Optogenetic control of ER Ca^{2+} release and development of a mammalian cell-based library screening platform for directed evolution

by

Shuce Zhang

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

University of Alberta

Abstract

Life on the earth's surface is bathed in the white light of the Sun. Biologists have been harnessing light as a research tool to understand and manipulate many biological processes inside the cells. Optogenetics — genetically encoded proteins that enable optical visualization or manipulation of physiological states — not only provides useful tools to understand our cells and body, but also has the potential to create revolutionary therapies for nervous system disorders in humans. Calcium signalling is among the fields that have benefitted most from the development of optogenetic tools. Over the past few decades, many tools have been engineered to sense the Ca^{2+} concentration (genetically encoded calcium ion indicator, GECI) and manipulate the Ca^{2+} handling inside the cell (genetically encoded calcium ion actuator, GECA). However, tools that specifically facilitate the study of Ca^{2+} release from the endoplasmic reticulum (ER) have not been extensively engineered. On the other hand, although most optogenetic tools are developed for application in mammalian cells or organisms, it is cumbersome and inefficient to screen these tools in mammalian cells, and thus the prokaryotic systems are still heavily relied on for the directed evolution of optogenetic tools. A solution to achieve clonal expression in mammalian cells would be greatly helpful in achieving mammalian cell-based library screening and directed evolution.

In **Chapter 1**, I will present a literature review of the methods for mammalian cell-based library screening. These methods are categorised as plasmid transfection, viral transduction, site-directed recombination, and *in situ* mutagenesis. Special attention is given to achieving variant separation in different cells when screening pooled libraries and the applications in engineering optogenetic tools, antibodies, membrane proteins, and other targets of interest for mammalian systems.

In **Chapter 2**, I will describe my original work of developing an optogenetic tool for light-induced ER Ca^{2+} release. This tool is a binary system consisting of an Orai channel rerouted to ER membrane by a C-terminal dilysine motif, and a STIM-based GECA that activates Orai in a light-dependent manner. We show that the ER-localised human Orai1 channels are functional and best activated by OptoCRAC (a LOV2 domain-based GECA) to mediate a mild ER Ca^{2+} release. This Ca^{2+} release can activate overexpressed RyR2 channels and elicit calcium oscillation via calcium-induced Ca^{2+} release (CICR). The binary system also has the potential of spatial targeting of Ca^{2+} release at ER-organelle and ER-plasma membrane (PM) junctions.

In **Chapter 3**, I will present my design of a mammalian cell-based library screening system for clonal expression. This system consists of an engineered HEK-293 FT landing pad stable cell line, and a pBAD-derived donor plasmid. With the help of Bxb1 recombinase, one molecule of the transfected donor plasmid in each cell is integrated into the genomic landing pad locus via specific irreversible *attP* × *attB* recombination, with the gene of interest (GOI) placed downstream of a mammalian promoter in the landing pad. This GOI thus becomes the only copy that is expressed in this cell. I report the creation and identification of a clone of stable cell harbouring a single copy of landing pad locus. I demonstrate that the variants of pooled plasmids are exclusively expressed in this stable cell line with overall expression efficiency of ~ 5 %. I will also present a design for a future generation of library screening system to facilitate screening of the linear donor DNA libraries.

In **Chapter 4**, I will describe my computational simulation on PhoCl1 dissociation pathway. Adaptive steered-molecular dynamics (ASMD) was used to sample the possible route of dissociation without presumptions of the dissociation direction. A possible route of dissociation and the dissociation-dependent pattern of conformational change of 201-207 loop

were identified through the ASMD simulation. The structural and mechanistic insights facilitated the directed evolution of PhoCl2 variants with improved dissociation efficiency and kinetics.

Preface

This thesis is an original work by Shuce Zhang.

Chapter 1 is my original literature review on mammalian cell-based library screening methodology.

Chapter 2 is my original work, except the experiments performed in the HEK-293 Orai1,2,3 triple knockout cell line (**Figure** 2.7 and **Figure** 2.8) which were performed by collaborators, Dr. Youjun Wang and colleagues, at Beijing Normal University. Eric Y. Fan-Lou assisted with calcium imaging experiments.

Chapter 3 is my original work, except the experiments performed in the LLP cell line (**Figure** 3.6**B,D**) which were performed by collaborators, Dr. François St-Pierre and colleagues, at Baylor College of Medicine.

Chapter 4 was published as part of the article Lu, X., Wen, Y., Zhang, S., Zhang, W., Chen, Y., Shen, Y., Lemieux, M.J. and Campbell, R.E., (2021). Photocleavable proteins that undergo fast and efficient dissociation. *Chemical Science*, *12*(28), 9658-9672. Adapted under the terms of under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence, Copyright 2021. I was responsible for designing, implementing, analysing, and writing all computational simulation experiments. Y.C. assisted with the computational simulation. X.L. assembled all constructs, screened new variants, performed protein characterizations and cell-based experiments as well as the manuscript composition. Y.W. was responsible for protein crystallization, X-ray data collection and structure determination. W.Z. conceived the idea of light-induced cell death experiment and performed the initial test. Y.S. contributed to manuscript edits. M.J.L. and R.E.C. supervised and acquired funding for this work.

Section 5.3.1 was my original work. Yoojin Choi assisted with *E. coli* expression and *in vitro* assays.

Section 5.3.2 was published as part of the article Tang, L., Zhang, S., Zhao, Y., Rozanov, N.D., Zhu, L., Wu, J., Campbell, R.E. and Fang, C., (2021). Switching between ultrafast pathways enables a green-red emission ratiometric fluorescent-protein-based Ca^{2+} biosensor. *International journal of molecular sciences*, *22*(1), 445. Adapted under the terms of under a Creative Commons Attribution (CC BY) license, Copyright 2021. I was responsible for performing, analysing, and writing of all live cell imaging experiments. L.T., N.D.R., L.Z. and J.W. performed the ultra-fast spectroscopy studies. C.F. and R.E.C supervised and acquired funding for this work.

Section 5.3.3 was published as part of the article Nasu, Y., Murphy-Royal, C., Wen, Y., Haidey, J.N., Molina, R.S., Aggarwal, A., Zhang, S., Kamijo, Y., Paquet, M.E., Podgorski, K. and Drobizhev, M. (2021). A genetically encoded fluorescent biosensor for extracellular L-lactate. *Nature communications*, *12*(1), 1-12. Adapted under the terms of under a Creative Commons Attribution (CC BY) license, Copyright 2021. I was responsible for performing and analysing the *in situ* pH titration experiment (Supplementary Figure 11) and expressed the protein for crystallography. Y.N. developed eLACCO1.1 and performed in vitro characterization. R.S.M. and M.D. measured one-photon absorbance spectra and two-photon excitation spectra. C.M.R. and J.H. performed acute brain slice imaging. Y.W. determined the crystal structure of eLACCO1. A.A. performed stopped-flow experiment. A.A. and K.P. performed the imaging of primary neurons and data analysis. Y.K. performed screening of leader sequence. M.-E.P produced AAV. M.J.L., K.P., J.S.B., G.R.G. and R.E.C. supervised research. Y.N. and R.E.C. wrote the manuscript.

Section 5.3.4 was a collaboration with Eric Y. Fan-Lou.

Dedication

Dedicated to my grandmother, Xiufen Lin （林秀芬）.

Acknowledgements

I am tremendously grateful for the mentorship and supervision of Dr. Robert E. Campbell. Throughout my graduate studies, I had the opportunity to explore my research interests and learn many new techniques with little restriction. My training at Campbell Lab has been invaluable.

My appreciation also goes to Dr. Julianne M. Gibbs and Dr. Michael J. Serpe, for serving on my supervisory committee, as well as Dr. Matthew S. Macauley, Dr. Howard S. Young and Dr. Corrie J.B. daCosta for serving as the arm's length examiners on my final exam and/or candidacy exam. Thanks to Dr. Charles A. Lucy and Dr. Ran Zhao for chairing the exams.

I am very grateful for my collaborators during my PhD studies: Dr. Youjun Wang at Beijing Normal University, Dr. Francois St-Pierre and Jihwan Lee at Baylor College of Medicine, Dr. Marie-Ève Paquet at CERVO Brain Research Center, Université Laval, and Dr. Jamal Daoud and Dr. Chengcheng Rao at Galenvs Sciences Inc.

I acknowledge the support of the grants from the Canadian Institutes of Health Research (FS-154310), Natural Sciences and Engineering Research Council of Canada (RGPIN- 2018-04364) and Resources for Research Groups from Compute Canada (rrg-recfp2) to my PhD supervisor Dr. Robert E. Campbell. Part of my funding was supported by the NSERC CREATE Advanced Protein Engineering Training, Internships, Courses, and Exhibition (APRENTICE) program and Mitacs Globalink Graduate Fellowship.

This work will not be possible without the access to research facilities in University of Alberta. Especially, I would like to thank G. Lambkin from Biological Services, Department of Chemistry, where I used the Nikon TIRF microscope regularly. I also thank Dr. X. Sun for the

access to Cell Imaging Facility, Department of Oncology, as well as the staff from Molecular Biology Service Unit (MBSU), Department of Biological Sciences. In addition, my special thanks go to Dr. M. S. Macauley and Dr. R. Derda for kindly allowing me to use their Zeiss LSM 700 confocal microscope, BD FACSMelody cell sorter, and KingFisher Duo robotic system.

I also thank my colleagues at Campbell lab: Dr. Y. Shen, Dr. S.-Y. Wu, Dr. X. Lu, Dr. Y. Nasu, Y. Li, Dr. Y. Zhao, Dr. F. Zheng, Dr. Y. Qian, Dr. R. Dalangin, Dr. L. Zarowny, Dr. W. Zhang, Dr. K. Takahashi-Yamashiro, A. Aggarwal, A. Kim, Dr. S. Khan. I also had the pleasure of mentoring three amazing undergraduate researchers: Y. Choi, Y. Chen, E. Y. Fan-Lou.

Finally, I would extend my thanks to my friends and family for their support.

