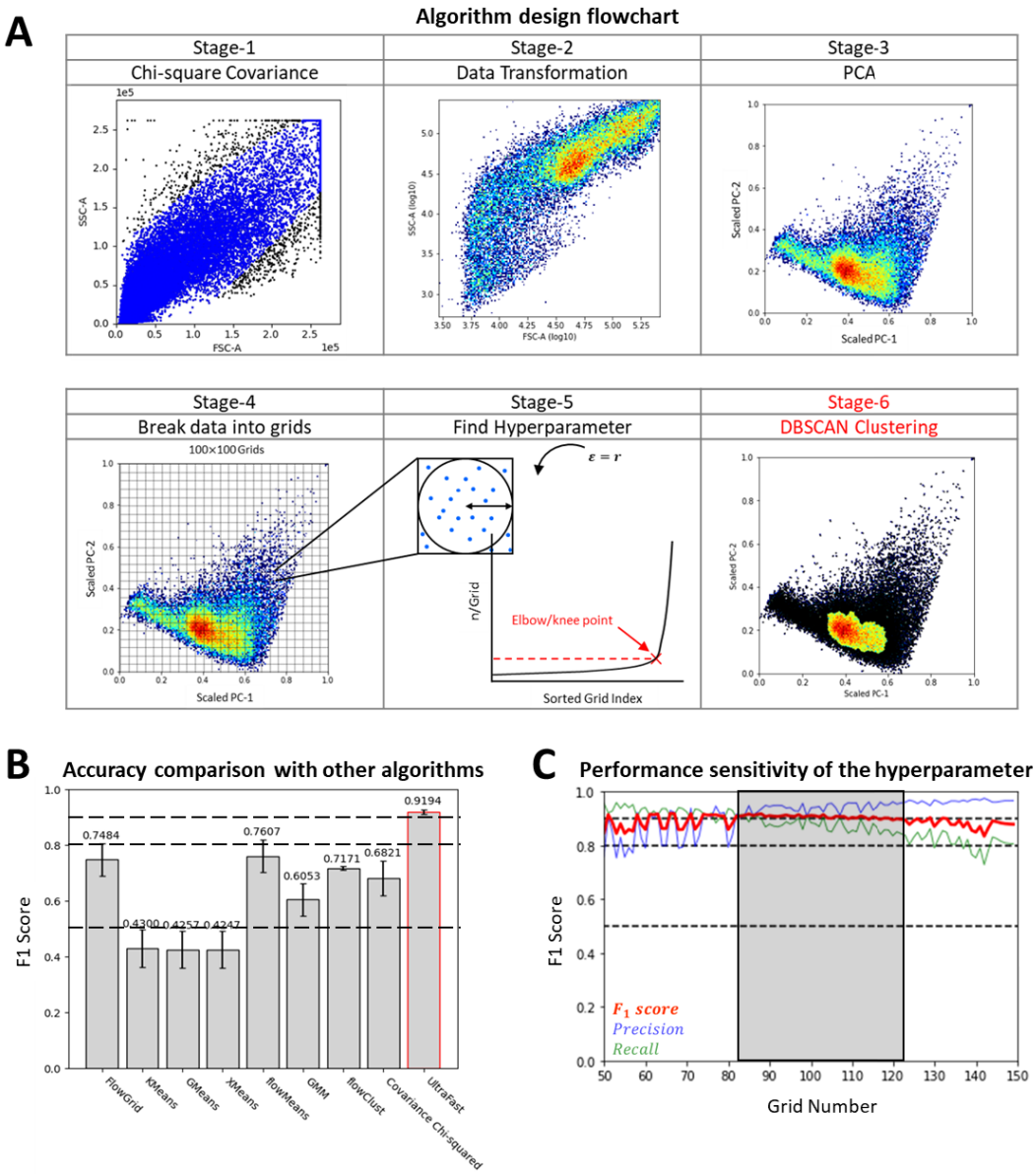


Figure 29 The unsupervised UltraFast singlet identification algorithm

The UltraFast algorithm's design consists of six stages (i) primary population identification (blue dots), (ii) data logarithmic transformation, (iii) PCA dimensionality reduction, (iv) generation of grid boxes, (v) finding the critical point using elbow method, and (vi) DBSCAN clustering (A). The F1 score of singlet identification results from different clustering algorithms, using the manual gated samples as the ground truth (B). The performance of UltraFast with different grid numbers (C).

4.2.2 The algorithm for single-cell level baseline subtraction-error correction

The baseline subtraction errors were corrected using the unsupervised collaborative filtering algorithm. This algorithm is currently used in many online streaming websites in their movie recommendation system. The algorithm learns from the original data containing rated and unrated movie scores from every user, and it generates attributes such as sci-fi, action, and romance (Figure 30 attributes in red). Those attributes are further used in model formulation to predict the unrated movie scores for each user. The prediction is achieved by minimising the cost function between predicted and known-rated scores using iterative gradient descent (see chapter two, section 2.7.3).

Collaborative filter algorism intuition

Movie Scores

	Movies				
Users	"Iron Man"	"Super Man"	"Transformer"	"Harry Potter"	"Titanic"
Alice	8.0	7.9	6.5	9.3	4.5
Bob	7.0	6.0	7.0	6.5	8.9
Carol	4.0	??	4.8	7.3	9.2
Dave	8.2	7.9	7.5	9.3	6.5
Eva	7.2	6.1	7.3	??	7.9
Fred	8.0	7.4	6.5	9.0	??
Grace	4.2	4.4	4.8	7.0	9.1

Predicted Movie Scores

	Movies				
Users	"Iron Man"	"Super Man"	"Transformer"	"Harry Potter"	"Titanic"
Alice	8.0	7.9	6.5	9.3	4.5
Bob	7.0	6.0	7.0	6.5	8.9
Carol	4.0	4.5	4.8	7.3	9.2
Dave	8.2	7.9	7.5	9.3	6.5
Eva	7.2	6.1	7.3	6.2	7.9
Fred	8.0	7.4	6.5	9.0	6.3
Grace	4.2	4.4	4.8	7.0	9.1

Low-rank matrix factorisation

$$J(x^{(1)},\dots,x^{(n_m)},\theta^{(1)},\dots,\theta^{(n_u)}) = \min_{\substack{x^{(1)},\dots,x^{(n_m)} \\ \theta^{(1)},\dots,\theta^{(n_u)}}} \frac{1}{2}\sum_{(i,j):r(i,j)=1}\left((\theta^{(j)})^T x^{(i)} - y^{(i,j)}\right)^2 + \frac{\lambda}{2}\sum_{i=1}^{n_m}\sum_{k=1}^{n}\left(x_k^{(i)}\right)^2 + \frac{\lambda}{2}\sum_{j=1}^{n_u}\sum_{k=1}^{n}\left(\theta_k^{(j)}\right)^2$$

$$Weight(\theta) \times Learned\ Attributes(X) \Rightarrow Movies\ Ratings(Y)$$

Attributes	"Iron Man"	"Super Man"	"Transformer"	"Harry Potter"	"Titanic"
Sci-fi	63%	60%	72%	0%	0%
Action	17%	15%	8%	8%	0%
Romance	5%	15%	7%	12%	94%
Fantasy	10%	5%	10%	70%	1%
Drama	5%	5%	3%	10%	5%
etc...	...	...	...	...	...

Figure 30 Collaborative filtering algorithm intuition

The collaborative filtering (CF) recommendation algorithm can perform unsupervised movie score prediction for unrated movies by users (red question marks and red scores). CF uses the existing data (black numbers) to learn attributes (red attributes) and use them to model and predict the move rating score while minimising the differences (costs) between the existing scores and predicted scores for rated movies. Meanwhile, unrated movie scores are also predicted.

Baseline subtraction errors can cause huge flow cytometry data loss since they lead to negative fluorescent readouts with no physical or biological meaning, and those values need to be removed before any statistical analysis. By examining the fluorescent intensity measured with seven different flow cytometers that used fluorescent detection channels ranging from 4 to 48, the average percentage of cells with negative fluorescent intensities was 10.53%-74.29% (Figure 31). Further, by examining both the fluorescent intensity heatmap and the fluorescent positive/negative binary plot generated from the full-spectrum flow cytometer data, even within the red-boxed primary detection channels, which had strong fluorescent signals, there are more than half of the population had negative value in at least one detector (Figure 32A and B).

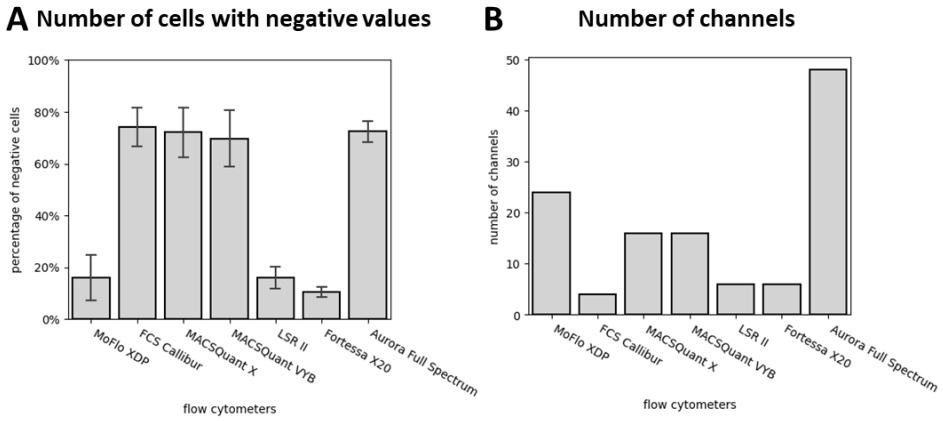


Figure 31 Number of cells with negative fluorescent intensities in seven flow cytometers
All seven flow cytometers examined here generated negative fluorescent intensity values. The percentage of cells with negative fluorescent values varied among different flow cytometers (A). The number of detection channels used in each flow cytometer was also different (B). Samples tested here were fluorescently tagged cell lines, including THP-1, HEK293T cell lines and human PBMCs. The fluorescent colours used here includes eGFP, eCFP, eYFP, mRFP, FITC, Violet-Blue, Aquamarine, and mNeoGreen.

The baseline subtraction-error correction was implemented with successfully identified singlet data using the UltraFast algorithm. Here we treat single cells as online streaming users, the detection channel as movies, the fluorescent intensities as movie rating scores, and negative data need to be predicted as the question marked unrated movie scores (Figure 30). The algorithm automatically learns pseudo-attributes from the original flow cytometry measurements, and it minimises the cost function between the predicted values and the good-quality data measurements to predict and recover the baseline over-subtracted fluorescent intensity values. The performance results demonstrated a 100% data recovery rate with an evaluation R^2 score > 0.998 and the mean absolute error (MAE) < 0.01% using the number of pseudo-attributes >= 40, where the algorithm performance became stable (Figure 32C). The data centres have been successfully recovered from the margin to the middle of the populations by examining the multichannel scatterplots and histogram overlays (Figure 33 and Appendix Figure 13), which achieved nearly error-free baseline subtraction correction with a 100% data recovery rate.

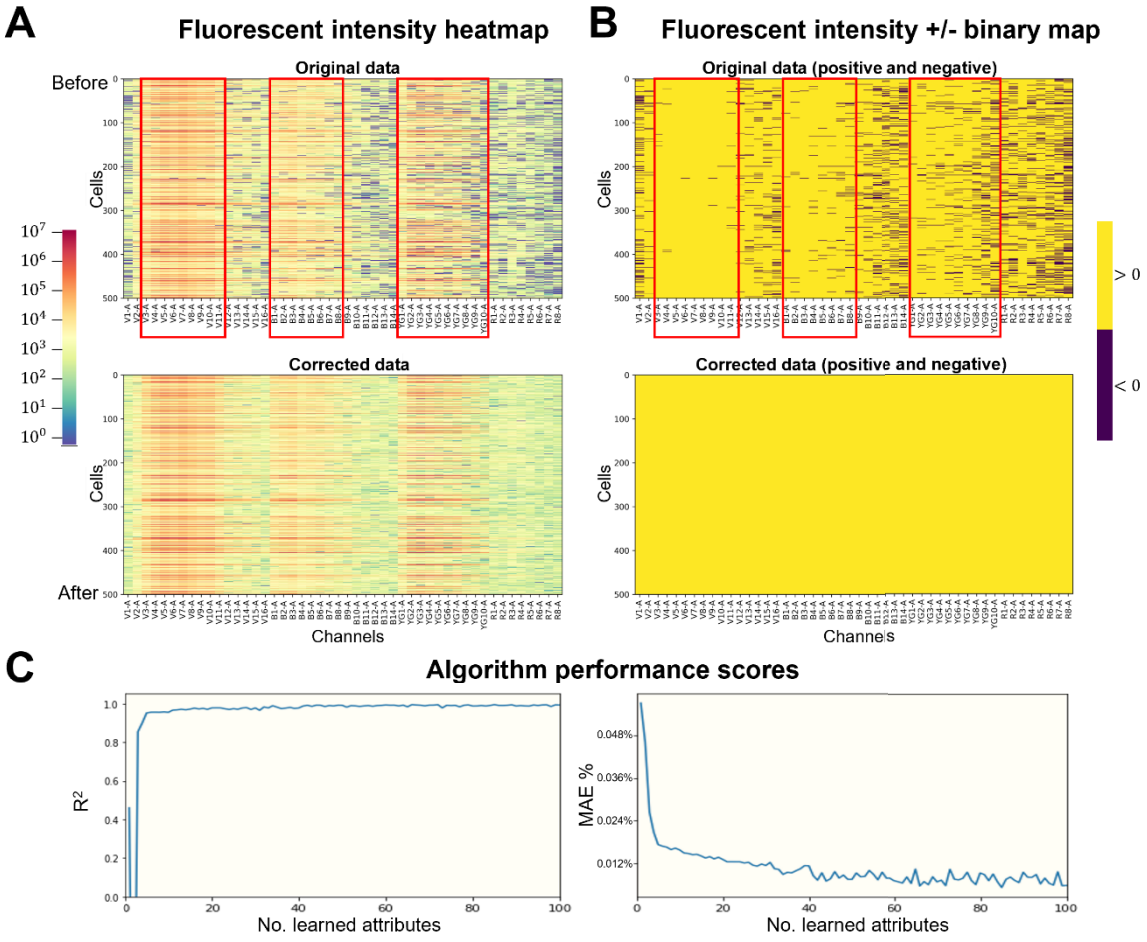


Figure 32 Performance of the baseline subtraction-error correction algorithm

The fluorescent intensity heatmaps (A) and positive/negative binary plots (B) indicate the fluorescent intensity levels and data recovery rate, respectively. Red boxes show the detection channels with strong signals. The profiles show a 100% data recovery rate after the baseline correction (B bottom) with highly identical fluorescent intensity profiles (A bottom). The performance of the baseline subtraction-error correction algorithm using R^2 and MAE%. The data in this figure was generated using the full-spectrum flow cytometer.

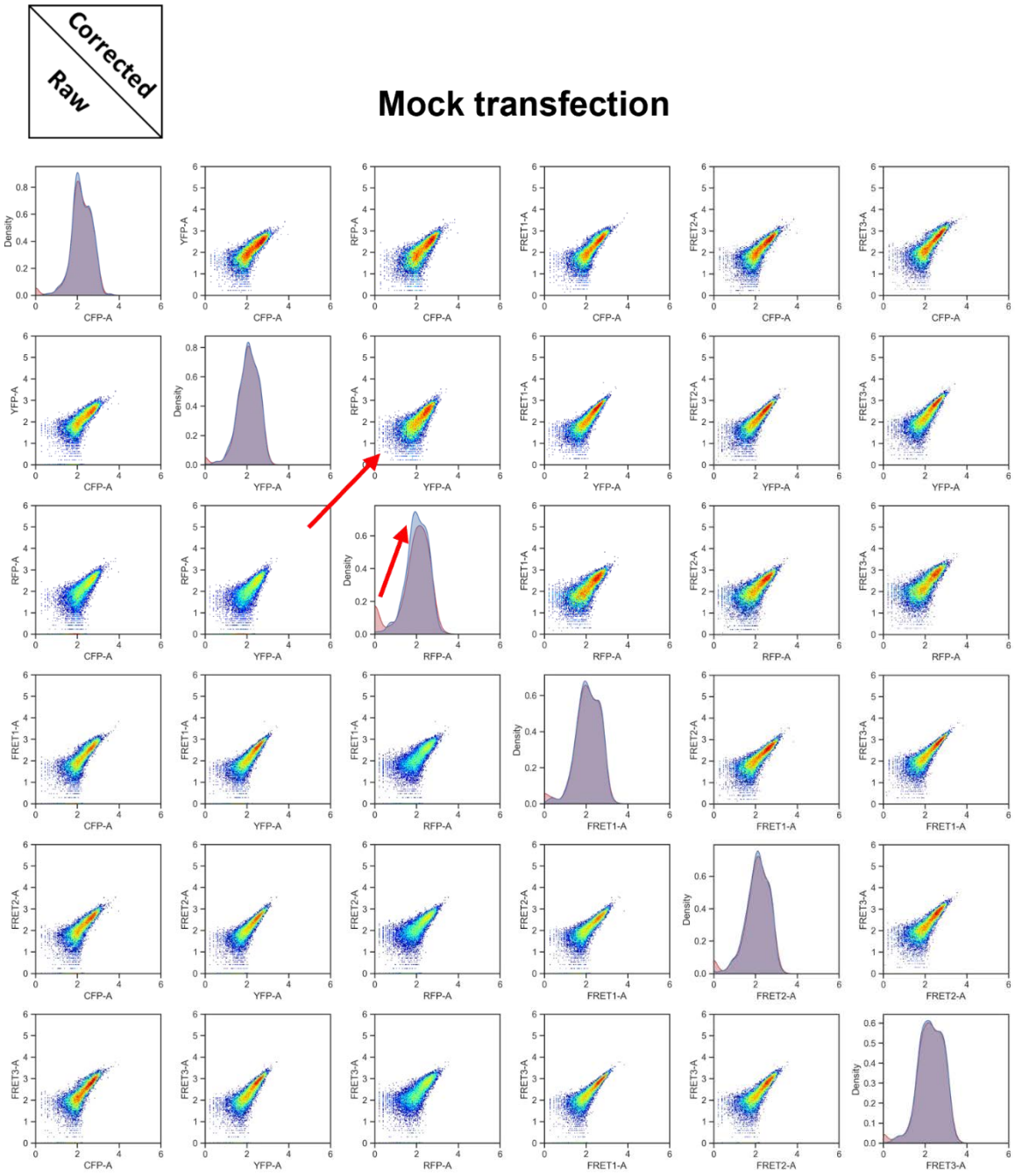


Figure 33 Baseline subtraction-error correction results for mock-transfected cells
Scatterplot and histogram overlays of the mock-transfected sample before (red histograms) and after (blue histograms) baseline subtraction-error correction. The scatter plots are symmetric along the diagonal histogram plots. The data centre returned to the main population from the edge (red arrows).

4.2.3 The algorithm for single-cell level autofluorescence prediction and removal

The unsupervised CF algorithm also achieves autofluorescence prediction and removal, and the algorithm can accomplish error-free results that generate pure autofluorescence-free fluorescent intensity readouts. The concept and algorithm formulation used here is the same as in the baseline subtraction correction section, but with differently manufactured structures (Figure 34A). Here, the bottom two blocks illustrate autofluorescence-only cells transfected with the pcDNA3 empty plasmid vector and have identical values for autofluorescence and total detected fluorescence. The top blocks are the test sample, with autofluorescence that needed to be predicted (on the left) and the total detected fluorescent (on the right) (Figure 34A). When the algorithm becomes stable (i.e., learned numbers of attributes > 20), it achieved R^2 score > 0.999 and percentage MAE < 0.001% indicating the near-perfect prediction accuracy (Figure 34B). The algorithm performance is more stable than the baseline correction step because the input data had already been cleaned during the baseline correction process. The outstanding performance was also evident by examining the multidimensional scatterplots and histogram overlays of the predicted autofluorescence over the known autofluorescence from the mock-transfected cells (Figure 35 and Appendix Figure 14). The algorithm learned from the separate pcDNA3 empty plasmid transfected sample and used to predict the mock transfection sample using only the transfection buffer and no plasmid DNA. Both the scatterplots and histograms show identical profiles tested by the Kolmogorov-Smirnov test with all p-values > 0.98. The results did not reject the null hypothesis of no significant difference and demonstrated that the predicted values and real values are from the same distribution, indicating the perfect accuracy produced by the autofluorescence prediction algorithm (Figure 35).

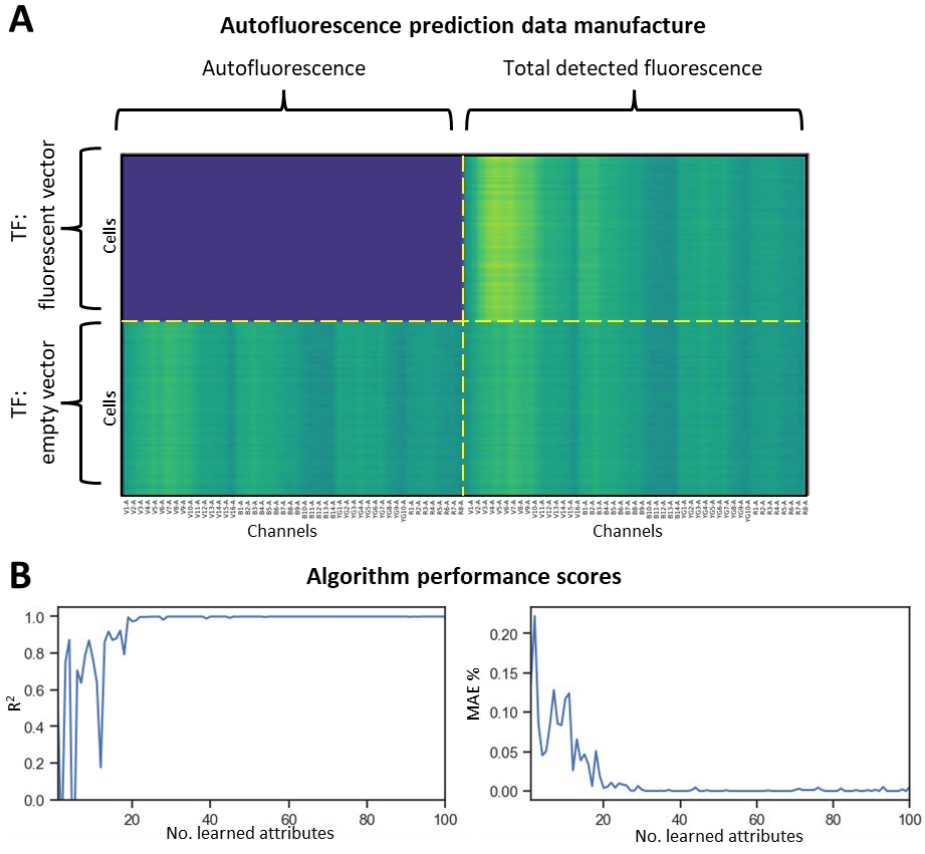


Figure 34 Performance of the autofluorescence prediction and removal algorithm
The data of the autofluorescence-only control sample and the test sample's data were concatenated into a new dataset (A). The performance of the autofluorescence prediction algorithm using the R^2 score and MAE% with numbers of attributes ranging from 1 to 100 (B).

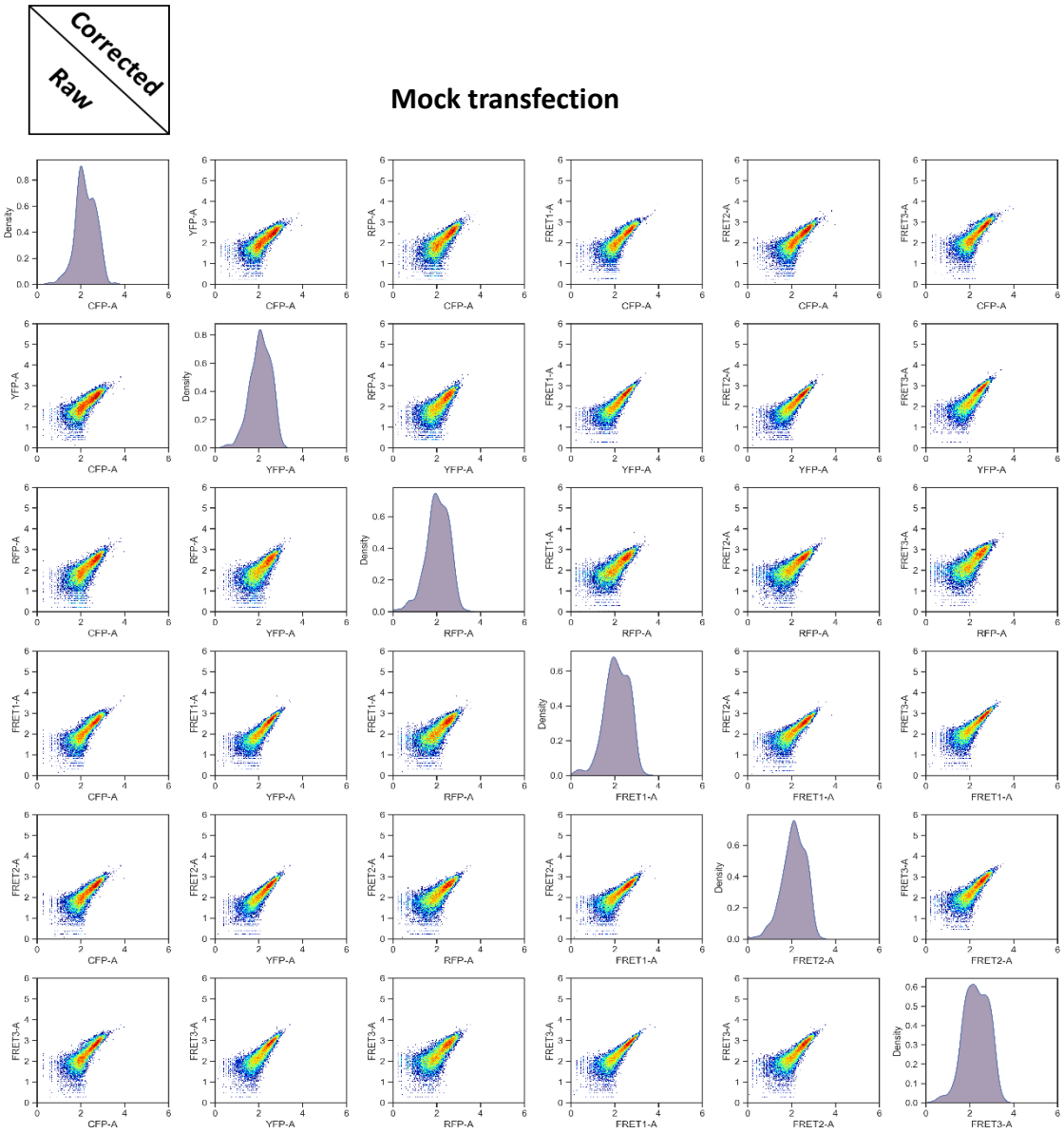


Figure 35 Autofluorescence prediction of the mock-transfected cells
Scatterplot and histogram overlays of the mock-transfected sample with the total fluorescence (red histograms) and autofluorescence (blue histograms) detection values. The scatter plots are symmetric along the diagonal histogram plots. Blue and red histograms are extremely identical.

4.2.4 The pre-processing expansion on the full-spectrum flow cytometer

Unlike the conventional flow cytometer which detects fluorescence emissions with the designated detection channels. The full spectrum detects normalised fluorescence signals using the entire detector array. For example, the 4L Cytek Aurora full-spectrum flow cytometer detects fluorescence signals using a total of 48 detection channels (see Figure 36 and Appendix Figure 1). For the mock transfection sample, it displayed the background autofluorescence profile across the entire 48 detection channels. For fluorescent protein-expressing cells, the primary detection channel can be chosen based on the fluorescent intensity histograms. In detail, the eCFP, eYFP and mRFP primary detection channels were the 405 nm violet laser-excited V5 channel, 488 nm blue lase-excited B3 channel, and the 561 nm laser-excited YG3 channel, respectively (Figure 36).

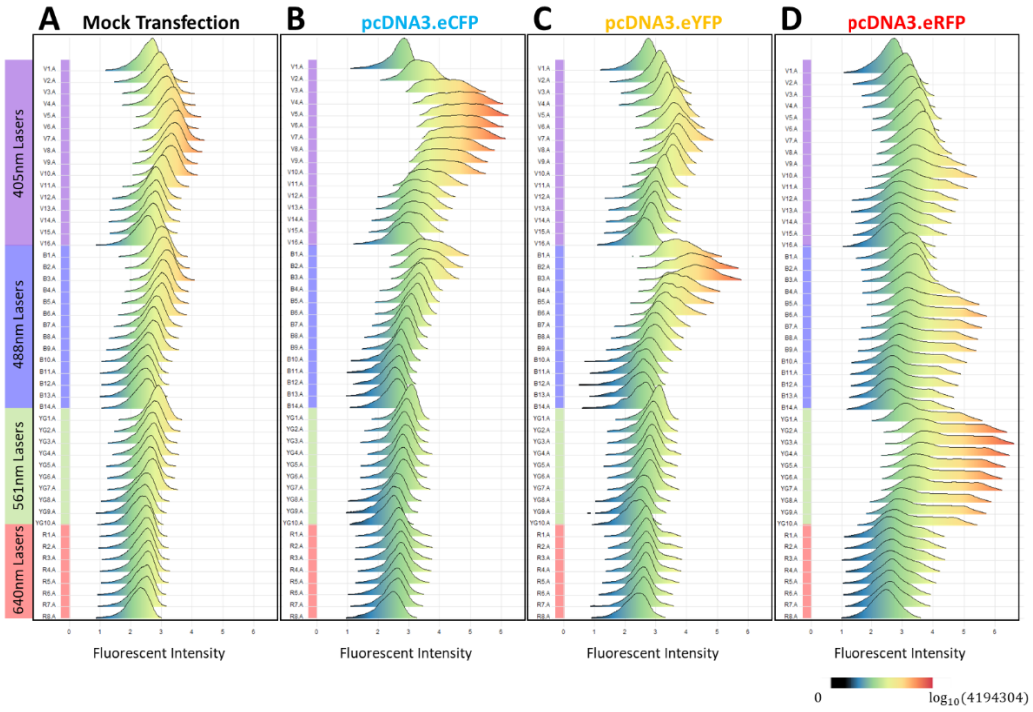


Figure 36 Fluorescence signals detected on the full-spectrum flow cytometer
A total of 48 detection channels recorded fluorescence signals for mock-transfected cells (A), eCFP expressing cells (B), eYFP expressing cells (C), and the mRFP expressing cells.

The process of Principal Component Analysis (PCA) in this study is primarily focused on the identification of optimal detection channels for fluorescent signals. Geometrically, PCA identifies the directions in the data that explain the maximal amount of variance. In this context, the first principal component (PC-1) is pivotal as it captures the most significant percentage of the variance, enabling the highest separation between fluorescence-positive and fluorescence-negative populations in fluorescent protein-expressing cells. This allows us to determine which detection channel contributes most significantly to the variance captured by PC-1, thereby identifying the primary detection channel with the best fluorescence positive-to-negative signal ratio.

However, it is crucial to differentiate between the role of PCA in detection channel identification and the process of FRET quantification. Once the optimal detection channels are identified (e.g., for FRET-donor, FRET-acceptor and FRET emission), the actual FRET quantification is performed using the direct fluorescent intensity measurements from these channels. This method ensures that the FRET quantification is based on actual detected intensities, rather than on compressed or transformed data. Thus, the PCA-based channel identification step does not introduce data compression in the quantification of FRET signals. It merely serves as a pre-processing step to enhance the accuracy and reliability of the subsequent FRET analysis by selecting the most appropriate channels for fluorescence detection.

The results demonstrated that the optimal primary detection channels for eCFP, eYFP and mRFP were V5 (508/20 nm), B3 (542/17 nm) and YG3 (615/20 nm) detection channels, respectively (Figure 37A). The results completely agreed with the decision made by investigating the fluorescent intensities manually. The FRET detection can be

automatically assigned to the same channels as the acceptor's filter range but on the donor-laser array. For example, the eCFP→eYFP (FRET1) detection channel can be assigned to V7 (542/17 nm) which is on the eCFP donor violet-laser detector array and with the same filter range as the eYFP acceptor using B3 (542/17 nm). Similarly, the eYFP→mRFP (FRET2) detection channel can be assigned to the blue-laser detection channel B6 (615/20 nm), and the eCFP→mRFP (FRET3) detection channel can be assigned to the violet detection channel V10 (615/20 nm), both were same as the mRFP detection channel YG3 (615/20 nm) off the yellow-green laser (Figure 37B).

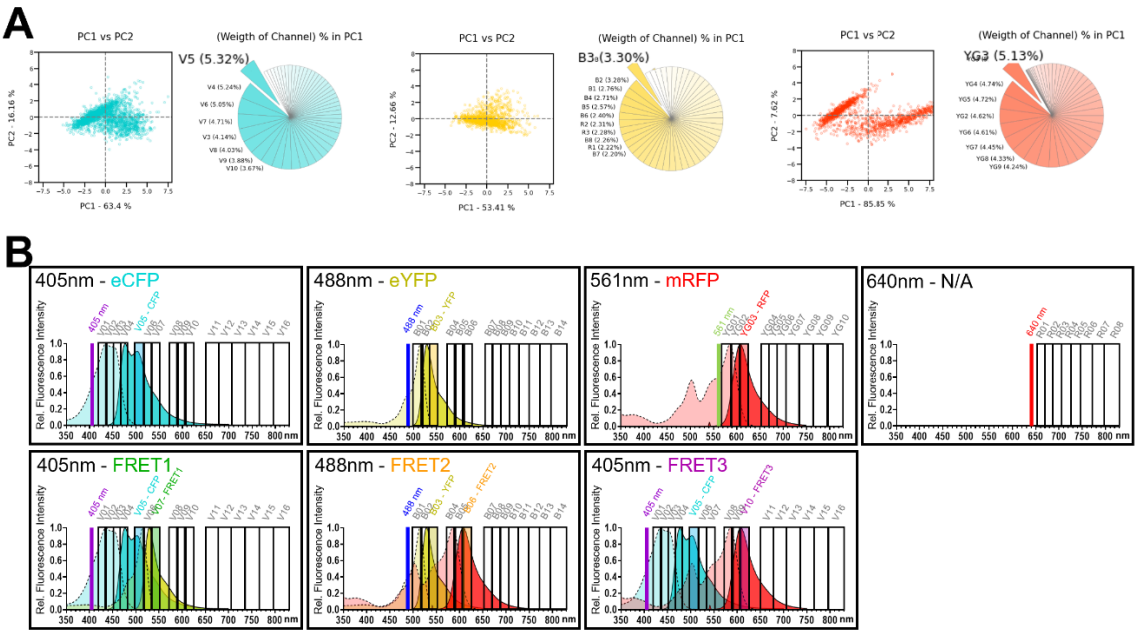


Figure 37 Automated optimal detection channel identification using PCA
PCA reveals the optimal detection channels as the V5, B3, and YG3 detection channels for eCFP, eYFP and mRFP expressing cells, respectively (A). The full-spectrum flow cytometer detection configuration for each of the fluorophores and the FRET signals, eCFP (V5), eYFP (B3), mRFP (YG3), eCFP→eYFP (FRET1, V7), eYFP→mRFP (FRET2, B6) and eCFP→mRFP (FRET3, V10) (B).

The full-spectrum microscopy MFI data were also used to validate the automated primary detection channel identification results generated using the full-spectrum flow

cytometry data (Figure 38). The detection was conducted using a 1 nm excitation step size ranging from 440 nm to 790 nm with a 5 nm emission detection step size ranging from 452/5 nm to 827/5 nm (Figure 38A). The optimal detection ranges for eCFP, eYFP and mRFP were 507/5 nm, 542/5 nm and 607/5 nm (Figure 38B). These results further supported the automated optimal detection channel identification results using PCA, which were 508/20 nm, 541/17 nm and 615/20 nm, overlapping with the full-spectrum microscopy results.

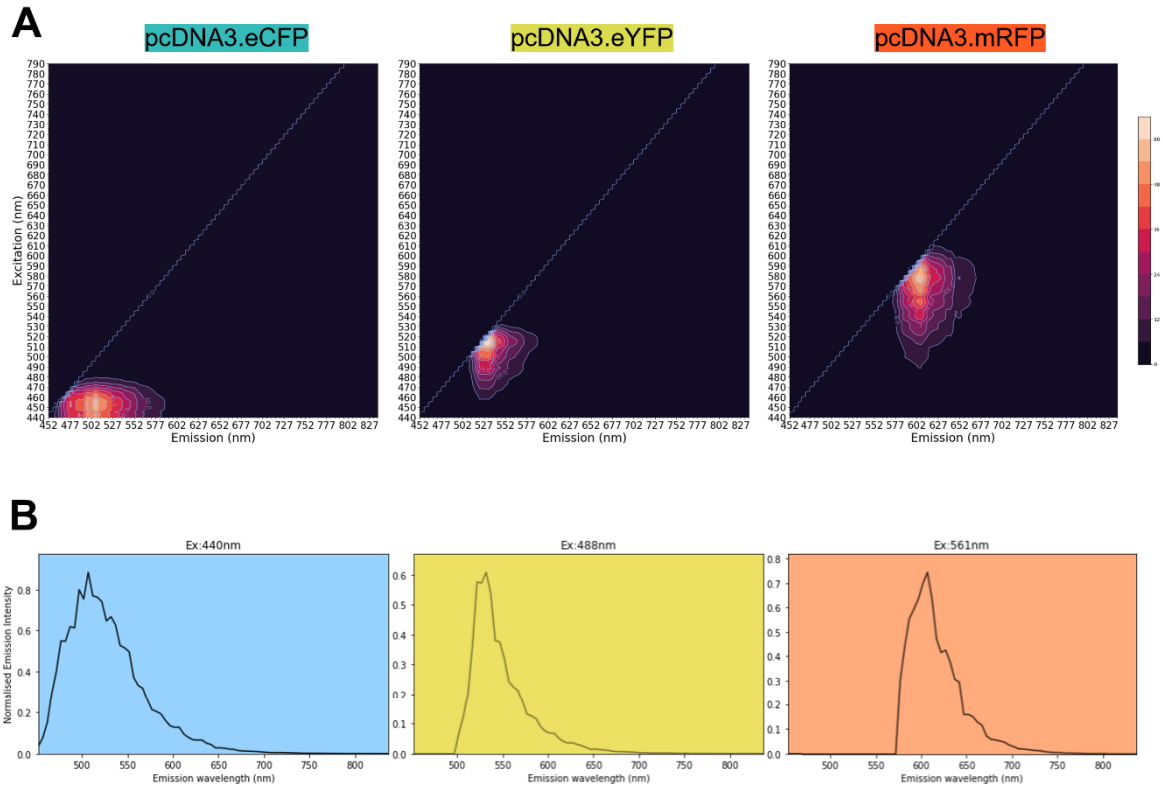


Figure 38 Full-spectrum microscopy for eCFP, eYFP and mRFP Ex/Em profiles
The contour plots for eCFP, eYFP and mRFP MFIs were calculated from the full-spectrum microscopy data (A). The excitation range was 440 nm to 790 nm with a 1 nm step size, and the emission range was 452/5 nm to 827/5 nm with a 5 nm step size. The emission profiles using fixed excitation, 440 nm for eCFP, 488 nm for eYFP and 561 nm for mRFP (B).

4.2.5 The performance of the spectral unmixing algorithm

First, this section examines the spillovers between the FRET donor and acceptor fluorophores, and it also demonstrates how the pre-processing steps mentioned above (i.e., baseline correction and autofluorescence removal) impacted the spectral unmixing results. For example, the spectral viewer illustrations display the spillovers between the donor and acceptor fluorophores for eCFP→eYFP FRET. On the one hand, upon 405 nm donor-laser excitation, eCFP is excited, and eYFP is also excited with suboptimal efficiency so that the eCFP primary emission can be detected by the eCFP filter along with some eYFP emission (Figure 39A red area). On the other hand, upon 488 nm acceptor-laser excitation, eYFP is excited, and eCFP is also excited with suboptimal efficiency so that the eYFP primary emission can be detected by the eYFP filter along with some spillovers from eCFP emission (Figure 39B red area).

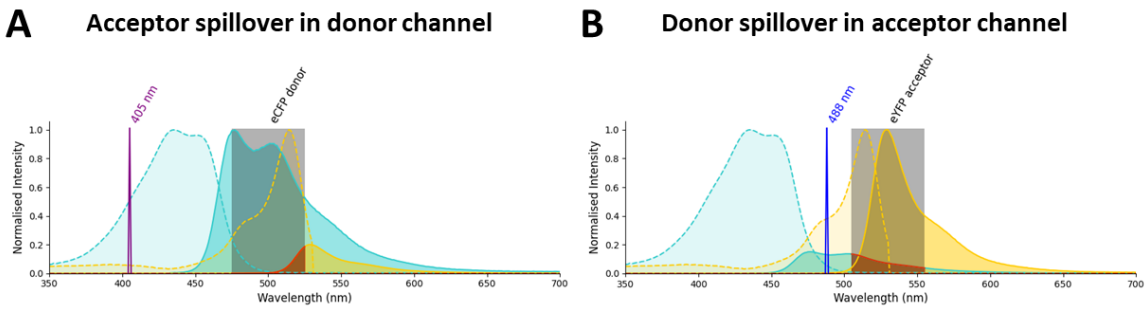


Figure 39 Spillover between FRET donor and acceptor fluorophores
The eYFP FRET acceptor can be suboptimally excited by the 405 nm donor laser, and the emission of 405 nm laser-excited eYFP can be detected in the donor channel as the spillover emission (A, red area). The eCFP FRET donor can be suboptimally excited by the 488 nm acceptor laser, and the emission of 488 nm laser-excited eCFP can be detected in the acceptor channel as the spillover emission (B, red area).

The 2-step collaborative filtering with the baseline correction and the autofluorescence removal steps has vastly enhanced the spectral unmixing result. Further, using the eCFP

and eYFP as an example, before spectral unmixing, the uncompensated emission of eCFP was detected with a MFI of 649.383 in the acceptor channel, and the eYFP was detected with an MFI of 617.518 in the donor channel (Figure 40A, dark grey bar). Here, the same spectral unmixing algorithm described in this study was used (see section 2.7.5) to compare the unmixed results for samples with different autofluorescence handling methods. The compared samples were (i) cells without autofluorescence removal, (ii) cells that used mean autofluorescence subtraction for autofluorescence removal, (iii) cells treated autofluorescence as an additional colour during autofluorescence removal, and (iv) cells cleaned with the collaborative filtering methods for autofluorescence removal before spectral unmixing. For cells without autofluorescence removal, spectral unmixing resolved fluorescence spillovers and demonstrated three-fold decreases in eCFP donor and eYFP acceptor spillovers with spreading errors (SE) = 426.868 for eCFP and SE = 581.453 for eYFP (Figure 40B). The SEs were the same for using the mean subtraction method since the data was moved downward by the mean value along the y-axis for each fluorophore. The residual spillovers decreased more than five times for the eCFP detected in the eYFP channel compared to the sample without autofluorescence removal. However, the eYFP detection demonstrated an overcompensation effect with the negative mean residual spillover value of eYFP detected in the eCFP channel (Figure 40C). Surprisingly, when performing spectral unmixing on samples that treated autofluorescence as a single colour for autofluorescence removal, the performance was worse than the other methods. Although the unmixed results demonstrated no positive spillover, they demonstrated around ten times worse overcompensations than the mean subtraction method for both eCFP and eYFP. In addition, the SE was also increased by around 2.5 times for eCFP and

around the same value for eYFP when compared to no autofluorescence removal and mean autofluorescence subtracted samples (Figure 40D). In contrast, the 2-step collaborative filtering achieved approximately zero spillovers for both the eCFP and eYFP. Moreover, it achieved around two thousand times reduced SE compared to the sample before spectral unmixing (Figure 40E).

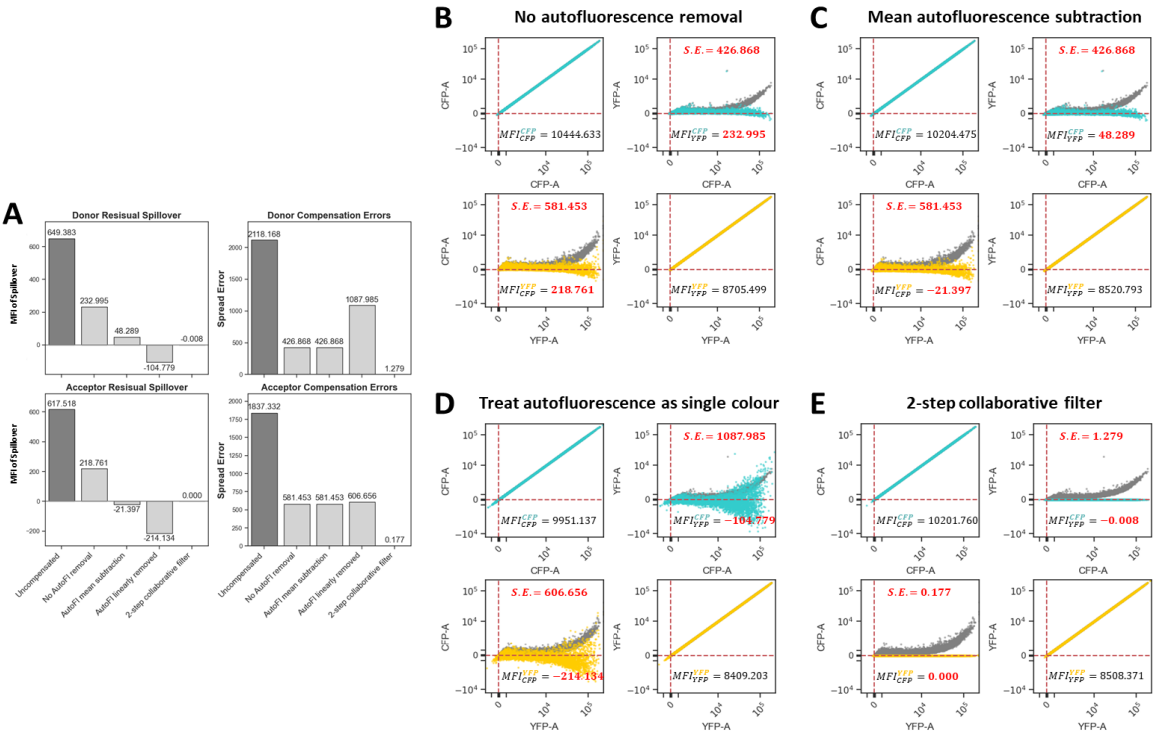


Figure 40 Spectral unmixing after autofluorescence removal with different methods
The summarised spillover and spread error results for spectral unmixing (A). The eCFP before (grey) and after (cyan) spectral unmixing and the eYFP before (grey) and after (yellow) spectral unmixing with different autofluorescence removal methods. Cells without autofluorescence removal (B). Data using the mean autofluorescence subtraction method (C). Data treated autofluorescence as a single colour and removed it during spectral unmixing (D). Data used the 2-step collaborative filtering method described in this study (D). Zeros are indicated as red dashed lines. SE and spillover MFIs are indicated in each graph.

Furthermore, when FRET is happening, the sensitised eCFP→eYFP FRET emission can be detected upon donor laser excitation (Figure 41 green histogram), which is the FRET emission is the same as the eYFP acceptor but with different intensities. Even using a 10

nm narrow range filter for FRET detection, the eCFP-donor and eYFP-acceptor will have high spillovers into the FRET detection channel due to the extensive spectrum overlaps between donor and acceptor fluorophores which are required for FRET to happen.

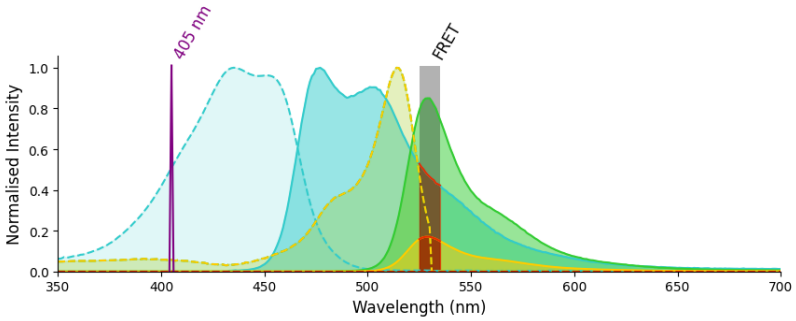


Figure 41 Donor and acceptor spillovers in the FRET channel
When FRET happens, the sensitised FRET emission (green histogram) can be detected through the FRET filter (black box). Even using a narrow bandpass filter, the eCFP donor spillover emission (red area under the cyan histogram) can be detected in the FRET channel. The eYFP acceptor spillover emission (red area under the yellow histogram) can also be detected in the FRET channel.

The FRET spectral unmixing was achieved using mathematical formulations to represent the detected fluorescent signals in each detector as the combined fluorescence from donor, acceptor and sensitised FRET emissions (see section 2.7.5). Here, the background-corrected and autofluorescence-removed data were utilised using eCFP→eYFP FRET as an example with the detection configuration shown (Figure 39A). After the two-step collaborative filtering pre-processing, the fluorescence correlations detected between each pair of the donor, acceptor and FRET channels increased near to one (Figure 42B). The linearities of each fluorophore detected among the three channels were vastly increased, and the spillover coefficients were calculated using the RANSAC linear fitting algorithm (Figure 42C and Appendix Figure 2). The spectral unmixing matrix was constructed for efficient single-cell level spectral unmixing (Figure 42D). The spectrally unmixed results indicated that the spillovers from eCFP and eYFP in

the FRET channel had been eliminated (Figure 42E, compare original data: grey dots, pre-processed data: black dots, and unmixed data: cyan and yellow dots).

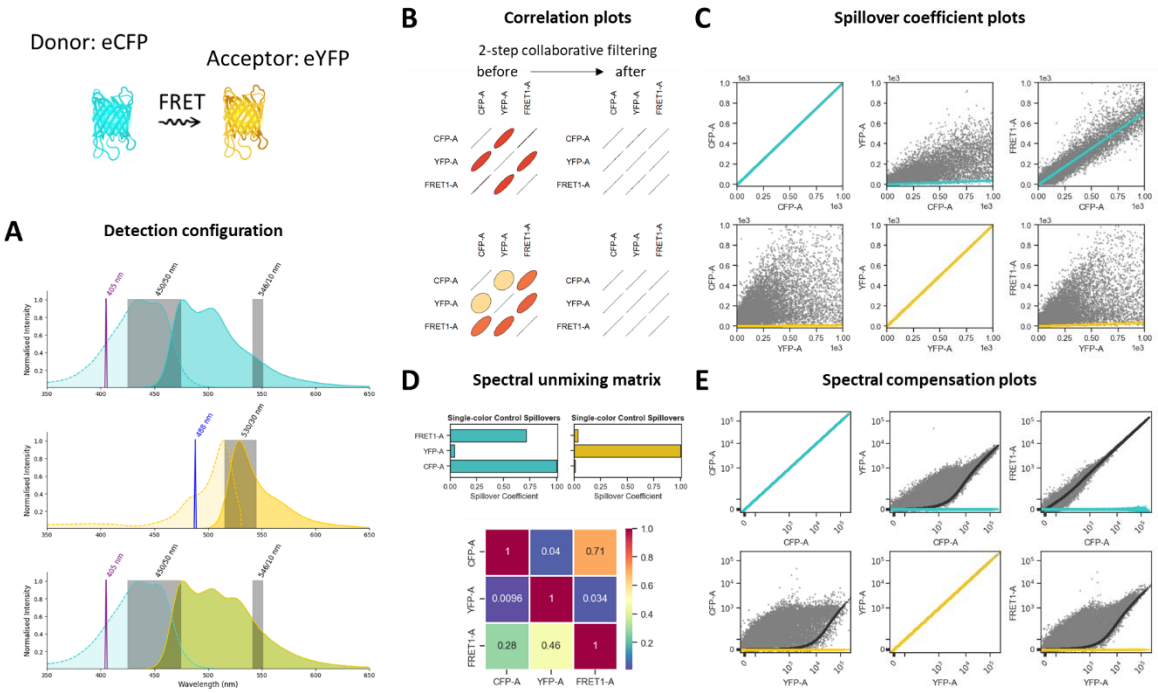


Figure 42 Spectral unmixing for eCFP→eYFP FRET
The spectral viewer diagrams and the detection configurations for eCFP donor, eYFP acceptor and sensitised FRET emissions (A). The correlation between each pair of the eCFP, eYFP, and the FRET channels before and after 2-step CF pre-processing (B). Spillover coefficient plots between each pair of the three detection channels (C), the data before (grey) and after (cyan and yellow) 2-step CF prep-processing are both displayed. The spectral unmixing matrix was generated using the spillover coefficients (D). The fluorescence detection scatter plots with uncompensated data before (grey) and (after) 2-step CF pre-processing and the spectrally unmixed data (yellow and cyan) after the 2-step CF pre-processing (E).

In this study, a total of 11 examinations were performed using five different FRET pairs and five different flow cytometers to evaluate the performance of the overall pre-processing and spectral unmixing procedures. In particular, I have used TNFR1-FP fusion constructs with eCFP→eYFP, eYFP→mRFP and eCFP→mRFP FRET pairs on LSR-II, Fortessa X20, and Aurora flow cytometers (total of nine experiments). I have also used eGFP→mRFP soluble FRET pairs on the FACS Calibur flow cytometer and violet blue→FITC conjugated antibody as FRET pairs on the MACSQuant X flow cytometer

(Figure 43A). Light and dark grey bars show the results before and after the overall pre-processing (Figure 43B and C). The performances were indicated by the channel-to-channel correlation and spillover linearity, which were all close to one (Figure 43B). In detail, the fluorescence correlations between donor and acceptor channels were improved from mean = 0.493 (SD = 0.353) to mean = 0.997 (SD = 0.010), the fluorescence correlations between donor and FRET channels were improved from the mean = 0.726 (SD = 0.281) to the mean = 1.000 (SD = 0.0002), and the fluorescence correlations between acceptor and FRET channels were improved from the mean = 0.692 (SD = 0.354) to the mean = 0.997 (SD = 0.010). The linearities were calculated using the R^2 scores. For FRET-donor fluorophore, the R^2 scores between donor and acceptor channels were improved from the mean = 0.465 (SD = 0.452) to the mean = 0.987 (SD = 0.028), the R^2 scores between donor and FRET channels were improved from the mean = 0.925 (SD = 0.114) to the mean = 1.000 (SD = 0.000007). For FRET-acceptor fluorophore, the R^2 scores between acceptor and donor channels were improved from the mean = 0.244 (SD = 0.280) to the mean = 1.000 (SD = 0.0007), the R^2 scores between acceptor and FRET channels were improved from the mean = 0.722 (SD = 0.385) to the mean = 1.000 (SD = 0.00002).

Furthermore, the errors were represented by model fitting errors and the spread errors after spectral unmixing, which are nearly zero (Figure 43C). The donor spillovers detected in acceptor and FRET channels were eliminated from the MFI = 148.010 (SD = 64.281) and the MFI = 168.192 (SD = 90.744) to the MFI = 1.278 (SD = 1.871) and the MFI = 0.412 (SD = 0.500) respectively. The acceptor spillovers detected in donor and FRET channels were eliminated from the MFI = 203.254 (SD = 197.167) and the MFI =

284.622 (SD = 365.516) to the MFI = 4.411 (SD = 9.485) and the MFI = 8.016 (SD = 19.952) respectively.

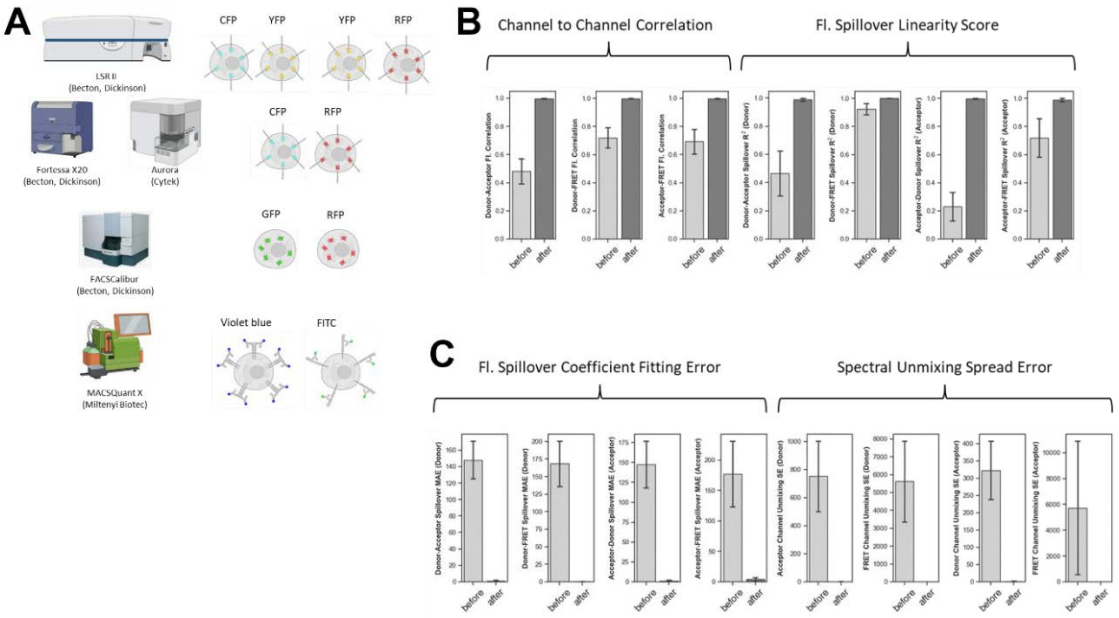


Figure 43 The performance evaluation of the entire pre-processing stage
The pre-processing stage includes singlet identification, baseline subtraction-error correction, autofluorescence prediction and removal, and spectral unmixing. The performance data were generated using a total of 11 experiments conducted on five different flow cytometers using five different FRET pairs (A). Channel-to-channel correlation and spillover coefficient linearity scores were reported for the performance evaluation (B). The spillover linear model fitting errors and the spectral unmixing spread errors were used to report errors (C).

The residual SEs of donor fluorophores calculated from the acceptor and FRET channels were decreased from the mean = 752.044 (SD = 706.322) and the mean = 5615.603 (SD = 6368.435) to the mean = 0.950 (SD = 1.390) and the mean = 0.620 (SD = 0.664) respectively. Lastly, the residual SE of acceptor fluorophores calculated from the donor and FRET channels decreased from the mean= 435.896 (SD = 416.894) and the mean = 7883.443 (SD = 15128.669) to the mean = 3.855 (SD = 8.508) and the mean = 6.612 (SD = 17.111) respectively. This level of accuracy and data cleansing has never been achieved previously, and it will allow investigators to perform the absolute data quantification of any flow cytometry analysis, including FRET experiments.

CHAPTER FIVE

TWO-PROTEIN THREE-COLOUR FLOW CYTOMETRY-BASED FRET INVESTIGATIONS

5.1 Introduction

Protein conformation and association state regulate biological processes, and the distance between two interacting protein molecules can be measured by FRET [28]. FRET can be qualified or semi-quantified with the physically altered emission spectral properties of the fluorescent tags [164, 165]. Upon donor-laser excitation, the quenched donor fluorescent intensity and sensitised FRET emission become evident (Figure 44). In order to fully quantify the FRET physical phenomenon, state of the art methods measure changes in fluorescence lifetime [166-168] or polarisation of emitted light [169-172], which require complex mathematical modelling and expensive specialised equipment. Another popular approach for FRET quantification uses the recovery of quenched donor fluorescent intensity upon photobleaching of the acceptor [173, 174]. It cannot be employed in flow cytometry because detecting donor fluorescence intensity before and after acceptor photobleaching is usually implausible due to the fast sample flow rate (millisecond per cell event). Therefore, simple fluorescent intensity-based methods are mostly used because of their fast, simple, non-destructive nature, with a standard hardware requirement referred to as "three-filter FRET" [114, 170, 175, 176]. The three filters measure (i) the donor-laser excited FRET-donor emission, (ii) the acceptor-laser excited FRET-acceptor emission, (iii) and the donor-laser excited FRET-donor to FRET-acceptor sensitised emission (Figure 44 grey boxes). However, the three-filter FRET emissions are not only sensitive to the protein-protein interaction levels but also impacted by the expression levels of interacting protein molecules [176].

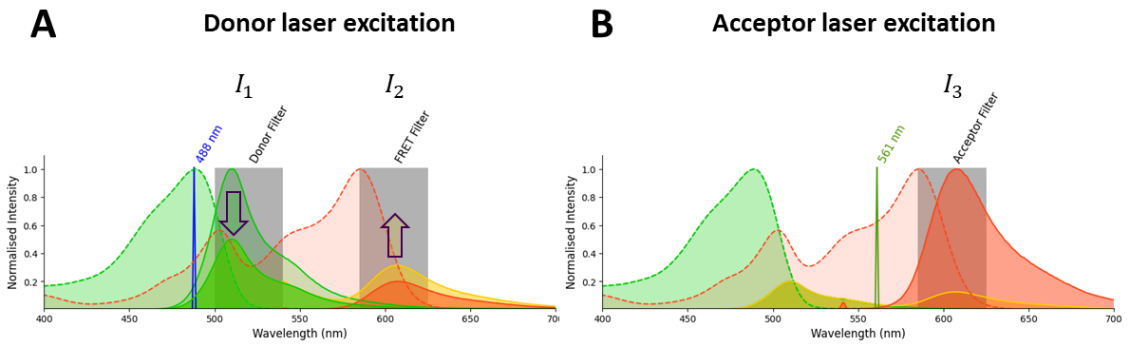


Figure 44 The fluorescence dynamics during FRET
An example of eGFP→mRFP FRET using the 488 nm donor laser and the 561 nm acceptor laser. When FRET happens, eGFP donor emission quenches in I_1 channel (the green histogram), and sensitised FRET emission increases in I_2 channel (the yellow histogram) (A). The donor emission is detected in the I_3 channel (B). Each of the I_1, I_2, and I_3 detection channel also picks up spillover emissions.

FRET efficiency is defined as the proportion of excited donor molecules undergoing energy transfer to the acceptor molecule [2]. The rate of FRET efficiency depends on the sixth power of the separation distance between the donor and the acceptor, thus providing a sensitive tool for measuring protein interactions within the 10-nm range. FRET efficiency is not impacted by the expression level of the protein molecules and can be derived from the fluorescent intensity, thus serving as the absolute quantitative solution for the detection of PPIs. Many approaches exist that use flow cytometry [118, 177] or microscopy [178, 179] to determine FRET efficiencies based on detecting fluorescent intensities. The fluorescent intensity-based FRET efficiency calculation requires a calibration factor, also known as the α factor [118, 129, 178] or the G factor [19, 179]. The α factor relates to the ratio of donor fluorescent intensity quenching to the sensitised acceptor fluorescent intensity gain during energy transfer. Because flow cytometry is statistically superior to fluorescence microscopy [180], it is an attractive method for FRET measurements using this approach. Once the α factor is determined, sensitized acceptor emission intensity can be converted to FRET efficiency (E).

Because protein expression of donor and acceptor varies among cells, the α factor requires the knowledge of the donor and acceptor expression ratios. The current methods are accomplished by varying the length and composition of the linker residues connecting the FRET donor and FRET acceptor. This way, the FRET-donor and FRET-acceptor always have a concentration ratio of one. However, implementing this method to non-linked, naturally free-interacting proteins is inaccurate. To achieve the calibration of unlinked FRET samples, I have established a single-cell level FRET efficiency algorithm together with the single-cell level α factor formulation that does not require chemically linked FRET-donor and -acceptor pairs. This chapter describes the flow cytometry single-cell level two-protein three-colour FRET efficiency algorithm with novel calibration components that can be utilised for both chemically linked and unlinked protein pairs. The algorithm's reliability has been tested with three sets of free-interacting human TNFR1-FPs FRET pairs on both a BD LSR-II and a BD Fortessa X20 flow cytometers with varied laser powers, filters, and photomultiplier tube (PMT) detector voltages. The current flow cytometry-based detection of FRET from TNFR1 homodimer interaction have already proven powerful in pharmaceutical therapeutics screening [181, 182]. The single-cell total quantification algorithm presented in this chapter will provide a detection platform with higher sensitivity and robustness. In addition, the performance of the two-protein three-colour FRET quantification algorithm will be demonstrated with two applications, which the data were collected from the online FlowRepository database: (i) small molecule Auxin detection in the plant *Arabidopsis thaliana* (Repository ID: FR-FCM-Z3FL) [34] and (ii) immune activation detection in human T-cells with TCR-CD4 interactions (Repository ID: FR-FCM-ZZR6) [126].

5.2 Results

5.2.1 Fully quantitative FRET efficiency algorithm with unlinked FRET calibration

Spectral unmixing using pre-processed flow cytometry data with formulated donor, acceptor, and FRET emission spillovers allows the absolute quantification of FRET efficiency, which reflects the PPI proximity. The FRET efficiency was computed using the standard linked FRET pairs with eGFP and mRFP as the FRET donor and acceptor. The pre-processed data showed significantly increased channel-to-channel fluorescence spillover correlations (Figure 45).

For eGFP→mRFP FRET, the spillover correlation increased from 0.037 to 0.994 between donor-to-acceptor channels, increased from 0.998 to 1.000 between donor-to-FRET channels, and increased from 0.055 to 0.994 between acceptor-to-FRET channels. All correlations were close to one, representing the actual physical properties of single-colour fluorescence emissions from either eGFP donor or mRFP acceptor with any impact from baseline subtraction-errors and autofluorescence completely eliminated. The clean fluorescence signals guarantee the accurate downstream formulation of the donor, acceptor, and the detection of FRET using these pure single-colour fluorescence signals and ensure accurate spectral unmixing outputs with minimal spread errors.

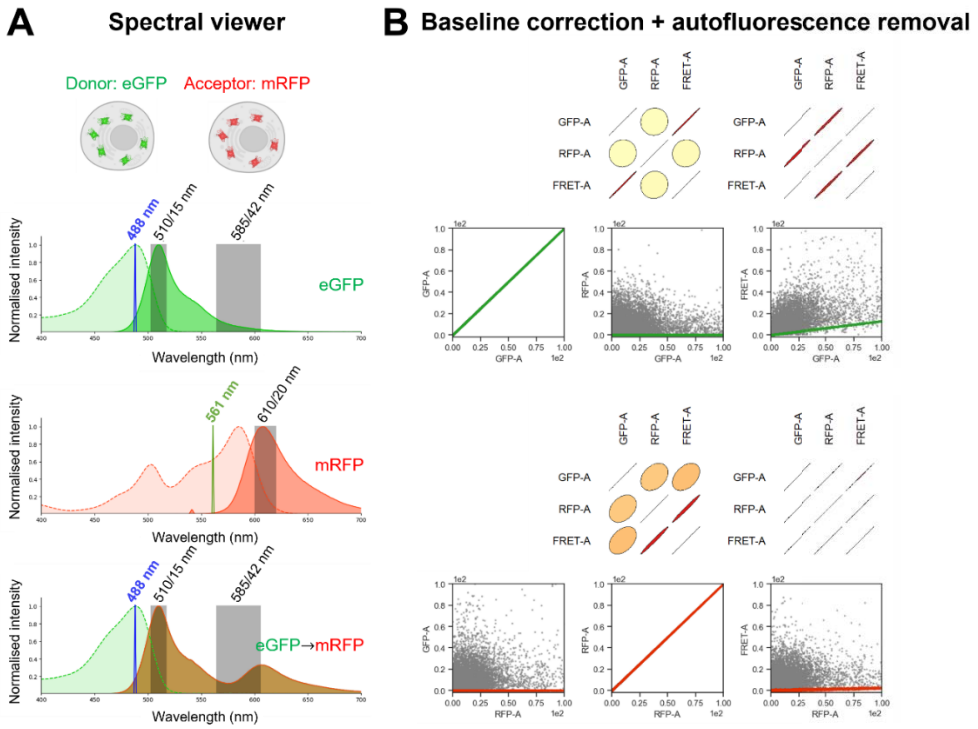


Figure 45 Pre-processing for eGFP→mRFP linked FRET pair

The spectral viewer diagrams of the detection configurations (A). The changes in channel-to-channel correlations and the spillover linearities after baseline correction and the autofluorescence removal (B). The strength of the correlation are displayed from weak to strong as from circles to linear lines, as well as from yellow to red. In each of the dot plots, the grey dots are data before pre-processing, with eGFP and mRFP single-colour controls at top and bottom, respectively. The coloured dots are data after pre-processing overlaid on top of the grey dots, where green colour dots represent the eGFP FRET donor, and the red colour dots represent the mRFP FRET acceptor.

Using the spillover coefficients, a formulated spectral unmixing matrix achieved single-cell level detection of pure single-fluorescence-components within each designated detection channel for each FRET donor, acceptor and sensitised FRET emissions (Figure 46A). After spectral unmixing, the spread error of the eGFP donor detected in the acceptor channel is 0.001, and in the FRET channel is 0.006, the spread error of mRFP FRET-acceptor detected in the donor channel is 0.118, and in the FRET channel is 0.443 (Figure 46B). Those near-zero spread errors guarantee the ultrasensitive detection of the sensitised FRET emissions and ensure accurate FRET efficiency quantifications. By

comparison, the previously published FRET efficiency calculations displayed huge variances and negative FRET efficiency values [16]. The limitations from the previously published data indicated the lack of accuracy because the linked system should have uniform donor-acceptor distances and FRET efficiencies. The negative FRET efficiency values from the original publication have no physical or biological meaning (Figure 46C). In contrast, the method presented in this chapter achieved all positive FRET efficiencies with more than 1000 fold decrease in SD and standard errors around the mean (SEM) (Table 2). Each linked standard displayed uniformed FRET efficiency reflecting the fixed amino acid linker length for each of the standard eGFP-AAs-mRFP FRET pairs (Figure 46).

Table 3 FRET efficiencies and alpha factors for eGFP→mRFP linked FRET pairs

SAMPLE	METHOD	FRET EFFICIENCY (%)			ALPHA FACTOR		
		MEAN	STD	SEM	MEAN	STD	SEM
GFP-7AA-RFP	Nedbal et. al.	37.1	10	0.2	0.1020	N.A.	N.A.
	Adv. Method 2	33.607280	0.012273	0.000203	0.1492	7.6566e-15	4.5684e-19
GFP-19AA-RFP	Nedbal et. al.	31.6	11	0.2	0.1020	N.A.	N.A.
	Adv. Method 2	27.059180	0.000444	0.000008	0.1492	9.8489e-15	5.9612e-19
GFP-32AA-RFP	Nedbal et. al.	27.4	12	0.2	0.1020	N.A.	N.A.
	Adv. Method 2	24.143029	0.302862	0.004800	0.1492	4.4916e-15	4.5740e-19

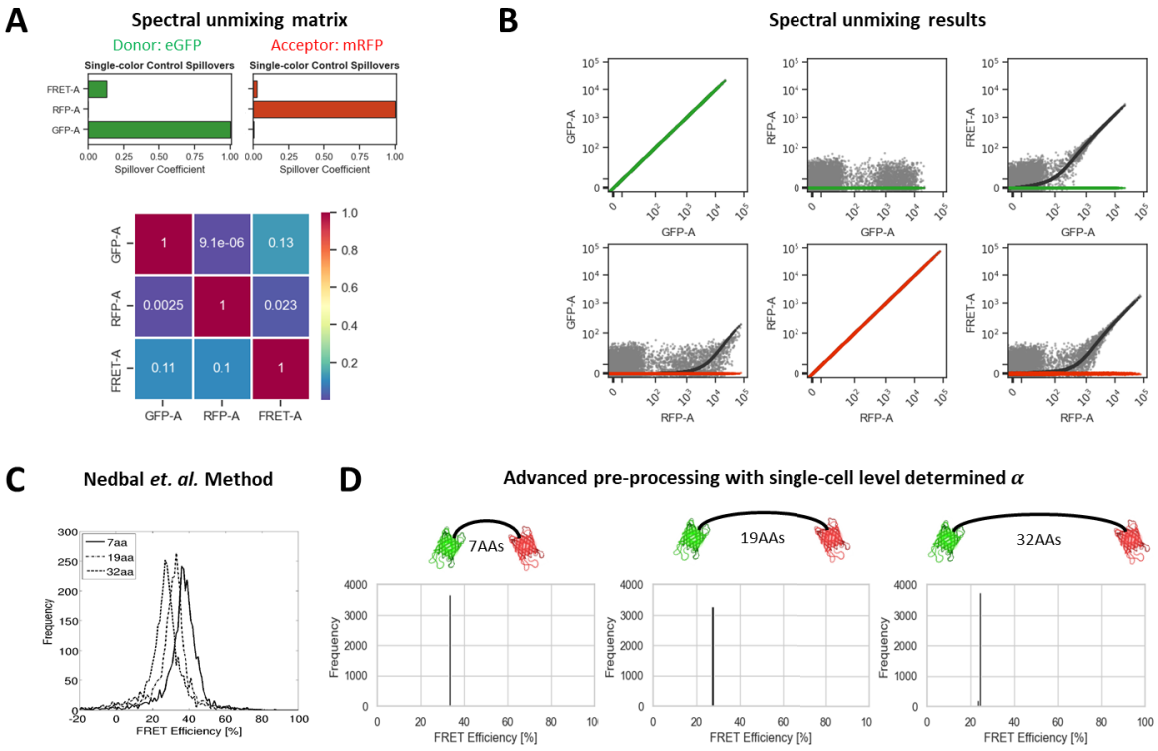


Figure 46 FRET efficiency quantification for linked eGFP→mRFP standards
The spillover bar plot and spectral unmixing matrix for eGFP FRET donor and mRFP FRET acceptor (A). The spectrally unmixing signals for eGFP and mRFP single-colour controls (B). The previously published FRET efficiency data [16] (C). Uniformed FRET efficiencies were achieved using the advanced flow cytometry-based FRET algorithms developed in this study for each linked FRET pair with fixed length (7AAs, 19AAs, and 32AAs) (D).

5.2.2 Reliability test using three FRET pairs on two different flow cytometers

The two-protein, three-colour FRET spectral unmixing formulation and FRET efficiency algorithm also showed excellent reliability when detecting PPIs for unlinked naturally existing free-interacting proteins. Homo-dimeric human TNFR1 interactions were detected by subcloning TNFR1 as fluorescent protein fusion constructs and expressing them in HEK293T cells. In total, three TNFR1-FP FRET pairs were utilised, including TNFR1-eCFP→TNFR1-eYFP, TNFR1-eYFP→TNFR1-mRFP, and TNFR1-eCFP→TNFR1-mRFP (Figure 47).

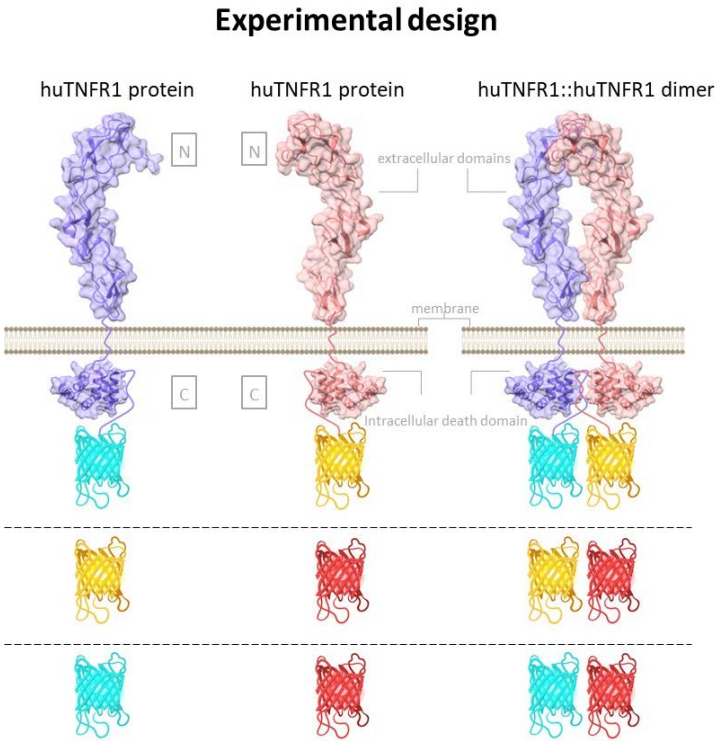


Figure 47 The detection of three pairs of TNFR1-FPs fusion proteins for unlinked FRET
The detection of unlinked FRET signals used human TNFR1 fluorescent protein fusion constructs with three different pairs of the FRET. The three FRET pairs include TNFR1-eCFP→TNFR1-eYFP, TNFR1-eYFP→TNFR1-mRFP, and TNFR1-eCFP→TNFR1-mRFP.

Each FRET pairs were separately detected on two distinctly configured flow cytometers, the LSR-II and the Fortessa X20 (Figure 48 and see section 2.4.2). In particular, the LSR-II had a 355 nm 20 mW UV laser trigon detector array, a 405 nm 25 mW violet laser trigon detector array, and a 635 nm 20 mW red laser trigon detector array, as well as a 488 nm 20 mW blue laser octagon detector array. While the Fortessa X-20 had three octagon laser detector arrays with a 405 nm 50 mW violet laser, a 488 nm 100 mW blue laser, and a 640 nm 40 mW red laser, but no trigon detector array.