Table of Contents

List of Tables

List of Figures

Glossary

2-APB	2-Aminoethoxydiphenyl borate
AAV	Adeno-associated virus
AdV	Adenovirus
AID	Activation-induced cytidine deaminase
ANSGA	Human Orai1 with L261A-V262N-H264G-K265A quadruple mutations
ASMD	Adaptive steered-molecular dynamics
attB	Bacterial attachment site
attP	Phage attachment site
BCR	B cell receptor
bGH	Bovine growth hormone
BLAST	The basic local alignment search tool
$[Ca^{2+}]_{cyt}$	Cytosolic Ca^{2+} concentration
$[Ca^{2+}]_{ER}$	Endoplasmic reticulum Ca^{2+} concentration
$[Ca^{2+}]_{o}$	Extracellular Ca^{2+} concentration
CAD	CRAC activation domain
CAG	Cytomegalovirus (CMV) enhancer, fused to the promoter, the first exon and the first intron of chicken beta-actin gene, and the splice acceptor of the rabbit beta-globin gene
Cas9	CRISPR-associated endonuclease 9
CC1	Coiled coil 1
CDR	Complementarity-determining region

ChR	Channelrhodopsin
ChRGR	Channelrhodopsin-green receiver
$ChRGR_{ER}$	Endoplasmic reticulum-localised channelrhodopsin-green receiver
CICR	Calcium-induced Ca^{2+} release
CMV	Cytomegalovirus
cpLOV2	Circularly permutated Light-oxygen-voltage-sensing domain 2
CRAC	Calcium release–activated channels
CRISPR	Clustered regularly interspaced short palindromic repeats
CRY2	Cryptochrome-2
CSR	Class switch recombination
dCas9	Catalytically dead Cas9
DHPR	Dihydropyridine receptor
DMEM	Dulbecco's Modified Eagle Medium
DMEM/F12	Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12
DNA	Deoxyribonucleic acid
DSB	Double-strand break
EBFP	Enhanced blue fluorescent protein
EF1α	Elongation factor-1 alpha
EGFP	Enhanced green fluorescent protein
EGTA	Ethylene glycol-bis(β-aminoethyl ether)-N,N,N’,N’-tetraacetic acid
EMA	European Medicines Agency
EP-PCR	Error-prone polymerase chain reaction
ER	Endoplasmic reticulum
FACS	Fluorescence-activated Cell Sorting
FBS	Fetal Bovine Serum

FDA	U.S. Food and Drug Administration
FISH	Fluorescence in situ hybridization
Flp	Flippase
FRT	Flippase recognition target
gDNA	Genomic deoxyribonucleic acid, a short synthetic RNA composed of a scaffold sequence necessary for Cas-binding and a user-defined ~20 nucleotide spacer that defines the genomic target to be modified.
GECA	Genetically encoded calcium channel actuators
GECI	Genetically encoded Ca^{2+} indicators
GEVI	Genetically encoded voltage indicators
GOI	Gene of interest
GPCR	G protein-coupled receptor
gRNA	Guide ribonucleic acid
HBSS	Hanks' Balanced Salt Solution
HEK-293 FT	Human embryonic kidney cell 293 FT strain
HO1	Heme oxygenase 1
Ig	Immunoglobulin
indel	Insertion and deletion mutation
IP_3	Inositol 1,4,5-trisphosphate
IP_3R	Inositol 1,4,5-trisphosphate receptors
IRES	Internal ribosome entry site
KD	Knock-down
KO	Knock-out
KKXX	Dilysine motif (lysine, lysine, 2 other amino acids, stop codon)
LDLR	Low-density lipoprotein receptor

LOV	Light-oxygen-voltage-sensing domain
LP	Long pass
LTCC	L-type calcium channel
LV	Lentivirus
mAb	Monoclonal antibody
MACS	Magnetic-activated cell sorting
MBSU	Molecular Biology Service Unit
MERFISH	Multiplexed error-robust fluorescence in situ hybridization
MHC	Major histocompatibility complex
miRFP	Monomeric near-infrared fluorescent protein
$mNG2_{1-10}$	Strand 1-10 of the split-mNeonGreen2 fluorescent protein
$mNG2_{1-11}$	Strand 11 of the split-mNeonGreen2 fluorescent protein
MOI	Multiplicity of infection
nCas9	Cas9 nickase
NGS	Next-generation sequencing
OptoCR	Optogenetic Ca^{2+} release
ORF	Open reading frame
PAM	Protospacer adjacent motifs
PAmCherry	Photoactivatable fluorescent protein mCherry
Ptet	Tet-responsive promoter
PCR	Polymerase chain reaction
PLC	Phospholipase C
PM	Plasma membrane
pMHC	Short peptide fragments presented by major histocompatibility complex proteins

POI	Protein of interest
poly(A)	Polyadenylation signal
Py	Polyomavirus
PyLT	Polyomavirus large T antigen
qPCR	Quantitative polymerase chain reaction
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
rAAV	Recombinant adeno-associated virus
RAG	Recombination activating genes
RDF	Recombination directionality factor
RMCE	Recombinase-mediated cassette exchange
RNAi	Ribonucleic acid interference
RNAP	RNA polymerase
RSS	Recombination signal sequences
RyR	Ryanodine receptors
scCRISPR-seq	Single-cell sequencing for CRISPR screening
SERCA	Sarco/endoplasmic reticulum Ca^{2+}-ATPase
SFFV	The spleen focus-forming virus
shRNA	Small hairpin ribonucleic acid
SHM	Somatic hypermutation
siRNA	Small interfering ribonucleic acid
smURFP	Small ultra-red fluorescent protein
SPA	Synthetic polyadenylation signal
STIM1	Stromal interaction molecule 1
STIM1ct	Stromal interaction molecule 1 cytosolic fragment
SOAR	STIM-Orai activating region

SOCE	Store-operated Ca^{2+} entry
SOICR	Store overload-induced Ca^{2+} release
SRE	Serum response element
SV40	Simian vacuolating virus 40
TCR	T cell receptor
TdT	Terminal deoxynucleotidyl transferase
$TetO_2$	Two tetracycline operator sequences
TetR	Tetracycline repressor protein
TG	Thapsigargin
TIRF(M)	Total internal reflection fluorescence (microscopy)
TK	Herpes simplex virus type 1 thymidine kinase
TMCO1	Transmembrane and coiled-coil domains 1
tTA	Tetracycline transactivator
t-tubules	Transverse tubules
VEGAS	Viral evolution of genetically actuating sequences
VSV	Vesicular stomatitis virus
VSV-G	Vesicular stomatitis virus G protein
WT	Wildtype
wtAAV	Wildtype adeno-associated virus

Chapter 1

1 Review of mammalian cell-based library screening methods

Chapter 1 is an original literature review by Shuce Zhang.

There has been tremendous progress and success in protein engineering and directed evolution over the past few decades, culminating with the 2018 Nobel Prize in Chemistry that recognized the seminal contributions of Frances H. Arnold, George P. Smith, and Sir Gregory P. Winter to this field[1]. There has been much effort invested in the use of mammalian cells as platforms for protein engineering for the development of protein-based tools for research applications and as therapeutics for human diseases. However, prokaryotic systems (e.g., *Escherichia coli* and phage), and lower-level eukaryotic organism (e.g., yeast), are still the go-to platforms for library screening, even when the main application of the evolved protein will involve expression in mammalian cells. Proteins evolved in non-mammalian platforms commonly exhibit issues such as protein misfolding, undesirable subcellular targeting, and unintentional post-translational modifications, when expressed in mammalian cells. Accordingly, proteins evolved in non-mammalian systems often require additional rounds of optimisation for better compatibility with mammalian cells. Therefore, it is of high relevance and interest for the protein engineering community to be able to evolve mammalian proteins, and proteins intended for use in mammalian systems using mammalian cell-based library screening systems.

An ideal mammalian cell-based screening system would exhibit the following features: a) different variants should be uniquely expressed in different cells with no cross-contamination between cells; b) libraries should be efficiently introduced into the cells such that a substantial proportion of the cells should express a member of the library; c) the time required from library preparation to cellular expression should be kept to a minimum; and d) the duration of the variant expression should be long enough to allow sequence recovery, further testing, and perhaps clonal expansion.

In this chapter, I will review the commonly used techniques for mammalian cell-based library screening. I will present the basic principles of these techniques, discuss a few important considerations when applying these techniques for screening protein variant libraries, and showcase some representative applications of these techniques.

1.1 Plasmid Transfection

Plasmids, the small, circular, self-replicating DNA molecules originally discovered in bacteria, are widely used vectors for cloning, engineering, and expressing genetic materials. Standardised protocols to construct, amplify, and purify plasmid DNA are well established and are typically performed with the help of engineered *E. coli* strains and commercially available reagents and kits. Once the plasmid has been purified, it can be delivered into a wide range of model organisms (e.g., yeast, mammalian cell culture, mouse embryo). In this chapter, I will focus only on methods for transfection of cultured mammalian cells. Plasmids that harbour mammalian promoters (e.g., CMV, SV40, EF1a promoters) can drive the expression of proteins or short RNA sequences for gene knockdown and knockout (e.g., H1, U6 promoters).

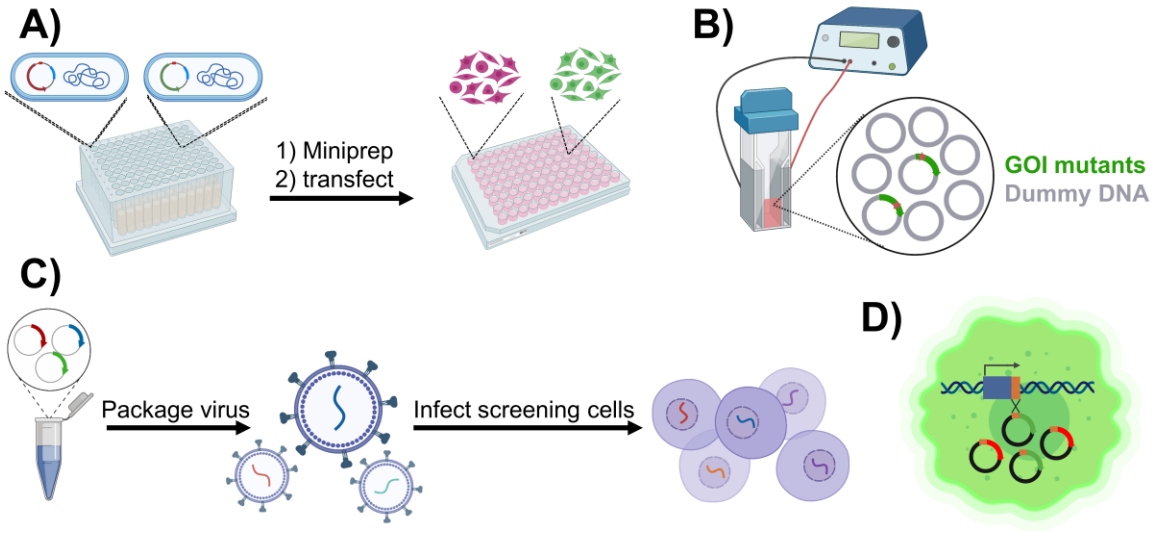

Figure 1.1 Schematics of strategies for delivering ex vivo generated libraries.

A| Transfection of arrayed libraries. Plasmids encoding different variants are purified and transfected separately. **B|** Transfection of pooled libraries. Plasmids encoding variants are diluted with dummy DNA and transfected in bulk. **C|** Virus-based gene delivery. Genes encoding the variants are first packaged in virus particles, which are used to transduce cells for screening. **D|** Recombinase-mediated clonal expression. Each cell integrates one molecule of the transfected DNA by site-directed recombination downstream of a promoter. All schematics in this chapter are created with BioRender.com unless otherwise stated.

1.1.1 Methods of delivery

Like other types of nucleic acid molecules, plasmid DNA is a highly negatively charged polymer. The hydrophilic nature of the plasmid and the electrostatic repulsion between DNA and the anionic plasma membrane (PM) results in the inefficient uptake of the native plasmid molecules. Introducing exogenous nucleic acid to cells in culture, termed “transfection”, therefore requires either chemical or physical method-assisted delivery[2]. Chemical transfection typically involves incubating the plasmid DNA with a cationic reagent carrier (e.g., calcium phosphate, Lipofectamine 2000, Polyethylenimine),

resulting in a positively charged complex that interacts with the PM. The DNA-carrier complex is generally believed to be taken up by endocytosis, followed by endosome rupture due to a proposed “proton sponge” mechanism of the cationic polymers[3]. Physical transfection methods include electroporation, gene gun, and microinjection, among which electroporation is the most commonly used for cultured mammalian cells. When short electrical pulses are applied, the external electrical field introduces transient reversible disruptions to the PM structure, allowing the plasmid DNA to reach the intracellular compartments, including the nucleus[4]. Simultaneous electroporation in 96-well format has been made commercially available by several manufacturers.

1.1.2 **Transient transfection and stable transfection**

Transfection can also be categorised into either transient transfection or stable transfection. For transient transfection, the plasmid DNA exists as an extrachromosomal DNA molecule and does not integrate into the host cell genome. The extrachromosomal DNA may be degraded or lost during cell division[5]. As a result, expression of genes encoded on the plasmid only lasts for a few days. Some cell lines have been engineered for prolonged expression of transiently transfected plasmid DNA. For example, the human embryonic kidney HEK-293T cell line stably expresses the Simian virus 40 (SV40) large-T antigen, allowing amplification of plasmids harbouring SV40 origins[6]. Similarly, HEK-293E stably expresses Epstein–Barr virus (EBV) nuclear antigen 1 (EBNA1), allowing episomal replication of plasmid harbouring EBV *Ori*P[7]. A Chinese hamster ovary (CHO) cell-based system, *Epi*-CHO[8] has been engineered for episomal expression of recombinant proteins and made commercially available by ACYTE Biotech. The *Epi*-CHO system consists of the CHO-K1 cell line stably expressing the polyomavirus (Py) large T

antigen (PyLT), and the expression vector pPyEBV harbouring Py origin, EBNA1 and *Ori*P, which can be amplified and retained in the CHO-K1 cell. However, transient gene expression does not require the nucleic acid to be transfected in the format of circular plasmid DNA. It is also possible to transfect RNA or ribonucleoprotein complex (RNP) for non-integrative transient gene expression, which has been commonly applied to the delivery of CRISPR/Cas9 system[9] and mRNA vaccines[10]. Linear DNA such as PCR products can also serve as the vector for transient expression[11].

Stable transfection requires the plasmid DNA being integrated into the genome of the host cell, which may spontaneously happen in a very small fraction of the transfected cells, and can be facilitated by linearising the plasmid DNA or using site-directed recombination systems (such as Flp-In and Jump-In cell lines, discussed later in **Section 1.3**). Stable transfection typically requires antibiotics selection, clonal isolation, and functional validation. Stable transfection is usually required for engineering monoclonal stable cell lines.

1.1.3 **Arrayed libraries and pooled libraries**

There are two formats of libraries used for mammalian cell-based screening: arrayed libraries and pooled libraries. In an arrayed library, each member of the library is spatially separated in different vials or different compartments of the plates, whereas in a pooled library all variants are mixed in bulk. **Table** 1.1 compares the differences between transfection using arrayed libraries and pooled libraries.

The format of libraries is nontrivial for mammalian cells. Unlike the transformation of *E. coli* cells, which results in homogenous plasmid uptake and clonal expression, each mammalian cell typically uptakes a large and variable number of plasmid molecules during transfection. This characteristic can be conveniently used for co-transfection and co-expression of multiple genes encoded on different plasmid molecules. On the other hand, genotype-phenotype linkage needs to be maintained without interference from other genotypes, which typically requires separation of the library members in different cells. When transfecting libraries of pooled variants, plasmid molecules encoding different variants are usually co-transfected and co-expressed.

Table 1.1 Comparison of types of libraries for transfection

Method	Pros	Cons	Examples
Arrayed library – Plasmid	• Variants are physically separated. • Purified plasmid can also be used for subcloning and archiving.	• Expensive and laborious. • Throughput is limited by purification capacity.	Ref [12]
Arrayed library – PCR product	• Faster and simpler procedure than plasmid purification. • Easy to implement on automation systems. • Inexpensive.	• Linear DNA has potential cytotoxicity.	Ref [13]
Pooled library	• Convenient library preparation. • Throughput is not limited by purification capacity.	• Co-transfect and co-expression is very likely.	Ref [14]

	• Requires extreme levels of dilution. • Likely bias and loss of diversity during library handling.

1.1.3.1 **Arrayed libraries**

One solution to the problem of multiple plasmid uptake in mammalian cells is to purify the plasmids encoding different variants separately (**Figure** 1.1**A**). For example, many G protein-coupled receptor (GPCR)-based optogenetic indicators have been engineered to sense a wide range of neurotransmitters, such as (dopamine[15,16], serotonin[17,18], norepinephrine[19], acetylcholine[20]). These sensors are typically chimeric proteins with a circularly permutated (cp) fluorescent protein (FP) inserted into the intracellular loop 3 between transmembrane (TM) helices 5 and 6, where the largest conformational change is believed to occur[21]. Since the linker sequences on both sides of the cp FP domain are critical for the sensor performance, optimising such sensors usually involves randomising the residues of the linker via site-directed mutagenesis with degenerate codons. *E. coli* colonies expressing such libraries are picked and cultured individually, followed by plasmid miniprep in 96-well format. Thus, the library members are separated into different wells/vials. For mammalian cell-based screening, cells are first plated in multi-well plates. Each well is transfected with a different variant of the library. Transfected cells are subjected to microscopy-based assays. For example, in Yulong Li's lab, the G protein-coupled receptor activation–based (GRAB) sensor variants are cloned into a customised pDisplay vector co-expressing a PM-tethered mCherry (encoded by

IRES-mCherry-CAAX), which serves to label the PM as well as to normalise the fluorescence of GRAB sensors. Fluorescent intensities of GRAB variants with and without the ligand are acquired by high-content imaging, from which the dynamic range ($\Delta F/F_0$) can be calculated. Membrane trafficking can also be evaluated from the images. Similar method has been applied to engineering the genetically-encoded voltage indicators (GEVI) JEDI-1P[22] and JEDI-2P[12].

Linear PCR amplicons have also been used in transfection as an alternative to transfecting circular plasmid DNA. Expression of exogenous genes does not require the circular topology of the DNA. The presence of a promoter sequence and a complete open reading frame (ORF) is generally sufficient for transcription and translation. Therefore, it is possible to perform PCR from an earlier stage of the cloning workflow and bypass the *E. coli* handling (ligation/assembly, transformation, clone picking, liquid culture, miniprep, etc) altogether. Some studies have even suggested that crude PCR reaction mixes can be directly used for transfection without compromising transfection efficiency[13]. For example, to engineer ASAP3, Michael Lin's lab used overlap-extension PCR to introduce mutations at specific positions of the template of ASAP2, while amplifying the full-length expression cassette (CMV promoter, ORF, poly-A signal)[13]. For library screening, a HEK-293 cell line overexpressing inwardly rectifying potassium channel Kir2.1 was used to mimic the resting membrane potential of neurons[23]. The HEK-293-Kir2.1 cells were plated in 384-well-plates and transfected with unpurified PCR reaction by lipofectamine 3000. After locating the field of view and the focus plane, time-lapse images were recorded at 100 Hz for 5 sec with a 150 V 10 μsec square pulse delivered at 3 sec into the recording to induce the depolarisation. An automated image analysis was performed to determine the pulse-induced fluorescence change of the GEVI variant.

1.1.3.2 **Pooled libraries**

Pooled libraries refer to a mixture of library members residing in the same vial without being physically compartmentalised (**Figure** 1.1**B**). Compared to the arrayed libraries where each variant is physically separated from one another and prepared individually, individual library members in pooled libraries are usually prepared in the same reaction or procedure. Handling pooled library requires special considerations. In particular, procedures that are known to introduce and amplify bias should be avoided or kept at a minimum. For example, liquid culture of *E. coli* cells carrying plasmid libraries should be avoided, and PCR amplification of pooled libraries should be performed only when necessary, using the fewest number of cycles possible. Routine quality control is also recommended to evaluate the library diversity.

During the transfection, each mammalian cell takes up a large number of plasmid molecules. It has been estimated that each transfected cell harbours on average ~10^5 copies of a plasmid 24 hr post transfection[24], with 75 to 50,000 copies in the nucleus[25]. Therefore, to deliver single variants into single cells in bulk, libraries need to be diluted to increase the probability that a cell is only transfected with a single library member. For chemical transfection, dummy DNA, such as a pUC19 plasmid which does not encode any mammalian expression elements, can be used to dilute the library (**Figure** 1.1**B**). This way, a variant-encoding plasmid molecule is more likely to be delivered with the dummy DNA molecules, rather than another library molecule encoding a different variant.

Piatkevich *et al.* have investigated how variants can be separated in different cells by transfecting pooled libraries in bulk[14]. To optimise the transfection protocol, a two-

member library encoding either a green (EGFP) or a red (mCardinal) fluorescent protein was co-transfected with pUC19. The population expressing both green and red fluorescence (double expressors) indicated at least two variant-encoding plasmid molecules were transfected. When the library was diluted less than 100 times by pUC19, double expressors always dominated the transfected population. With a dilution factor of 100, the frequency of double expressor is approximately one half that of single expressors, with an overall transfection efficiency ~ 5 %. With a dilution factor of 1,000 where only 0.35 % of the cells were transfected, double expressors still represented ~ 15 % of the transfected cells[14]. To balance the transfection efficiency and variant separation, the authors transfected pooled GEVI variants at a dilution factor of 100. This screening campaign lead to a high-performance opsin-based GEVI Archon1[14]. Later in **Section 1.3**, I will introduce a recombinase-facilitated method for screening the pooled libraries by conventional transfection without diluting with dummy DNA.

Screening of pooled libraries in mammalian cell often necessitates phenotype and genotype determination at single-cell level. For phenotype measurement, unlike the average of multiple cells when screening arrayed plasmid libraries, phenotypes presented by single cells may depend on the protein expression level, which can vary considerably from one cell to the next. Cells exhibiting phenotypes of interest often need to be picked or sorted for enrichment or isolation. For genotyping and sequence recovery, single-cell techniques such as barcoding, whole genome amplification, and single-cell sequencing may be required. Although preparation of pooled library may be much cheaper than preparing arrayed libraries with the same diversity, cell isolation and sequence recovery downstream of pooled library can be expensive and nontrivial.

1.2 Viral transduction

Viral vector is another popular and efficient method for delivery genetic material. The virus-mediated gene delivery, also known as transduction[5], is especially useful for delivering genes to hard-to-transfect cells and living animals, and also when sustained expression is desired. In this section, I will briefly review the basis of lentivirus (LV) and adeno-associated virus (AAV) transduction (a comparison is provided in **Table** 1.2), followed by the discussion of representative examples of utilising viral transduction for library screening purposes.

1.2.1 Lentivirus (LV)

Lentivirus (LV) is a subtype of retrovirus. Like other retroviruses, LV carries a single-stranded RNA (ssRNA) genome. Upon infection, the RNA is reverse transcribed into DNA and integrated into the host genome. However, unlike typical retroviruses that only infect dividing cells, lentivirus can infect both dividing and non-dividing cells[26].

Most LV vectors used for research purposes are derived from human immunodeficiency virus (HIV). To produce LV, the packaging cell line, usually HEK-293 T or HEK-293 FT, is co-transfected with a transfer plasmid, 1 or 2 packaging plasmid(s), and an envelope-encoding plasmid. The transfer plasmid harbours an expression cassette of up to 8.5 kb[27] flanked by 5'- and 3'-long terminal repeats (LTR). The region between the LTRs is transcribed into ssRNA and packaged into the virion. The packaging plasmids encode the necessary genes for virus replication (*Gag*, *Pol*, *Rev*). These genes are encoded on the same plasmid for the 2nd generation packaging system, whereas the 3rd

generation encodes them on two plasmids. Because these genes are only supplied in trans when packaging, the LV particles do not carry these genes. As a result, the packaged LV virions can integrate their genome into the host genome, but the infected cells cannot produce new virions. The envelope plasmid encodes a glycoprotein, usually the vesicular stomatitis virus (VSV) G protein (VSV-G), which binds to the low-density lipoprotein receptor (LDLR) on the outer surface of host cells as the receptor for cell entry[28]. Compared to the native HIV that only infects CD4+ T cells[29], the VSV-G pseudotyped LV can infect a wide range of cell types[30]. LV can also be pseudotyped by many other envelope proteins[31], in order to infect specific tissues or cell types, study specific envelope-receptor interactions (e.g., Spike protein-ACE2 receptor[32]), and retroviral display of proteins of interest (POI) (see **Section 1.2.4**).