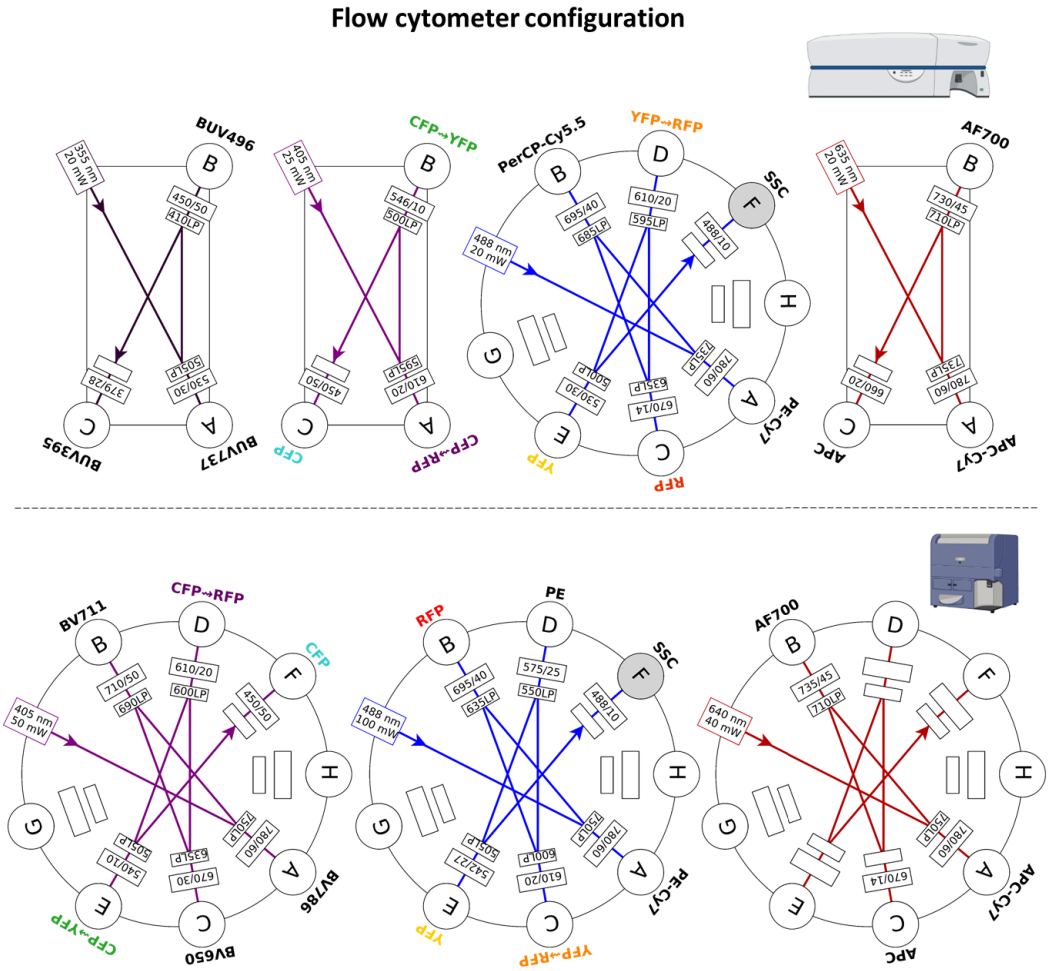


Figure 48 LSR-II and Fortessa X20 flow cytometer configurations
LSR-II with three trigon detector arrays (UV, violet, and red) and one octagon detector array (blue) (top). Fortessa X20 with three octagon detector arrays (violet, blue and red) (bottom).

The FRET efficiencies of each FRET pair were calculated using two approaches. One used the linked calibration method with sample means, i.e., the linear relation between quenched donor fluorescent intensity and increased sensitised FRET fluorescent intensity to conduct population mean-based calibration using the α factors. The other one used newly established single-cell level FRET efficiency calculation in this chapter (see section 2.7.6). Moreover, the linked calibration method requires FRET-donor to FRET-acceptor concentration ratios C_D/C_A is fixed, so it assumes $C_D = C_A$. The single-

cell level FRET efficiency calculation developed in this chapter has no assumption on the FRET-donor and FRET-acceptor concentrations. Both approaches used the same pre-processing steps to achieve spectrally unmixed fluorescence readouts for a fair comparison. To find the FRET baseline of random PPIs, the human CD27-FPs fluorescently tagged receptor was used as the FRET acceptor, with the human TNFR1-FPs as the FRET donor. CD27 is a structurally similar receptor protein to TNFR1, and they are from the same human TNFRSF, but CD27 does not interact with TNFR1. By comparison, the new method achieved significantly reduced baseline signals with mean baseline values decreased from 0.442 (SD = 0.205) to 0.132 (SD = 0.085) on LSR-II (p<0.001) and from 0.435 (SD = 0.216) to 0.052 (SD = 0.043) on Fortessa X20 (p<0.001) for TNFR1-eCFP→CD27-eYFP (Figure 49A and B). Similar results were also observed with TNFR1-eYFP → CD27-mRFP and TNFR1-eCFP → CD27-mRFP. For TNFR1-eYFP → CD27-mRFP, the mean baseline FRET efficiencies dropped from 0.410 (SD = 0.165) to 0.102 (SD = 0.043) on LSR-II (p<0.001) and from 0.561 (SD = 0.143) to 0.091 (SD = 0.058) on Fortessa X20 (p <0.001). For TNFR1-eCFP → CD27-mRFP, the mean baseline FRET efficiencies dropped from 0.237 (SD = 0.127) to 0.067 (SD = 0.025) on LSR-II (p<0.001) and from 0.258 (SD = 0.133) to 0.038 (SD = 0.011) on Fortessa X20 (p<0.001) (Figure 49A and B).

For the detection of TNFR1 interaction FRET-positive signals, the results demonstrated improved consistency between the FRET efficiencies calculated from two flow cytometers' data. In detail, comparing the improvement from the linked population-wise calibration method to the single-cell level calibration method, the TNFR1-eCFP→TNFR1-eYFP efficiency calculations had similar differences of 0.008 and 0.015

between the detected signals from the LSR-II and the Fortessa X20 flow cytometers with total variance declined from 0.044 to 0.018 (see Figure 49 green histograms and bars). Meanwhile, for the detection of TNFR1-eYFP → TNFR1-mRFP FRET, the difference decreased from 0.065 to 0.015, with the total variance dropping from 0.011 to 0.005 (see Figure 49 orange histograms bars). Similar results were also observed from TNFR1-eCFP→TNFR1-mRFP FRET efficiency calculations, in which the difference decreased from 0.083 to 0.025 with the total variance dropped from 0.029 to 0.006 (see Figure 49 purple histograms and bars).

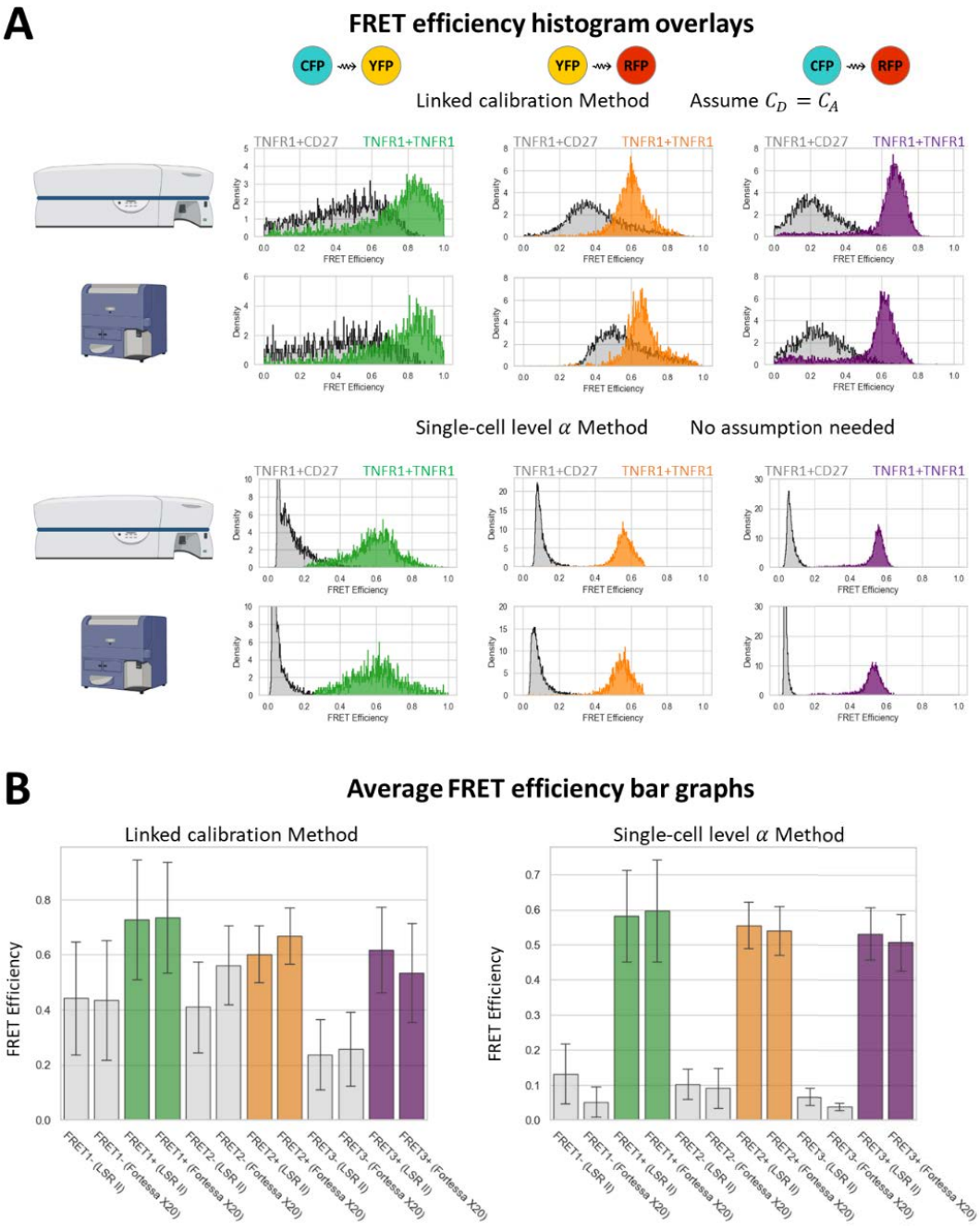


Figure 49 FRET algorithm reliability test using three FRET pairs on two flow cytometers
Histogram overlays of the single-cell level distributions of each FRET pair TNFR1-eCFP→TNFR1-eYFP (green), TNFR1-eYFP→TNFR1-mRFP (orange), and TNFR1-eCFP→TNFR1-mRFP (purple) tested on LSR-II and Fortessa X20 (A). The FRET efficiencies were calculated using the linked population mean approach for α factor calibration (top) and the single-cell level α factor calibration method developed in this study for unlinked FRET pairs (bottom). Negative controls (grey) used CD27 as the FRET acceptor proteins. The mean FRET efficiencies were calculated for statistical comparisons and displayed as bar graphs, and the error bars are standard deviations of the FRET efficiencies (B).

5.2.3 An application demonstration of small molecule (Auxin) detection in plants

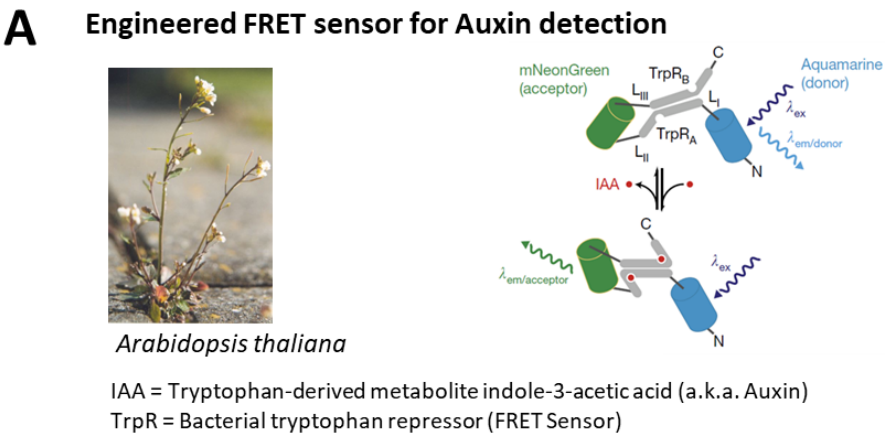
The two-protein three-colour FRET algorithm introduced in this study also demonstrated more robust and sensitive small molecule detection capability. Here, a bioengineered FRET sensor detected the small molecule tryptophan-derived metabolite indole-3-acetic acid (IAA, also known as Auxin). IAA/Auxin is vital in plant bioprocessing and gravity sensing, and tracking the IAA/Auxin concentration in plant samples is crucial to study plant physiology [183]. The data was collected from online FlowRepository database (Repository ID: FR-FCM-Z3FL) [34], where the FRET sensor was engineered using two linked tryptophan receptors with the FRET donor Aquamarine and FRET acceptor mNeoGreen also linked to the tryptophan receptors (Figure 50A). Together, the overall linked FRET sensor forms a unique fusion protein in which the tryptophan receptors undergo conformational changes upon IAA/Auxin binding and bring the Aquamarine → mNeoGreen FRET pairs in proximity. In other words, the higher the concentration of IAA/Auxin in the sample, the higher the average FRET emission can be detected for each individual cell until binding saturation is reached. This time, the newly developed FRET efficiency-based two-protein three-colour FRET algorithm was compared to the FRET quantification method introduced in the original publication. The original method utilised the FRET emission detection ratio of the detected sensitised FRET emission in the FRET channel over the quenched donor emission detected in the donor channel [34].

The originally published approach has limited detection sensitivity. It uses raw detection readouts which did not resolve the fluorescence spillover. Another factor that caused the sensitivity limitation was the population-wise calculation of fluorescence ratios. It

lacks the single-cell level resolution, which could not perform quality assessments using single-cell level distribution. Moreover, the fluorescence ratio approach is only semi-quantifiable and can only be used to predict the small molecule concentration from the same experiment. When conducting small molecule concentration prediction from separately acquired data, the fluorescence ratio will be highly impacted by the hardware configurations, i.e., the laser power, filter bandpass range, and detector voltages, since the fluorescence ratio does not directly represent protein-protein proximity as defined in FRET efficiency.

Therefore, using the FRET efficiency quantification approach, the sensitivity of IAA/Auxin detection has been notably enhanced. Originally, the detection range for IAA/Auxin was between 10^{-4} M and 10^{-5} M, now encompassing lower concentrations from 10^{-5} M to 10^{-6} M (Figure 50B). In addition, even with more concentration data within the linear concentration detection range, a higher R^2 value was achieved (Figure 50B). In details, R^2 = 0.982 with three linear concentration data using the FRET fluorescence ratio approach, while R^2= 0.993 with four linear concentration data using the single-cell level FRET calibration and FRET efficiency quantification algorithm. Moreover, the single-cell level absolute quantification using the FRET efficiency reflects the physical interaction proximities of the bioengineered FRET-donor FRET-acceptor molecules. Therefore, it allows the FRET detection quantity assessment across different experiments along with the negative control sample values using the single-cell level FRET efficiency distribution (Appendix Figure 18). The improved small molecule FRET quantification allows more sensitive concentration detection that is not limited to IAA/Auxin. The same approach can also detect small molecule diagnostic markers

related to human diseases and monitor the concentration of small molecule therapeutics for effective cure of human diseases.



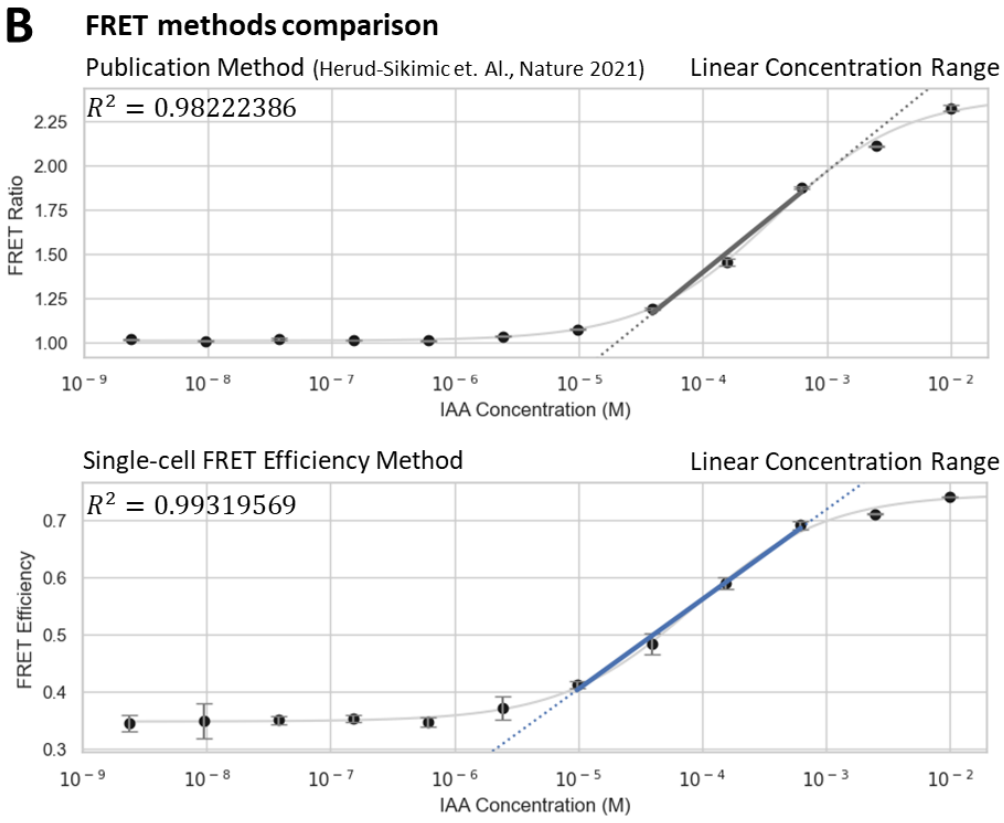


Figure 50 Small molecule Auxin/IAA detection using bioengineered FRET sensor
The Aquamarine → mNeoGreen FRET sensor detected the Auxin/IAA concentration by conformational changes induced upon Auxin/IAA binding to the tryptophan repressors (A). The detection of Auxin/IAA concentration using FRET ratios method from the original publication (black, top) [34] and new FRET efficiencies calculation in this study (blue, bottom) (B).

5.2.4 An application demonstration of detecting T-cells response to antigen

The two-protein three-colour FRET algorithm also demonstrated improved detection of the human immunological TCR to CD4 receptor interactions that can completely distinguish the positive interacting samples from the background controls. When B cells present foreign antigens to T cells, the TCR of T cells recruits CD4 to form a complex that activates immune responses [126]. The original study used the staphylococcal enterotoxin B (SEB) superantigen to activate the TCR-CD4 interactions (data collected from the FlowRepository database, Repository ID: FR-FCM-ZZR6). TCR and CD4 receptors were tagged with antibodies conjugated using Violet blue and FITC. Here, TCR-Violet blue functioned as the FRET donor and CD4-FITC as the FRET acceptor. When SEB was added to the sample, TCR and CD4 interacted at proximity and produced FRET emissions. However, TCR-CD4 interaction also responds to many other factors, including environmental triggers during sample handling and can produce high background FRET emissions. This was observed in the original publication [126], where the SEB-positive sample heavily overlapped with the SEB-negative sample. In addition, the original published FRET efficiency results have physically and biologically meaningless negative FRET efficiencies. By comparison, the two-protein, three-colour FRET algorithm introduced in this study accomplished a complete separation of the FRET efficiencies detected and calculated from the SEB-positive and SEB-negative samples (Figure 51B). In addition, the negative FRET efficiency is also eliminated, instead the positive FRET efficiencies represents the actual physical behaviours bewteen the FRET donor and FRET acceptor fluorophores and biological interactions between the CD4 and TCR molecules. The SD reduced from 0.083 to 0.010, decreasing the coefficient of variation (CV) from 0.666 to 0.078. In addition, the mean signal-to-noise ratio also increased from 1.304 to

1.324. The improved detection results indicated that the newly developed FRET algorithm is more suitable for clinical applications which require higher sensitivities, such as screening for drugs that modulate immune cell receptor proximities and disease diagnosis for immunocompromised individuals with impaired receptor interactions.

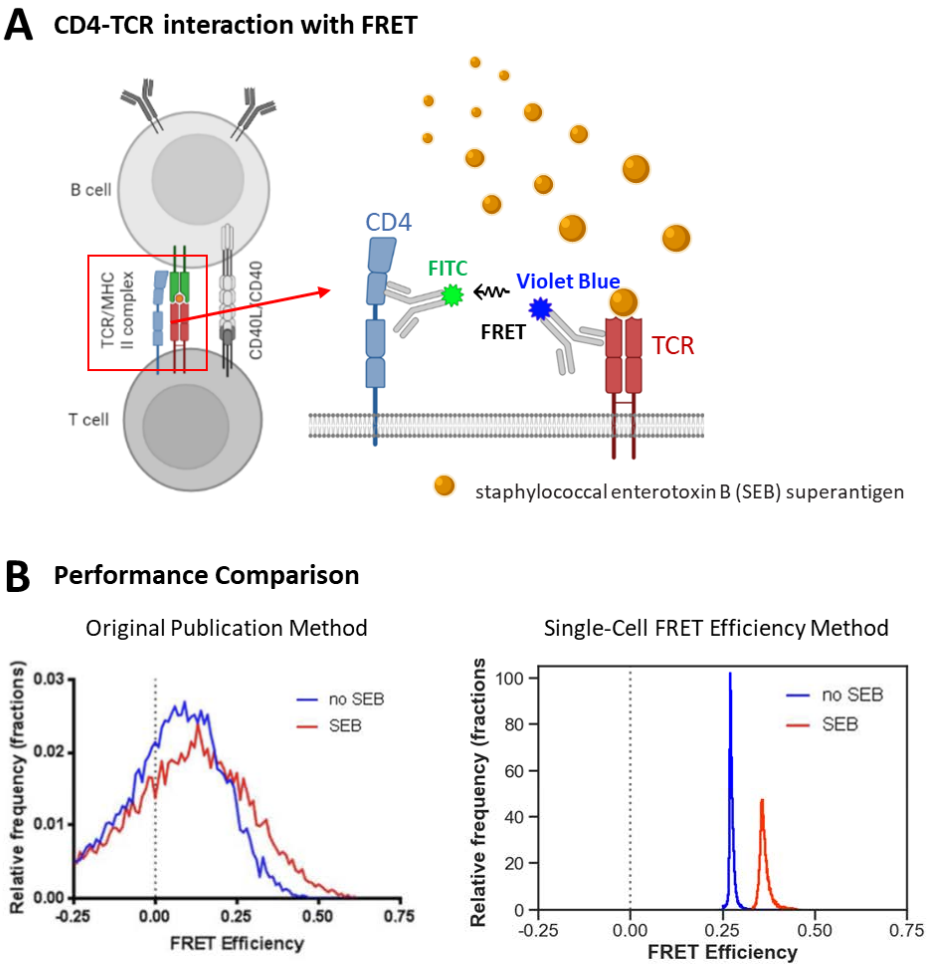


Figure 51 CD4-TCR interaction detection using flow cytometry-based FRET
CD4 interacted with TCR when SEB was added to the sample. CD4 was tagged using a FITC-conjugated antibody, and TCR was tagged using a violet-blue-conjugated antibody. When CD4 interacts with TCR, FRET happens between violet-blue and FITC (A). FRET efficiencies were calculated for samples with (red) and without (red) SEB. The comparison was made between the original published data and the data calculated using FRET algorithms developed in this study (B). The left figure of panel B is adapted from the original publication of von Kolontaj et al. [126], allowing for a direct comparison with the FRET quantification method developed in this study.

CHAPTER SIX

THREE-PROTEIN SIX-COLOUR FLOW CYTOMETRY-BASED FRET INVESTIGATIONS

6.1 Introduction

Many biological processes arise from complex protein-protein interactions with more than just two protein entities. For example, TNFR1 has an inherent capacity to form transmembrane homodimers through the pre-ligand-binding assembly domain (PLAD) (Figure 52A) [71]. The previous chapter has demonstrated the detection of interactions between two protomers within a TNFR1 homodimer complex using the two-protein three-colour FRET assays. Moreover, TNFR1 interacts with the symmetrical trimeric ligand TNF [138], and each TNF can recruit three TNFR1 homodimers at the interface between every two TNF protomers (Figure 52B) [137]. Therefore, the TNFR1 dimers are arranged in a tightly packed hexagonal arrangement with room only for ligand binding and the formation of a large hexagonal network through the [homotrimer ligand]-to-[homodimer receptor]-to-[homotrimer ligand] interactions (Figure 52C) [120].

Within each set of dimeric TNFR1s interaction with the same trimeric TNF ligand, a three-protein six-colour FRET can be detected with the combination of all three pairs of the two-protein three-colour FRETs (Figure 52). In detail, the three-protein six-colour FRET consists of the simultaneous detection of eCFP→eYFP FRET1, eYFP→mRFP FRET2, eCFP→mRFP FRET3, and the two-step eCFP→eYFP→mRFP FRET4. The FRET4 is detected on the same lasers and detectors as the three two-protein FRETs. Together, there are three protein molecules (TNFR1-eCFP, TNFR1-eYFP and TNFR1-mRFP) and six detection colours (eCFP, eYFP, mRFP, FRET1, FRET2, and FRET3, to note: FRET4 = FRET1+FRET2+FRET3) (Figure 53).

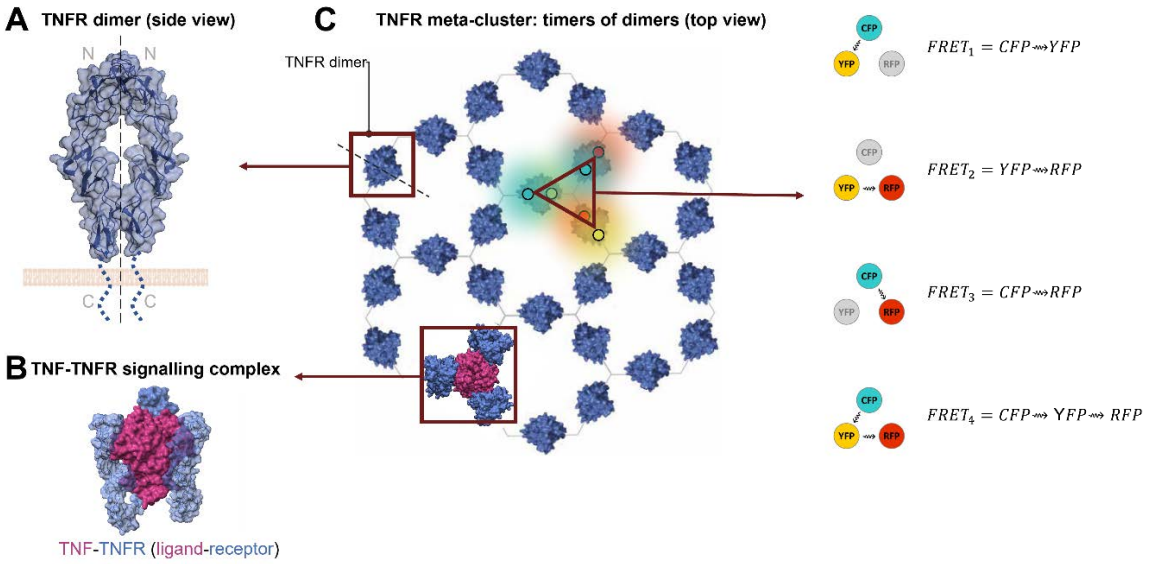


Figure 52 FRET signals of the TNFR hexagonal meta-cluster structure

TNFR can form a homodimer through the PLAD (A). Each TNFR homodimer interacts with the trimeric TNF cytokine at the interface of each two TNF protomers (B). The TNF recruits three TNFR homodimers and forms the hexagonal meta-cluster (C). The meta-cluster allows the detection of three sets of two-protein three-colour FRETs and the two-step three-protein six-colour FRET.

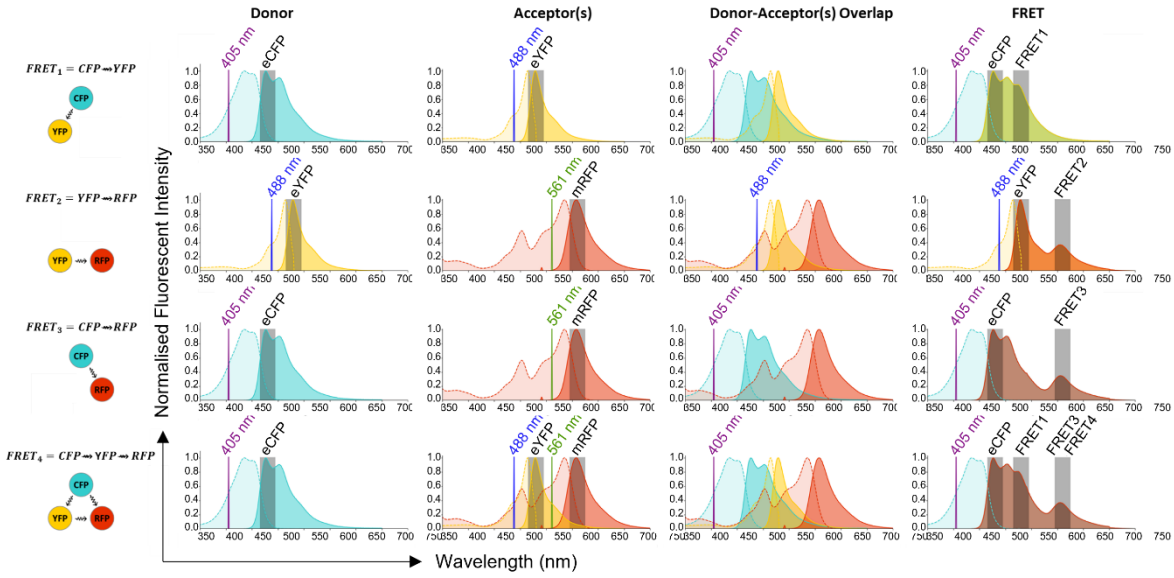


Figure 53 Three-protein six-colour FRET spectral viewer diagrams

The spectral viewer diagrams for eCFP→eYFP FRET1 (first row), eYFP→mRFP FRET2 (second row), eCFP→mRFP FRET3 (third row), and the two-step eCFP→eYFP→mRFP FRET4 (last row).

The complexity of protein-protein interactions does not stop at the number of interacting proteins; this complexity is manifested in the mechanism of interaction of biological phenomena at multiple levels. Therefore, achieving FRET detection with more than two protein entities opens the gate for more sophisticated biological investigations. On the one hand, take human TNFR1 as an example; germline mutations in TNFR1 cause the autoinflammation disease termed TNFR1-associated periodic syndrome (TRAPS) [184-186]. TRAPS is clinically significant due to its long-term recurrent inflammation symptoms that cause tissue and organ damage [152, 187, 188], but the exact cause of TRAPS has not been fully understood to date [184, 189, 190]. On the other hand, many viruses from the *poxviridae* encode human viral TNFR1 homologies (vTNFR, also called cytokine response modifiers, CRM) that can interact with human TNFR1 to evade TNFR1-related immune response [191-196]. The most infamous variola poxvirus, causes smallpox with up to 30% mortality during the Smallpox era [197]. With such strong natural selection, it is possible to hypothesise that humans and poxviruses are evolutionary rivals. Therefore, utilising the three-protein FRET platform can simultaneously reveal biological interactions involving human TNFR1 with at least five folds, (i) the wildtype TNFR1 dimeric interactions, (ii) the wildtype TNFR1 trimeric interactions, (iii) the wildtype TNFR1 and TRAPS mutant interactions, (iv) the human wildtype TNFR1 and poxvirus vTNFR interactions, and lastly (v) the impact of how TRAPS mutants alter the interactions between human wildtype TNFR1 and the poxvirus vTNFR.

The existing 3-protein FRET methods rely on the calibration of linked FRET pairs and artificially linked target protein samples [17-19, 75, 198]. Calibration using linked FRET pairs to determine the α factor is calculated at the population level and has a fixed FRET-

donor FRET-acceptor concentration and interaction ratio. This approach has low accuracy, poor representation, and lack of the capacity for the detection of unlinked free-interacting proteins. Using linked detecting protein samples for FRET assays is also inappropriate since the protein entities are brought in proximity in advance by amino acid linkers, which makes it prone to false-positive FRET results. Notably, few of the existing flow cytometry-based methods for analysing FRET between a single donor and multiple acceptors account for the change in relative energy transfer dynamics. Such dynamics result from competing interactions between fluorescent protein energy transfers when a single FRET donor provides energy to multiple fluorescent acceptors. To achieve the sensitive and accurate quantification of complex interactions among three unlinked free-interacting protein entities, I have formulated a three-protein six-colour spectral unmixing algorithm that resolves the absolute unmixed clean fluorescent readouts for single-cell level efficiency quantifications. In addition, I have also resolved the energy transfer dynamics among the same donor different acceptor FRET system using the CoDA technique. This chapter presents the formulation of the three-protein six-colour FRET algorithms and demonstrates the application in detecting the aberrant PPIs between the wildtype human TNFR1 and six prototypical TRAPS mutant receptors from the 62 SDM-generated TRAPS-YFP constructs based on the real-life patients' sequencing data [199]. In addition, this chapter also shows that myxoma poxvirus encodes MT2 (a vTNFR) that interacts with human TNFR1 and provides evidence for proving that TRAPS mutations are evolutionary trade-offs that can disarm the poxvirus immune evasion mechanisms act on human TNFR1.

6.2 Results

6.2.1 Investigation of the TNFR1 trimeric structures using FRET algorithms

While the upgraded three-protein six-colour FRET algorithm represents a significant advancement over the traditional two-protein three-colour FRET algorithm, it is imperative to acknowledge the increased complexity and challenges it presents. This algorithm facilitates the study of more intricate molecular interactions, such as confirming the existence of TNFR1 hexagonal meta-clusters in HEK293T cells co-transfected with pcDNA3.huTNFR1-eCFP, pcDNA3.huTNFR1-eYFP, and pcDNA3.huTNFR1-mRFP overexpression plasmids. However, managing the more complex spectrum components in the six-colour FRET system poses substantial challenges.

The complexity arises not only from the increased number of fluorescent proteins involved but also from the intricate spectral overlap and cross-excitation issues inherent to a six-colour system. These factors significantly complicate the process of accurately quantifying FRET efficiency and necessitate more advanced and precise control and calibration techniques.

To illustrate these complexities, in this chapter, I investigated human TNFR1 trimeric interactions, using structurally similar but non-interacting human TNFR2 and CD27 as background controls (Figure 54, shown in grey and pink). The experimentation involved replacing the cyan, yellow, and red positions separately with the non-interacting TNFR2 or CD27 receptors or replacing them with the pcDNA3 empty vector plasmid as the third component. This approach was necessary to manage the complexities of the three-way

FRET system and to effectively isolate and remove specific donor and/or acceptor interactions (Figure 54).

These experiments underscore the heightened level of difficulty in implementing a three-protein six-colour FRET quantification approach compared to its two-protein three-colour counterpart. The increased number of fluorescent proteins and their interactions require meticulous experimental design and analysis, highlighting the significant advancement yet complexity of this technique.

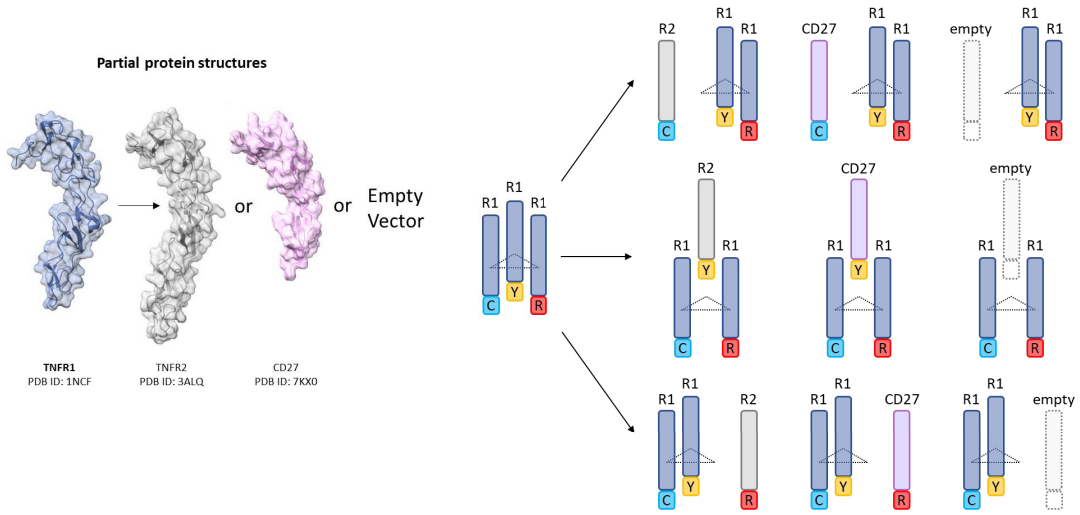
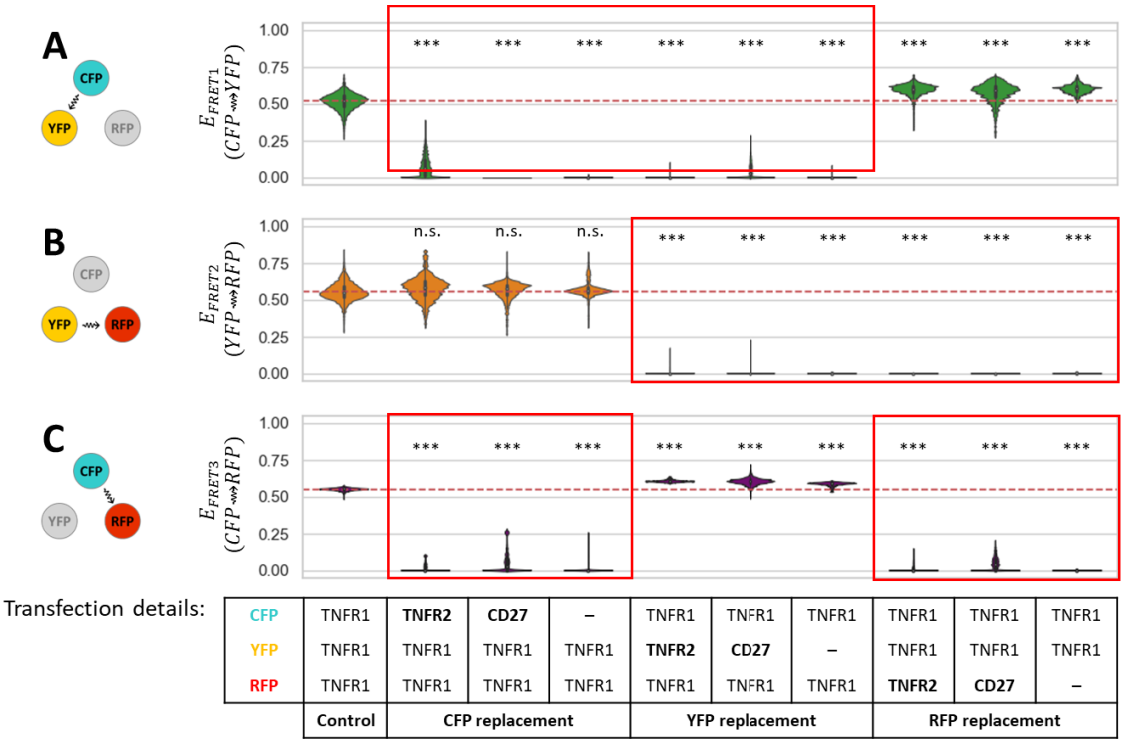


Figure 54 Receptor replacement FRET experiment design for trimeric TNFR1
TNFR1 (blue), TNFR2 (grey), and CD27 (pink) are structurally similar and non-interacting proteins to each other. The eCFP, eYFP or mRFP position of the TNFR1 trimer was replaced by the TNFR2, CD27 component or replaced with the empty pcDNA3 vector.

Triple transfected HEK293T cells expressing TNFR1-eCFP, TNFR1-eYFP and TNFR1-mRFP were analysed. The first genuine indication of the TNFR1 molecular trimerization interaction producing FRET emissions was demonstrated by examining the FRET efficiency of the TNFR1-eCFP, TNFR1-eYFP, and TNFR1-mRFP triple-expressing samples. All three TNFR1 FRET pairs demonstrated FRET emissions irrespective of the FRET pair

combination: eCFP → eYFP, eYFP → mRFP or eCFP → mRFP, with the average FRET efficiency for eCFP→eYFP was 0.514 with SD = 0.056, the average FRET efficiency for eYFP→mRFP was 0.554 with SD = 0.062, and with the mean = 0.550 and SD = 0.010 for eCFP→mRFP FRET (Figure 55).

When the FRET-donor and/or FRET-acceptor was replaced with the non-interacting receptors or empty pcDNA3 vector, the FRET efficiency that involves the particular FRET-donor and/or FRET-acceptor dramatically decreased close to zero, indicating the disappearance of the FRET event (Figure 55, red boxes). In detail, when replacement happened at the eCFP position since eCFP is the donor for eCFP→eYFP and eCFP→mRFP FRETs, the FRET efficiency near completely diminished (second, third and fourth samples in Figure 55A and C) but remained positive for eYFP→mRFP FRET which did not involve eCFP (second, third and fourth samples in Figure 55B). Similarly, when replacement happened at the eYFP position since eYFP is the FRET-acceptor for eCFP→eYFP FRET and is the FRET donor in eYFP→mRFP FRET, the FRET efficiency reduced to near zero (fifth, sixth, seventh samples in Figure 55A and B) but remained positive for eCFP→mRFP FRET which did not involve eYFP (fifth, sixth, seventh samples in Figure 55C). Lastly, the same phenomenon was observed in mRFP replacement samples, with the positive FRET efficiency remaining for eCFP→eYFP FRET-positive samples since it did not involve mRFP (eighth, ninth, tenth samples in Figure 55A) but produced FRET-negative results for eYFP→mRFP and eCFP→mRFP FRETs (eighth, ninth, tenth samples in Figure 55B and C).



Transfection details:

	Control	CFP replacement			YFP replacement			RFP replacement		
CFP	TNFR1	**TNFR2**	**CD27**	–	TNFR1	TNFR1	TNFR1	TNFR1	TNFR1	TNFR1
YFP	TNFR1	TNFR1	TNFR1	TNFR1	**TNFR2**	**CD27**	–	TNFR1	TNFR1	TNFR1
RFP	TNFR1	TNFR1	TNFR1	TNFR1	TNFR1	TNFR1	TNFR1	**TNFR2**	**CD27**	–

Figure 55 FRET efficiencies for different TNFR receptor combinations
The eCFP → eYFP FRET efficiencies (A), the eYFP → mRFP FRET efficiencies (B), and the eCFP → mRFP FRET efficiencies (C) for TNFR1-eCFP + TNFR1-eYFP + TNFR1-mRFP triple-transfected cells. In addition, TNFR2, CD27 and empty pcDNA3 vectors were used to replace the fluorescent components in the TNFR1 trimeric FRET system to compare alterations to the FRET efficiencies. The median FRET efficiencies of the control sample transfected with three TNFR1-FPs expressing plasmids are indicated as the red dashed lines. *** indicates $p<0.01$ from the z-test result, and n.s. indicates not significant from the z-test result, when compares with the control samples. The table below indicates the details of the plasmid used for cell transfection align with the violin plot in panel A, B and C.

When the same eCFP FRET-donor donated energy to both the eYFP FRET-acceptor and the mRFP FRET-acceptor simultaneously in the three-way FRET scenario, the FRET efficiencies for eCFP→eYFP and eCFP→mRFP FRETs are lower than those samples having only the eCFP→eYFP or eCFP→mRFP FRET in the two-protein independent interacting manner (Figure 55 and Appendix Figure 19). The lower FRET efficiencies further prove the genuine detection of the three-way trimeric interactions among TNFR1 receptors and prove the existence of the TNFR1 hexagonal meta-cluster conformation. The

median FRET efficiencies for the TNFR1-eCFP, TNFR1-eYFP, and TNFR1-mRFP triple-expressing samples were shown with the dashed red lines in each violin plot (Figure 55).

Furthermore, an interesting pattern emerges when altering the acceptors in the FRET system. When the mRFP acceptor was replaced with non-interacting proteins CD27 or TNFR2, the eCFP⟶eYFP FRET efficiencies consistently showed higher values compared to the control group with both TNFR1-eYFP and TNFR1-mRFP acceptors ($p<0.01$ for all mRFP-replaced or mRFP-removed samples). This trend is evident in the green-coloured group in Figure 55 and Appendix Figure 19. A similar increase in efficiency was noted for eCFP⟶mRFP FRET when the eYFP acceptor was replaced or removed ($p<0.01$ for all eYFP-replaced or eYFP-removed samples), as shown in the purple-coloured group in the same figures.

In contrast, the FRET efficiencies for eYFP⟶mRFP interactions remained consistent across samples, regardless of the presence or absence of the eCFP component. This is represented in the orange-coloured group in Figure 55 and Appendix Figure 19. The consistency of eYFP ⟶ mRFP FRET efficiencies across different scenarios can be attributed to the fact that this interaction does not involve eCFP, which in the three-way FRET scenario, eCFP donates energy to two acceptors simultaneously.

6.2.2 investigation of WT TNFR1 and TRAPS TNFR1 interactions

In the last section, the FRET efficiency results have demonstrated the outstanding formulation of the three-protein six-colour three-way FRET algorithm. However, the energy competition dynamics were not fully analysed with the successful detection of the same eCFP FRET-donor simultaneously donating energy to both eYFP and mRFP

FRET-acceptors. Here, I utilised the CoDA algorithms, treating the entire eCFP energy as one or 100%, and each composition that involves the eCFP energy dynamics were used to constitute the overall 100% of the total eCFP energy. In detail, the total eCFP energy contains three compositions (i) quenched eCFP donor fluorescent intensity after FRET, (ii) eCFP energy donated to eYFP results in sensitised eCFP→eYFP FRET emission, and (iii) eCFP energy donated to mRFP results in sensitised eCFP→mRFP FRET emission. Each composition's percentage value can be calculated by dividing the individual components by the sum of all, so increasing one composition cause reduction in the others. All individual cells can be further visualised in a triangular simplex space with the total sum distance to the three edges always equal to one, and each distance reflects each of the eCFP compositional values. With further logarithmic-ratio transformation, it achieves sub-compositional coherence that ensures the data is not impacted by any undiscovered hidden composition [124]; this also brings the data to a comparable interval scale that permits direct comparison and mathematic modelling between the energy dynamics and the FRET efficiencies.

Here, six TNFR1 TRAPS mutations were used as examples, all located within the CRD1/PLAD of the TNFR1 (Figure 56) and critical for receptor::receptor interactions. In addition, the six mutations are Y20H, C29Y, C30S, C30R, C43R, and P46L, where the Y20H represents the non-critical mutation that has little impact on the protein structure, the cysteine-involved mutations (C29Y, C30S, C30R, C43R) represent the structural-critical mutations since cysteine forms the disulphide bond responsible for the stable protein folding. The proline-involved mutation P46L represents another structural-critical mutation since proline is responsible for the structure rigidity. Moreover, the C30S and

C30R served as the same location mutations to different amino acids, and the C30R and C43R mutations showed the same mutations at different locations (Figure 56). Each of the six TRAPS mutant receptors were generated as the -eYFP fusion construct and co-expressed at the centre eYFP position with its eCFP FRET-donor and mRFP FRET-acceptor as the WT TNFR1-FP fusion proteins.

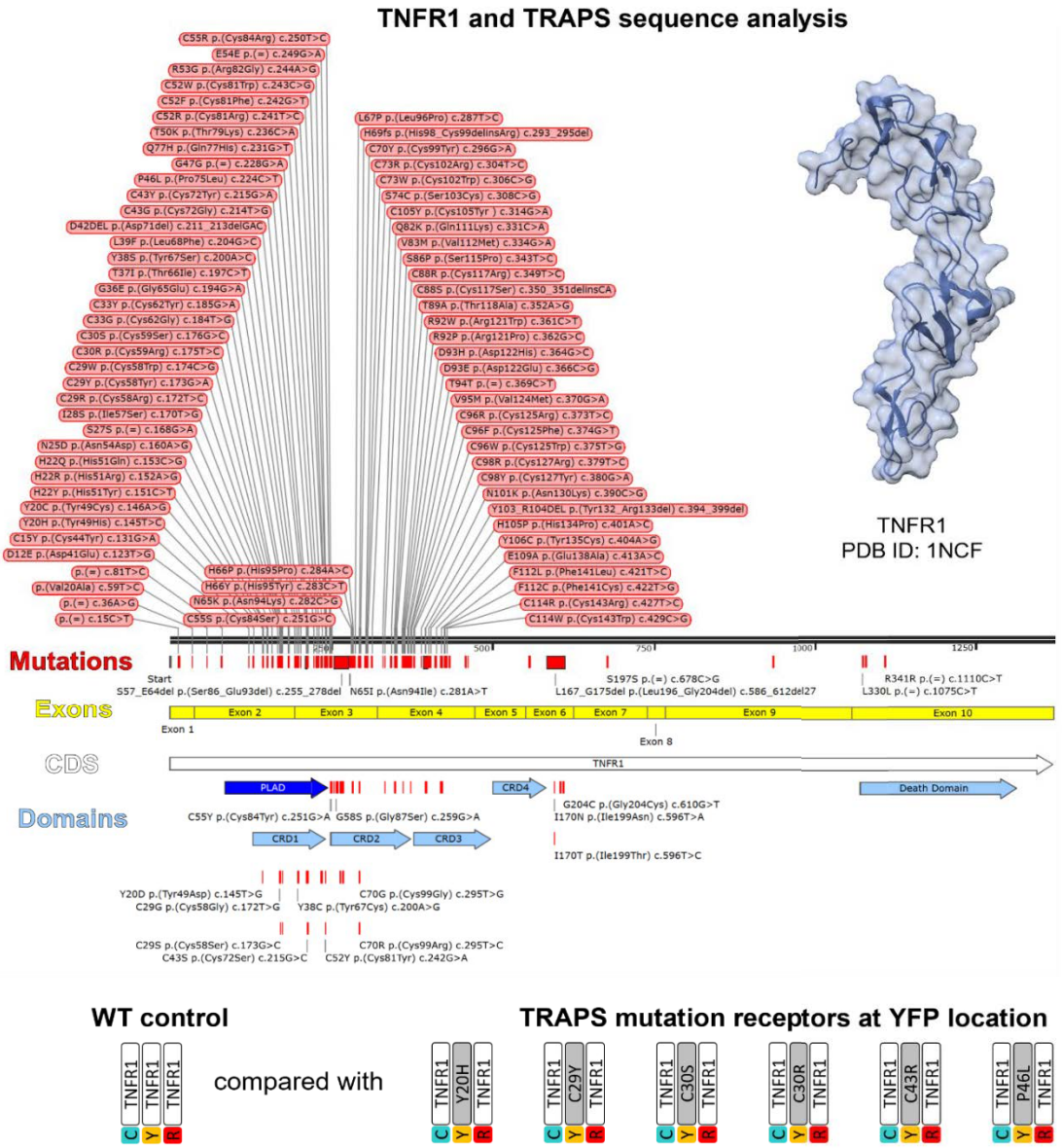


Figure 56 TNFR1 TRAPS mutations

TNFR1 CDS (white arrow) contains ten exons, four CRDs and one death domain. TRAPS mutations were found across the whole TNFR1 CDS, but most TRAPS mutations were found in the four CRDs. The WT TNFR1 trimeric FRET was compared to the same samples with replacement at the eYFP position with the TRAPS TNFR1 mutated receptors (grey).

Within the triangular CoDA simplex space, the mean position of the Y20H, C30S, C43R and P46L are close to the WT control sample with slight mean value shifts to the right, and the C29R and C30R TRAPS mutations are further away from the WT control with more shift to the right. The shift to the right indicates there are relatively higher interaction preference between the WT TNFR1-eCFP to the WT TNFR1-mRFP and relatively less interaction preference between the WT TNFR1-eCFP to the TRAPS TNFR1-mRFP, meaning the C29R and C30R TRAPS mutations produced mutated TNFR1 receptor conformations that have reduced interactions to the WT TNFR1 receptor compared to Y20H, C30S, C43R and P46L (Figure 57A). However, when examining the single-cell level data within the triangular CoDA simplex space, the C43R showed a very different distribution to the WT control even though their mean values are close (Figure 57B, blue colour vs grey colour). Within each of the single-cell level plots, as the data distributed from left to right the relative CoDA-resolved eCFP energy shifted from eCFP→eYFP to eCFP→mRFP. When examining the FRET efficiency heatmaps overlays onto the single-cell CoDA data (Figure 57B last 4 columns), negative correlations were observed between the eCFP→eYFP FRET efficiency and the eCFP energy as the data distributed from left to right. Meanwhile, positive correlations were observed between the eYFP→mRFP FRET efficiencies and the relative CoDA-resolved eCFP energy flow from eCFP→eYFP to eCFP→mRFP (from left to right within each single-cell level plot), as well as between the eCFP→mRFP FRET efficiencies and the same relative CoDA-resolved eCFP energy flow.

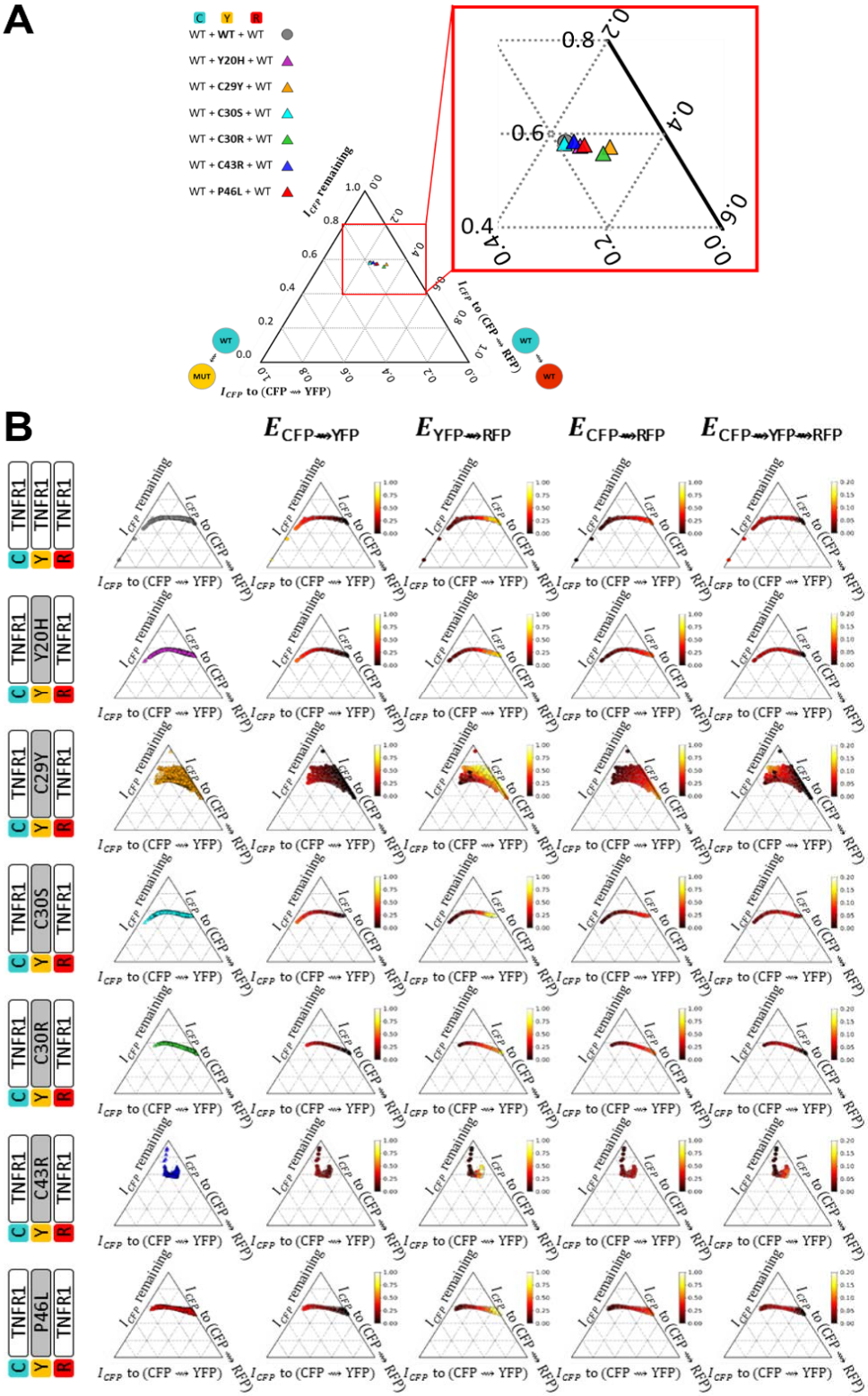


Figure 57 CoDA of WT TNFR1 and TRAPS TNFR1 interactions

The CoDA triangular simplex space has three corners, and each indicates the three compositions of the eCFP energy: (i) the quenched remaining eCFP, (ii) eCFP donated to eYFP, and (iii) eCFP donated to mRFP. The data includes the WT control sample and six TRAPS mutation samples. The mean values of the compositional eCFP energy are displayed in the large triangle (A). The single cell level data (first column) and heatmap overlays of each of the involved FRET efficiencies are displayed for all samples (B).

PCA produces dimensionality-reduced two-dimensional plots that allow direct visualisation of all of the WT and TRAPS mutation data in one plot. The eYFP→mRFP sensitised FRET emission and quenched eYFP donor emission constitute a two-composition CoDA space. Together with the three-composition eCFP data, the comprehensive three-way FRET results revealed that only the Y20H structurally non-critical TRAPS mutation has the near-identical profile as the WT controls (Figure 58A, pink and grey overlays), all the other TRAPS mutations showed different profiles on the PCA plot when compared to the WT control sample (Figure 58).

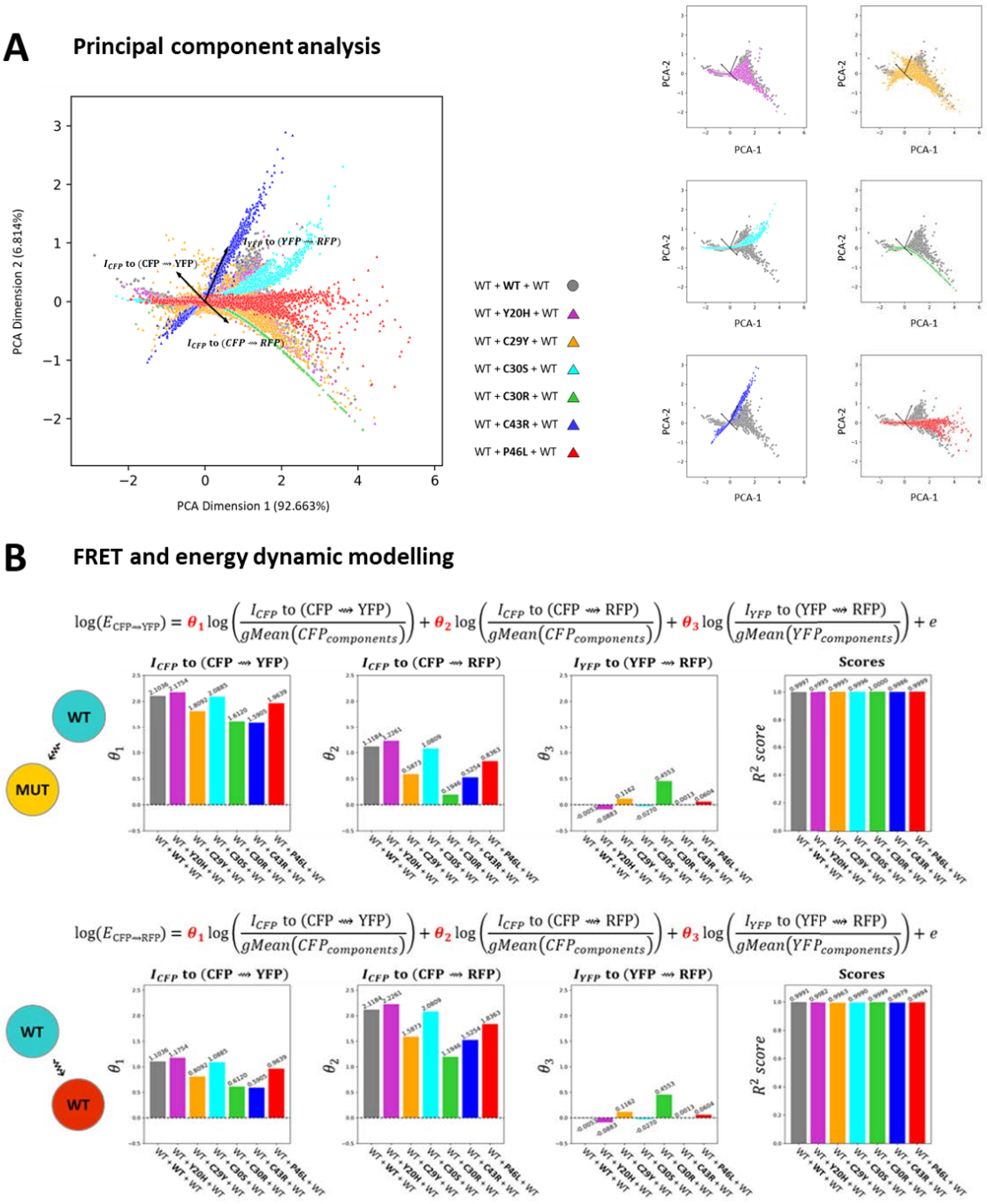


Figure 58 PCA and multivariate models of the TRAPS TNFR1 and WT TNFR1 interactions

PCA dimensionality reduction using standardised CLR-transformed data of eCFP energy to eYFP, eCFP energy to mRFP, and eYFP energy to mRFP (A). Multivariate modelling using logarithmic eCFP→eYFP and eCFP→mRFP FRET efficiencies as dependent variables and the CLR-transformed three compositions in PCA as independent variables (B). The first three bar graphs in each group indicate the coefficient of the multivariate models, and the last bar graph in each group indicates the R^2 values. The colours of the bar graphs correspond to those used in the PCA plots.

Interestingly, based on the multivariate model, both Y20H and C30S have highly similar patterns as the WT control sample in terms of explaining the FRET efficiencies with the impacts from energy dynamic from each of the compositional energy flows eCFP→eYFP, eCFP→mRFP, and eYFP→mRFP. Surprisingly, the C30S has different profiles from the WT control in the PCA plot (Figure 58). The close-to-one R^2 values indicated the perfect model explanation, and it shows that the energy flow from eYFP to mRFP has little impact on the eCFP→eYFP or eCFP→mRFP FRET efficiencies (Figure 58B) and reveals the contribution to the profile difference between C30S and WT control was ignored by pure eCFP energy calculated CoDA results but was captured by the PCA after CoDA transformation which included the eYFP energy calculated CoDA results. The distinct interaction mechanism of each TRAPS mutant receptor to the WT TNFR1 can potentially explain the causation of TRAPS autoinflammation, which has not been fully explored to date. The algorithm can also be used as a high-performance screening tool to find TRAPS therapeutics that can recover the aberrant interactions.

6.2.3 investigation of human TNFR1-targetted human-poxvirus interactions

This section tests the hypothesis that human and poxvirus are evolutionary competitors in that poxviruses encode vTNFR1, which target human TNFR1 to evade TNFR1-related immune response, and human TRAPS mutations are evolutionary results that also target TNFR1, which subvert the poxvirus TNFR1-related immune evasion. In other words, the periodic inflammatory symptoms from TRAPS are evolutionary trade-offs when they prevent the more severe consequence, such as mortality caused by smallpox variola poxvirus. Sequence alignment showed that the investigated poxvirus members, myxoma virus, variola virus and monkeypox virus all encode human TNFR-homologous

receptors that share highly conserved sequences within the PLAD receptor::receptor interaction-critical domain (Figure 59A). For biosafety concerns, the non-permissive myxoma virus was used to replace the huTNFR1-mRFP in the tree-way FRET, together with the WT huTNFR1-eCFP and TRAPS huTNFR1-eYFP (Figure 59A). Due to the lack of protein structure study on the MT2 (vTNFR) encoded by the myxoma virus, the MyxMT2 protein structure was predicted using AlphaFold (for method see section 2.6.7) [127]. The prediction results generated a protein structure with reference sequence coverage ranging from 5 to 6746 sequences for each amino acid position, 95.183 prediction score in the PLAD region (67.866 mean prediction score for the entire sequence) (Figure 59B). In addition, the structure alignment of predicted MyxMT2 to the TNFR1 dimeric complex demonstrated the interaction possibility between the myxoma virus MT2 and human TNFR1 proteins with alignment score = 229.9 and root-mean-square deviation (RMSD) = 5.929 (for method see section 2.6.7) (Figure 59B).

The CoDA results demonstrated the different impacts of TRAPS TNFR1 mutated receptors on the interaction between myxoma poxvirus MT2 and WT human TNFR1. The CoDA plots revealed great shifts of the mean values to the left-side triangle edge compared to the sample without myxoma MT2 protein (Figure 60A, mean data in the black box are from the TRAPS experiment in the last section for comparison). From the single-cell level distribution, only the Y20H sample has obvious interactions between WT huTNFR1-eCFP and myxoma poxvirus MyxMT2-mRFP with a reduced amount of interaction compared to the WT control sample with WT human TNFR1 as both the eCFP and eYFP FRET components (Figure 60B).

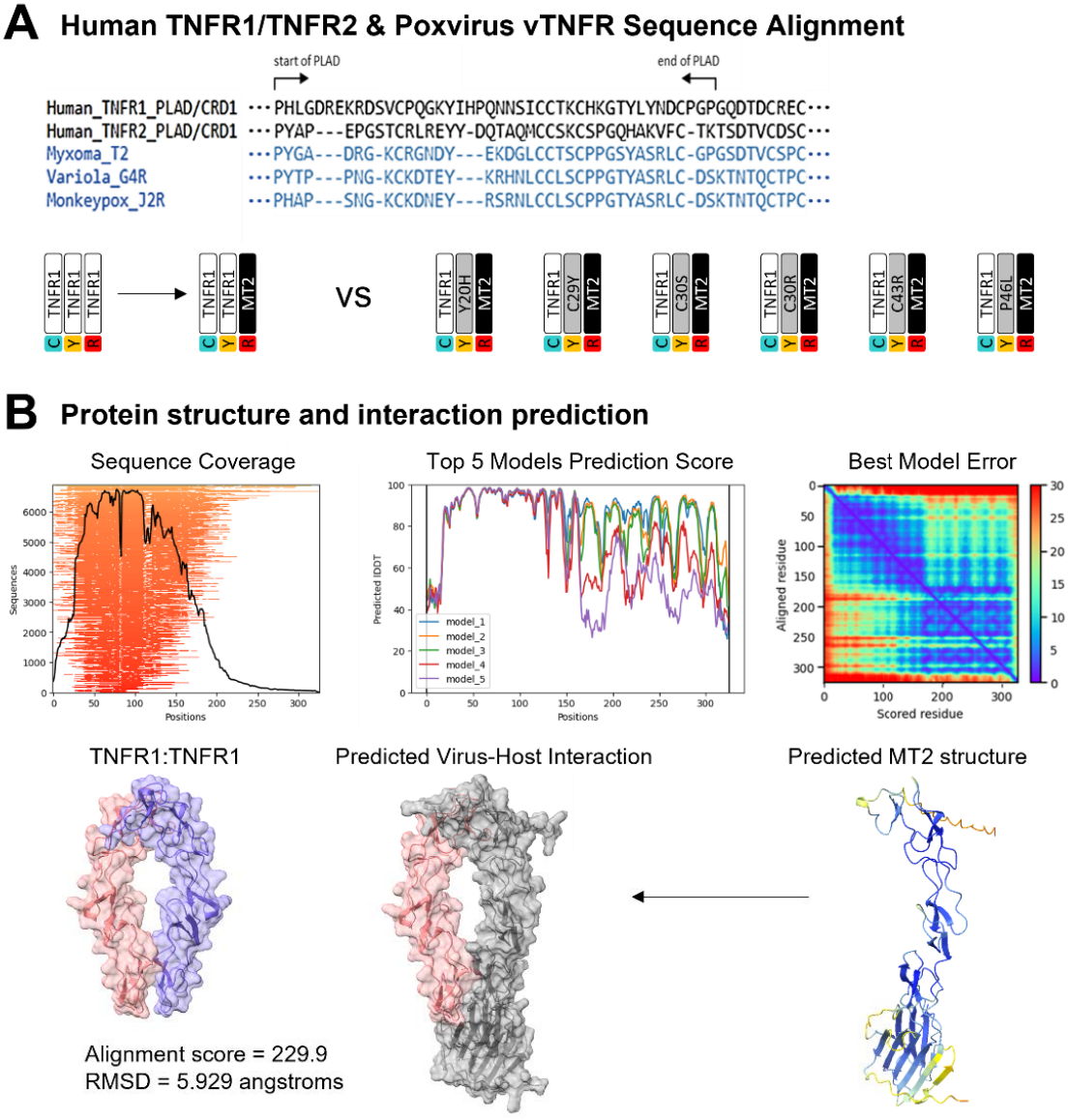


Figure 59 vTNFR sequence alignment and structure predictions

The human TNFR1, human TNFR2, and poxviruses-encoded MyxMT2, VarG4R, and MpvJ2R have highly conservatived sequences within the PLAD as demonstrated with the sequence alignment (A). MyxMT2 was selected as a safe prototype vTNFR to replace the mRFP FRET component in TNFR1 trimeric FRET system. The protein structure prediction and amino acid alignment of MyxT2 to the TNFR1 self-interacting dimer structure (B). The sequence coverage plot shows how many sequence entries from the database were used for the prediction at each amino acid position of the query protein amino acid sequence. The prediction plot score displays the IDDT scores of the top 5 ranked prediction models. The model error plot demonstrates the aligned residue errors from each amino acid position from the best prediction model. The TNFR1 dimer (pink and purple) is displayed at the bottom left, and the grey protein 3D structure is the predicted MyxMT2 protein structure rendered using the ribbon structure on the right. Within the predicted ribbon structure, the prediction errors from high to low are indicated from red to blue colours.

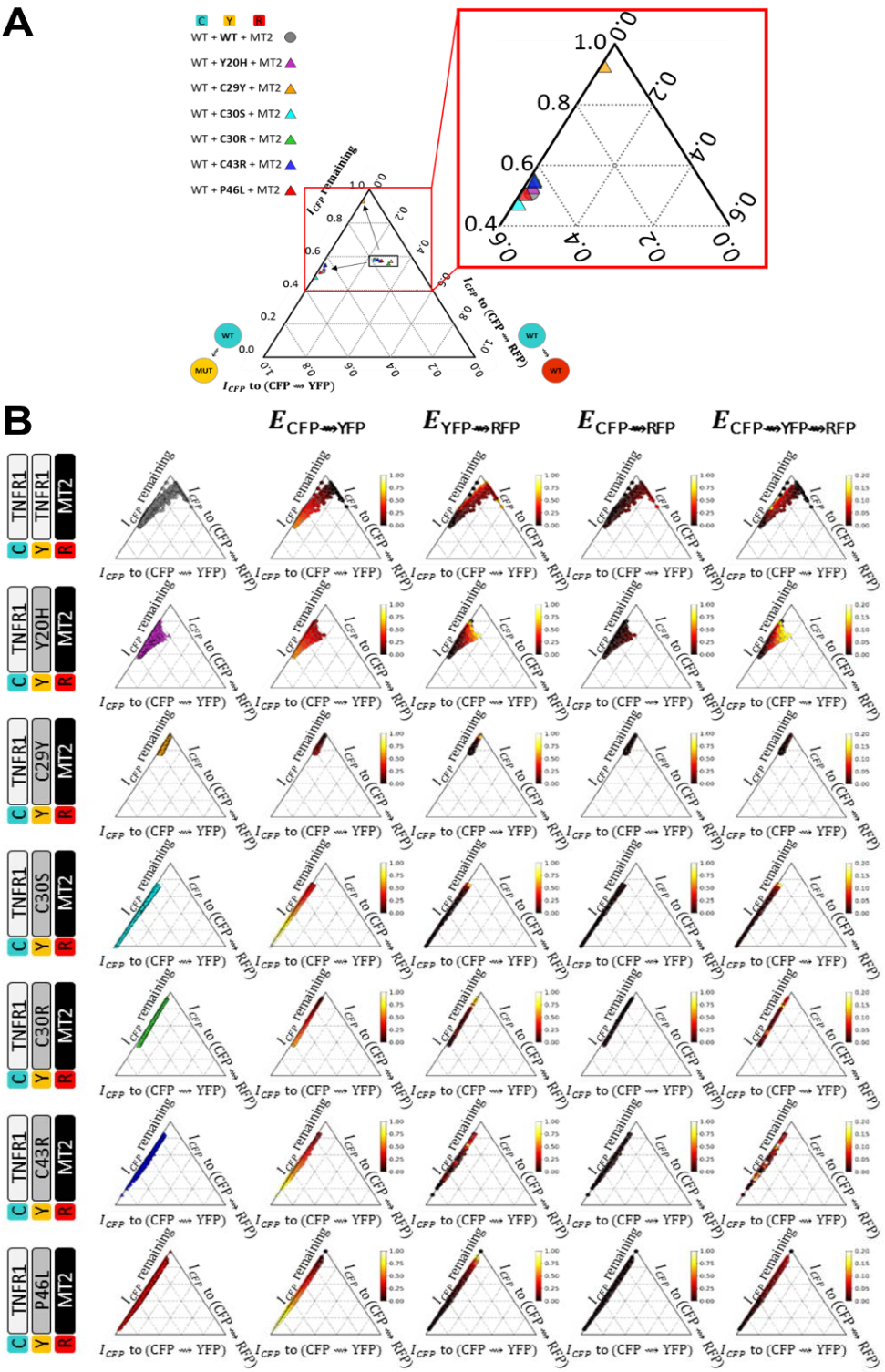


Figure 60 CoDA of poxvirus-human interactions that target TNFR1 interactions

The CoDA triangular simplex space has three corners, and each indicates the three compositions of the eCFP energy: (i) the quenched remaining eCFP, (ii) eCFP donated to eYFP, and (iii) eCFP donated to mRFP. The data includes the WT TNFR1 at the eCFP position, the WT TNFR1 and six TRAPS TNFR1 at the eYFP position, and MyxMT2 at the mRFP position. The mean values of the compositional eCFP energy are displayed in the large triangle (on the left). The single cell level data (first column) and heatmap overlays of each of the involving FRET efficiencies are displayed for all samples (on the right). The mean data in the black box are from the TRAPS experiment in the last section for comparison.

All other TRAPS TNFR1-eYFP expressing samples have near complete reduction between WT huTNFR1-eCFP and myxoma poxvirus MyxMT2-mRFP and with varied degrees of interactions between WT huTNFR1-eCFP and TRAPS TNFR1-eYFP (Figure 60B). Interestingly, when myxoma MT2 protein presents, the C29Y mutant demonstrated the most decreased interaction between WT huTNFR1-eCFP and TRAPS TNFR1-eYFP. At the same time, other TRAPS mutations displayed a broad range of interaction levels between WT huTNFR1-eCFP and TRAPS TNFR1-eYFP and overall similar mean values compared to the control sample (Figure 60). The two-dimensional reduced visualisation from the PCA plots indicated the Y20H had the most similar profile to the WT control sample when myxoma MT2 protein presented with the P46L showed heterogenous patterns, indicating more than one interaction mechanisms of P46L to both WT huTNFR1 and MyxMT2 vTNFR (Figure 61).

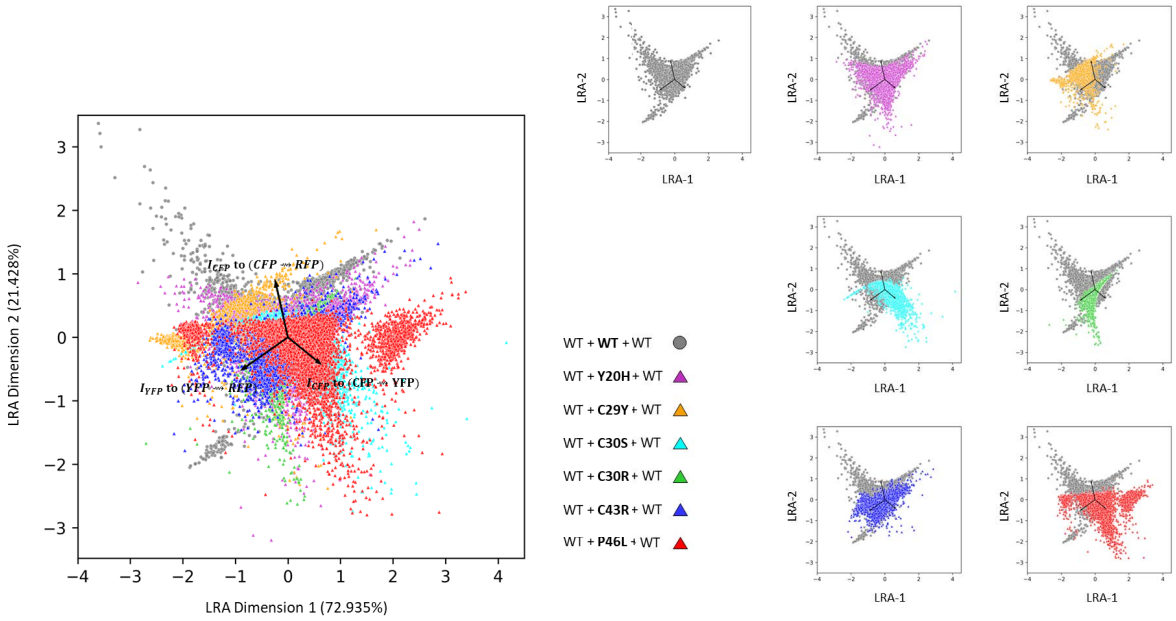


Figure 61 PCA of poxvirus-human interactions that target TNFR1 interactions
PCA dimensionality reduction using standardised CLR-transformed data of eCFP energy to eYFP, eCFP energy to mRFP, and eYFP energy to mRFP. Different TRAPS mutations are coloured.

The three-protein, six-colour FRET algorithm demonstrated different virus-host interactions and TRAPS mutation protective mechanisms. This was demonstrated by further summarization of results from the two studies investigated (i) WT TNFR1-eCFP + TRAPS TNFR1-eYFP + WT TNFR1-mRFP (ii) WT TNFR1-eCFP + TRAPS TNFR1-eYFP + MyxMT2-mRFP and presented the data together using radar plots. Six components on the left and right indicate with or without the virus vTNFR. On the left side, the middle component indicates the WT TNFR1::WT TNFR1 interaction, and the top and bottom both symmetrically indicate the WT TNFR1::TRAPS TNFR1 interactions (Figure 62). The three components on top to bottom (on the right) display the interactions between WT TNFR1::TRAPS TNFR1, human WT TNFR1::MyxMT2, and TRAPS TNFR1::MyxMT2, respectively. The distance from each corner to the plot centre denotes the extent of the FRET signals and interaction levels. The Y20H mutation showed the most similar interaction pattens for all six PPIs compared to the WT control sample (Figure 62 magenta sample). The C29Y mutation showed a more obvious evolutionary advantage with a decoy protection model, and it demonstrated reduced myxoma MT2 to WT TNFR1 interactions and vastly increased preference towards interactions between TRAPS TNFR1 and myxoma MT2 (Figure 62 yellow sample). In addition, the C30S also displayed a different non-decoy protective mechanism, and it demonstrated overall reduced human-viral interactions between C30S TNFR1 and myxoma T2, as well as between WT TNFR1 and myxoma T2. In addition, the interactions between the C30S TNFR1 and WT TNFR1 remained high (Figure 62 cyan sample), unlike in the C29Y TRAPS mutation sample. The other TRAPS mutations indicated different extents of the combined mechanisms as the C29Y and C30S. This summarised result further demonstrated the robustness of the three-protein six-colour FRET algorithm. Together

with the FRET detection pipeline, this chapter established a prototypic study for screening mutant-protein-based biosimilar drugs that can disarm viral immune evasion and for finding therapeutic candidates against virus infections.