Packaged LV virions are released from host cells into the supernatant of cell culture. Filtered, sterile supernatant containing LV virions can be used for infection directly. Purification of LV particles can also be performed, if necessary, by ultracentrifugation, ultrafiltration, and ion exchange chromotography[33,34]. Several methods can be used to determine the titre of LV, such as flow cytometry-based functional titration, quantitative reverse transcription polymerase chain reaction (qRT-PCR) measurement of LV genome RNA, and detection of the HIV p24 capsid protein by ELISA or test paper[35]. Many commercially solutions (such as the Lenti-X system from TaKaRa Bio) are available for optimised LV packaging, purification, and titration.

Lentivirus is particularly useful for introducing exogenous sequences to primary cell culture and cell lines that are difficult to transfect. For example, Janelia Reseach Campus houses a high throughput platform that tests genetically-encoded Ca^{2+} indicators

(GECIs) in primary rat hippocampal neuron culture[36]. In this system, GECI variants are arrayed and individually packaged into LV. The resulting arrayed lentiviral GECI libraries are used to transduce primary hippocampal cultures. Like the assay introduced in **Section 1.1.3.1**, the neurons are subjected to field stimulation to induce action potentials with time-lapse images acquired for downstream analysis[36].

Lentivirus is commonly used for library screening in the format of pooled library. For example, antibody libraries in the single-chain variable fragment (ScFv) format can be packaged as LV and used to transduce cells at a low multiplicity of infection (MOI, the ratio of infectious virions to cells). Transduced cells display the ScFv variants on their cell surface via a transmembrane domain as a PM anchor[37]. Camino *et al.* reported a function-based antibody display assay where a ScFv library (diversity ~ 1.5×10^5) was cloned to form fusion proteins with a T cell receptor ζ (TCRζ) domain[37]. The resulting TCRζ-based chimeric antigen receptor (CAR) library in LV was used to transduce Jurkat T cells[38]. T cell activation, indicated by CD69 expression, was used as the reporter for antigen binding for fluorescence-activated cell sorting (FACS) [37].

Table 1.2 Comparison of LV and AAV transduction

	Lentivirus (LV)	**Adeno-associated virus (AAV)**
Viral genome	(+) ssRNA	(+) or (-) ssDNA
Copies of genome packaged each virion	2	1 - 2
Packaging capacity	~ 8.5 kb	~ 4.5 kb

Plasmids for packaging	• Transfer plasmid (1) • Packaging plasmid (1 or 2) • Envelope plasmid	• Transfer plasmid (1) • Helper plasmid (1) • Rep/Cap plasmid (1)
Packaging cell line	HEK-293 T or derivatives	HEK-293 or derivatives
Genome integration	Yes	• wtAAV integrate specifically at AAVS1 locus on Chr 19. • rAAV usually exists as circular dsDNA concatenated episomes, but may integrate at random sites at 0.1 % frequency.
Expression	Stable.	Potentially long lasting.
Cells infected	Dividing and non-dividing.	Dividing and non-dividing.
Tropism alteration	By pseudotyping with different envelope proteins	By choosing different serotypes, and/or pseudotyping with different capsid proteins from other serotypes.
Immunogenicity	Low	Very low
Biosafety containment	BSL-2	BSL-1

1.2.2 **Adeno-associated virus (AAV)**

Adeno-associated virus (AAV), a subtype of Parvoviridae, is a small, non-enveloped virus. Although wild-type AAV (wtAAV) is endemic to human[39], it is believed that AAVs do not cause any human diseases. AAV is incapable of self-replication. The replication of AAV requires the presence of a helper virus, such as Adenovirus (AdV). By

virtue of the non-pathogenicity and very low immunogenicity profile, AAV is considered a particularly promising viral vector for gene therapy[40]. Recombinant AAV (rAAV) has been widely used for *in vivo* experiments. So far, three rAAV-based therapies have been approved by U.S. Food and Drug Administration (FDA) or European Medicines Agency (EMA) [40].

Similar to LV (discussed in the previous section), production of rAAV requires triple transfection of a transfer plasmid encoding the transgene, a Rep/Cap plasmid encoding the AAV Rep and capsid proteins, and a helper plasmid encoding AdV E4, E2a and VA proteins[41]. The packaging cell line, usually HEK-293 generated by AdV transformation, provides two other necessary AdV proteins E1a and E1b for AAV production[42]. Depending on the choice of the Rep/Cap plasmid, rAAV can be packaged as different serotypes, as well as pseudotypes with capsid proteins from other serotypes or engineered capsid proteins for altered tropism[43,44].

In each rAAV particle, the genome exists as single-stranded DNA (ssDNA). AAV capsids have a capacity of ~ 4.5 kb. Mass spectroscopy analysis has shown that each rAAV8 viron primarily packages one copy of genome. A second copy of genome, complete or incomplete, may be packaged to fill the remaining room of the capacity[45]. After infection, wtAAV integrates its genome specifically to the AAVS1 locus on human chromosome 19[46]. For rAAV, since the *Rep* gene is removed from the transfer vector and supplied in trans, site-specific genome integration does not happen. rAAV genome predominantly exists as circularised monomer and concatemer episomes and drives long-term expression[47]. ~ 0.1 % frequency of random integration of rAAV genome has been observed[48].

1.2.3 **Functional genomic screening**

Functional genomics screening is a powerful method to identify genes involved in certain cellular/molecular processes by systematic genome-wide knock-out (KO) or knock-down perturbations of mammalian cells. Cells that exhibit a pre-determined phenotype are selected or enriched, and the genes associated with the phenotype can be identified from these cells. Notably, the key players in mediating the store-operated Ca^{2+} entry (SOCE, discussed in greater detail in **Chapter 2**), STIM[49,50] and Orai[51,52], were both identified by genome-wide RNAi screenings.

Nowadays, functional gene KO screens are most commonly achieved using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR-associated protein (Cas) 9 technology. Following the cleavage of dsDNA at the specific locus by the endonuclease Cas9, non-homologous end joining (NHEJ) is activated and repairs the double-strand break (DSB) in an error-prone manner. Insertion and deletion mutations (indels) are typical outcomes of NHEJ, resulting in functional KO of genes due to frameshift. Gene KD is usually achieved by RNA interference (RNAi) using small interfering RNA (siRNA), small hairpin RNA (shRNA), or CRISPR interference (CRISPRi) with a catalytically dead Cas9 (dCas9). In particular, shRNA screening and CRISPR-based screening (KO or CRISPRi-mediated KD) are usually administered in the format of pooled LV for scope and scale considerations[53]. Arrayed screens are more often used for smaller scale validation and characterisation. The rest of this section will be focused on the discussion of pooled CRISPR screening using LV (**Figure 1.2**).

CRISPR-Cas9 system has been engineered and widely used as a binary system composed of the Cas9 endonuclease (or dCas9 for RNAi) and a single-guide RNA (sgRNA). The sgRNA contains both structural regions of constant sequences as well as a ~ 20-nt targeting sequences that defines the specificity of Cas9 binding and cleavage[54]. The sgRNA library is introduced into cells by pooled LV. Following infection, the sgRNA sequence is integrated into the genome, allowing downstream sequence recovery and sequencing. The use of LV also allows screening in hard-to-transfect cells such as primary cell culture. To ensure that most cells only receive one sgRNA, cells need to be transduced at an MOI of 0.5 or lower[53]. The second component of the CRISPR screening, Cas9 or dCas9 protein, can be transiently expressed from a plasmid, or expressed from the same LV, or stably expressed in the cell. When working with a Cas9-expressing cell line, several clones may need to be tested to eliminate the possibility of a clonal artifact[53].

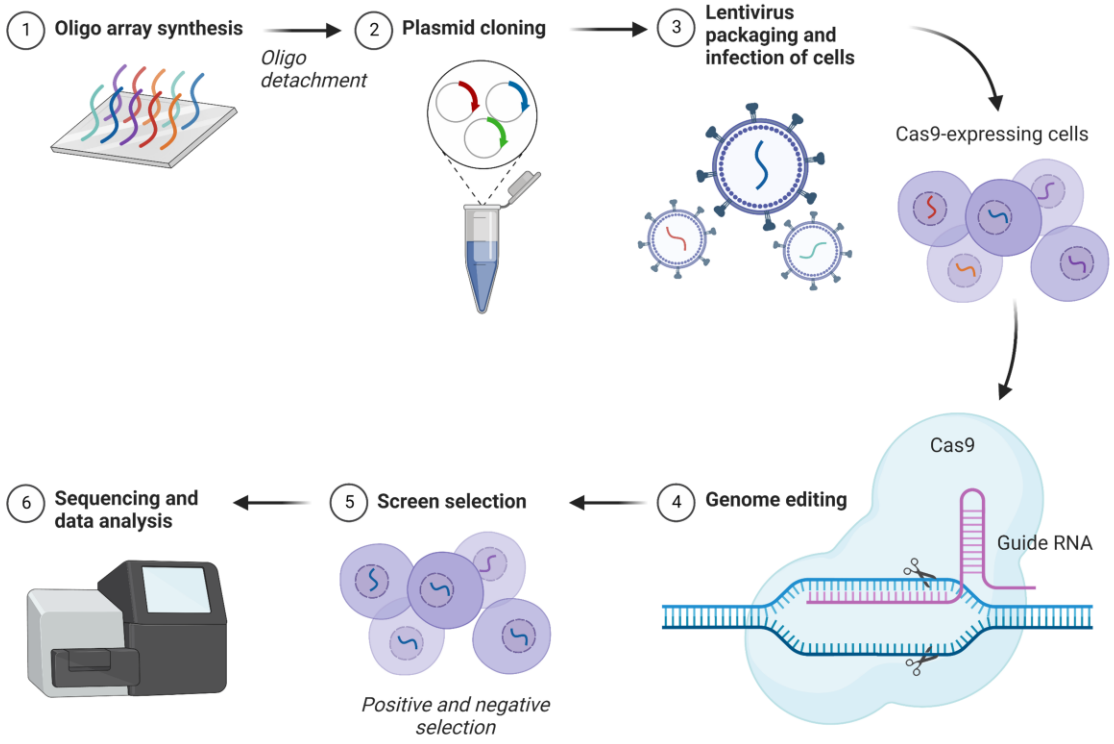

Figure 1.2 Schematics of CRISPR screening workflow.

A CRISPR screening typically starts with synthesising the oligonucleotides, which will be cloned into a LV vector to package a pooled LV sgRNA library. The pooled LV library encoding sgRNAs is used to infect cells expressing Cas9 or its variants, where functional KO or KD (CRISPRi) will take place. Following the appropriate selection methods to enrich the cells with desired phenotypes, the targets of gene manipulation can be determined by sequencing the sgRNA locus of the enriched cells by NGS.