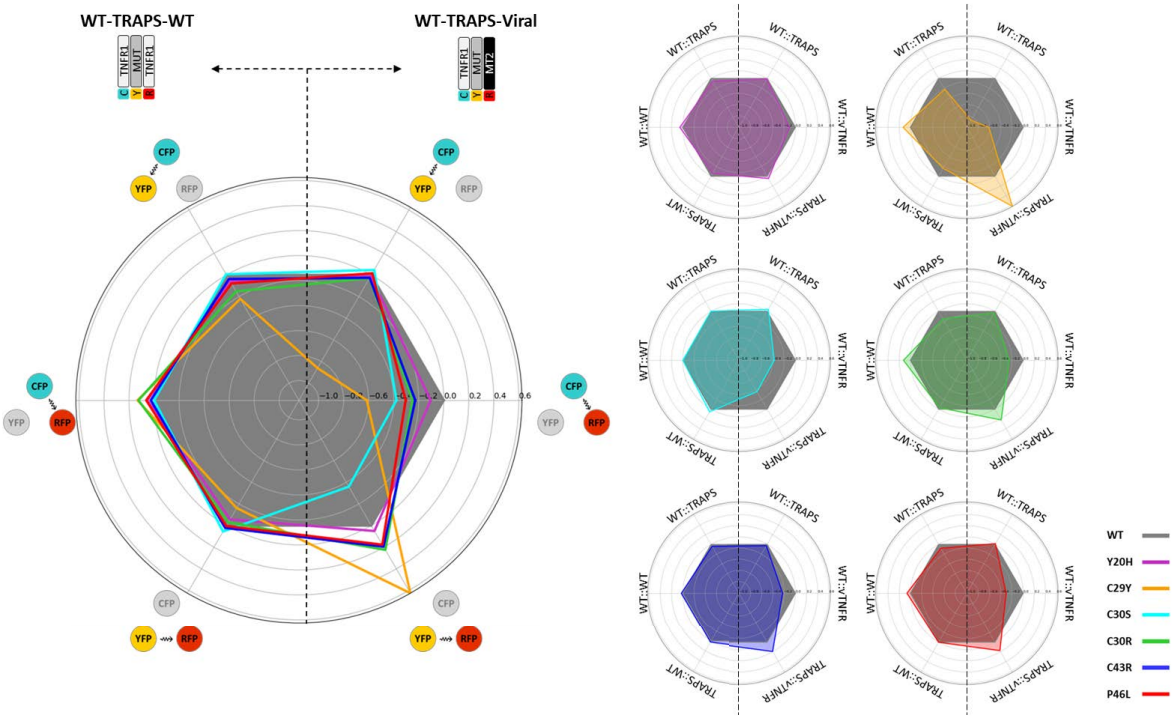


Figure 62 Poxvirus-human interactions together with TRAPS mutations
The radar plots show summarised results for the WT TNFR1 and TRAPS TNFR1 interaction experiments and the WT TNFR1 (left three corners), TRAPS TNFR1 and MyxMT2 interaction experiments (right three corners). Each radar plot shows the interaction between WT TNFR1 and TRAPS TNFR1 (top, both sides), between two WT TNFR1 (middle, left side) or between WT TNFR1 and MyxMT2 (middle, right side), and between TRAPS TNFR1 and MyxMT2 (bottom, both sides). Samples with different TRAPS mutations are coloured.

CHAPTER SEVEN

BEYOND FRET:

THE SIMULTANEOUS DETECTION OF FRET

AND CELLULAR SIGNALLING PATHWAYS

7.1 Introduction

Protein-protein interactions often result in the conformational change of the interacting proteins and lead to the association or dissociation of other protein complexes to activate cellular signalling pathways. Therapeutics target the protein-protein interactions that can regulate the downstream cellular signalling pathways and provide treatments for human diseases caused by abnormal activation of cellular signalling. However, current therapeutics block or promote cellular signalling in a broad, and non-specific way. Indeed, the 'holy grail' of pharmaceutical research into signalling pathway regulatory therapeutics is to identify a signalling pathway-specific inhibitory drug molecule. Therefore, the simultaneous detection of the protein-protein interactions and how changes in PPIs regulate multiple related cellular signalling pathways is crucial for finding the next-generation, pathway-specific therapeutics to treat and cure human diseases.

For example, TNFR1 interactions cause the disassociation of the silencer of death domain (SODD) protein from the TNFR1 intracellular death domain (DD) [200]. The DD of TNFR1 is then permitted to recruit and form four protein complexes, namely complex-I, -IIa, -IIb, and -IIc, for different TNFR1 signalling pathways (Figure 63). Complex-I is composed of TNFR-associated death domain (TRADD) [201], receptor-interacting serine/threonine protein kinase 1 (RIPK1) [202], TNFR-associated factor 2 or 5 (TRAF2/5) [203], the cellular inhibitor of apoptosis protein 1 or 2 (cIAP1/2) and linear ubiquitin chain assembly complex (LUBAC) [204], which leads to NF-κB activation and proinflammatory response [205]. Complex-IIa composed of FAS-associated death domain (FADD), TRADD, RIPK1, and pro-caspase-8, complex IIb composed of RIPK1,

RIPK3, and pro-caspase-8 and the formation of both complexes can lead to cellular apoptosis [153, 206, 207]. When RIPK1 and RIPK3 accumulate, they form complex-IIc, which leads to mixed lineage kinase domain-like (MLKL) protein activation, and cause cell necroptosis [201, 208].

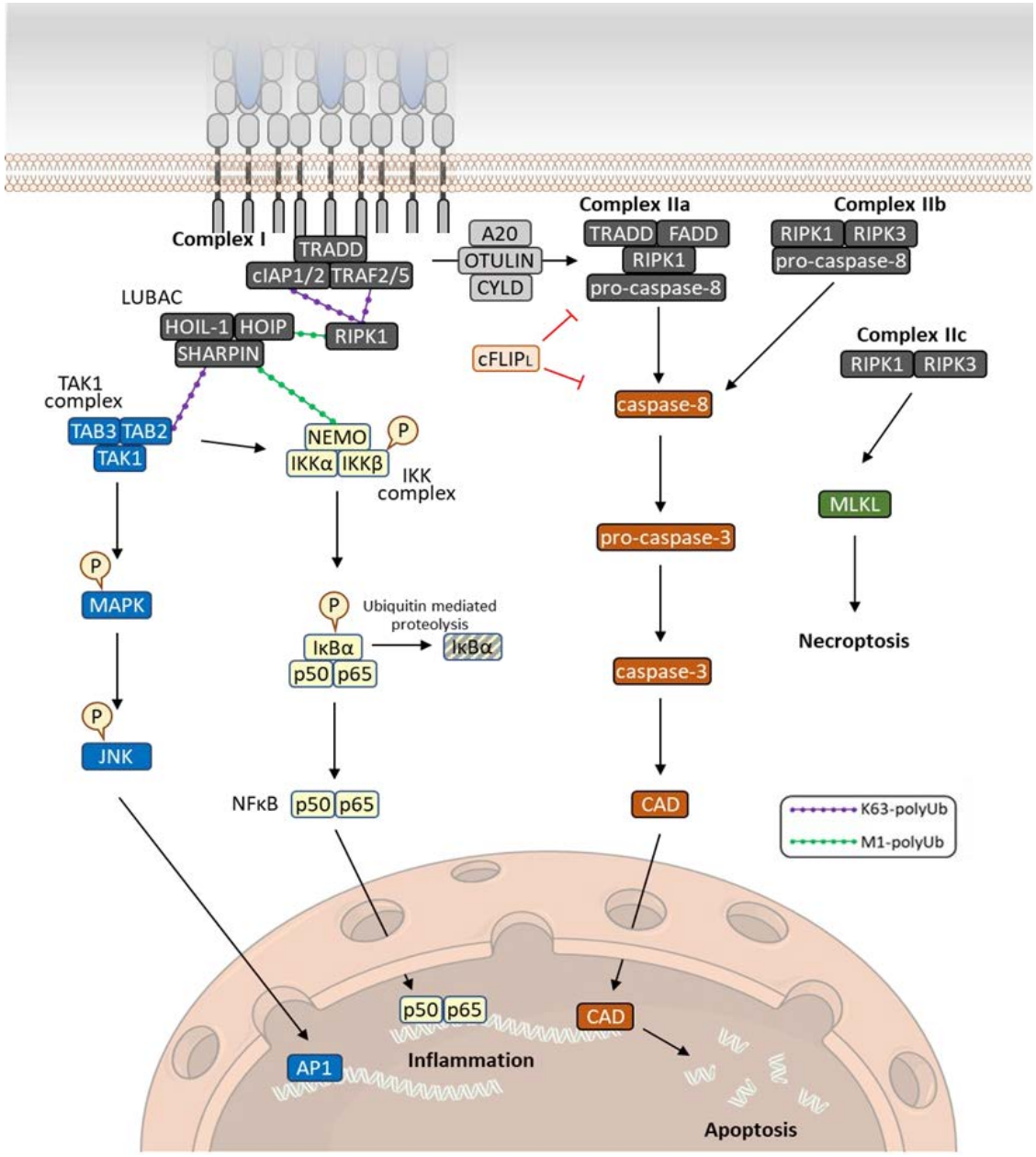


Figure 63 The complex TNFR1 intracellular signalling pathways
TNFR1 interactions recruit different protein complexes. Complex-I activates the pro-inflammatory signalling pathway through JNK-AP1 (blue) and NF-κB (yellow). Complex-IIa and complex-IIb activate apoptotic cell death signalling pathways through the caspase cascade (orange). Complex-IIc activates necroptotic cell death (green).

Aberrant TNF/TNFR signalling, such as increased TNF or altered TNFR expression, can cause many inflammatory diseases, including rheumatoid arthritis, multiple sclerosis, ankylosing spondylitis, Crohn's disease, and inflammatory bowel disease [143]. TNF/TNFR signalling-related gene mutations can also cause autoinflammation [151]. Therapeutic targeting of TNF/TNFR signalling has been developed to involve TNF-specific monoclonal antibodies (e.g., infliximab, adalimumab, Humicade, Golimumab, Certolizumab Pegol) and recombinant fusion proteins (Etanercept) that function by sequestering TNF and inhibiting ligand binding to the receptor [209]. However, TNF-targeting therapeutics have been found to be associated with immunosuppressive adverse effects, including reactivation of chronic or latent bacterial and virus infections, such as tuberculosis [210], adenoviral pneumonia [211, 212], and varicella-zoster [213, 214]. Other adverse effects include an increased risk of demyelinating disorders [215] and lymphomas [207], which are believed to be mostly caused by unselectively blocking all TNF-TNFR1 biological activities.

Current detection methods are still limited to separately detecting the PPIs using FRET and then investigating the signalling pathway activations using total cell lysate and other molecular biology approaches or detecting the single-cell level signalling requires a separate independent experiment [71, 216]. In addition, many investigations ignore the fact that cells contribute energy to activate multiple cellular signalling pathways simultaneously and treat different cellular signalling results independently. This chapter further upgraded the spectral unmixing methodology to include two cellular signalling fluorescent reporters and integrated the three-protein six-colour FRET, two cellular signalling fluorescent reporters. Together with the CoDA technique for dynamic

compositional interpretations, it allows simultaneous detection of FRET for PPI and the dynamic changes of the multiple signalling pathways activities. This chapter demonstrates the flow cytometry analysis of the simultaneous detection of TNFR1 protein interactions in correlation with the relative changes between the NF-κB activated inflammation and the cell death reporters.

7.2 Results

7.2.1 Detection of cell death using the fixable UV-blue live/dead stain

Utilising the fixable UV-blue live/dead stain with the unsupervised GMM clustering algorithm has accomplished automatic unbiased distinguishment of unstained background noise from the DMSO buffer, the stained live population and the stained dead population (Figure 64A). The results demonstrated the successful detection of the cell death signalling pathways' activities without impact on detection of the eCFP, eYFP or mRFP fluorescent components and *vice versa*. The positive control for cell death was generated by five minutes of 50 °C heat shock to induce partial population cell death. The medium-only sample treated with DMSO (Figure 64A, black histogram), the mock transfection sample (Figure 64A, green histogram) and the positive control sample (Figure 64A, purple histogram) stained with the fixable UV-blue live/dead stain were concatenated to train the unsupervised GMM. The test samples are cells transfected with human CD27, TNFR1 or TNFR2 receptor expressing plasmids, to examine whether the live/dead stain can be used for TNFR1-induced cell death detection. In order to investigate the whether the pre-processing steps can resolve the independent fluorescent signals of the live/dead, eCFP, eYFP, mRFP and FRET emissions, HEK-293T cells were transfected with human CD27, TNFR1 or TNFR2 receptor expressing plasmids

with or without the fused fluorescent protein components. In detail, for data analysis, the concatenated data was transformed using bi-exponential transformation to achieve better normality and then fitted to the GMM algorithm with the cluster number set to three. The trained GMM automatically generated thresholds for the concatenated data to distinguish each of the control samples (Figure 64A) and then was used to fit the HEK-293T cell samples transfected with the pcDNA3 overexpression plasmids to independently express human CD27, human TNFR2 and human TNFR1 as the non-fluorescently tagged full-length receptors (Figure 64B, left). The GMM was also used to fit the sample co-expressed the same human receptors as the fluorescent protein-tagged -eCFP, -eYFP and -mRFP fusion proteins, as mentioned in the previous chapters (Figure 64B, right). By comparison, there is no significant difference between the fluorescently tagged fusion protein expressing sample and the non-tagged full-length receptor sample, for CD27-FPs and CD27 t = -0.114 p = 0.920, for TNFR2-FPs and TNFR2 t = -0.112 p = 0.921, and for TNFR1-FPs and TNFR1 t = -0.297 p = 0.795 (Figure 64B). The result indicated (i) the C-terminus fused fluorescent component does not interfere with the receptor biology compared with the original non-tagged full-length receptors, (ii) the whole analytical pipeline has outstanding pre-processing procedures that eliminated the fluorescence spillover effects of all three fluorescent proteins, i.e., eCFP, eYFP and mRFP from the live/dead signals (as no significant difference between samples with and without fused fluorescent protein). In addition, the results showed that CD27 does not cause cell death compared to the mock transfection negative control sample (F = 0.009, p = 0.991) (Figure 64C). Both TNFR2 and TNFR1 can cause cell death with significant differences from the negative control sample, in which TNFR2 caused around one-fold more cell death than the negative control (F = 11.619, p = 0.022), and TNFR1

caused around five to six folds more cell death than the negative control (F = 63.016, p = 0.0009) (Figure 64C).

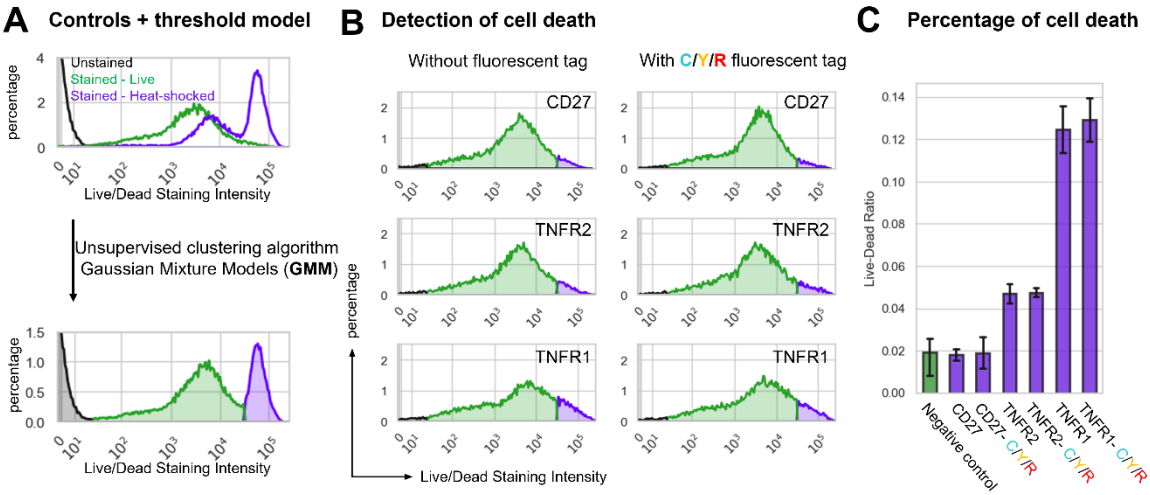


Figure 64 TNFR1 cell death signalling detection
Unstained control, cell death negative control and cell death positive control were concatenated for GMM training and unsupervised thresholds establishment (A). The GMM thresholds were applied to CD27-, TNFR1-, and TNFR2-expressing cells with or without the fluorescent fusion proteins (B). The percentage of cell death detected using live/dead stain and GMM thresholds (C).

7.2.2 Detection of TNFR1-induced NF-κB activated proinflammatory signalling

This section reports the newly constructed NF-κB inflammation reporter achieved TNFR1 receptor-specific detection of the inflammatory signalling pathway activation and demonstrated full compatibility with the eCFP, eYFP, mRFP three-way FRET system, as well as the cell death reporter. In detail, the robustness of the inflammatory reporter was tested to deliver four layers of outcomes, (i) the reporter is TNFR1 specific, (ii) the eCFP, eYFP, mRFP and FRET spillovers were completely unmixed from the reporter emissions, (iii) identification of the optimal copies of the NF-κB response element that can produce the highest signal-to-background ratio, and (iv) the reporter and the live/dead stain are compatible (Figure 65).

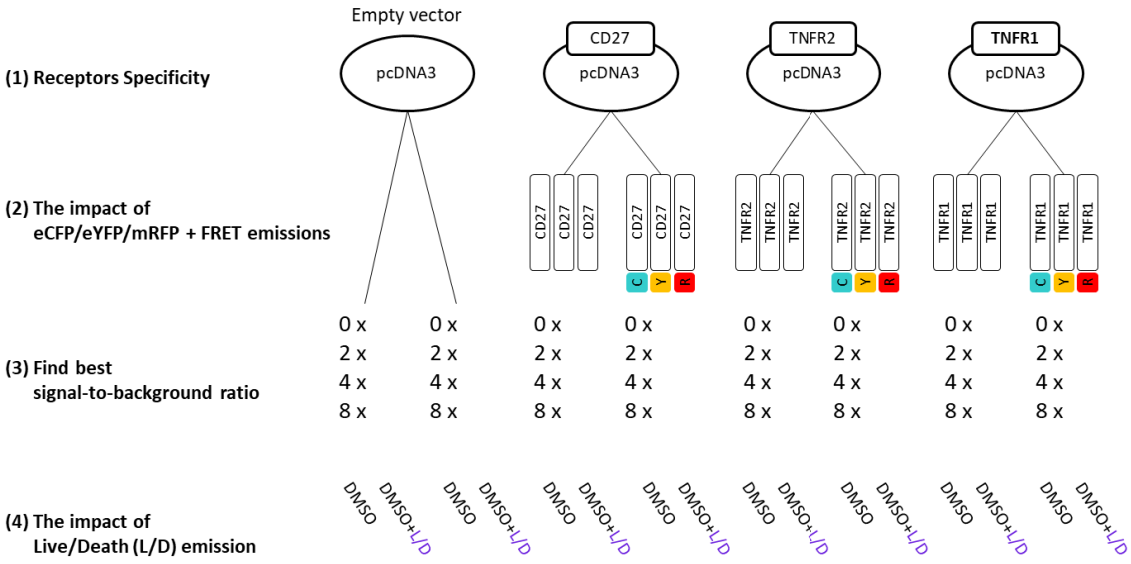


Figure 65 The proinflammatory NF-κB reporter valiation experiment design
Four layers of the experiment to test (1) whether the reporter is TNFR1 specific, (2) whether the fluorescent protein tags will impact the reporters, (3) find how many copies of NF-κB response elements achieve the best signal-to-background ratio, and (4) whether the reporter is compatible with the cell death detection.

The results showed that the pcDNA3.empty plasmid transfected negative control sample produced minimal miRFP703 emission with the reporter having NF-κB response element ranging from 0, 2, 4 and 8 copies (Figure 66A). The result also showed that neither CD27 nor TNFR2 expressing samples activated the inflammation signalling pathway significantly above the negative control samples, tested using ANOVA with p = 0.321 (Figure 66, top and middle group). Meanwhile, TNFR1 group samples produced no significant detection of inflammation signalling with zero or two copies of NF-κB response element in the reporter plasmids compared to the negative control samples (p = 0.646) (Figure 66, bottom group top two rows). However, significant NF-κB reporter emissions was detected for reporter plasmid having four copies of the NF-κB response

element (p = 1.334e-7) and with eight copies of the NF-κB response element (p = 2.864e-8) (Figure 66, bottom group bottom two rows). In addition, the samples with four copies of the NF-κB response elements delivered the highest signal-to-background ratios (Figure 66C).

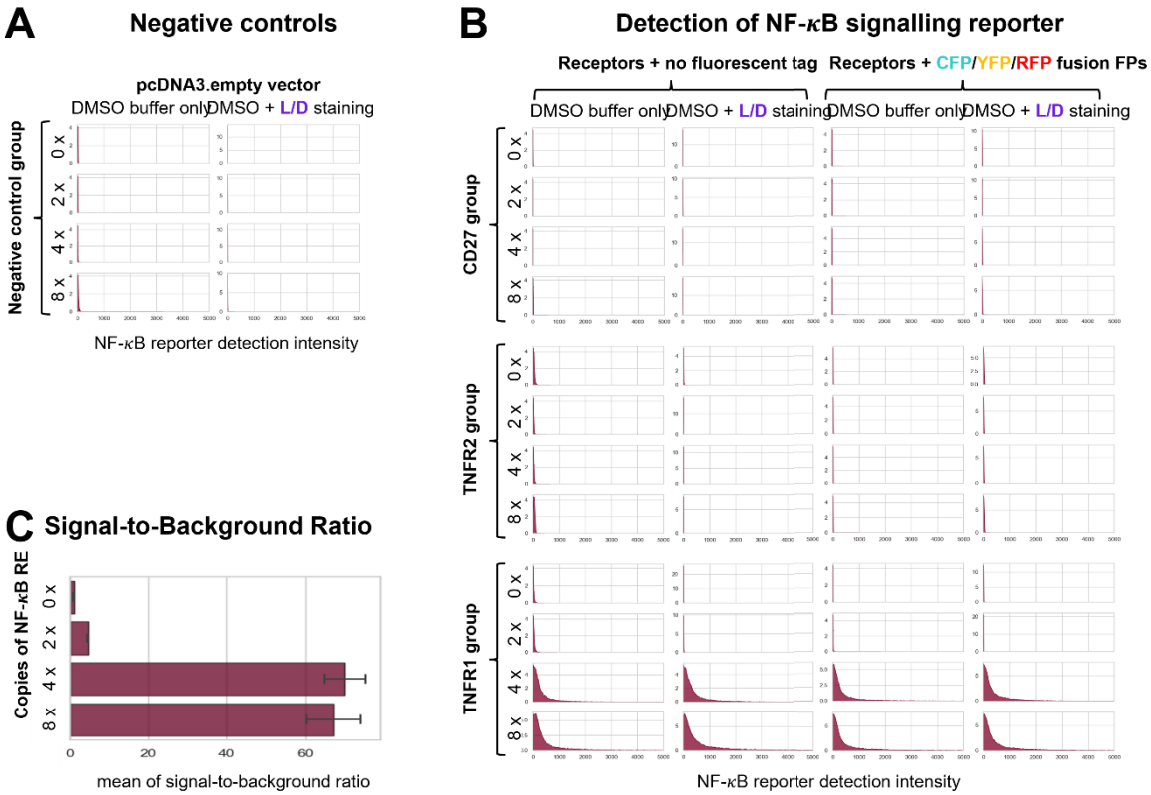


Figure 66 NF-κB proinflammatory reporter validation
Negative controls that transfected with proinflammatory reporters with different copies of NF-κB response elements and no receptor plasmid, with or without live/dead stain (A). Cells co-transfected with proinflammatory reporter plasmids having different copies of the NF-κB response elements and the receptor plasmids. The receptor plasmids were CD27, TNFR2, and TNFR1 with or without fluorescent protein fusion components. The samples were treated with or without live/dead staining (B). Signal-to-background ratios were detected using the proinflammatory reporters with different copies of NF-κB response elements (C).

7.2.3 Detection of the compositional dynamics of multiple signalling pathways

Furthermore, integrating the CoDA algorithm into the signalling reporter fluorescent emissions revealed the balance between cell death and inflammation signalling pathway activities. It also allows the detection of the relative dynamics of the two signalling pathways regarding TNFR1 receptor interactions at the single-cell level. When directly correlating the logarithmic transformed cell death reporter emission with the TNFR1 FRET reporter emissions, a highly scattered weak positive correlation was observed with $corr$ = 0.235, 0.522, 0.186, 0.526 and p = 2.578e-35, 4.2311e-188, 1.630e-22, 3.730e-192 for eCFP→eYFP, eYFP→mRFP, eCFP→mRFP, eCFP→eYFP→mRFP respectively (Figure 67A, purple). The inflammation reporter also demonstrated weak negative correlations to the TNFR1 FRET reporter emissions with $corr$ = -0.434, -0.029, -0.485, -0.170 and p = 1.968e-124, 0.128, 2.933e-159, 6.065e-19 for eCFP→eYFP, eYFP→mRFP, eCFP→mRFP, eCFP→eYFP→mRFP respectively (Figure 67A, dark red). After performing closure and logarithmic ratio transformation using CoDA, the relative compositional intensity of the cell death reporter showed an increased positive correlation with the TNFR1 receptor::receptor interactions, and the relative compositional intensity of the inflammation reporter presented an increased negative correlation with the normalised TNFR1 FRET emissions (Figure 67B). In detail, the correlation was 0.575 (p = 2.814e-237) between CoDA-transformed cell death and TNFR1 eCFP→eYFP FRET, and the correlation was 0.483 (p = 1.636e-157) between cell death and eYFP→mRFP FRET, with correlation was 0.561 (p = 2.997e-223) between cell death and eCFP→mRFP FRET, and correlation equalled 0.606 (p = 9.299e-270) between cell death and eCFP→eYFP→mRFP two-step FRET (Figure 67B). Meanwhile, because the LR balanced between cell death and inflammation, the correlations are -0.575, -0.483, -0.561 and -0.606 between CoDA-

transformed inflammation reporter emissions and each of the eCFP→eYFP FRET, eYFP→mRFP FRET, eYFP→eYFP FRET and eCFP→eYFP→mRFP two-step FRET with the same levels of significance, respectively. Thus, the algorithm provided insights into relative signalling activities regarding receptor::receptor interactions and provided solutions to design and screen drugs that can fine-tune the balance between multiple signaling pathways. They can help us find the 'Holy Grail' of the next-generation cellular signalling pathway-specific therapeutics for human diseases.

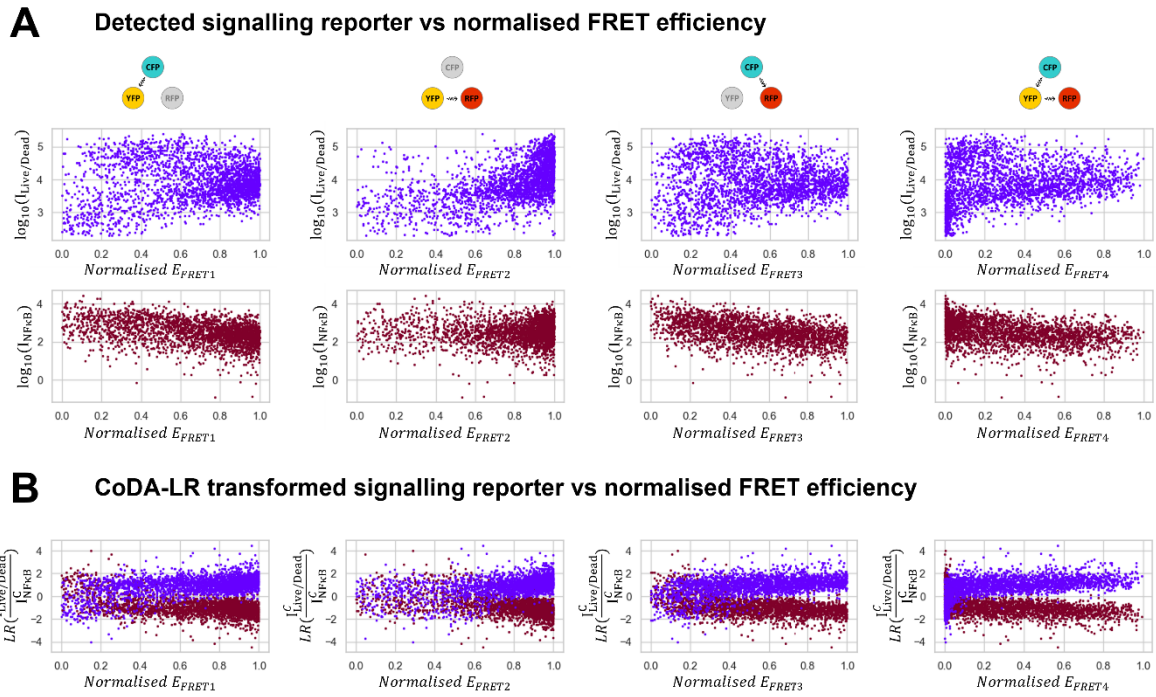


Figure 67 Correlation between signalling pathways and TNFR1 interactions
Correlations between logarithmic normalised cellular signalling pathway readouts and normalised FRET efficiencies (A). Correlations between CoDA-transformed signalling pathway readouts and normalised FRET efficiencies (B).

SUMMARY AND PERSPECTIVE

This research project aimed to develop a comprehensive and fully automated analytical workflow that can detect multi-protein interactions with FRET and assess the activities of various cellular signalling pathways at a single-cell level. The workflow relies on a combination of high-throughput flow cytometry techniques and advanced machine-learning algorithms. With the presented approach, researchers can accurately quantify cytometry data and achieve more efficient and reliable results. The analytical workflow covers every single stage of the multiparameter flow cytometry analysis together with the FRET investigation. The DBSCAN-based UltraFast singlet identification algorithm achieved outstanding accuracy compared to many existing unsupervised flow cytometry clustering algorithms and turned the previously unaddressed labour-intensive, subjective, time-consuming process into an easy, objective, and rapid automated process. The collaborative filtering recommendation-based algorithms accomplished error-free baseline subtraction-error correction, and autofluorescence prediction and removal, which recovered the pure single fluorescence emission of each individual fluorescent component through simple linear matrix-based spectral unmixing, therefore allowing the absolute quantification of flow cytometry analysis, including FRET experiments.

This study accomplished the detection of sensitive, robust, and true single-cell level FRET signals of two-protein, three-colour FRET emissions. Using the chemically linked eGFP-mRFP FRET pair, this study validated the FRET calibration and spectral unmixing formulation developed for naturally existing free-interacting PPIs. The algorithms were proved to be highly reliable using three unliked FRET pairs (TNFR-eCFP→TNFR-eYFP, TNFR-eYFP→TNFR-mRFP, and TNFR-eCFP→TNFR-mRFP) on two distinctly configured

flow cytometers (LSR-II and Fortessa X20). After demonstrating the usage with IAA/Auxin concentration detection and T cell antigen response assay, the FRET analytical pipeline was validated to be suitable for both small molecule detection and immunological diagnosis.

The performance of further upgraded, more complex three-protein six-colour FRET detection was evaluated using expression vectors that encode human TNFR1, TNFR2, and CD27 as -eCFP, -eYFP and -mRFP fusion proteins. In addition, a panel of 62 TRAPS TNFR1 mutant receptors (six were used as a prototype) as -eYFP fusion constructs, as well as vTNFR-mRFP fusion proteins from the poxvirus – myxoma virus (MT2 protein) were also used in this study. The three-way single-cell level FRET analysis established the possibility of TNFR1 forming hexagonal meta-clusters and revealed aberrant receptor interactions that potentially caused the human autoinflammatory disease TRAPS. The results also uncovered the human TNFR1-targeted mechanism of poxvirus immune evasion using vTNFR and ultimately discovered that human TRAPS mutations are a possible evolutionary trade-off in exchange for the disarming of poxvirus immune evasion to protect humans from poxviruses. In summary, these data demonstrated a set of prototype studies as a high-performance screening tool to find autoinflammation therapeutics that can recover the aberrant protein-protein interactions and for screening mutant-protein-based bio-similar drugs that can disarm viral immune evasion and for uncovering therapeutic candidates against virus infections.

Utilising CoDA and LR-based algorithms with the cell live/dead reporter and newly constructed NF-κB miRFP inflammatory reporter, the correlations among multiple intracellular signalling activities and the energy dynamics within the three-way FRET

system became evident. Thus, this approach provides insights into relative signalling activities concerning receptor::receptor interactions and provides solutions to design and screen drugs that can fine-tune the balance between multiple signal pathways. The entire ML-powered FRET analytical pipeline can help us find the 'Holy Grail' of next-generation cellular signalling pathway-specific therapeutics for human diseases.

When conducting the *in-situ* PPI detection in laboratories, the first step of using FRET is to achieve fluorescence labelling of the proteins of interest. There are currently two commonly used approaches, one uses fluorophore-conjugated antibodies for protein labelling as demonstrated by the TCR-CD4 experiment, and the other uses subcloning techniques to bioengineer the expression of the target proteins as fluorescent protein fusion entities as presented in chapter 3. There are many limitations in the antibody staining approach in FRET studies. Several reports have shown that there lacks an absolute quantification of the fluorophore-to-antibody conjugation ratio or the antibody-to-protein binding ratio, which heavily relies on the mean value calculated from the whole population perspective and sacrifices the single-cell level resolution accuracy and sensitivity [17]. In addition, antibody-targeting epitopes can also be PPI-critical domains and lead to false-negative FRET detection if the antibody works as antagonist molecules that effectively block PPI. Antibodies can also lead to false-positive FRET detection if the antibody cross-links two proteins and artificially brings them in proximity.

Moreover, if the target proteins express intracellularly, i.e., not on the cell surface, this will require permeabilization of the cell membrane before antibody staining, which can also cause unpredicted biologically irrelevant results. Therefore, the construction of a

fluorescent fusion is a superior solution for labelling the target proteins for FRET-based PPI investigations, as it always labels the protein in a one-to-one ratio. The entire fusion protein is also expressed as a single entity throughout the entire *in-situ* protein manufacture, assembly and trafficking processes. However, Chan et al. evaluated TNFR1 as fluorescent protein fusion constructs and observed inconsistent results on the FRET detection capacities when placing the CFP and YFP moieties at the N-terminus or C-terminus of the receptor protein [70]. In this study, the C-terminus TNFRSF-eCFP, -eYFP, and -mRFP fusion constructs were subcloned and were found to achieve the successful detection of FRET signals and have no significant difference in biological signalling activations when compared to the original unlabelled receptor proteins. This finding is consistent with that of Chan (2001), who described that the fluorescence of fusions at the N-terminal is sensitive to the linker length, and the intracellular C-terminus fusions of CFP and YFP were found to be far less sensitive to the length of the linker between fluorescent protein and receptor. Therefore, when conducting FRET investigations in the future, the biological function of the newly constructed fusion proteins must be tested before FRET experiments.

Furthermore, the selection of the fluorescent protein is also critical for the success of the detection of FRET signals. The FRET-donor and FRET-acceptor must have sufficient quantum yield QY for efficient energy transfer. Excitation coefficients ε, especially for the FRET-acceptor, play an important role in resonance FRET emission. The protein maturation rate can also impact the abundance of stable fluorophores. One surprising variable in this study was that TNFR1-eYFP has a more rapid expression decline observed with the live-cell imaging compared to other TNFR-FP fusion constructs prepared in

chapter three. A potential explanation is that the different combinations of fluorescent protein and receptor sequences may have varying codon biases, contributing to the variable expression dynamics. The tRNA pool from HEK-293T cells, which produces those fusion proteins, may also play a role. The cell death caused by TNFR1 signalling is another cause of the early decrease in TNFR1-eYFP. Therefore, this study also demonstrated that characterising each FRET component's expression and fluorescent emission profiles is crucial for FRET experiment design and have important implications for developing the FRET assay. Moreover, there is abundant room for further exploration with the newly developed UV-excited violet fluorescent proteins and red laser-excited infrared fluorescent proteins, as well as the potential of utilising upconverting fluorescent particles in the FRET system to free up the availability within the blue-green-yellow-red spectrum range for utilising widely well-established commercial dyes for the simultaneous detection of other biological activities.

Regarding the ML-powered flow cytometry singlet identification, this study aimed to provide an automated, unbiased, efficient solution based on the unsupervised DBSCAN algorithm. Therefore, the UltraFast algorithm developed in this study further enhanced the power of DBSCAN as the density-based clustering algorithm. The most satisfactory performance evaluation results demonstrated that UltraFast handles flow cytometry data with irregular shapes and outcompeted many other unsupervised clustering algorithms developed or utilised for flow cytometry data. In this study, I utilised the elbow method to automatically choose the hyperparameters ε and $MiniPts$, and this mirrors those of the previous studies that have utilised such application for more efficient automated hyperparameter tuning [100, 101, 217]. In addition, converting the

data into grids increases the computational efficiency and allows the tuning of ε and $MiniPts$ becomes tuning of a physically meaningful "grid number" hyperparameter, approximately equivalent to the strictness of the singlet gating. However, the density-based property of DBSCAN also becomes the most challenging aspect of implementing DBSCAN for singlet identification since DBSCAN cannot cluster populations with large differences in densities well. This is particularly problematic since the data collected in many flow cytometry studies displays small cell debris and large cell aggregates along with various background noises, demonstrating dramatically different densities as the true singlet population. Small debris can often be eliminated by applying a threshold during data collection, and I have used a covariance and chi-squared test 2D gaussian algorithm to exclude the majority of those large aggregates and background noise.

In contrast to the previous study, which used this approach with a more stringent threshold to consequently exclude 60% population to achieve singlet data [128], this study only applies a permissive threshold to exclude 5% of those non-singlet data to assist the proper function of DBSCAN. In fact, the 2D gaussian step can be optional if the data quality is good from the beginning. Undoubtedly, it is always more import to maintain a high-standard laboratory sample handling and harvesting protocol that can effectively prevent background noise, small debris, and large cell aggregates. This will also improve the efficiency of the DBSCAN-based UltraFast singlet identification algorithm. Furthermore, in this study, the true ground of the singlet and non-singlet events were established using the current gold-standard manual gating strategy using FSC and SSC signals. In addition, the results were double confirmed by cell sorting followed by widefield microscopy. One unanticipated finding was that the non-singlet

population sorted outside the manual singlet gate also has around 10% singlets. Instead of challenging the gold-standard manual gating strategy, this is more likely to be that the non-singlets were shredded into singlets during cell sorting by the high-speed fluidic flow rate and vibration at the sorting point. However, these results need to be further validated using the flow cytometry-based technique that allows imaging outputs without cell sorting in place. This is an important issue for future research to utilise techniques such as imaging flow cytometry to confirm the true ground of the manual gating identified singlet and non-singlet population. Regardless, the singlet was validated to be nearly 100% accurate from the manual gating and still guaranteed the singlet identification algorithm performance using $F1$ scores, which should help us to maintain the confidence of the robustness and usefulness of the UltraFast algorithm for singlet identification.