A variety of methods for assaying the phenotypes have been developed. In some cases, the desired cells can be enriched or depleted. For example, cells with better fitness of proliferating are enriched whereas those with worse fitness are depleted[55]. Other phenotypes do not confer the cells with advantages or disadvantages in proliferation, but drives the expression of fluorescent reporters or cell surface antigens. Alternatively, cells can be manually tagged by photoactivatable fluorescent proteins (such as PAmCherry)[56,57] or cell labelling via photobleaching (CLaP)[58]. These phenotypes allow cells to later be

enriched by FACS or magnetic-activated cell sorting (MACS). Next-generation sequencing (NGS) is usually employed to quantify the abundance of different sgRNA sequences by counting the number of reads[55]. By comparing the presentation of sgRNAs with different conditions, treatments or time points, genes associated with the sgRNA sequences that are enriched or depleted can be identified, indicating selective advantage or sensitivity of these genes[53]. In addition to quantifying the enrichment of sgRNAs of a pool of selected cells, the single cell-level CRISPR screening (scCRISPR-seq), which profiles the transcriptome and/or proteome on a single cell level and associates with the sgRNA expressed in that cell, has also become more accessible and affordable to researchers.

For some types of assays, cells cannot be enriched by fitness of growing or cell sorting. Accordingly, methods are needed to identify the sgRNAs transduced in the cells of interests *in situ*. One elegant solution is to barcode the sgRNA constructs. For example, Xiaowei Zhuang's lab pioneered the development of methods to associate the sgRNAs with unique 12-digit ternary barcodes. The barcodes each cell carries can be read out using the multiplexed error-robust fluorescence in situ hybridization (MERFISH) technique with 18 rounds of FISH. In situ genotyping of sgRNA not only allows for an all-imaging based workflow, but also addresses the limitation that sgRNAs are not readily picked up during scCRISPR-seq due to the lack of polyadenylation[59].

LV systems do have several limitations that can complicate their application for pooled library screening. LV virions carry two copies of ssRNA genome. Recombination may happen when the transfer plasmid is being transcribed into ssRNA, known as template switching[60], which may break the association between sgRNA and the

barcodes[61]. In some cases the two copies of ssRNA may encode different members of library, known as pseudodiploidity[62], which can compromise the variant separation ability. Aggregation of the viral particles[63] is another cause of delivery of multiple variants into the same cell.

1.2.4 **Retroviral/Lentiviral display**

LV and other retroviruses are enveloped with a PM-derived membrane[64]. Overexpressed PM-localised integral membrane proteins in the packaging cell are likely to be incorporated into the viral particle (**Figure 1.3A**)[65]. For example, LV used in research labs is typically enveloped with VSV-G due to overexpression of VSV-G in the packaging cell. VSV-G is a VSV-originated protein, yet it is properly functional and mediates the cell entry of LV into a large variety of cell types[28]. When SARS-CoV-2 Spike protein is overexpressed in the packaging cell line in lieu of VSV-G, LV virions are pseudotyped with Spike protein and recapitulate many important steps of cell entry of ACE2-expressing cell lines. The Spike-pseudotyped LV is a safe, nonreplicating alternative to the SARS-CoV-2 virions for studying Spike-ACE2 interaction and antibody neutralisation without the need for Biosafety level 3 facilities[32].

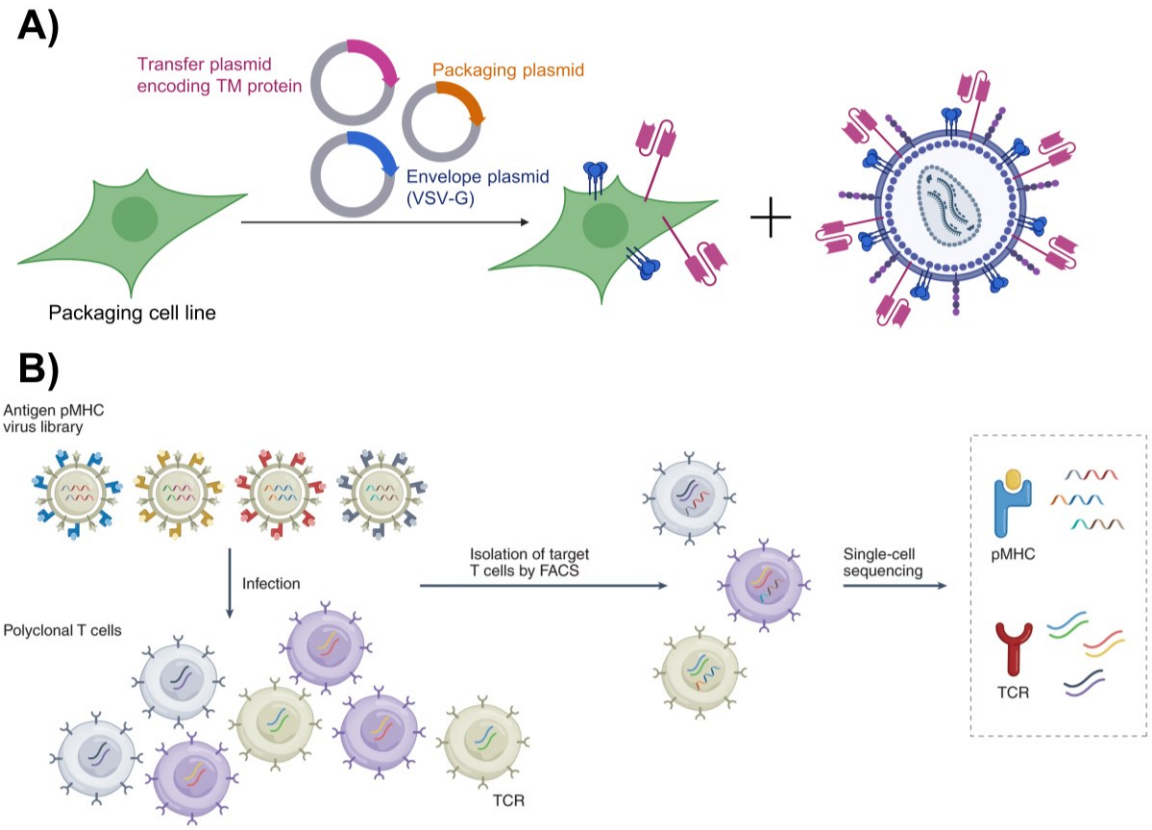

Figure 1.3 Overview of retroviral/lentiviral display technique.

A| Schematics of the molecular biology of lentiviral display. When packaging LV using a packaging cell line such as HEK-293 FT, a membrane protein encoded by the transfer plasmid is expressed in the packaging cell alongside with or in place of VSV-G. Both VSV-G and the membrane protein can be incorporated into the envelope membrane of LV and displayed on its surface. **B|** Schematics of TCR-pMHC screening using RAPTR system. Pooled VSV-G mutant-pseudotyped LV particles displaying a library of pMHCs are used to infect polyclonal primary T cells in a library-on-library fashion. The cell entry is mediated by the specific binding of the TCR-pMHC pair. The infected cells are isolated and subjected to single-cell sequencing to identify the sequences of the peptide and TCR. Adapted with permission from the Licensor: Springer Nature Nature Methods, Yuqian Wang et al, *Engineered retroviruses map ligand–receptor interactions*[66], Copyright 2022.

The ability to incorporate and display membrane proteins has inspired researchers to harness retroviruses as a display platform for screening and directed evolution (**Figure**

1.3A). Retroviral/lentiviral display is conceptually similar to its Nobel Prize-winning prokaryotic counterpart, phage display. However, expression of eukaryotic proteins, which typically need chaperones to facilitate folding and may have posttranslational modifications, is not appropriate for phage display. As LV infect mammalian cells rather than bacteria, they can be used as a mammalian production platform for proteins[67]. Taube et al. demonstrated that antibodies could be displayed both on the PM and LV particles in single-chain variable fragment-constant fragment (ScFv-Fc) format[68]. A series of transmembrane anchor domains of both viral and non-viral origins were tested for their ability to display ScFv-Fc. With one round of selection using magnetic beads and FACS, antibody expressing cells were enriched up to 10^6 times[68]. In another example, Merten and co-workers used retrovirus display for directed evolution of an enzyme[69]. The virions were compartmentalised in water-in-oil emulsions. Each emulsion was tested for the presence of functional enzymes using a microfluidic device. Through a single round of screening, active wild-type tissue plasminogen activator was enriched by more than 1,300 times[69]. Recently, a “receptor–antigen pairing by targeted retroviruses” (RAPTR) system was reported to profile T cell receptors (TCR) and B cell receptors (BCR) with their cognate antigens (**Figure 1.3B**)[70]. In this work, LV particles were produced by co-expressing the VSV-G K47Q R354A mutant with ablated affinity with its native receptor LDLR, and the antigen candidates encoded by the LV genome. Since the VSV-G mutant is non-functional, cell entry is mediated by the specific interaction of the displayed antigen and the cell surface receptor (**Figure 1.3B**). Specifically, for TCR-based screening, short peptide fragments presented by major histocompatibility complex (MHC) proteins (pMHCs) served as the antigen candidates displayed on the LV envelope. Cells infected by TCR-pMHC-mediated entry were labelled by both transduction marker GFP encoded

by LV, as well as activation marker CD69. RAPTR system is particularly useful for profiling TCR-pMHC pairing in human primary T cells in a library-versus-library format (**Figure 1.3B**)[70].

To date, the applications of retroviral/LV display are relatively limited. One reason is that the molecular mechanisms of membrane protein incorporation are not fully understood[71]. In particular, antibodies displayed in IgG format has been achieved and widely used for cell surface display[72,73], but IgG displayed on viral particles has not been reported. There is still a great deal of potential of LV display to be explored in accelerating membrane protein engineering for their binding properties.

1.3 Recombinase-mediated genome integration

Bacteriophage- and yeast-derived recombinases have been used to mediate sequence-specific recombination between two DNA molecules and also within one DNA molecule[74,75]. These recombinases can be used to achieve a variety of genome manipulations, including integration/excision of circular DNA molecules, inversion, translocation, and cassette exchange[74,75]. Unlike the programmable endonucleases (such as Cas 9, TALENs and zinc fingers), DNA recombinases do not produce permanent double-stranded breaks (DSB)[76]. Many of the DNA recombinases remain active not only in their native host cells, but also in cell free systems, *E. coli,* and other eukaryotic cells when heterologously expressed[75,77]. Therefore, DNA recombinases represent a class of genome editing tools that yields highly specific and predictable outcomes, making them

useful tools for achieving conditional knock-in and knock-out *in vitro* and *in vivo*. For example, mouse strains that carry Cre recombinase under different promoters can express Cre recombinase in a tissue-specific manner. When crossed with another strain carrying *lox*P loci, gene expression can be selectively turned on or off in a specific tissue[78]. Similar systems with several different recombinases are also available for other model organisms for developmental studies. Split versions of recombinases have also been engineered that can reconstitute function upon chemical- or light-induced dimerization, facilitating more flexible and precise control of gene expression in mammalian cells[79]. In addition, various applications have been made possible with serine recombinases, such as plasmid construction using λ integrase (Gateway cloning)[80–82], genetic logic gate computation[83–87], and biological event recording[88].

The first identified recombinases were the tyrosine recombinases, such as, Flp[89–91] and Cre[92–94]. Tyrosine recombinases recognise and catalyse the reversible recombination between identical sequences using an enzymatic mechanism that involves a nucleophilic tyrosine hydroxyl group[95–98]. More recently, serine recombinases, another class of recombinases, have been reported. Unlike tyrosine recombinases, serine recombinases recognise non-identical DNA sequences and catalyse directional recombination via a mechanism that involves a nucleophilic serine hydroxyl group[99–101]. The reverse reactions of serine recombinases typically require an additional recombination directionality factor (RDF). In the absence of the RDF, the forward reactions catalysed by serine recombinases are effectively irreversible[102–105]. There is great potential for serine recombinases to facilitate the next generation of gene therapies[106,107].