This study's collaborative filtering (CF) recommendation algorithm is a novel approach to correct the baseline subtraction error and predict autofluorescence. Although the CF algorithm was first used in 1992 [115] and has been widely used in websites such as Netflix and Amazon with large user databases, this is the first time CF has been used in flow cytometry analysis for data pre-processing. The most significant finding in this study is the power of CF in predicting the nonlinear single-cell level of fluorescent signals, such as baseline and autofluorescence. The error-free performance of CF in flow cytometry data analysis is a remarkable finding and provides new insights into the potential and direction of developing flow cytometry-based algorithms. The CF algorithm outperforms many recently developed linear-based algorithms in removing autofluorescence [90-92, 218].

Moreover, many iterative algorithms have been developed for the downstream spectral unmixing procedure to minimise the spread error caused by spectral compensation, such as the spectral compensation method developed by Roederer (2001) [219, 220], the spillover spreading matrix-based method developed by Nguyen et al. (2013) [87], the generalized unmixing model developed by Novo et al. (2013) [221], and the latest AutoSpill algorithm developed by Roca et al. (2021) [88]. However, all above mentioned approaches ignores the influence of baseline subtraction error and autofluorescence to the fluorescence spillover linearity. Therefore, the errors are always embedded in the fluorescence intensity calculations used in these methods that limits the degree of accuracy they can achieve. In this study, the pre-processing steps exceptionally improved the quality of the fluorescence intensity data by correcting the baseline subtraction error and removing the autofluorescence prior to spectral unmixing, so a simple single-step matrix-based noniterative linear unmixing can fully resolve the fluorescence spillover issue. Thus, it saves computational time without the need for iterative computations. In general, therefore, the results suggest that when predicting flow cytometry data influenced by many factors, the non-linear CF algorithm is preferable to the linear algorithms and opens the possibility of examining many nonlinear-based algorithms in flow cytometry data pre-processing for future studies. One major limitation of the CF algorithm that did not emerge in this study but should be mentioned is that it may face the challenge of data sparsity. As the CF predicts data based on existing high-quality data, a poor flow cytometry dataset can make CF unreliable. Luckily, this can be prevented by performing hardware QA and configuration baseline checking prior to data acquisition with flow cytometry QA beads. If, in rare situations, when an experiment cannot be conducted repetitively, such as using precious

clinical samples, or the machine has sudden abnormal performance, flowAI and flowClean can be integrated into the pre-processing pipeline to remove data with aberrant flow rate and fluorescence signals, therefore decreasing the relative data sparsity.

Furthermore, the CF algorithm will suffer serious scalability problems as the number of cells and detection parameters grow. For example, in this study, a sample with 30,000 cells and eight fluorescence parameters takes around one minute of the CF processing time for baseline correction and autofluorescence removal. Current spectral flow cytometers have 64 detection channels and can perform staining of cell surface markers simultaneously using 40 colours in a single tube. The CF processing time will dramatically increase when using full-spectrum flow cytometry. This raises intriguing questions regarding how to improve the CF algorithm's efficiency or perhaps find a more efficient strategy for applying the CF algorithm. In future research, users can first apply unsupervised clustering algorithms to perform cell subtype identification and then only carry out the CF processing for this cell population as a potential solution to maintain efficiency if the cell number is increased. Meanwhile, users can apply the linear parametric dimensionality reduction algorithm PCA to compress the high-parameter flow cytometry data into a dataset with lower parameters and perform CF processing, followed by uncompressing data to the original high-parameter/dimensional space. These investigations are required to evaluate the data information loss and CF accuracy by combining with the PCA or finding a more efficient non-linear algorithm with the same level of accuracy.

The present study was designed to determine the α FRET calibration factor using the unlinked FRET data for the detection of two-protein three-colour FRET signals and the three-protein six-colour FRET signals. The current result indicates that the unlinked single-cell level alpha calibration can achieve higher accuracy and detection sensitivity than the existing calibration method using chemically linked FRET pairs. Surprisingly, the unlinked calibration algorithm formulation has not been fully proposed in previous studies, and most existing single-cell level flow cytometry-based FRET still use chemically linked controls. The chemically linked controls produce artificial readouts and can never represent the naturally existing free-interacting PPIs. This study confirms that the single-cell level unlinked FRET calibration algorithm shares the same physical properties with the previously published method, which used linked controls, and can produce higher FRET accuracy, detection sensitivity, lower background random PPIs readout, and higher reproducibility.

Interestingly, the successful formulation of the unlinked calibration is achieved by and highly relies on the quality of the data pre-processing. It is important to consider the possible restrictions in these approaches, including both previously mentioned linked or newly proposed unlinked calibration methods, as both require the knowledge of the excitation coefficients of the fluorophores used in the FRET analysis. This information can be obtained from many previous publications, and the FPbase online database [117] has a large collection of the property information for many fluorophores, including the eCFP, eGFP, eYFP, mRFP, Violet-Blue, FITC, Aquamarine, NeoGreen, as well as the UV-violet live/dead stain and the miRFP703 for inflammation reporters. Of course, the excitation coefficient can also be measured using a spectrometer.

Moreover, unlike the linked calibration method, the unlinked calibration method removed the requirement of pre-established knowledge or assumption on FRET-donor and FRET-acceptor concentrations. This is a rather useful outcome since the concentration of the FRET-donor and FRET-acceptor cannot be quantitated for most of the naturally existing free-interacting proteins. This is because the donor emission no longer solely correlates to the protein expression level, and it is also impacted by PPI levels that cause FRET-donor emission quenching effects.

Furthermore, the FRET results generated with the flow cytometers in this study should only represents the averaged PPIs at the single-cell level instead of the single FRET-pair level. This is because the unlinked fluorescently tagged proteins might not all participate in the FRET pairs, which means they may not interact with each other, or a proportion of the interacting complex may only have the donor or acceptor fluorophores, i.e., eCFP::eCFP and eYFP::eYFP in dimeric interactions. This interpretation is consistent with our previous observations in which the single-cell level unlinked FRET represent the average interaction level of all fluorescently tagged proteins with different PPI degrees at the single-cell level [32]. Despite many available single-molecule level FRET detection techniques, flow cytometry-based single-cell level FRET detection demonstrates significantly high sampling power and throughput that guarantees unexcelled statistical power.

The compositional data analysis (CoDA) algorithm has recently emerged as a new paradigm for the analysis of omics data, including genomic next-generation sequencing data, proteomic data, and transcriptomics [222]. One of the initial objectives of this study was to use CoDA to describe the three-protein six-colour competitive energy

dynamics among the same FRET donor to two different FRET acceptors. Another objective is to use the CoDA algorithm to resolve the compositional cellular signalling pathway activities activated by the same receptor::receptor interactions in the same cell. Compositional data are multivariate data where relative values rather than absolute values are essential. A compositional sample's constituents, or parts, are often stated as ratios adding up to one or 100% as percentage values. The most obvious finding to emerge from the analysis is that the data can reveal the compositional impacts of each three-way FRET component to the rest of them and maintain sub-compositional coherence using LR transformation. Although the CoDA application and LR transformation has been mentioned in many biological investigations, this study innovatively resolved the FRET energy competition dynamics and relative cellular singling correlations using the CoDA algorithm. The CoDA results demonstrated the capacity to understand the complex PPIs such as TNFR1 trimeric meta-clusters, complicated biological phenomena of human TRAPS autoinflammatory disease, human-poxvirus interactions, and understand the balance between cell death and inflammation signalling pathway activations upon TNFR1 interactions. Despite these promising results, zeros in compositional data are a major issue since taking a logarithmic ratio on zeros is impossible. Although the zero-value issue did not emerge in this study, various zero substitution algorithms have been proposed [223-225], and the impact of any of these on the FRET results can be investigated in the future. In addition to the purely statistical approach to solve the zero-value issue, using domain knowledge to combine components to produce new components can also solve the issue.

Without imaging outputs, flow cytometry measures the fluorescence intensity at the single-cell level in a fluidic system. As a result, it has little information about the morphology of individual cells, and it lacks spatial information that is important for the understanding of living tissues and animals. The cellular morphological data can be investigated for future study using high-throughput multi-channel imaging flow cytometers. It exposes detailed information on the cell size, shape, and circularity, as well as the localisation and colocalization of multiparameter fluorescently tagged markers, including FRETs. The imaging flow cytometry data can also be used to increase singlet identification accuracy without the need for cell sorting.

Moreover, combining the FRET algorithm with the imaging flow cytometry technique exposes the signalling pathways' morphological details, such as the cell's location/organelle where signalling is initiated. More intriguingly, examining the doublet events may permit the examination of the intercellular protein-protein interactions. For spatial information, there is a currently developed live mouse, fluorescent lifetime (FLIM) assay for the detection of the FRET signals using GFP-RFP FRET pair [226]. For future studies, it is also worth exploring the possibility of directly applying the fluorescent intensity-based algorithm to the live animal FRET system because it can be upgraded to more than two protein molecules, and the data acquisition is more rapid than the FLIM assay. Furthermore, utilising the FLIM technique to evaluate each two-protein FRET control can further confirm the fluorescent-based result to achieve better algorithm quality assessment. With the potential to achieve the simultaneous detection of cellular signalling activities and three-protein FRET signals as presented in this study, if integrated with the live animal model, it is possible to monitor malfunctioning PPIs in

organ-specific malignancies and measure the course of the disease in real-time while also visualising the spatial information from many fluorescence readouts. It will provide more biologically relevant information on the connections between FRET and cellular signals, resulting in more effective diagnostic and pharmacological screening tools in animal model-based investigations.

Another direction to expand the future flow cytometry technology is combining it with Raman spectroscopy. In addition to flow cytometry which measures the forward and side scattered light from a suspension of cells or particles as they pass through a laser beam, Raman spectroscopy is a technique that measures inelastic scattered light from a sample [227]. In flow cytometry, FSC measures the size of the cell or particle, while the SSC measures the granularity or internal complexity of the cell or particle. These measurements provide information about the physical properties of the cells or particles and are used to identify and analyze different cell types or subpopulations. On the other hand, Raman spectroscopy provides information about the sample's chemical composition by measuring the inelastic scattered light from the sample's vibrational modes and spectrum shifts [228]. Raman spectroscopy and flow cytometry are both powerful techniques in the field of biotechnology and medical research [229, 230]. Combining these two techniques has the potential to provide even more detailed and accurate information about cells and other biological samples. By integrating Raman spectroscopy with flow cytometry, researchers can simultaneously measure a wide range of properties of individual cells, including their size, shape, and chemical composition. One major advantage of this approach is that it allows for high-throughput analysis of large numbers of cells. The flow cytometry component of the system allows

for the rapid sorting and analysis of cells, while the Raman spectroscopy component provides detailed chemical information. This can be particularly useful for applications such as cancer research, where large numbers of cells need to be analysed to identify specific markers or characteristics associated with the disease. Another advantage of this approach is that Raman spectroscopy can provide information on the chemical composition of cells, which is impossible with traditional flow cytometry methods. This can be useful for identifying specific biomolecules or for detecting changes in cell metabolism. However, one limitation of this approach is that Raman spectroscopy is typically a relatively low signal-to-noise technique, making it difficult to obtain accurate and reliable data in some cases. Additionally, integrating Raman spectroscopy with flow cytometry can be a complex and technically challenging task, requiring specialised equipment and expertise. Overall, the integration of Raman spectroscopy with flow cytometry has the potential to provide powerful new insights into the biology of cells and other biological samples. However, it is important to keep in mind the limitations of this approach and to carefully consider the specific experimental goals and requirements when planning and executing these types of experiments.

In the foreseeable future, the detection parameters of flow cytometry-based technology will grow by expanding the number of fluorescence detection channels and the combination of other technologies. Meanwhile, the complexity of data visualisation will also grow exponentially as the detection parameter increases, this phenomenon is known as "the curse of dimensionality". Currently, the most advanced full spectral flow cytometer has 64 fluorescence detection channels, which requires 2,016 two-dimensional scatterplots to examine the data, if all channels are used adequately. Finally,

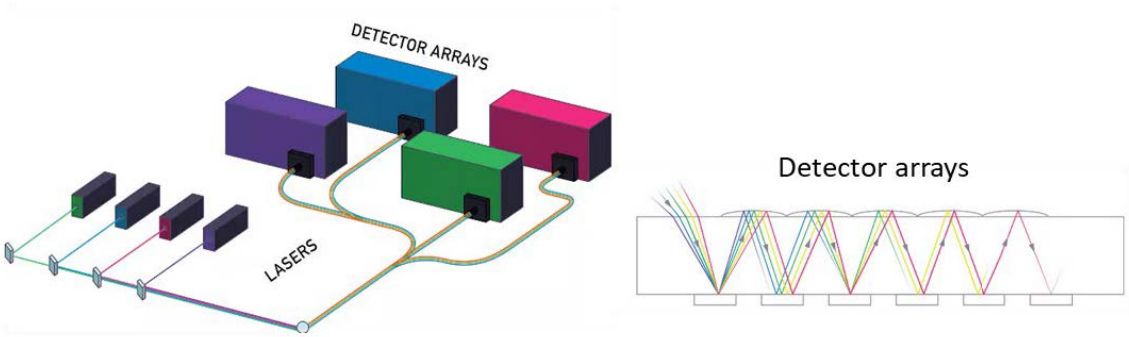
to resolve the visualisation complexity when the detection parameter expands, there are many parametric and non-parametric dimension reduction algorithms. PCA, t-distributed stochastic neighbour embedding (t-SNE) [231], uniform manifold approximation and projection (UMAP)[232], the dimensionality reduction technique based on triplet constraints (TriMap) [233], and pairwise controlled manifold approximation (PaCMAP) [234], are all dimensionality reduction techniques available to compress the high parameter/dimensional data in a single reduced two-dimensional scatter plot. PCA is a linear parametric technique that seeks to find the principal components of the data, which are the directions of greatest variance. It is simple to implement and computationally efficient, but it can be sensitive to outliers and may not preserve the local structure of the data. The t-SNE algorithm is a non-parametric technique that seeks to preserve the local structure of the data by minimising the divergence between probability distributions of similar data points. UMAP, like t-SNE, also preserves the local structure of the data, but also preserves the global structure by approximating the underlying manifold of the data. TriMap and PaCMAP are also non-parametric techniques, but they are based on triplet constraints and pairwise controlled manifold approximation, respectively. These non-parametric techniques can preserve both the local and global structure of the data by hyperparameter tuning, but they can be more computationally expensive than PCA. In summary, the choice of dimensionality reduction technique will depend on the specific characteristics of the cells and sample populations and the goals of the biological investigations.

Despite the limitations of flow cytometry-based assay, it still has distinct advantages over many other technologies. It enables the user to carry out single-cell level cell-

sorting, which opens doors to deeper investigations, including single-cell DNA and RNA sequencing, single-cell proliferation and progression modelling on selected and sorted cell subtypes. This will help researchers better understand diseases like cancer, which can develop from a single mutated cell. Together with the FRET and cellular signalling reporters, flow cytometry still holds an unbeatable sample acquisition rate, sampling power, statistical robustness, detection sensitivity, and high-parameter capacity. Linking flow cytometry data with genomic, proteomic, and transcriptomic data permits multi-level investigation of the causation for physiological conditions or biological phenomena and allows the high-throughput diagnosis of human diseases and screening of personalised therapeutics that can contribute greatly to human health.

Appendix

Appendix Figure 1 Full-spectrum flow cytometer configuration
The violet, blue, yellow-green and red lasers and their detector arrays. The detector arrays each have 16, 14,10, and 8 detector channels.

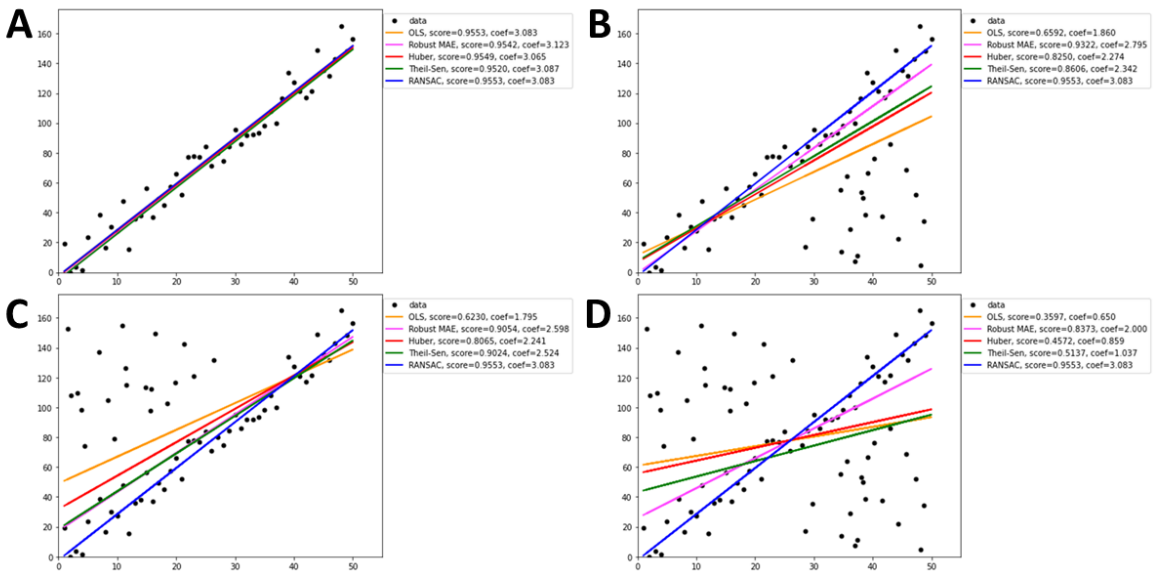


Laser	Channel	Center Wavelength (nm)	Bandwidth (nm)	Wavelength Start (nm)	Wavelength End (nm)
Violet	V1	428	15	420	435
	V2	443	15	436	451
	V3	458	15	451	466
	V4	473	15	466	481
	V5	508	20	498	518
	V6	525	17	516	533
	V7	542	17	533	550
	V8	581	19	571	590
	V9	598	20	588	608
	V10	615	20	605	625
	V11	664	27	651	678
	V12	692	28	678	706
	V13	720	29	706	735
	V14	750	30	735	765
	V15	780	30	765	795
	V16	812	34	795	829
Red	R1	660	17	652	669
	R2	678	18	669	687
	R3	697	19	688	707
	R4	717	20	707	727
	R5	738	21	728	749
	R6	760	23	749	772
	R7	783	23	772	795
	R8	812	34	795	829

Laser	Channel	Center Wavelength (nm)	Bandwidth (nm)	Wavelength Start (nm)	Wavelength End (nm)
Blue	B1	508	20	498	518
	B2	525	17	516	533
	B3	542	17	533	550
	B4	581	19	571	590
	B5	598	20	588	608
	B6	615	20	605	625
	B7	660	17	652	669
	B8	678	18	669	687
	B9	697	19	688	707
	B10	717	20	707	727
	B11	738	21	728	749
	B12	760	23	749	772
	B13	783	23	772	795
	B14	812	34	795	829
Yellow Green	YG1	577	20	567	587
	YG2	598	20	588	608
	YG3	615	20	605	625
	YG4	660	17	652	669
	YG5	678	18	669	687
	YG6	697	19	688	707
	YG7	720	29	706	735
	YG8	750	30	735	765
	YG9	780	30	765	795
	YG10	812	34	795	829

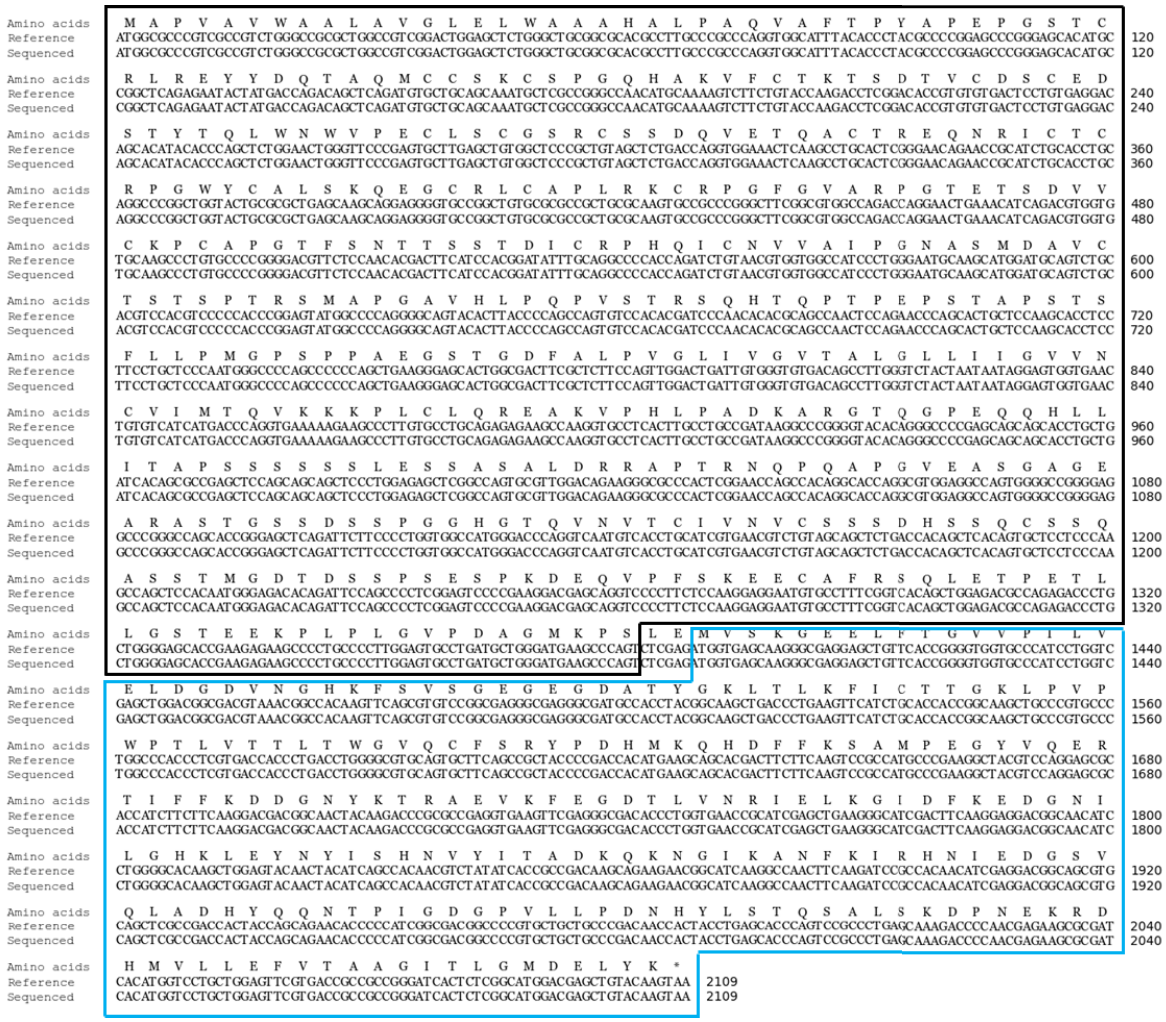
Appendix Figure 2 The RANSAC robust linear regression

OLS (orange lines), robust MAE (pink lines), Huber (red lines), Theil-Sen (green lines), and RANSAC (blue lines) were compared using data with different outliers. The ground truth coefficient was 3, with no outlier (A), outliers at the x-axis direction (B), outliers at the y-axis direction (C), and outliers at both the x-axis and y-axis directions (D).



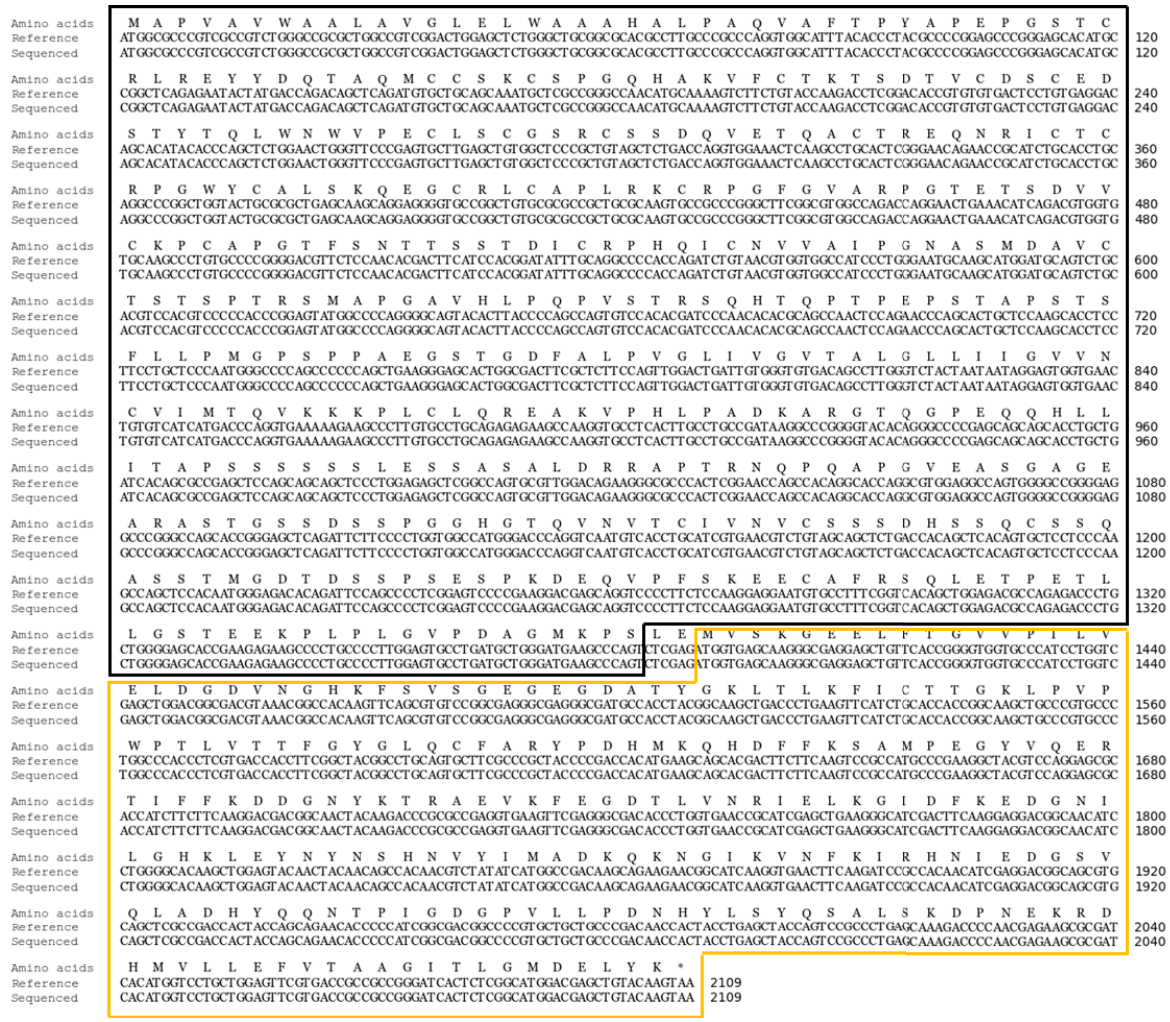
Appendix Figure 3 TNFR2-eCFP sequencing result

Human TNFR2 (black box) subcloned with the eCFP (cyan box) to form a fusion fluorescence receptor TNFR2-eCFP. The sequencing result showed no mutation.



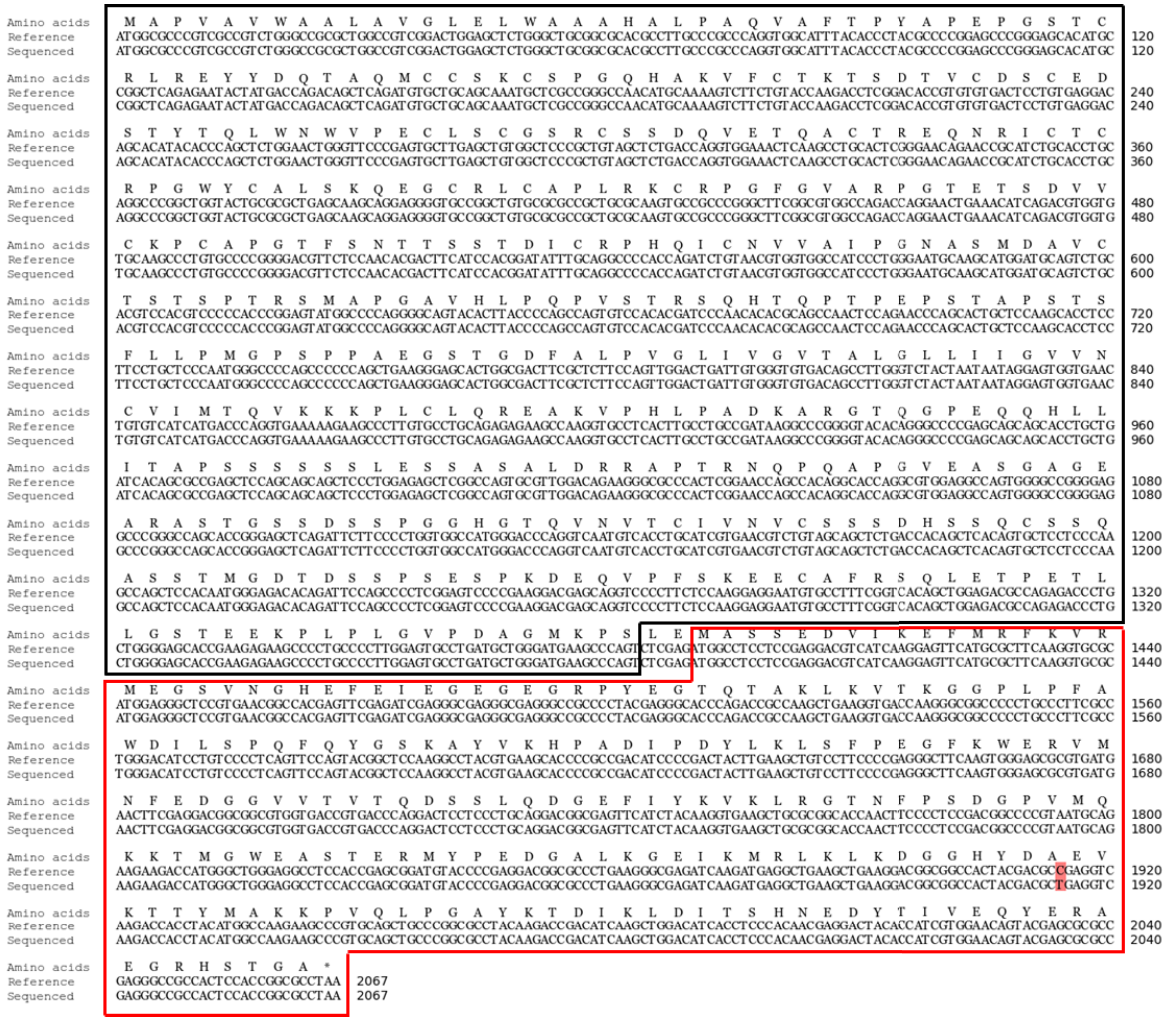
Appendix Figure 4 TNFR2-eYFP sequencing result

Human TNFR2 (black box) subcloned with the eYFP (yellow box) to form a fusion fluorescence receptor TNFR2-eYFP. The sequencing result showed no mutation.



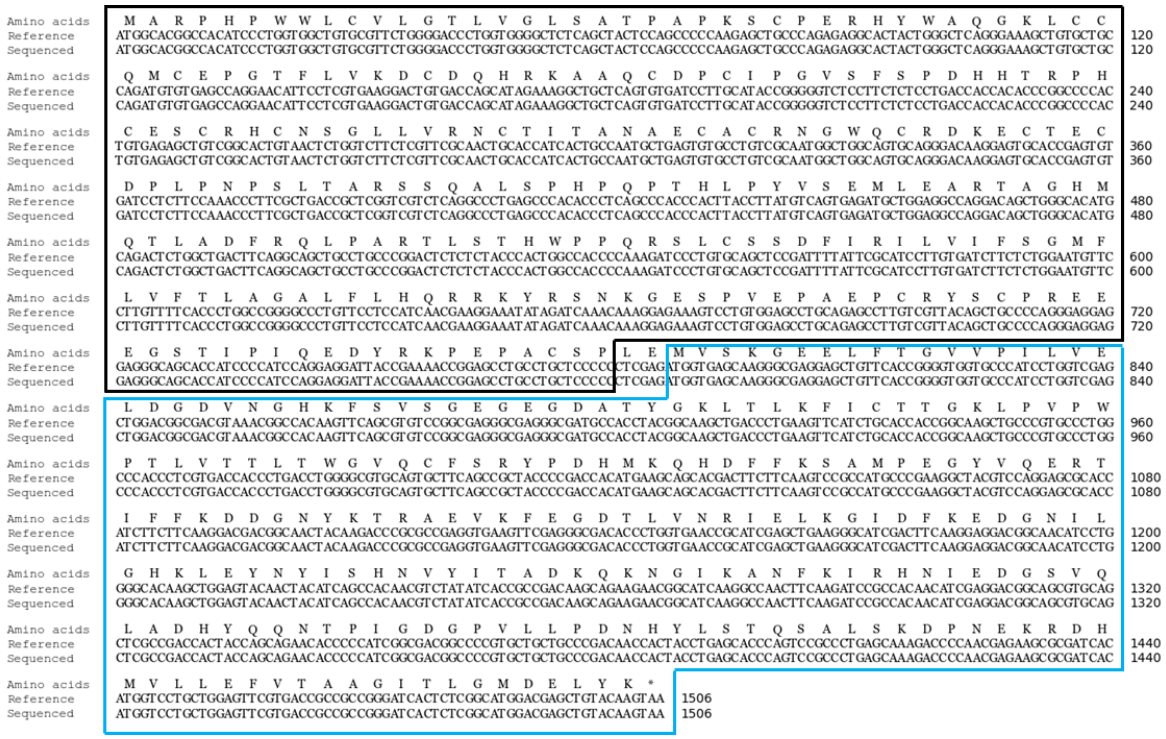
Appendix Figure 5 TNFR1-mRFP sequencing result

Human TNFR2 (black box) subcloned with the mRFP (red box) to form a fusion fluorescence receptor TNFR2-mRFP. No mutation in the TNFR2 sequence. One silent mutation was found in the mRFP ORF: GCC→GCT, alanine→alanine (red) at the 525th nucleotide.



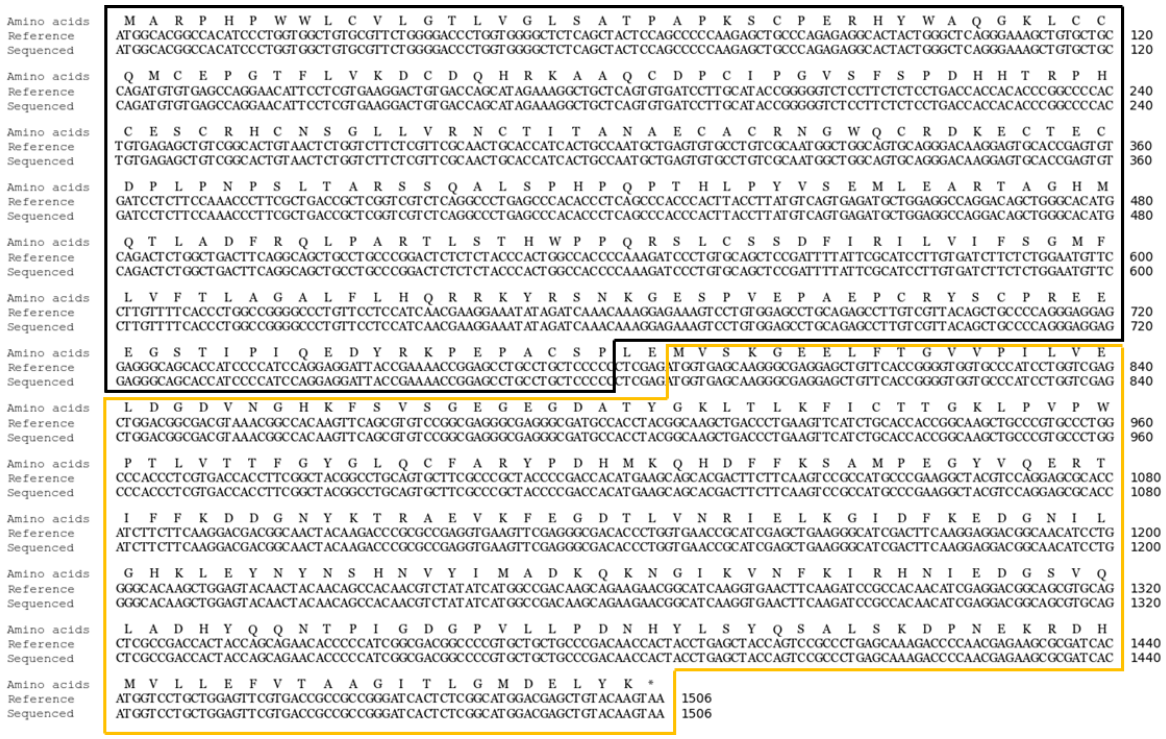
Appendix Figure 6 CD27-eCFP sequencing result

Human CD27 (black box) subcloned with the eCFP (cyan box) to form a fusion fluorescence receptor CD27-eCFP. The sequencing result showed no mutation.



Appendix Figure 7 CD27-eYFP sequencing result

Human CD27 (black box) subcloned with the eYFP (yellow box) to form a fusion fluorescence receptor CD27-eYFP. The sequencing result showed no mutation.