1.3.1 **Commercially available platform cell lines for genome engineering**

Recombinase-mediated DNA incorporation has been used to facilitate the establishment of stable cell lines. Generally, in this strategy a platform cell line is first established by introducing a DNA sequence (colloquially called a “landing pad”) harbouring a recognition sequence of the recombinase into the host genome. This platform cell line is then used to generate various expression cell lines by co-transfecting a donor plasmid and a plasmid encoding the recombinase. The transfected cells will express the recombinase, which, in turn, will mediate the recombination between the landing pad and the donor plasmid. For example, the Flp-In™ systems, which were first demonstrated by O’Gorman et al[108] and later became commercially available from Thermo Fisher, are cell lines stably transfected with p*FRT*/lacZeo that provides a Flippase recognition target (*FRT*)-containing landing pad for Flp recombinase. The *FRT* sequence on the landing pad is downstream of an SV40 promoter. To create isogenic stable cell lines, the commercially available pcDNA5/*FRT* donor plasmid is co-transfected with the Flp-encoding pOG44 plasmid. The donor plasmid contains a CMV cassette that drives the constitutive expression of the gene of interest (GOI), and a promoter-less ORF encoding aminoglycoside phosphotransferase following a *FRT* sequence. Upon recombination of the *FRT* sequences between the host genome and donor plasmid, the aminoglycoside phosphotransferase is placed downstream of the SV40 promoter, conferring the host cell hygromycin resistance, which can be used to select cells that have undergone successful recombination.

There are two major limitations of the Flip-In™ system. The first limitation is that the GOI expression under a CMV promoter is constitutive and independent of the host

genome integration. That is, both transiently transfected donor plasmid molecules and the stably transfected donor plasmid molecules will contribute to the GOI expression. The second limitation is that the tyrosine recombinase Flp catalyses the reversible recombination of the same *FRT* sequences. As a result, excision of circular DNA is favoured over the integration into the genome. Recombination between donor plasmid molecules is also possible, which may result in complicated recombination outcome and unpredictable behaviour. Therefore, isolation of monoclonal stable cell lines is necessary for Flp-In™ system. In a representative example, Grimson and colleagues used the Flp-In™ T-REx 293 cell line and a pcDNA5/*FRT*-based poly-cistronic vector to screen the regulation sequences of 3' UTR. The authors claimed that the Flp-In™ system was advantageous by yielding uniform single-copy integrations positioned at the same gene locus, removing the dosage effect and location effect, thus improving the SNR[109]. From each 10 cm petri dish, the authors were able to harvest an average of 173 ± 49 colonies surviving the antibiotics selection[109], suggesting a decent recombination efficiency for stable cell line isolation, but far less than ideal for library screening.

Another commercially available cell engineering system is the Jump-In™ cell lines developed by Thermo Fisher[110,111]. In this system, a landing pad plasmid (pJTI™) is introduced by the ΦC31-mediated recombination into a pseudo-*attP* site in the host genome. This landing pad carries an *attP*-site for R4 recombinase proceeding a promoter-less antibiotics resistance gene. Similar to the Flp-In™ system, the donor plasmid for Jump-In™ system (pJTI™ R4 DEST CMV-TO pA) encodes the GOI cassette under the CMV promoter. Upon recombination, a EF1α promoter is placed in front of the antibiotic resistance gene on the landing pad, conferring the host cell resistance to the corresponding selection agent. Consistent with the Flp-In™ system, the GOI under a CMV

promoter can be expressed from both integrated and un-integrated donor plasmid molecules. The major advantage of the Jump-In™ system is the unidirectional and irreversible integration by virtue of the serine recombinase R4. Notably, the landing pad of Jump-In™ cell lines does not provide a promoter that drives the ORF on the donor plasmid, making this system unsuitable for screening protein libraries. However, this unique design of incorporating a promoter to drive the expression a selection marker system could be potentially used to evolve an engineered promoter.

1.3.2 **Bxb1 recombinase and recent progress in mammalian cell screening system**

Recently, the Bxb1 serine recombinase has received increasing attention as a new tool for genome engineering (**Figure 1.1D**). As with other serine recombinases, Bxb1 is highly unidirectional. Unlike ΦC31, there are no known pseudo-*attP*/*B* sites present in human genome for Bxb1[112]. That is, Bxb1 is specific towards the Bxb1 recognition sites that are introduced in the landing pad, and unreactive towards all native human genome loci. Indeed, systematic comparisons have demonstrated Bxb1 to have the highest efficiency and accuracy[113] with favourable enzymatic properties at low expression level in mammalian cells[114]. These desirable features have made Bxb1 recombinase an attractive tool for genome engineering and library screening.

Some efforts have been made to utilise Bxb1 for cell engineering. For example, Leonhardt and colleagues used CRISPR/Cas9 to introduce in-frame *attP* sites at various genomic loci after the start codon[115]. A donor plasmid can then provide a fluorescent or enzymatic reporter gene, or mutants of the open reading frame (ORF), that will be

expressed from the native gene locus. The authors also claimed that the in-frame *attP* site introduced to the genome may also serve as an epitope tag for antibody-based detection and precipitation[115].

In another example, Matreyek and colleagues engineered a series of tetracycline-inducible *attP* landing pad HEK-293 cell lines for assessing protein variant libraries encoded on a promoter-less *attB* donor plasmid[116–118]. This system was intended for barcoded deep mutational scanning by next-generation sequencing (NGS) and achieved a recombination efficiency of 8%[116] and has been applied to screening mutants of the human ACE2 receptor[118].

While Leonhardt's and Matreyek's system uses a single *attP* × *attB* recombination for circular plasmid integration, other researchers have developed host cell lines that use two orthogonal *attP* × *attB* pairs for donor DNA integration. For example, researchers from Pfizer created a CHO cell line for stable expression of monoclonal antibodies (mAb) by recombinase-mediated cassette exchange (RMCE) with donor DNA encoding the mAb ORF[119]. This system relied on antibiotic resistance selection and the recombination efficiency was not reported. In another example, the Pedersen lab created a virus-free genome-wide functional knock-out screening system in CHO cells by RMCE with libraries of CRISPR/Cas9 gRNA[120]. The integrated landing pad provides an EF1a promoter and two orthogonal *attP* sites for recombination, while the donor plasmid encodes a promoter-less puromycin resistance gene and the gRNA under a U6 promoter flanked by two orthogonal *attB* sites. The authors reported an RMCE efficiency of 4% as measured by the loss of mCherry fluorescence due to inactivation by recombination. Although this system ensures the incorporation of one gRNA in each cell, the transiently transfected

unintegrated donor plasmid can still express the gRNA under the U6 promoter. Expression of Cas9 protein at the presence of unintegrated donor plasmid might lead to off-target knockout. On the other hand, in principle the RMCE design allows the utilization of non-plasmid linear libraries as donor DNA, though neither of the previous reports explored this possibility.

In **Chapter 3**, I will describe my original work on a mammalian cell-based library screening system for clonal expression of protein engineering libraries. This system also facilitates seamless prokaryotic expression and is compatible with direct screening of PCR libraries.

1.4 *In situ* mutagenesis

The discussion in the previous sections focused on mammalian cell-based screening methods for libraries that are generated outside of living cells and introduced into the cells for expression as is. In this section, I will discuss the generation of mutations in mammalian cells (*in situ* mutagenesis) and their applications.

1.4.1 Lymphocyte-inspired *in situ* mutagenesis

The human body has a diverse (~ 10^{12}) repertoire of unique, naïve antibodies[121]. The diversity of antibodies is not encoded in the germline genome. Rather, the diversity is generated through a series of intricately regulated diversification processes of immunoglobulin (Ig) genes in the somatic lymphocytes.

Human Igs consist of a heavy chain (encoded by IgH genes) and a light chain (IgL, encoded by Igκ or Igγ). Each chain has a N-terminal variable (V) region and a C-terminal constant (C) region. The variable region is highly variable and responsible for antigen-binding[122]. The V region of IgH gene locus contains multiple variable (V_H), diversity (D_H) and joining (J_H) segments, whereas the V region of IgL loci contain multiple V_L and J_L segments. When B cells are developing in the bone marrow, one segment from each of the V_H, D_H and J_H are joined to form the V region of IgH while one segment from each of the V_L and J_L segments are joined to form the V region of IgL. The bespoke process, called V(D)J recombination (**Figure 1.4A**)[123,124], is mediated by the recombination activating genes 1 (RAG1) and RAG2 recombinases. RAGs are specifically expressed in lymphocytes and recognises the recombination signal sequences (RSS) that are present on the 3' side of V_H and V_L segments, both sides of D_H segments, and 5' side of J_H and J_L segments. In addition to the combinatorial diversity from the different segments, further diversification of the sequence is introduced terminal deoxynucleotidyl transferase (TdT) that adds template-independent nucleotides[123,125], as well as NHEJ DNA repair pathway[126] following the DSB by RAG proteins. At this stage, the naive antibody repertoire can bind a diverse range of antigens, but the affinity is not yet optimal for protection.

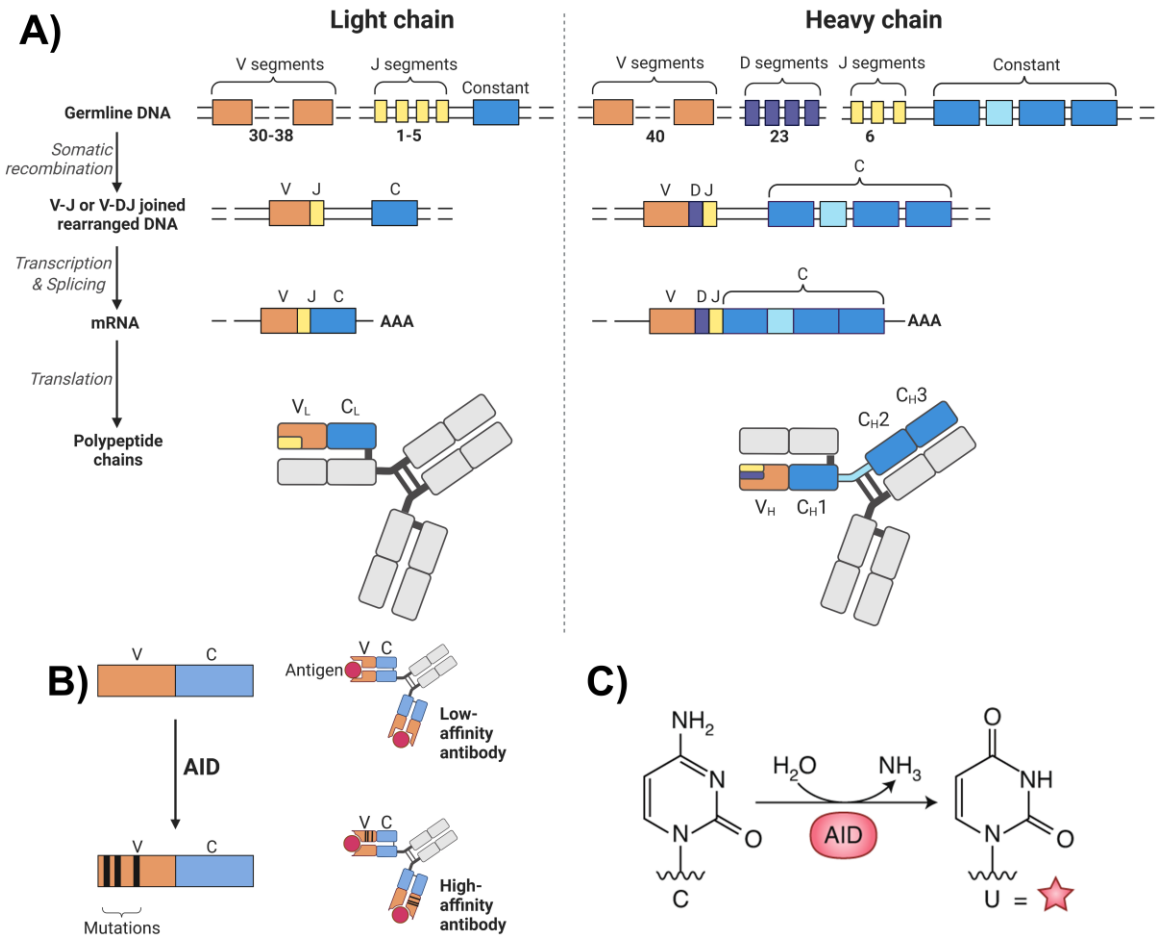

Figure 1.4 Schematics of antibody diversification in B cells.

A| V(D)J recombination in B cells joins one segment from each of the V_H, D_H and J_H segments to form the V region of the heavy chain, and one segment from each of the V_L and J_L segments to form the V region of the light chain. This process is mediated by RAG1/2 via a DSB intermediate, followed by TdT-mediated addition of non-templated nucleotides and the error-prone NHEJ DNA repair. Both the combination of V(D)J segments and the mutations introduced by TdT and NHEJ contribute to the antibody diversity at this stage. **B|** SHM in B cells takes place via the AID-mediated deamination of bases in the Ig loci, followed by the error-prone DNA repair mechanism. SHM results in further diversification of antibody sequence and increase the antigen-specific affinity of the antibody. **C|** AID hydrolyses the 4-amino group on cytosine (C) and converts it into uracil (U).

Following antigen stimulation of B cells in the periphery, a second stage of antibody diversification takes place by somatic hypermutation (SHM) of the V regions[127] (**Figure 1.4B**). SHM is mediated by the activation-induced cytidine deaminase (AID), which deaminates cytidine to uridine of the transient ssDNA stage during transcription on both sense and anti-sense strands (**Figure 1.4C**)[128,129]. Although non-Ig genes may also be deaminated during transcription, AID preferentially targets the cytidine in the WRCH ([AT][AG]C[TAC]) motif, which is also well recognised by the error-prone DNA repair pathways, resulting in several possible repair outcomes[130,131].

Many of the abovementioned processes can be recapitulated in cultured cells. To start with, B cell-derived cell lines are capable of mutate to endogenous Ig sequences as well as exogenous genes via SHM. For example, the Ramos cell line, derived from human Burkitt's Lymphoma, was shown to constitutively hypermutates its V regions of its genes during culture. The resulting IgM antibody is expressed and displayed on its cell surface[132]. This trait has made Ramos a useful tool for generating and maturating antibodies[133]. Through iterative culture and screening, the endogenous IgM of Ramos was evolved to recognise streptavidin with nanomolar K_d[133]. Similarly, the avian B cell line DT40 has also been used for antibody discovery with shorter generation time and higher mutation rates. Since AID also may also mutate non-Ig genes, the Ramos cell line has been reported to mutate and evolve exogenous non-Ig proteins like fluorescent proteins[134] and anti-apoptosis protein Bcl-xL[135]. Non-lymphocyte mammalian cell lines, like CHO cells and non-small cell lung carcinoma H1299 cells, are also shown to be capable of SHM when transfected with AID and mutate the stably expressing Ig genes[136,137]. Unfortunately, the ability to mutate random genes makes AID cytotoxic when expressed at high levels. Moderation of AID expression level may be required for directed evolution[138]. Alternatively,

restricting the access of AID to only the regions close to the GOI is beneficial for reducing the off-target mutation. For example, in the T7 polymerase-driven continuous editing (TRACE) system (**Figure 1.5A**)[139], AID is fused with the T7 bacteriophage RNA polymerase (T7 RNAP). T7 RNAP-mediated transcription is initiated at T7 promoter, both of which are orthogonal to the mammalian molecular machineries. Once transcription is initiated, T7 RNAP is highly processive and can travel for several kb[140], while the AID fused to the T7 RNAP mutates this part of gene. Another targeted AID design is to fuse AID with the Cas9 protein in a base editor system, which will be discussed in the next section.

Another design is to recapitulate the mutagenesis following V(D)J recombination (**Figure 1.4A**). In the HuTARG system[141], a plasmid is constructed to encode both the IgH and IgL expression cassettes of antibody 731 (Ab731). A pair of a 23-bp RSS and a 12-bp RSS separated by a 1.4 kb space holder was introduced to a series of positions within complementarity-determining region 1 (CDR1), CDR2 and CDR3 for each V, D, and J segment with the appropriate RSS signal for recombination. The resulting plasmid library was introduced in a HuTARG cell line stably transfected with TdT, RAG1 and RAG2 via Cre-*lox*P recombination in a one-cell-one-copy fashion. Expression of RAG1 and RAG2 was induced by tetracycline to initiate RSS-mediated recombination, resulting in the removal of RSS pair and the spacer. Following the recombination and DNA repair, full IgG proteins comprising of both chains were displayed on the cell surface. By staining with fluorophore-conjugated ligand followed by FACS, the authors were able to evolve IgGs with picomolar affinity towards IL-13[141]. The HuTARG system provides a flexible platform that combines *in vitro* and *in situ* mutagenesis and library screening within a fully

humanised background. It is also possible to introduce multiple options of V(D)J segments[142] and evolve T cell receptor (TCR)-based antibodies[143].

1.4.2 **CRISPR-mediated mutagenesis**

As is described in **Section 1.2.3**, the typical outcome of DNA repair followed by CRISPR/Cas9-mediated DSB is highly error prone and may serve as a way of introducing mutations. Some efforts have been made to screen for in-frame mutants to create drug-resistant variants of endogenous gene loci[144] and exogenously expressed enzymes[145] followed by Cas9-induced DSB and NHEJ. However, NHEJ is much more likely to result in indels that are out of frame and lead to loss of function due to a premature stop codon. Therefore, CRISPR-based methodologies that do not involve NHEJ pathway are highly desired for gene diversification purpose.

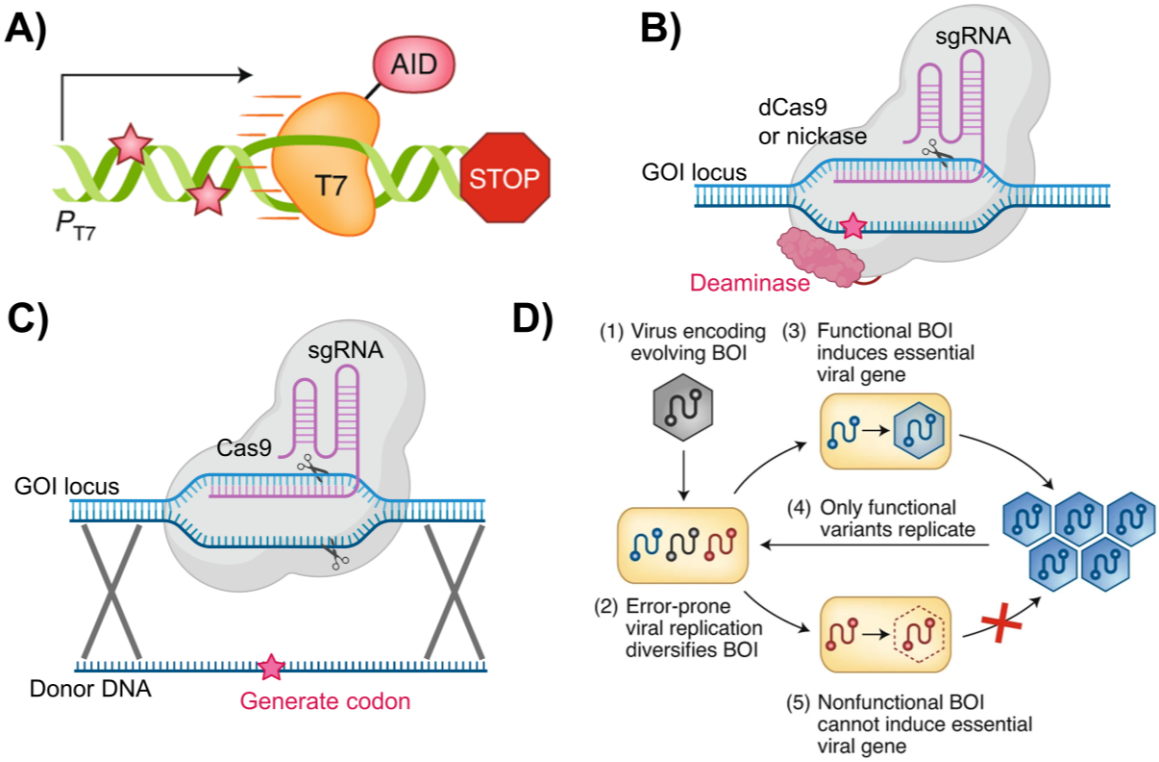

Figure 1.5 Examples of *in situ* mutagenesis techniques.

A| Schematics of TRACE system[139] where AID is fused with T7 RNAP that recognises the T7 promoter preceding the GOI and transcribes the GOI. AID introduces mutations by deaminating the bases on the partially winded DNA during transcription. Adapted from Figure 4 of ref[146] **B|** CRISPR/Cas9-based diversifying base editor (in this example, CRISPR-X system[147]) where the GOI is targeted by the gRNA with the deaminase acting on the ssDNA portion of the Cas9 complex. **C|** Gene diversification by CRISPR/Cas9-mediated HDR. Unlike NHEJ, HDR takes place in the presence of homologous template and has less chance of introducing indel mutations. Candidate variants can be encoded by degenerate codons on the template DNA, which is usually ssDNA harbouring 5'- and 3'-phosphorothioate linkages. **D|** General mechanism for virus-assisted continuous evolution in mammalian cells. When the activity of biomolecule of interest (BOI) is coupled to the fitness of viral replication, virus encoding the active variants will be enriched over the generations. Adapted with permission from the Licensor: Springer Nature Nature Methods, Samuel J. Hendel et al, *Directed evolution in mammalian cells*[146], Copyright 2021.

One such method that avoids the NHEJ pathway involves promoting homology-directed repair (HDR, **Figure 1.5C**). HDR requires the presence of a donor DNA homologous to the sequence around the DSB. In CRISPR/Cas9 system, synthetic ssDNA harbouring 5'- and 3'-phosphorothioate linkages is a good option for donor DNA with up to 35 % HDR efficiency[148]. Degenerate codons can be incorporated into the ssDNA donor to generate site-saturated randomisation libraries at specific positions[149]. More recently, Erdogan *et al.* reported a CRISPR-HDR-based system for engineering fluorescent proteins[150]. The cell line was stably transfected with a single copy of mRuby3-mTagBFP2 parent construct via Flp-FRT recombination, with a 59 bp deletion in mRuby3 gene where many residues are involved in interactions with the chromophore. The deletion rendered mTagBFP out of frame and non-fluorescent, but fluorescence could be rescued upon HDR with WT mRuby3 donor which served as an in-frame reporter. In this design, targeted mutations can be introduced to the deleted region via a combination of degenerate codons and the *in situ* error-prone repair processes. The single-copy genomic incorporation of the parent construct also ensures that the library could be screened in a single-variant-per-cell manner. With the parent construct mRuby3-mTagBFP fused to the lysosome-associated membrane glycoprotein 3, the authors identified a long Stokes shift, pH-resistant mRuby3 variant compatible with the lysosome context[150].