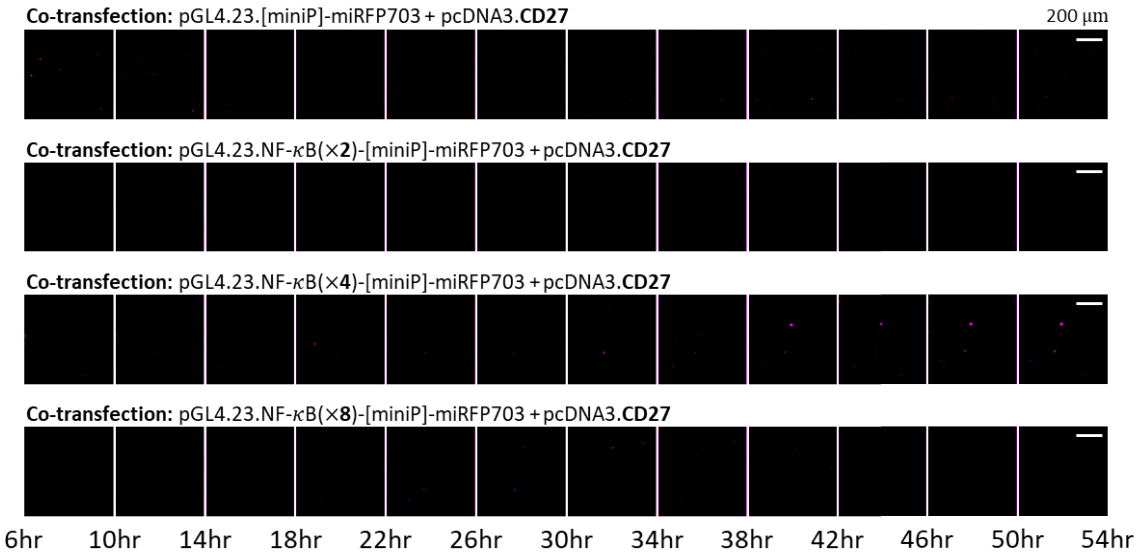


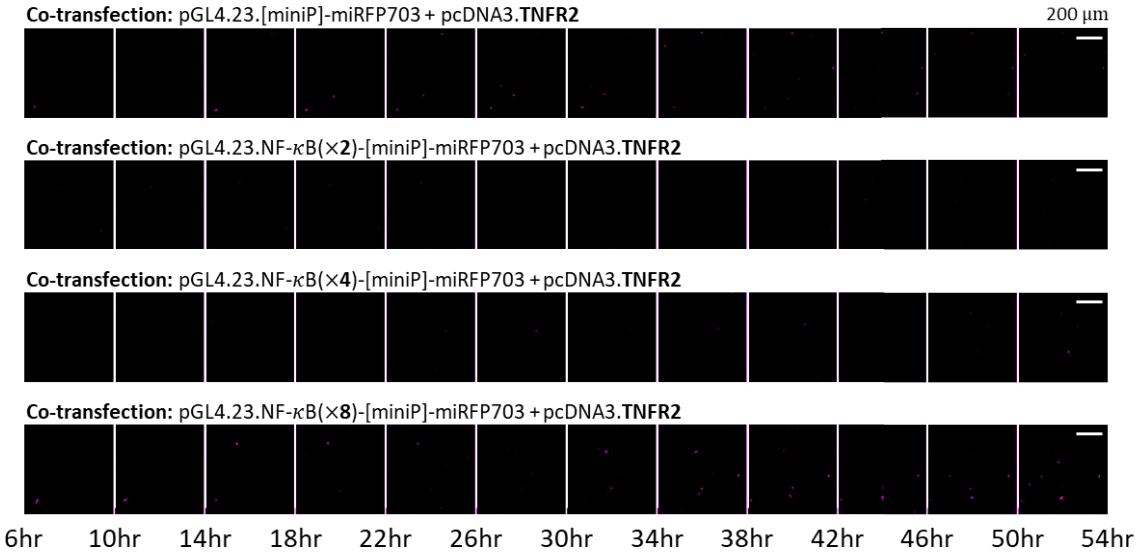
Appendix Figure 8 CD27-mRFP sequencing result

Human CD27 (black box) subcloned with the mRFP (red box) to form a fusion fluorescence receptor CD27-mRFP. No mutation in the CD27 sequence. One silent mutation was found in the mRFP ORF: GCC→GCT, alanine→alanine (red) at the 525th nucleotide.

Appendix Figure 9 Live-cell imaging of NF-κB reporter and TNFR2 or CD27 co-transfected HEK-293T cells

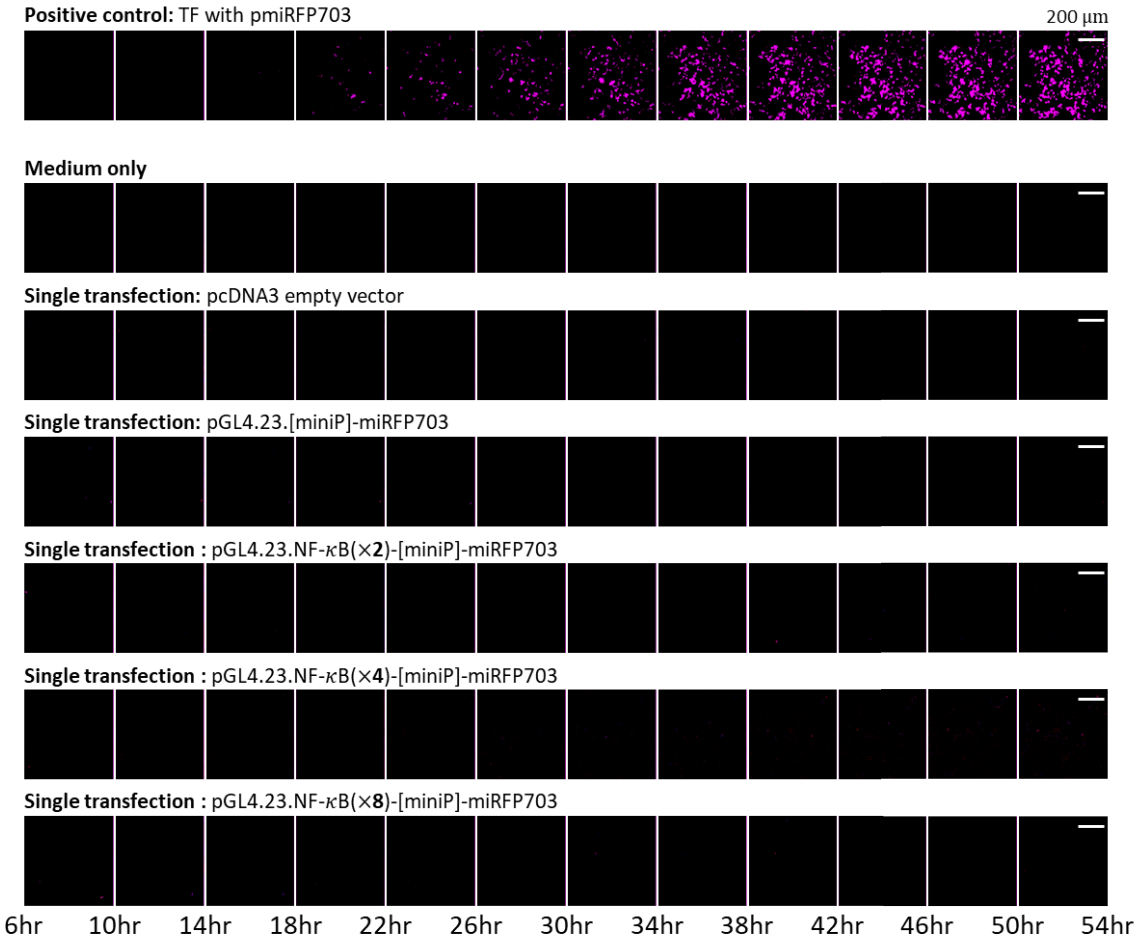
No significant expression of miRFP703 was detected for HEK-293T cells co-transfected with the TNFR2 or CD27 expressing plasmids and the NF-κB reporter plasmids with 0×, 2×, 4×, or 8× response elements.





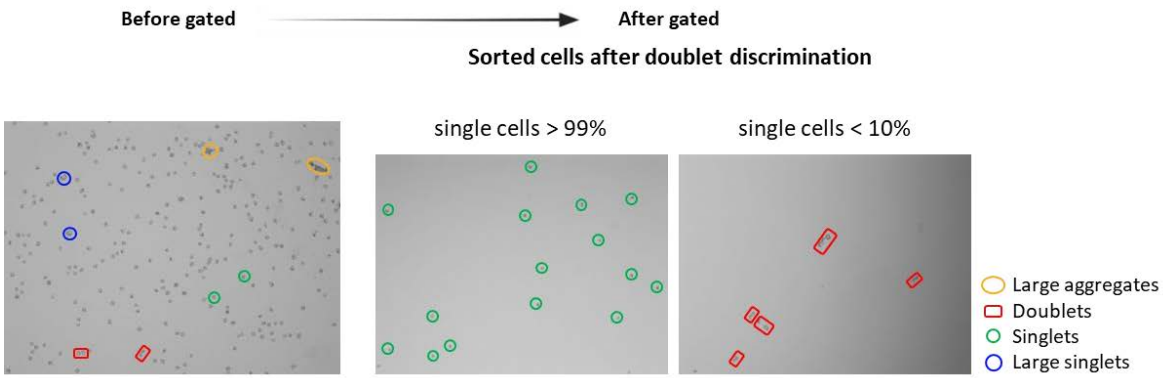
Appendix Figure 10 Live-cell imaging of NF-κB reporter control samples

No significant expression of miRFP703 was detected for untransfected or mock-transfected HEK-293T cells. No significant expression of miRFP703 was detected for HEK-293T cells transfected with only the NF-κB reporter plasmids (i.e., not co-transfection with receptor plasmid). Positive expression of miRFP703 was detected for cells transfected with pmiRFP703 plasmid with the constitutive CMV promoter.



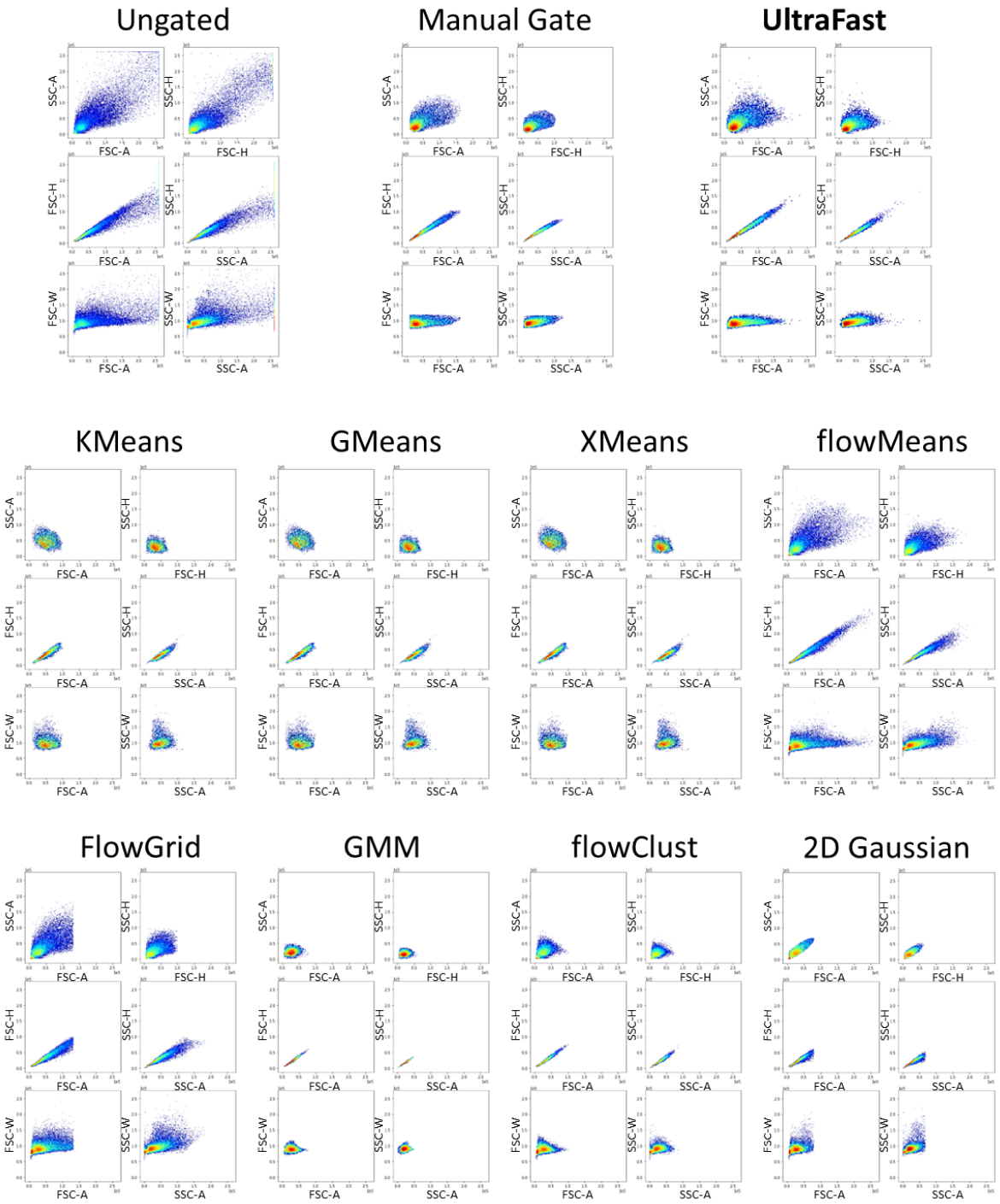
Appendix Figure 11 Manual gated singlet examination

Harvested cells were sorted into singlets and non-singlets and examined by widefield microscopy.

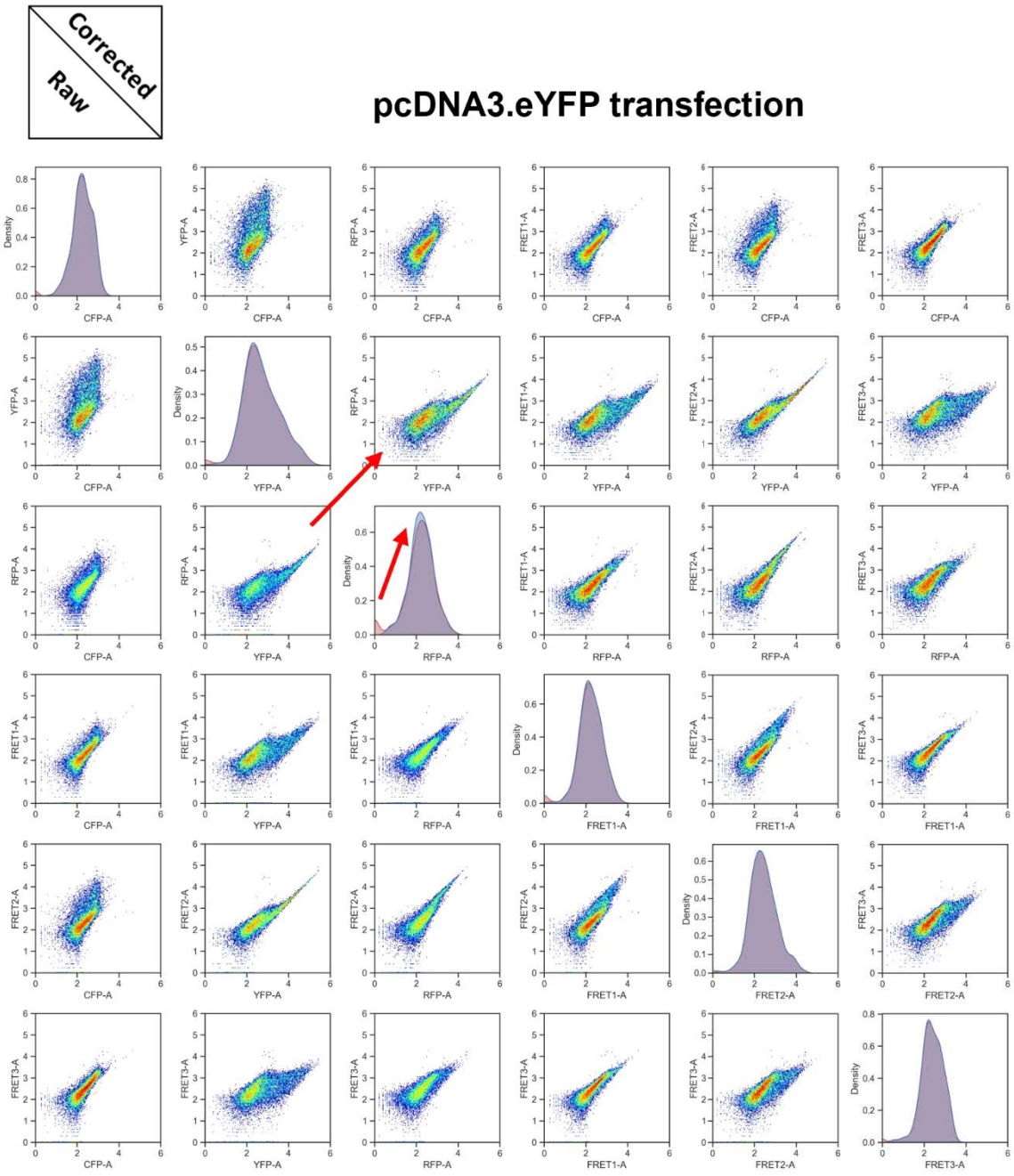


Appendix Figure 12 Singlet identification using different algorithms

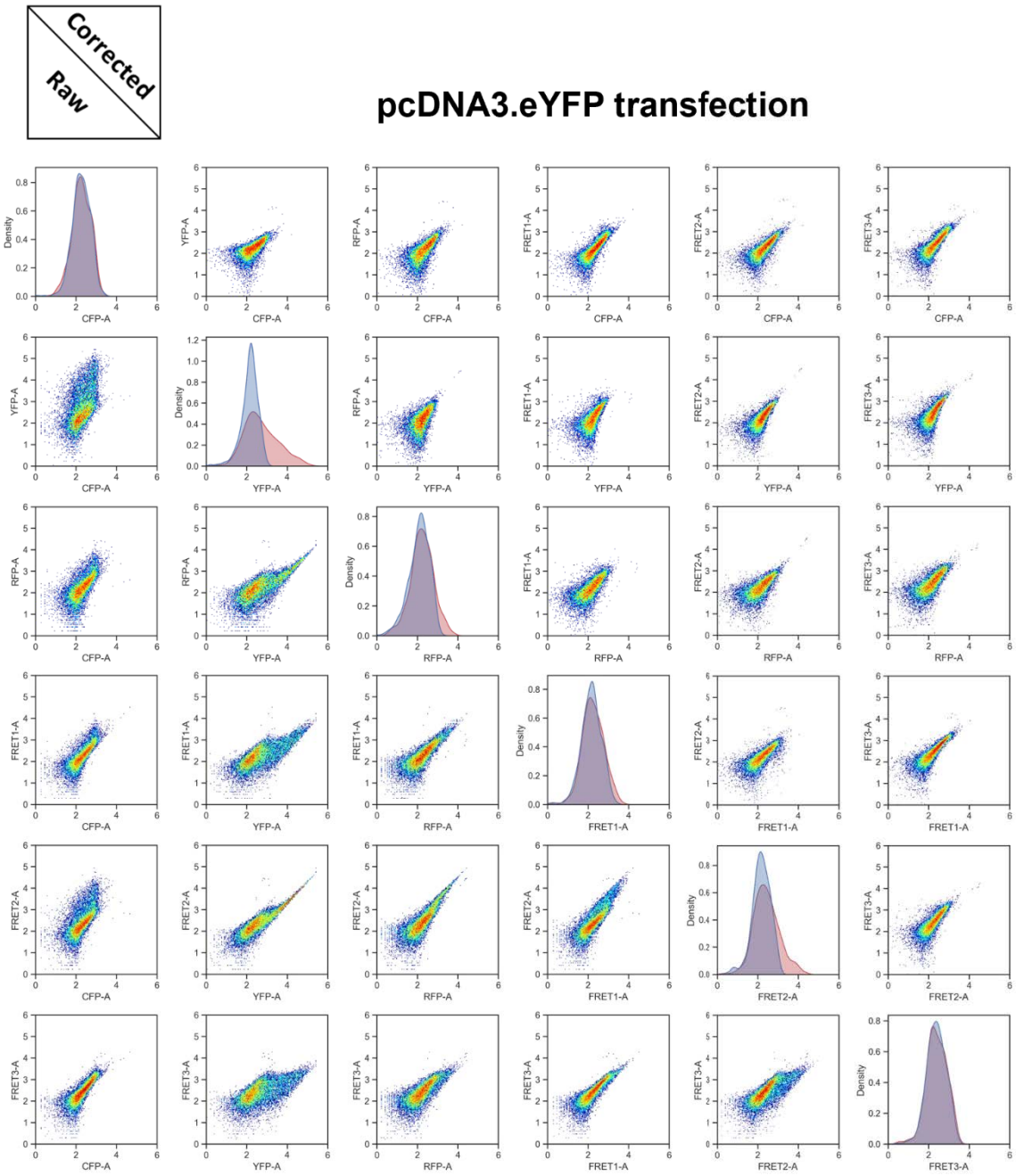
The scatter plot profiles of the ungated cells, manually gated singlets, and singlets were identified with different algorithms. The six scatter plots are the same ones described in the well-established manual gating approach.



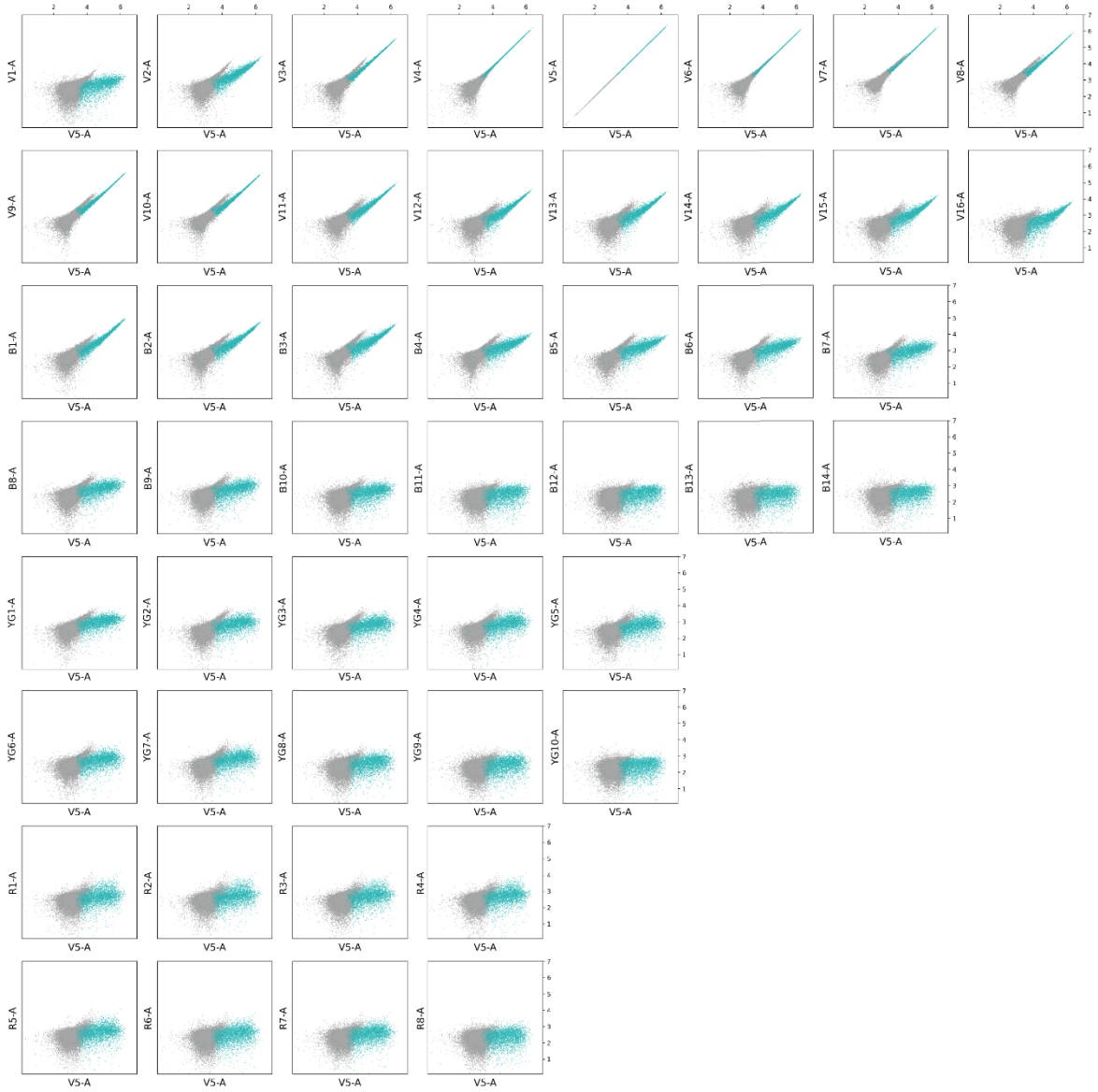
Appendix Figure 13 Baseline subtraction-error correction for eYFP expressing cells
Scatterplot and histogram overlays of the eYFP expressing sample before (red histograms) and after (blue histograms) baseline subtraction-error correction. The scatter plots are symmetric along the diagonal histogram plots.



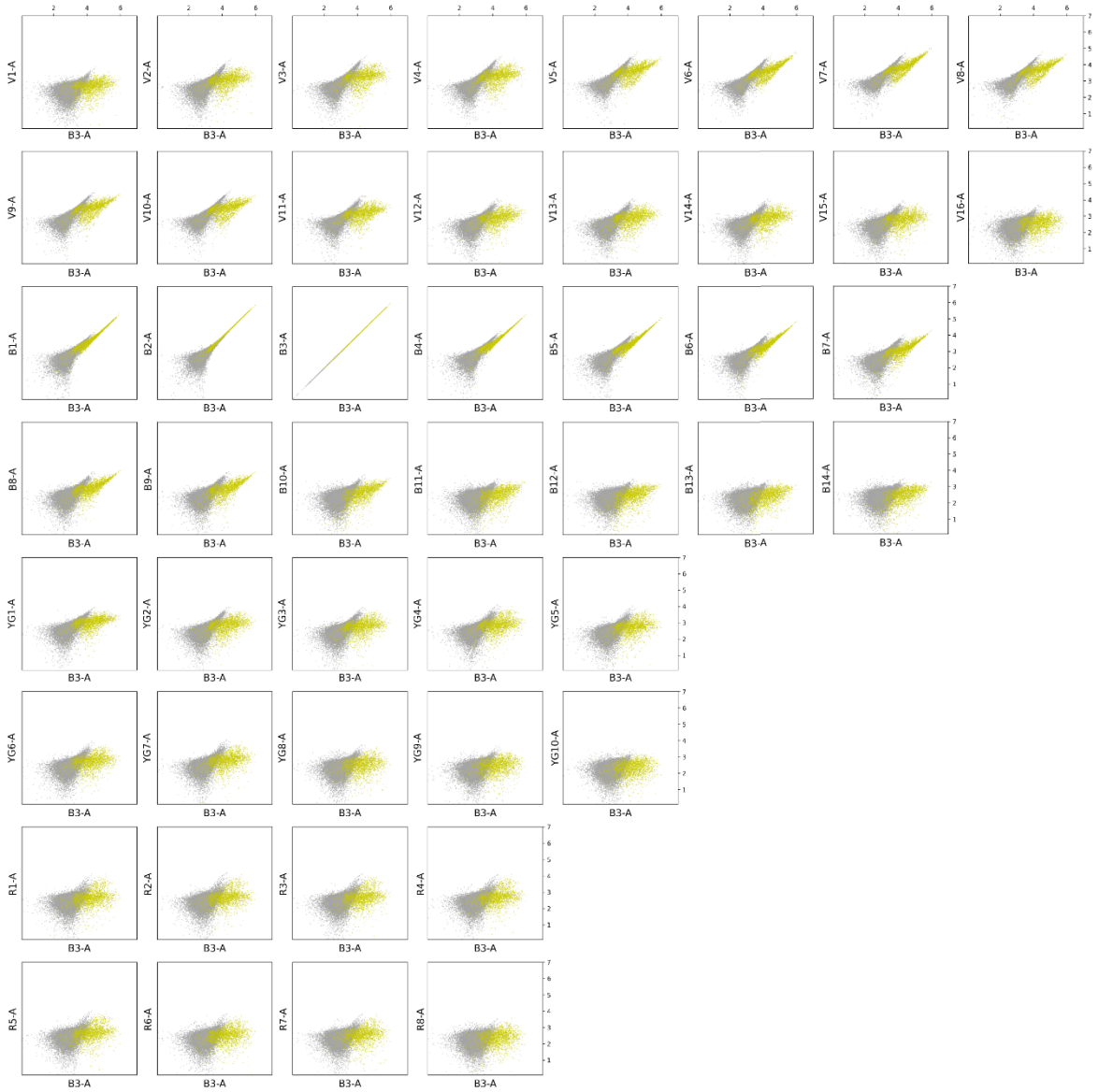
Appendix Figure 14 Autofluorescence prediction for eYFP expressing cells
Scatterplot and histogram overlays of the eYFP expressing sample with the total fluorescence (red histograms) and autofluorescence (blue histograms) detection values. The scatter plots are symmetric along the diagonal histogram plots.



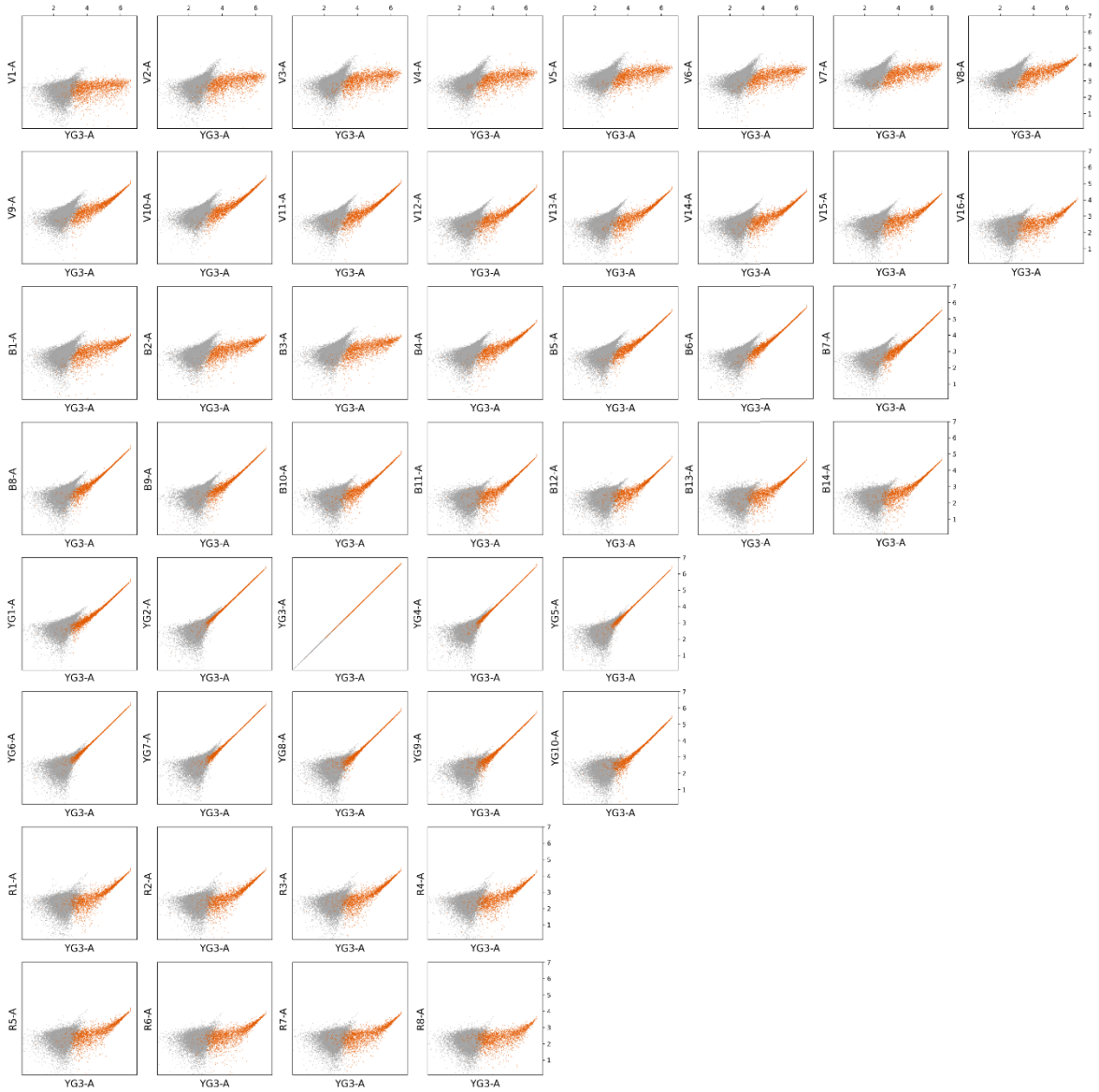
Appendix Figure 15 Unsupervised cell subtype clustering for eCFP transfected cells
The GMM unsupervised algorithm identified the eCFP-positive and eCFP-negative populations within the sample. A total of 48 channels from the full-spectrum flow cytometers were used.



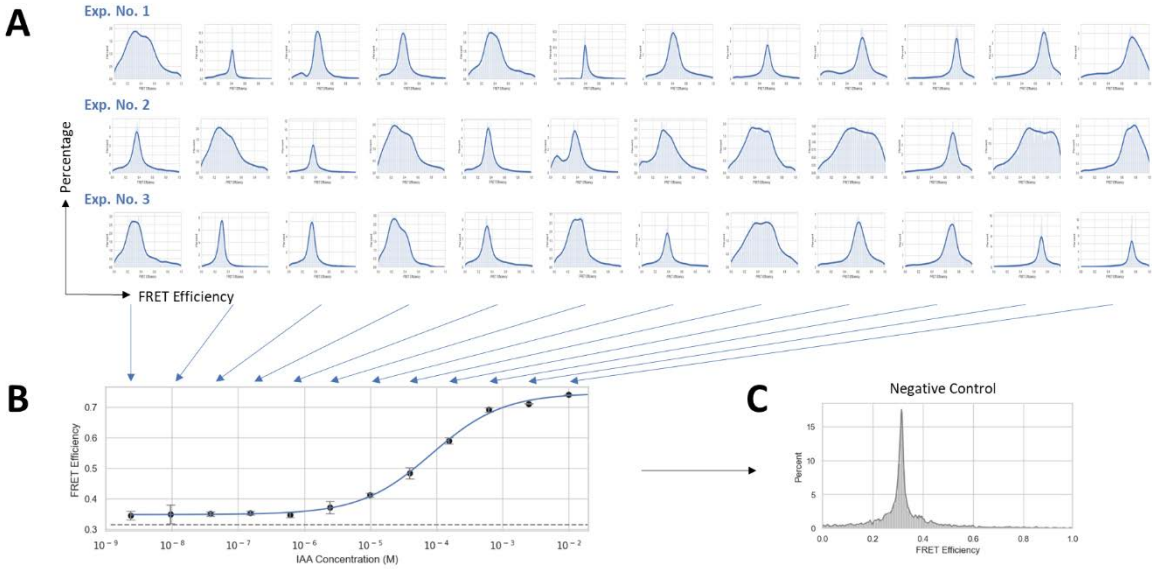
Appendix Figure 16 Unsupervised cell subtype clustering for eYFP transfected cells
The GMM unsupervised algorithm identified the eYFP-positive and eYFP-negative populations within the sample. A total of 48 channels from the full-spectrum flow cytometers were used.



Appendix Figure 17 Unsupervised cell subtype clustering for mRFP transfected cells
The GMM unsupervised algorithm identified the mRFP-positive and mRFP-negative populations within the sample. A total of 48 channels from the full-spectrum flow cytometers were used.

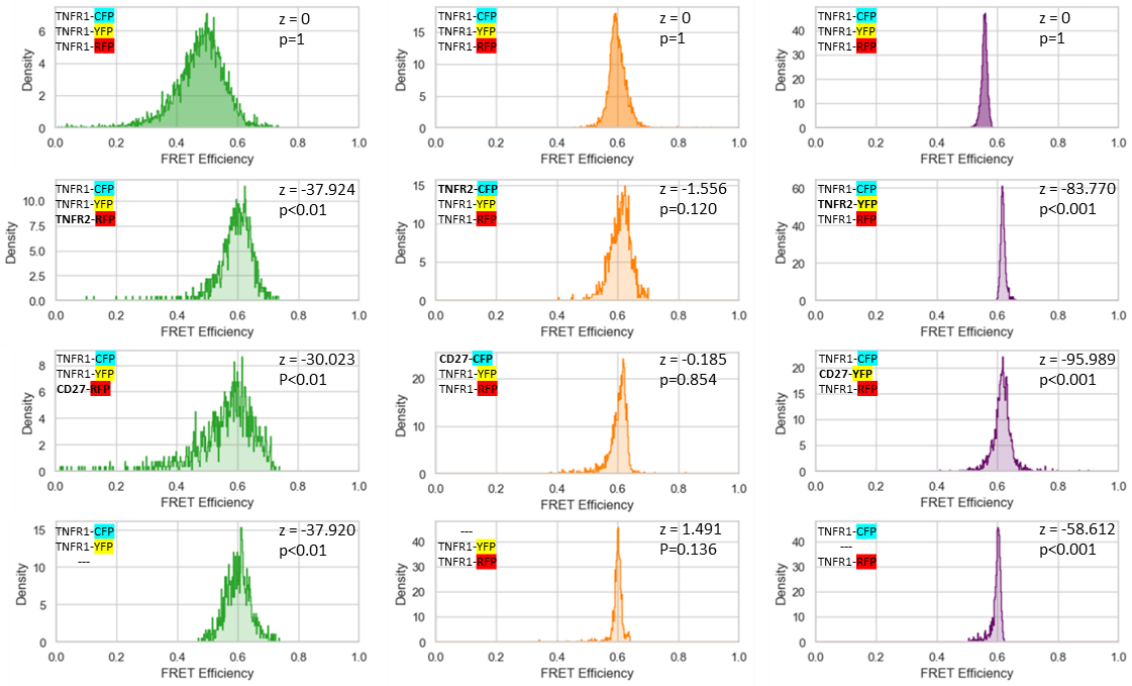


Appendix Figure 18 Single-cell level FRET distributions of the Auxin detection
Histograms demonstrate the distributions of calculated FRET efficiencies from three replicate experiments (A). Each can be used for data quality assessment for the small molecule Auxin concentration detection profile (B). Meanwhile, the single-cell level data also allowed the quality assessment of the negative control sample (C).



Appendix Figure 19 Histogram of FRET efficiencies for different TNFR receptor combinations

Histograms demonstrate the distributions of calculated FRET efficiencies with CFP→YFP FRET as green, YFP→RFP FRET as orange, and CFP→RFP FRET as purple. Cells transfected with TNFR1-CFP + TNFR1-YFP + TNFR1-RFP are as control sample (top row). The z values and p values are results from z-test for each sample compared with the relative control sample.



References

1. PubMed. *National Library of Medicine (US), National Center for Biotechnology Information (NCBI)* 2023 [cited 2023 Janurary 1st]; Available from: https://pubmed.ncbi.nlm.nih.gov.
2. Förster, T., *Energiewanderung und fluoreszenz.* Naturwissenschaften, 1946. **33**(6): p. 166-175.
3. Stryer, L. and R.P. Haugland, *Energy transfer: a spectroscopic ruler.* Proc Natl Acad Sci USA, 1967. **58**(2): p. 719-26.
4. Yang, Z., et al., *Super-resolution fluorescent materials: an insight into design and bioimaging applications.* Chem Soc Rev, 2016. **45**(17): p. 4651-67.
5. Matko, J., et al., *Analysis of cell surface molecular distributions and cellular signaling by flow cytometry.* J Fluoresc, 1994. **4**(4): p. 303-14.
6. Nagy, P., et al., *EGF-induced redistribution of erbB2 on breast tumor cells: flow and image cytometric energy transfer measurements.* Cytometry, 1998. **32**(2): p. 120-31.
7. Nagy, P., et al., *rFRET: A comprehensive, Matlab-based program for analyzing intensity-based ratiometric microscopic FRET experiments.* Cytometry A, 2016. **89**(4): p. 376-84.
8. Ujlaky-Nagy, L., et al., *Flow cytometric FRET analysis of protein interactions.* Methods Mol Biol, 2018. **1678**: p. 393-419.
9. Cossarizza, A., et al., *Guidelines for the use of flow cytometry and cell sorting in immunological studies (third edition).* Eur J Immunol, 2021. **51**(12): p. 2708-3145.
10. Nagy, P., et al., *Novel calibration method for flow cytometric fluorescence resonance energy transfer measurements between visible fluorescent proteins.* Cytometry A, 2005. **67**(2): p. 86-96.
11. Roszik, J., et al., *Evaluation of intensity-based ratiometric FRET in image cytometry--approaches and a software solution.* Cytometry A, 2009. **75**(9): p. 761-7.
12. Banning, C., et al., *A flow cytometry-based FRET assay to identify and analyse protein-protein interactions in living cells.* PLoS One, 2010. **5**(2): p. e9344.
13. He, L., et al., *A flow cytometric method to detect protein-protein interaction in living cells by directly visualizing donor fluorophore quenching during CFP-->YFP fluorescence resonance energy transfer (FRET).* Cytometry A, 2003. **55**(2): p. 71-85.
14. Nguyen, A.W. and P.S. Daugherty, *Evolutionary optimization of fluorescent proteins for intracellular FRET.* Nat Biotechnol, 2005. **23**(3): p. 355-60.
15. Ohashi, T., et al., *An experimental study of GFP-based FRET, with application to intrinsically unstructured proteins.* Protein Sci, 2007. **16**(7): p. 1429-38.
16. Nedbal, J., et al., *Time-domain microfluidic fluorescence lifetime flow cytometry for high-throughput Förster resonance energy transfer screening.* Cytometry A, 2015. **87**(2): p. 104-18.
17. Fábián, Á., et al., *TripleFRET measurements in flow cytometry.* Cytometry A, 2013. **83**(4): p. 375-85.

18. He, L., et al., *Determination of tumor necrosis factor receptor-associated factor trimerization in living cells by CFP->YFP->mRFP FRET detected by flow cytometry.* Nucleic Acids Res, 2005. **33**(6): p. e61.
19. Galperin, E., V.V. Verkhusha, and A. Sorkin, *Three-chromophore FRET microscopy to analyze multiprotein interactions in living cells.* Nat Methods, 2004. **1**(3): p. 209-17.
20. *Cytek Aurora - Technology*. 2023 [cited 2023 Janurary 1st]; Available from: https://www.cytekbio.com/pages/aurora.
21. Zhou, H., et al., *Accurate investigation on the fluorescence resonance energy transfer between single organic molecules and monolayer WSe2 by quantum coherent modulation-enhanced single-molecule imaging microscopy.* J Phys Chem Lett, 2019. **10**(11): p. 2849-2856.
22. Miyawaki, A., et al., *Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin.* Nature, 1997. **388**(6645): p. 882-7.
23. Zhang, J., et al., *Creating new fluorescent probes for cell biology.* Nat Rev Mol Cell Biol, 2002. **3**(12): p. 906-18.
24. Bene, L., et al., *Depolarized FRET (depolFRET) on the cell surface: FRET control by photoselection.* Biochim Biophys Acta, 2016. **1863**(2): p. 322-34.
25. Förster, T., *10th Spiers Memorial Lecture. Transfer mechanisms of electronic excitation.* Discussions of the Faraday Society, 1959. **27**: p. 7-17.
26. Haugland, R.P., J. Yguerabide, and L. Stryer, *Dependence of the kinetics of singlet-singlet energy transfer on spectral overlap.* Proc Natl Acad Sci USA, 1969. **63**(1): p. 23-30.
27. *FRET tells us about proximities, distances, orientations and dynamic properties.* Rev Mol Biotechnol, 2002. **82**(3): p. 177-179.
28. Stryer, L., *Fluorescence energy transfer as a spectroscopic ruler.* Annu Rev Biochem, 1978. **47**: p. 819-46.
29. Dale, R.E., J. Eisinger, and W.E. Blumberg, *The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer.* Biophys J, 1979. **26**(2): p. 161-93.
30. Woehler, A., *Simultaneous quantitative live cell imaging of multiple FRET-based biosensors.* PLoS One, 2013. **8**(4): p. e61096.
31. Sun, Y., et al., *Three-color spectral FRET microscopy localizes three interacting proteins in living cells.* Biophys J, 2010. **99**(4): p. 1274-83.
32. Ni, Z., et al., *Analysis insights for three FRET pairs of chemically unlinked two-molecule FRET cytometry.* Cytometry A, 2022. **101**(5): p. 387-399.
33. Szabo, A., et al., *Quantitative characterization of the large-scale association of ErbB1 and ErbB2 by flow cytometric homo-FRET measurements.* Biophys J, 2008. **95**(4): p. 2086-96.
34. Herud-Sikimic, O., et al., *A biosensor for the direct visualization of auxin.* Nature, 2021. **592**(7856): p. 768-772.
35. Chen, H., et al., *FRET-based sensor for visualizing pH variation with colorimetric/ratiometric strategy and application for bioimaging in living cells, bacteria and zebrafish.* Analyst, 2020. **145**(12): p. 4283-4294.
36. Horvath, G., et al., *Selecting the right fluorophores and flow cytometer for fluorescence resonance energy transfer measurements.* Cytometry A, 2005. **65**(2): p. 148-57.

37. Sebestyen, Z., et al., *Long wavelength fluorophores and cell-by-cell correction for autofluorescence significantly improves the accuracy of flow cytometric energy transfer measurements on a dual-laser benchtop flow cytometer.* Cytometry, 2002. **48**(3): p. 124-35.
38. Shrestha, D., et al., *Understanding FRET as a research tool for cellular studies.* Int J Mol Sci, 2015. **16**(4): p. 6718-56.
39. Roy, R., S. Hohng, and T. Ha, *A practical guide to single-molecule FRET.* Nat Methods, 2008. **5**(6): p. 507-16.
40. Bolbat, A. and C. Schultz, *Recent developments of genetically encoded optical sensors for cell biology.* Biol Cell, 2017. **109**(1): p. 1-23.
41. Shaner, N.C., et al., *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.* Nat Biotechnol, 2004. **22**(12): p. 1567-72.
42. Gregory, A.P., et al., *TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis.* Nature, 2012. **488**(7412): p. 508-511.
43. Rohilla, S., et al. *Multi-target immunofluorescence using spectral FLIM-FRET for separation of undesirable antibody cross-labeling*. in *European Conference on Biomedical Optics*. 2019. Optical Society of America.
44. Dey, A., et al., *A multi-colour confocal microscopy method for identifying and enumerating macrophage subtypes and adherent cells in the stromal vascular fraction of human adipose.* J Immunol Methods, 2021. **491**: p. 112988.
45. Shimomura, O., F.H. Johnson, and Y. Saiga, *Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea.* J Cell Comp Physiol, 1962. **59**: p. 223-39.
46. Prasher, D.C., et al., *Primary structure of the Aequorea victoria green-fluorescent protein.* Gene, 1992. **111**(2): p. 229-33.
47. Chalfie, M., et al., *Green fluorescent protein as a marker for gene expression.* Science, 1994. **263**(5148): p. 802-5.
48. Inouye, S. and F.I. Tsuji, *Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein.* FEBS Lett, 1994. **341**(2-3): p. 277-80.
49. Sapsford, K.E., L. Berti, and I.L. Medintz, *Materials for fluorescence resonance energy transfer analysis: beyond traditional donor-acceptor combinations.* Angew Chem Int Ed Engl, 2006. **45**(28): p. 4562-89.
50. Jones, A.M., D.W. Ehrhardt, and W.B. Frommer, *A never ending race for new and improved fluorescent proteins.* BMC Biol, 2012. **10**: p. 39.
51. Davidson, M.W. and R.E. Campbell, *Engineered fluorescent proteins: innovations and applications.* Nat Methods, 2009. **6**(10): p. 713-17.
52. Miyawaki, A. and R.Y. Tsien, *Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein.* Methods Enzymol, 2000. **327**: p. 472-500.
53. Campbell, R.E., et al., *A monomeric red fluorescent protein.* Proc Natl Acad Sci USA, 2002. **99**(12): p. 7877-82.
54. Scott, B.L. and A.D. Hoppe, *Optimizing fluorescent protein trios for 3-Way FRET imaging of protein interactions in living cells.* Sci Rep, 2015. **5**: p. 10270.
55. Nagai, T., et al., *A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications.* Nat Biotechnol, 2002. **20**(1): p. 87-90.

56. Rekas, A., et al., *Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity.* J Biol Chem, 2002. **277**(52): p. 50573-8.
57. Murakoshi, H., S.J. Lee, and R. Yasuda, *Highly sensitive and quantitative FRET-FLIM imaging in single dendritic spines using improved non-radiative YFP.* Brain Cell Biol, 2008. **36**(1-4): p. 31-42.
58. Bajar, B.T., et al., *Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting.* Sci Rep, 2016. **6**: p. 20889.
59. Betolngar, D.B., et al., *pH sensitivity of FRET reporters based on cyan and yellow fluorescent proteins.* Anal Bioanal Chem, 2015. **407**(14): p. 4183-93.
60. Spiess, E., et al., *Two-photon excitation and emission spectra of the green fluorescent protein variants ECFP, EGFP and EYFP.* J Microsc, 2005. **217**(Pt 3): p. 200-4.
61. Rizzo, M.A., et al., *An improved cyan fluorescent protein variant useful for FRET.* Nat Biotechnol, 2004. **22**(4): p. 445-9.
62. Kremers, G.J., et al., *Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius.* Biochemistry, 2006. **45**(21): p. 6570-80.
63. Heim, R. and R.Y. Tsien, *Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer.* Curr Biol, 1996. **6**(2): p. 178-82.
64. Tsien, R.Y., *The green fluorescent protein.* Annu Rev Biochem, 1998. **67**: p. 509-44.
65. Lim, J., et al., *Flow cytometry based-FRET: basics, novel developments and future perspectives.* Cell Mol Life Sci, 2022. **79**(4): p. 217.
66. Lam, A.J., et al., *Improving FRET dynamic range with bright green and red fluorescent proteins.* Nat Methods, 2012. **9**(10): p. 1005-12.
67. McRae, S.R., C.L. Brown, and G.R. Bushell, *Rapid purification of EGFP, EYFP, and ECFP with high yield and purity.* Protein Expr Purif, 2005. **41**(1): p. 121-7.
68. Bajar, B.T., et al., *A guide to fluorescent protein FRET pairs.* Sensors (Basel), 2016. **16**(9).
69. Crivat, G. and J.W. Taraska, *Imaging proteins inside cells with fluorescent tags.* Trends Biotechnol, 2012. **30**(1): p. 8-16.
70. Chan, F.K.-M., et al., *Fluorescence resonance energy transfer analysis of cell surface receptor interactions and signaling using spectral variants of the green fluorescent protein.* Cytometry A, 2001. **44**(4): p. 361-368.
71. Chan, F.K., et al., *A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling.* Science, 2000. **288**(5475): p. 2351-4.
72. Chan, F.K.M., et al., *A domain in TNF receptors that mediates ligand-independent receptor assembly and signaling.* Science, 2000. **288**(5475): p. 2351-4.
73. Balatskaya, M.N., et al., *Analysis of GPI-anchored receptor distribution and dynamics in live cells by tag-mediated enzymatic labeling and FRET.* Methods Protoc, 2020. **3**(2).
74. Giraldez, T., T.E. Hughes, and F.J. Sigworth, *Generation of functional fluorescent BK channels by random insertion of GFP variants.* J Gen Physiol, 2005. **126**(5): p. 429-38.

75. He, L., et al., *TRAF3 forms heterotrimers with TRAF2 and modulates its ability to mediate NF-kappaB activation.* J Biol Chem, 2004. **279**(53): p. 55855-65.
76. Long, Y., et al., *In vivo FRET-FLIM reveals cell-type-specific protein interactions in Arabidopsis roots.* Nature, 2017. **548**(7665): p. 97-102.
77. Harkes, R., et al., *Dynamic FRET-FLIM based screening of signal transduction pathways.* Sci Rep, 2021. **11**(1): p. 20711.
78. Kalinin, S., et al., *Probability distribution analysis of single-molecule fluorescence anisotropy and resonance energy transfer.* J Phys Chem B, 2007. **111**(34): p. 10253-62.
79. Monaco, G., et al., *flowAI: automatic and interactive anomaly discerning tools for flow cytometry data.* Bioinformatics, 2016. **32**(16): p. 2473-80.
80. Fletez-Brant, K., et al., *flowClean: Automated identification and removal of fluorescence anomalies in flow cytometry data.* Cytometry A, 2016. **89**(5): p. 461-71.
81. Fry, J.M., T.R.L. Fry, and K.R. McLaren, *Compositional data analysis and zeros in micro data.* Appl Econ, 2000. **32**(8): p. 953-959.
82. Killick, R., P. Fearnhead, and I.A. Eckley, *Optimal detection of changepoints with a linear computational cost.* JASA, 2012. **107**(500): p. 1590-1598.
83. *FlowJo™ Software [software application] Version 10.8.1*. 2023, Becton, Dickinson and Company.
84. Ross, D., *Automated analysis of bacterial flow cytometry data with FlowGateNIST.* PLoS One, 2021. **16**(8): p. e0250753.
85. Castillo-Hair, S.M., et al., *FlowCal: a user-friendly, open source software tool for automatically converting flow cytometry data from arbitrary to calibrated units.* ACS Synth Biol, 2016. **5**(7): p. 774-80.
86. Beal, J., et al., *TASBE flow analytics: a package for calibrated flow cytometry analysis.* ACS Synth Biol, 2019. **8**(7): p. 1524-1529.
87. Nguyen, R., et al., *Quantifying spillover spreading for comparing instrument performance and aiding in multicolor panel design.* Cytometry A, 2013. **83**(3): p. 306-15.
88. Roca, C.P., et al., *AutoSpill is a principled framework that simplifies the analysis of multichromatic flow cytometry data.* Nat Commun, 2021. **12**(1).
89. Tung, J.W., et al., *New approaches to fluorescence compensation and visualization of FACS data.* Clin Immunol, 2004. **110**(3): p. 277-83.
90. Roederer, M., *Distributions of autofluorescence after compensation: Be panglossian, fret not.* Cytometry A, 2016. **89**(4): p. 398-402.
91. Roederer, M. and R.F. Murphy, *Cell-by-cell autofluorescence correction for low signal-to-noise systems: Application to epidermal growth factor endocytosis by 3T3 fibroblasts.* Cytometry, 1986. **7**(6): p. 558-565.
92. Woolfe, F., et al., *Autofluorescence removal by non-negative matrix factorization.* IEEE Trans Image Process, 2011. **20**(4): p. 1085-1093.
93. Monici, M., *Cell and tissue autofluorescence research and diagnostic applications.* Biotechnol Annu Rev, 2005. **11**: p. 227-56.
94. Murphy, R.F., *Automated identification of subpopulations in flow cytometric list mode data using cluster analysis.* Cytometry, 1985. **6**(4): p. 302-9.
95. Aghaeepour, N., et al., *Rapid cell population identification in flow cytometry data.* Cytometry A, 2011. **79**(1): p. 6-13.

96. Ge, Y. and S.C. Sealfon, *flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding.* Bioinformatics, 2012. **28**(15): p. 2052-8.
97. Hamerly, G. and C. Elkan, *Learning the k in k-means.* Adv Neural Inf Process Syst, 2003. **16**.
98. Pelleg, D., *X-means: Extending k-means with efficient estimation of the number of clusters.* Proceedings of the Seventeenth International Conference on Machine Learning, 2000.
99. Qian, Y., et al., *Elucidation of seventeen human peripheral blood B-cell subsets and quantification of the tetanus response using a density-based method for the automated identification of cell populations in multidimensional flow cytometry data.* Cytometry B Clin Cytom, 2010. **78 Suppl 1**: p. S69-82.
100. Rahmah, N. and I.S. Sitanggang, *Determination of optimal epsilon (eps) value on DBSCAN algorithm to clustering data on peatland hotspots in sumatra.* IOP Conference Series: Earth and Environmental Science, 2016. **31**.
101. Schubert, E., et al., *DBSCAN Revisited, Revisited.* ACM T DATABASE SYST, 2017. **42**(3): p. 1-21.
102. Ye, X. and J.W.K. Ho, *Ultrafast clustering of single-cell flow cytometry data using FlowGrid.* BMC Syst Biol, 2019. **13**(Suppl 2): p. 35.
103. Cron, A., et al., *Hierarchical modeling for rare event detection and cell subset alignment across flow cytometry samples.* PLoS Comput Biol, 2013. **9**(7): p. e1003130.
104. Finak, G., et al., *Merging mixture components for cell population identification in flow cytometry.* Adv Bioinformatics, 2009: p. 247646.
105. Lo, K., et al., *flowClust: a Bioconductor package for automated gating of flow cytometry data.* BMC Bioinformatics, 2009. **10**: p. 145.
106. Pyne, S., et al., *Automated high-dimensional flow cytometric data analysis.* Proc Natl Acad Sci U S A, 2009. **106**(21): p. 8519-24.
107. Rubbens, P., et al., *PhenoGMM: gaussian mixture modeling of cytometry data quantifies changes in microbial community structure.* mSphere, 2021. **6**(1).
108. Qiu, P., et al., *Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE.* Nat Biotechnol, 2011. **29**(10): p. 886-91.
109. Levine, J.H., et al., *Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis.* Cell, 2015. **162**(1): p. 184-97.
110. Günter, S. and H. Bunke, *Self-organizing map for clustering in the graph domain.* Pattern Recognition Letters, 2002. **23**(4): p. 405-417.
111. Van Gassen, S., et al., *FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data.* Cytometry A, 2015. **87**(7): p. 636-45.
112. Wehrens, R. and L.M.C. Buydens, *Self- and super-organizing maps in R: the kohonen package.* J Stat Softw, 2007. **21**(5).
113. Jones, D.M. and S. Padilla-Parra, *Imaging real-time HIV-1 virion fusion with FRET-based biosensors.* Sci Rep, 2015. **5**(1): p. 1-10.
114. Gordon, G.W., et al., *Quantitative fluorescence resonance energy transfer measurements using fluorescence microscopy.* Biophys J, 1998. **74**(5): p. 2702-13.
115. Kofoed, E.M., M. Guerbadot, and F. Schaufele, *Dimerization between aequorea fluorescent proteins does not affect interaction between tagged estrogen receptors in living cells.* J Biomed Opt, 2008. **13**(3): p. 031207.

116. Wu, P. and L. Brand, *Resonance energy transfer: methods and applications.* Anal Biochem, 1994. **218**(1): p. 1-13.
117. Lambert, T.J., *FPbase: a community-editable fluorescent protein database.* Nat Methods, 2019. **16**(4): p. 277-278.
118. Szöllősi, J., et al., *Fluorescence energy transfer measurements on cell surfaces: a critical comparison of steady-state fluorimetric and flow cytometric methods.* Cytometry A, 1984. **5**(2): p. 210-6.
119. Goldberg, D., et al., *Using collaborative filtering to weave an information tapestry.* Communications of the ACM, 1992. **35**(12): p. 61-70.
120. Vanamee, E.S. and D.L. Faustman, *Structural principles of tumor necrosis factor superfamily signaling*, in *Sci Signal*. 2018.
121. Camerini, D., et al., *The T cell activation antigen CD27 is a member of the nerve growth factor/tumor necrosis factor receptor gene family.* J Immunol, 1991. **147**(9): p. 3165-9.
122. Mukai, Y., et al., *Solution of the structure of the TNF-TNFR2 complex.* Sci Signal, 2010. **3**(148): p. ra83.
123. Gale, A., *Interactions of viral and cellular tumour necrosis factor receptor molecules (UTS Thesis)*. 2016.
124. Sherwood, S., *Expression and functional characterisation of Variola and Monkeypox virus tumour necrosis factor receptor (TNFR) proteins (UTS Thesis)*. 2012.
125. Graham, F.L., et al., *Characteristics of a human cell line transformed by DNA from human adenovirus type 5.* J Gen Virol, 1977. **36**(1): p. 59-72.
126. von Kolontaj, K., et al., *Automated nanoscale flow cytometry for assessing protein-protein interactions.* Cytometry A, 2016. **89**(9): p. 835-43.
127. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
128. Razo-Mejia, M., et al., *Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction.* Cell Syst, 2018. **6**(4): p. 456-469 e10.
129. Trón, L., et al., *Flow cytometric measurement of fluorescence resonance energy transfer on cell surfaces. Quantitative evaluation of the transfer efficiency on a cell-by-cell basis.* Biophys J, 1984. **45**(5): p. 939-46.
130. Greenacre, M., *Compositional data analysis.* Annu Rev Stat Appl, 2021. **8**(1): p. 271-299.
131. Kruskal, W.H. and W.A. Wallis, *Use of ranks in one-criterion variance analysis.* J Am Stat Assoc, 1952. **47**(260): p. 583-621.
132. Croft, M. and R.M. Siegel, *Beyond TNF: TNF superfamily cytokines as targets for the treatment of rheumatic diseases.* Nat Rev Rheumatol, 2017. **13**(4): p. 217-233.
133. Marsters, S.A., et al., *Identification of cysteine-rich domains of the type 1 tumor necrosis factor receptor involved in ligand binding.* J Biol Chem, 1992. **267**(9): p. 5747-50.
134. Smith, C.A., T. Farrah, and R.G. Goodwin, *The TNF receptor superfamily of cellular and viral proteins: Activation, costimulation, and death*, in *Cell*. 1994. p. 959-962.
135. Chan, F.K., *The pre-ligand binding assembly domain: a potential target of inhibition of tumour necrosis factor receptor function.* Ann Rheum Dis, 2000. **59 Suppl 1**: p. i50-3.

136. Chan, F.K.M., *Three is better than one: pre-ligand receptor assembly in the regulation of TNF receptor signaling*, in *Cytokine*. 2007. p. 101-7.
137. McMillan, D., et al., *Structural insights into the disruption of TNF-TNFR1 signalling by small molecules stabilising a distorted TNF.* Nat Commun, 2021. **12**(1): p. 582.
138. Banner, D.W., et al., *Crystal structure of the soluble human 55 kd TNF receptor-human TNFβ complex: Implications for TNF receptor activation.* Cell, 1993. **73**(3): p. 431-445.
139. Naismith, J.H., et al., *Crystallographic evidence for dimerization of unliganded tumor necrosis factor receptor.* J Biol Chem, 1995. **270**(22): p. 13303-7.
140. Stafford, W.F., et al., *Calcium-dependent structural changes in scallop heavy meromyosin.* J Mol Biol, 2001. **307**(1): p. 137-47.
141. Loening, A.M., T.D. Fenn, and S.S. Gambhir, *Crystal structures of the luciferase and green fluorescent protein from Renilla reniformis.* J Mol Biol, 2007. **374**(4): p. 1017-28.
142. Selvin, P.R., *Fluorescence resonance energy transfer*, in *Methods Enzymol*. 1995. p. 300-34.
143. Sedger, L.M. and M.F. McDermott, *TNF and TNF-receptors: From mediators of cell death and inflammation to therapeutic giants - past, present and future*, in *Cytokine Growth Factor Rev*. 2014. p. 453-72.
144. Smith, C.A., et al., *A receptor for tumor necrosis factor defines an unusual family of cellular and viral proteins.* Science, 1990. **248**(4958): p. 1019-23.
145. Zhang, F., et al., *Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry.* Nat Immunol, 2019. **20**(7): p. 928-942.
146. Macaubas, C., et al., *Oligoarticular and polyarticular JIA: epidemiology and pathogenesis.* Nat Rev Rheumatol, 2009. **5**(11): p. 616-26.
147. McInnes, I.B. and G. Schett, *Cytokines in the pathogenesis of rheumatoid arthritis.* Nat Rev Immunol, 2007. **7**(6): p. 429-42.
148. Mellins, E.D., C. Macaubas, and A.A. Grom, *Pathogenesis of systemic juvenile idiopathic arthritis: some answers, more questions.* Nat Rev Rheumatol, 2011. **7**(7): p. 416-26.
149. Aksentijevich, I. and D.L. Kastner, *Genetics of monogenic autoinflammatory diseases: past successes, future challenges.* Nat Rev Rheumatol, 2011. **7**(8): p. 469-78.
150. Hull, K.M., et al., *The TNF receptor-associated periodic syndrome (TRAPS): emerging concepts of an autoinflammatory disorder.* Medicine (Baltimore), 2002. **81**(5): p. 349-68.
151. Jarosz-Griffiths, H.H., et al., *TNF receptor signalling in autoinflammatory diseases.* Int Immunol, 2019. **31**(10): p. 639-648.
152. Ozen, S. and Y. Bilginer, *A clinical guide to autoinflammatory diseases: familial Mediterranean fever and next-of-kin.* Nat Rev Rheumatol, 2014. **10**(3): p. 135-47.
153. Micheau, O. and J. Tschopp, *Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes.* Cell, 2003. **114**(2): p. 181-90.
154. Paolini, A., et al., *Cell death in coronavirus infections: uncovering its role during Covid-19.* Cells, 2021. **10**(7).

155. Ruby, J., H. Bluethmann, and J.J. Peschon, *Antiviral activity of tumor necrosis factor (TNF) is mediated via p55 and p75 TNF receptors.* J Exp Med, 1997. **186**(9): p. 1591-1596.
156. Sedger, L., *Viral inhibition of tumour necrosis factor-α (TNFα) and TNF-receptor induced apoptosis and inflammation.* Curr Med Chem Anti Inflamm Anti Allergy Agents, 2005. **4**(6): p. 597-615.
157. Wong, G.H. and D.V. Goeddel, *Tumour necrosis factors alpha and beta inhibit virus replication and synergize with interferons.* Nature, 1986. **323**(6091): p. 819-22.
158. Upton, C., et al., *Myxoma virus expresses a secreted protein with homology to the tumor necrosis factor receptor gene family that contributes to viral virulence.* Virology, 1991. **184**(1): p. 370-382.
159. Schreiber, M. and G. McFadden, *The myxoma virus TNF-receptor homologue (T2) inhibits tumor necrosis factor-alpha in a species-specific fashion.* Virology, 1994. **204**(2): p. 692-705.
160. Moss, B., *Poxviridae: the viruses and thier replication.* Fields virology, 1996: p. 2637-2671.
161. Moss, B., *Poxvirus DNA replication.* Cold Spring Harb Perspect Biol, 2013. **5**(9).
162. Pickup, D., *Poxviral modifiers of cytokine responses to infection.* Infect Agents Dis, 1994. **3**(2-3): p. 116-127.
163. Grant, B.D. and J.G. Donaldson, *Pathways and mechanisms of endocytic recycling.* Nat Rev Mol Cell Biol, 2009. **10**(9): p. 597-608.
164. Jares-Erijman, E.A. and T.M. Jovin, *FRET imaging.* Nat Biotechnol, 2003. **21**(11): p. 1387-95.
165. Patterson, G.H., D.W. Piston, and B.G. Barisas, *Forster distances between green fluorescent protein pairs.* Anal Biochem, 2000. **284**(2): p. 438-40.
166. Roszik, J., J. Szollosi, and G. Vereb, *AccPbFRET: an ImageJ plugin for semi-automatic, fully corrected analysis of acceptor photobleaching FRET images.* BMC Bioinformatics, 2008. **9**: p. 346.
167. Becker, W., *Fluorescence lifetime imaging--techniques and applications.* J Microsc, 2012. **247**(2): p. 119-36.
168. Hoppe, A., K. Christensen, and J.A. Swanson, *Fluorescence resonance energy transfer-based stoichiometry in living cells.* Biophys J, 2002. **83**(6): p. 3652-64.
169. Tramier, M. and M. Coppey-Moisan, *Fluorescence anisotropy imaging microscopy for homo-FRET in living cells.* Methods Cell Biol, 2008. **85**: p. 395-414.
170. Rocheleau, J.V., M. Edidin, and D.W. Piston, *Intrasequence GFP in class I MHC molecules, a rigid probe for fluorescence anisotropy measurements of the membrane environment.* Biophys J, 2003. **84**(6): p. 4078-86.
171. Lidke, D.S., et al., *Imaging molecular interactions in cells by dynamic and static fluorescence anisotropy (rFLIM and emFRET).* Biochem Soc Trans, 2003. **31**(Pt 5): p. 1020-7.
172. Squire, A., et al., *Red-edge anisotropy microscopy enables dynamic imaging of homo-FRET between green fluorescent proteins in cells.* J Struct Biol, 2004. **147**(1): p. 62-9.
173. Karpova, T.S., et al., *Fluorescence resonance energy transfer from cyan to yellow fluorescent protein detected by acceptor photobleaching using confocal microscopy and a single laser.* J Microsc, 2003. **209**(Pt 1): p. 56-70.

174. Bastiaens, P.I., et al., *Imaging the intracellular trafficking and state of the AB5 quaternary structure of cholera toxin.* EMBO J, 1996. **15**(16): p. 4246-53.
175. Woehler, A., J. Wlodarczyk, and E. Neher, *Signal/noise analysis of FRET-based sensors.* Biophys J, 2010. **99**(7): p. 2344-54.
176. Zal, T. and N.R. Gascoigne, *Photobleaching-corrected FRET efficiency imaging of live cells.* Biophys J, 2004. **86**(6): p. 3923-39.
177. Szöllősi, J., et al., *Flow cytometric resonance energy transfer measurements support the association of a 95-kDa peptide termed T27 with the 55-kDa Tac peptide.* Proc Natl Acad Sci U S A, 1987. **84**(20): p. 7246-50.
178. Nagy, P., et al., *Intensity-based energy transfer measurements in digital imaging microscopy.* Eur Biophys J, 1998. **27**(4): p. 377-89.
179. Chen, H., et al., *Measurement of FRET efficiency and ratio of donor to acceptor concentration in living cells.* Biophys J, 2006. **91**(5): p. L39-41.
180. Black, C.B., et al., *Cell-based screening using high-throughput flow cytometry.* Assay Drug Dev Technol, 2011. **9**(1): p. 13-20.
181. Lo, C.H., et al., *An Innovative High-Throughput Screening Approach for Discovery of Small Molecules That Inhibit TNF Receptors.* SLAS Discov, 2017. **22**(8): p. 950-961.
182. Chan, F.K., *Monitoring molecular interactions in living cells using flow cytometric analysis of fluorescence resonance energy transfer.* Methods Mol Biol, 2004. **261**: p. 371-82.
183. Paque, S. and D. Weijers, *Q&A: Auxin: the plant molecule that influences almost anything.* BMC Biol, 2016. **14**(1): p. 67.
184. Pettersson, T., et al., *Setting up TRAPS.* Ann Med, 2012. **44**(2): p. 109-18.
185. Kimberley, F.C., et al., *Falling into TRAPS--receptor misfolding in the TNF receptor 1-associated periodic fever syndrome.* Arthritis Res Ther, 2007. **9**(4): p. 217.
186. Aggarwal, B.B., *Signalling pathways of the TNF superfamily: a double-edged sword.* Nat Rev Immunol, 2003. **3**(9): p. 745-56.
187. Nedjai, B., et al., *Proinflammatory action of the antiinflammatory drug infliximab in tumor necrosis factor receptor-associated periodic syndrome.* Arthritis Rheum, 2009. **60**(2): p. 619-25.
188. Masson, C., et al., *Tumor necrosis factor receptor-associated periodic syndrome (TRAPS): definition, semiology, prognosis, pathogenesis, treatment, and place relative to other periodic joint diseases.* Joint Bone Spine, 2004. **71**(4): p. 284-90.
189. Bachetti, T., et al., *Autophagy contributes to inflammation in patients with TNFR-associated periodic syndrome (TRAPS).* Ann Rheum Dis, 2013. **72**(6): p. 1044-52.
190. Turner, M.D., A. Chaudhry, and B. Nedjai, *Tumour necrosis factor receptor trafficking dysfunction opens the TRAPS door to pro-inflammatory cytokine secretion.* Biosci Rep, 2012. **32**(2): p. 105-12.
191. Pontejo, S.M., A. Alejo, and A. Alcami, *Comparative biochemical and functional analysis of viral and human secreted tumor necrosis factor (TNF) decoy receptors.* J Biol Chem, 2015. **290**(26): p. 15973-84.
192. Pontejo, S.M., A. Alejo, and A. Alcami, *Poxvirus-encoded TNF decoy receptors inhibit the biological activity of transmembrane TNF.* J Gen Virol, 2015. **96**(10): p. 3118-23.
193. Benedict, C.A., P.S. Norris, and C.F. Ware, *To kill or be killed: viral evasion of apoptosis.* Nat Immunol, 2002. **3**(11): p. 1013-8.

194. Xu, X., P. Nash, and G. McFadden, *Myxoma virus expresses a TNF receptor homolog with two distinct functions.* Virus Genes, 2000. **21**(1/2): p. 97-109.
195. Sedger, L. and G. McFadden, *M-T2: a poxvirus TNF receptor homologue with dual activities.* Immunol Cell Biol, 1996. **74**(6): p. 538-45.
196. Gooding, L.R., *Virus proteins that counteract host immune defenses*, in *Cell*. 1992. p. 1-4.
197. *World Health Organization (WHO) - Smallpox*. 2023 [cited 2023 1st January]; Available from: https://www.who.int/health-topics/smallpox.
198. Hochreiter, B., et al., *Advanced FRET normalization allows quantitative analysis of protein interactions including stoichiometries and relative affinities in living cells.* Sci Rep, 2019. **9**(1): p. 8233.
199. *Infevers Database: TNFRSF1A (NM_001065.3) sequence variants*. 2018 [cited 2018 04 September]; Available from: https://infevers.umai-montpellier.fr/web/.
200. Takada, H., et al., *Role of SODD in regulation of tumor necrosis factor responses.* Mol Cell Biol, 2003. **23**(11): p. 4026-33.
201. Hsu, H., J. Xiong, and D.V. Goeddel, *The TNF receptor 1-associated protein TRADD signals cell death and NF-kappa B activation.* Cell, 1995. **81**(4): p. 495-504.
202. Hsu, H., et al., *TRADD-TRAF2 and TRADD-FADD interactions define two distinct TNF receptor 1 signal transduction pathways.* Cell, 1996. **84**(2): p. 299-308.
203. Tada, Y., et al., *Collagen-induced arthritis in TNF receptor-1-deficient mice: TNF receptor-2 can modulate arthritis in the absence of TNF receptor-1.* Clin Immunol, 2001. **99**(3): p. 325-33.
204. Hayden, M.S. and S. Ghosh, *Shared principles in NF-kappaB signaling.* Cell, 2008. **132**(3): p. 344-62.
205. Haas, T.L., et al., *Recruitment of the linear ubiquitin chain assembly complex stabilizes the TNF-R1 signaling complex and is required for TNF-mediated gene induction.* Mol Cell, 2009. **36**(5): p. 831-44.
206. Kovalenko, A., et al., *The tumour suppressor CYLD negatively regulates NF-kappaB signalling by deubiquitination.* Nature, 2003. **424**(6950): p. 801-5.
207. Wang, L., F. Du, and X. Wang, *TNF-alpha induces two distinct caspase-8 activation pathways.* Cell, 2008. **133**(4): p. 693-703.
208. Sun, L., et al., *Mixed lineage kinase domain-like protein mediates necrosis signaling downstream of RIP3 kinase.* Cell, 2012. **148**(1-2): p. 213-27.
209. Sedger, L.M., et al., *Therapeutic antibody ‐based drugs in the treatment of human inflammatory disorders*, in *Immunotherapy - Myths, Reality, Ideas, Future*. 2017.
210. Lawrance, I.C., et al., *Serious infections in patients with inflammatory bowel disease receiving anti-tumor-necrosis-factor-alpha therapy: an Australian and New Zealand experience.* J Gastroenterol Hepatol, 2010. **25**(11): p. 1732-8.
211. Ahmad, N.M., K.M. Ahmad, and F. Younus, *Severe adenovirus pneumonia (AVP) following infliximab infusion for the treatment of Crohn's disease.* J Infect, 2007. **54**(1): p. e29-32.
212. Kang, M.-J., et al., *Adenoviral pneumonia during etanercept treatment in a patient with rheumatoid arthritis.* Korean J Intern Med, 2007. **22**(1).
213. Ma, C., B. Walters, and R.N. Fedorak, *Varicella zoster meningitis complicating combined anti-tumor necrosis factor and corticosteroid therapy in Crohn's disease.* World J Gastroenterol, 2013. **19**(21): p. 3347-51.

214. Tresch, S., et al., *Disseminated herpes zoster mimicking rheumatoid vasculitis in a rheumatoid arthritis patient on etanercept.* Dermatology, 2009. **219**(4): p. 347-9.
215. Andreadou, E., et al., *Demyelinating disease following anti-TNFa treatment: a causal or coincidental association? Report of four cases and review of the literature.* Case Rep Neurol Med, 2013. **2013**: p. 671935.
216. Lo, C.H., E.C. Huber, and J.N. Sachs, *Conformational states of TNFR1 as a molecular switch for receptor function.* Protein Sci, 2020. **29**(6): p. 1401-1415.
217. Sander, J., et al., *Density-based clustering in spatial databases: The algorithm gdbscan and its applications.* Data Min Knowl Discov, 1998. **2**(2): p. 169-194.
218. Alberti, S., D.R. Parks, and L.A. Herzenberg, *A single laser method for subtraction of cell autofluorescence in flow cytometry.* Cytometry, 1987. **8**(2): p. 114-9.
219. Roederer, M., *Compensation in flow cytometry.* Curr Protoc Cytom, 2002. **Chapter 1 Unit 1 14**: p. 22 (1).
220. Roederer, M., *Spectral compensation for flow cytometry: Visualization artifacts, limitations, and caveats.* Cytometry, 2001. **45**(3): p. 194-205.
221. Novo, D., G. Grégori, and B. Rajwa, *Generalized unmixing model for multispectral flow cytometry utilizing nonsquare compensation matrices.* Cytometry Part A, 2013. **83A**(5): p. 508-520.
222. Quinn, T.P., et al., *Understanding sequencing data as compositions: an outlook and review.* Bioinformatics, 2018. **34**(16): p. 2870-2878.
223. Martín-Fernández, J.A., Mathematical Geology, 2003. **35**(3): p. 253-278.
224. Martín-Fernández, J.A., et al., *Model-based replacement of rounded zeros in compositional data: Classical and robust approaches.* Comput Stat Data Anal, 2012. **56**(9): p. 2688-2704.
225. Palarea-Albaladejo, J., J.A. Martín-Fernández, and J. Gómez-García, *A parametric approach for dealing with compositional rounded zeros.* Mathematical Geology, 2007. **39**(7): p. 625-645.
226. Nobis, M., et al., *A RhoA-FRET biosensor mouse for intravital imaging in normal tissue homeostasis and disease contexts.* Cell Rep, 2017. **21**(1): p. 274-288.
227. Raman, C.V., *A new radiation.* P INDIAN AS, 1953. **37**(3): p. 333-341.
228. Jones, R.R., et al., *Raman techniques: fundamentals and frontiers.* Nanoscale Res Lett, 2019. **14**(1): p. 231.
229. Ember, K.J.I., et al., *Raman spectroscopy and regenerative medicine: a review.* NPJ Regen Med, 2017. **2**: p. 12.
230. Brown, M. and C. Wittwer, *Flow cytometry: principles and clinical applications in hematology.* Clin Chem, 2000. **46**(8): p. 1221-1229.
231. Van der Maaten, L. and G. Hinton, *Visualizing data using t-SNE.* J Mach Learn Res, 2008. **9**(11).
232. McInnes, L., et al., *UMAP: uniform manifold approximation and projection.* J Open Source Softw, 2018. **3**(29): p. 861.
233. Amid, E. and M.K. Warmuth, *TriMap: Large-scale dimensionality reduction using triplets.* arXiv preprint arXiv:1910.00204, 2019.
234. Wang, Y., et al., *Understanding how dimension reduction tools work: an empirical approach to deciphering t-SNE, UMAP, TriMAP, and PaCMAP for data visualization.* arXiv, 2020.