Another method is to use the base editor systems (**Figure 1.5B**). Base editors typically use Cas9 variants with one (Cas9 nickase, nCas9) or both (dCas9) catalytic residues mutated, and which are therefore incapable of producing DSB. These Cas9 variants either are directly fused with a deaminase (designs similar to ref[151]), or recruit a deaminase via the binding of MS2 loop and MS2 domain fused to the deaminase (e.g., CRISPR-X[147]). The deaminase chemically modifies the bases close to the target site.

While most versions of base editors were designed to produce precise and specific editing outcomes[151–153], several diversifying base editors[147,154–156] have been engineered to facilitate targeted library generation.

One limitation of the CRISPR-based systems is that the window for mutation is rather narrow (4 – 50 bp). Therefore, to mutate a larger area of the gene, a pooled sgRNA may be used. During iterative directed evolution, new sgRNA need to be synthesized to accommodate the mutations introduced in the previous rounds. More importantly, some regions may not be accessible to CRISPR-based tools due to the lack of appropriate protospacer adjacent motifs (PAM)[146]. Tools with wider tolerance of PAM sequences, larger mutation windows, and more diverse mutation profile will be highly desirable for *in situ* mutagenesis.

1.4.3 **Continuous evolution in mammalian cells**

In directed evolution with libraries prepared *ex vivo*, mutagenesis and phenotype screening or selection are separated spatially and temporally. In contrast, *in situ* mutagenesis makes it possible to generate mutations and screen or select the mutations in the same cell. In this way, it becomes possible to implement directed evolution in a continuous manner (**Figure 1.5D**). The most successful continuous evolution system is an *E. coli*-phage system developed by David Liu's lab and called phage-assisted continuous evolution (PACE)[157]. In the PACE system, mutations are continuously introduced by highly error-prone DNA polymerases[158]. The activity of the GOI (such as an

RNAP) is linked to the expression of Gene III of M13 phage in *E. coli*, and thus also to the fitness of the phage. The winner variants will have greater infectious progeny, which in turn infect more *E. coli* cells. PACE is a highly efficient system. After ~200 rounds over 8 days, the T7 RNAP was evolved to recognise the T3 promoter with >200-fold increase of activity[157].

Two mammalian cell-based systems that implement PACE-like virus-assisted continuous evolution have been reported (**Figure 1.5D**). The system published by Matthew Shoulder's lab[159] used an engineered AdV. AdV has a dsDNA genome that is conveniently compatible with standard molecular biology procedures. Two essential genes, the DNA polymerase (AdPol) and protease (AdProt), were removed from the AdV vector and replaced with the sequence encoding the GOI. An engineered error-prone AdPol variant named EP-Pol (AdPol-F421Y-D827A) was stably expressed from a genomic locus in the HEK-293 A host cell genome and used to replicate and mutate the AdV genome at a rate of 3.7×10^{-5} mutations per base per passage. AdProt, which plays essential roles from cell entry to DNA replication, was used as a fitness gene and placed downstream of the tetracycline transactivator (tTA) operator in the host cell. The trans-complementation design ensured that AdV was only capable of replication in the engineered host cell line[159]. As a proof of principle, tTA was cloned into the AdV vector as the GOI. In the presence of doxycycline, an allosteric inhibitor of tTA[160], only the cells infected with doxycycline-resistant mutants were able to express AdProt, thus increasing the progeny of such mutants. Continuous evolution was performed by transferring AdV-containing cell culture supernatant to the fresh host cell culture. Sequence analysis suggested that the AdV-based continuous evolution was able to enrich the reported doxycycline-insensitive tTA mutants[161] E47K and H100Y to more than 70% frequency by

the fifth passage[159]. In addition, the authors also demonstrated the use of AdV-based system to evolve the Cre recombinase and leucyl-tRNA synthetase[159].

Another mammalian cell-based continuous evolution system, the “viral evolution of genetically actuating sequences” (VEGAS) system, was published by Bryan Roth, Justin English and coworkers[162]. In this system, the Alphavirus Sindbis[163] was used as the vector. The Sindbis RNA-dependent RNA replicase specifically replicates the ssRNA genome, primed by conserved cis-acting 5-3’ sequences[164]. Like most ssRNA viruses, the Sindbis virus replicates with a high mutation rate due to the absence of proof-reading ability[165,166]. The obligate RNA-only life cycle and specificity of the replicase makes Sindbis an orthogonal mutagenesis system in the mammalian context without mutating the host genome. The mutation rate of the VEGAS system was estimated to be 1.0×10^{-5} ($\pm 0.4\times 10^{-5}$) mutations per base per hour[162]. In the VEGAS system, the expression of Sindbis structural genes (E1, E2 and Caspid) was used as the selective pressure. In addition to evolving the tTA variants, the authors demonstrated the development of a genetic circuit to couple the G protein-coupled receptor (GPCR) activation with structural gene expression by using the serum response element (SRE) minimal promoter. With this circuit, the authors evolved constitutively active mutants of an opioid-recognition receptor MRGPRX2, as well as nanobodies stabilizing the active-state of the GPCRs serotonin 2A (5-HT_{2A}), dopamine-D2 (DRD2), and pH-sensing (GPR68) receptors, each within a week[162].

Continuous evolution is a promising directed evolution strategy that leverages the power of evolution principles at the population genetics level[167,168]. It also represents a convenient and efficient solution for directed evolution without iterative molecular cloning

or artificial selection. However, careful design and extensive optimisation are required for such systems to behave properly. In particular, creative designs to couple the desired phenotype to the fitness of viral propagation is key in continuous evolution systems (**Figure 1.5D**). Notably, both of the mammalian continuous evolution systems reported to date rely on the transcriptional regulation of viral essential genes as the selection method. The development of a broader range of phenotype-fitness coupling methods will greatly expand the versatility of such systems for engineering a wide variety of GOIs.

1.5 Scope of thesis

In Chapters 2, 3 and 4, I will describe my original work on the development of an optogenetic endoplasmic reticulum Ca^{2+} release system, a mammalian cell-based library screening system for directed evolution, and a computational study on the dissociation mechanism of photocleavable protein (PhoCl). These projects will be presented as standalone chapters for two reasons: 1) They are the most innovative achievements of my PhD studies and represent my intellectual and scientific contributions, rather than routine, technical contributions. 2) They are the more conclusive and complete works over my other contributions.

For the completeness of the presentation, I will briefly describe the key findings of the other projects during my PhD in **Section 5.3**.

Chapter 2

2 Orai channel-based light-activated Ca^{2+} release

*Chapter 2 is an original work by Shuce Zhang, unless otherwise stated. S.Z. was responsible for formed the idea, designed the study, established the methodology, constructed all plasmids, and performed all live cell imaging experiments except the experiments performed in the HEK-293 Orai1,2,3 triple knockout cell line (**Figure 2.7** and **Figure 2.8**) which were performed by collaborators, Dr. Youjun Wang and colleagues, at Beijing Normal University. Eric Y. Fan-Lou assisted with calcium imaging experiments under the direct supervision of S.Z.*

2.1 Introduction

Calcium ion (Ca^{2+}) is the most ubiquitous and versatile second messenger in cells. Starting from the moment of fertilization, Ca^{2+} signalling takes place throughout the entire life course and regulates a wide range of physiological processes in both excitable and non-excitable cells[169,170]. Ca^{2+} signalling can be as transient as milliseconds to seconds for muscle contraction and neurotransmitter release, or as sustained as hours or days in the case of gene transcription[171,172]. Ca^{2+} signalling can also happen as local events of [Ca^{2+}] elevation (Ca^{2+} puffs), global elevations in [Ca^{2+}] that propagate within the cell or transmit to the neighbouring cells (Ca^{2+} waves), or even as [Ca^{2+}] oscillations[173,174]. The versatility of Ca^{2+} signalling is thus encoded in such spatiotemporal dynamics of the simple metal ion to regulate a diverse myriad of cellular processes.

Under resting conditions, the cytosolic $[Ca^{2+}]_{cyt}$ is ~ 100 nM, or about 1/10,000 of the concentration outside the cell (~ 1 mM). The endoplasmic reticulum (ER) has a $[Ca^{2+}]_{ER}$ ~ 0.5 – 1 mM and serves as the intracellular Ca^{2+} store. The elevation of cytosolic Ca^{2+} concentration typically stems from two major sources: 1) Ca^{2+} influx from extracellular Ca^{2+} pool and 2) Ca^{2+} release from the ER Ca^{2+} store[173]. The maintenance of ER-Ca^{2+} homeostasis is essential for both protein folding and chaperone function and involves many ER-resident proteins. Abnormalities of Ca^{2+} homeostasis of ER Ca^{2+} store are associated with several cardiac and neurological conditions[175].

A complex set of pathways and molecular machineries is dedicated to maintaining the homeostasis of Ca^{2+} and achieving precise regulation of the Ca^{2+} signal on different temporal and spatial scales. Ca^{2+} is typically released from the ER Ca^{2+} store via IP_3 receptors (IP_3R) in response to increased IP_3 following stimulation and activation of phospholipase C (PLC), or via ryanodine receptors (RyR) in response to the elevation of $[Ca^{2+}]_{cyt}$, known as Ca^{2+}-induced Ca^{2+} release (CICR)[176,177]. More recently, a novel ER Ca^{2+} release channel encoded by the transmembrane and coiled-coil domains 1 (TMCO1) gene was identified to mediate the store overload-induced Ca^{2+} release (SOICR)[178]. Uptake of cytosolic Ca^{2+} into the ER Ca^{2+} store is mediated by sarco/endoplasmic reticulum Ca^{2+} ATPase (SERCA).

Experimentally, ER Ca^{2+} release is generally induced by pharmacological treatments. Thapsigargin (TG) almost irreversibly inhibits SERCA and depletes the ER Ca^{2+} store[179]. Ionomycin is an ionophore that renders the plasma membrane (PM) and ER membrane permeable to Ca^{2+}, thus reversibly depleting the ER Ca^{2+} store[180]. Agonists of G-protein coupled receptors such as histamine and carbachol are also used to induce ER

Ca^{2+} release and Ca^{2+} oscillations[180]. These pharmacological treatments usually produce rather complete and sustained ER Ca^{2+} depletion, which is rarely seen under physiological conditions.

Optogenetic tools are considered to be particularly promising and useful for cellular manipulation and perturbation by virtue of their reversibility, non-invasiveness, specificity, and spatiotemporal resolution[181]. Some optogenetic tools have been developed for the purpose of inducing ER Ca^{2+} release. Examples include receptor tyrosine kinases (RTKs) such as OptoFGFR[182], and G-protein coupled receptors (GPCRs) such as OptoXR[183] and Gαq[184], which when activated, further activate phospholipase C and the IP_3-induced Ca^{2+} release. A disadvantage of these designs is that they not only activate Ca^{2+} signalling, but also involve second messengers that participate in other pathways. Another type of design is to engineer a light-gated Ca^{2+}-permeable channel. One of such effort was an engineered ER-localised channelrhodopsin-green receiver (ChRGR) as the first light-induced ER Ca^{2+} channel that mediates subtle and transient Ca^{2+} release[185]. Another design of channelrhodopsin 2 (ChR2) – ryanodine receptor 2 (RyR2) claimed to have achieved light-induced Ca^{2+} release from the ER via calcium-induced Ca^{2+} release (CICR)[186].

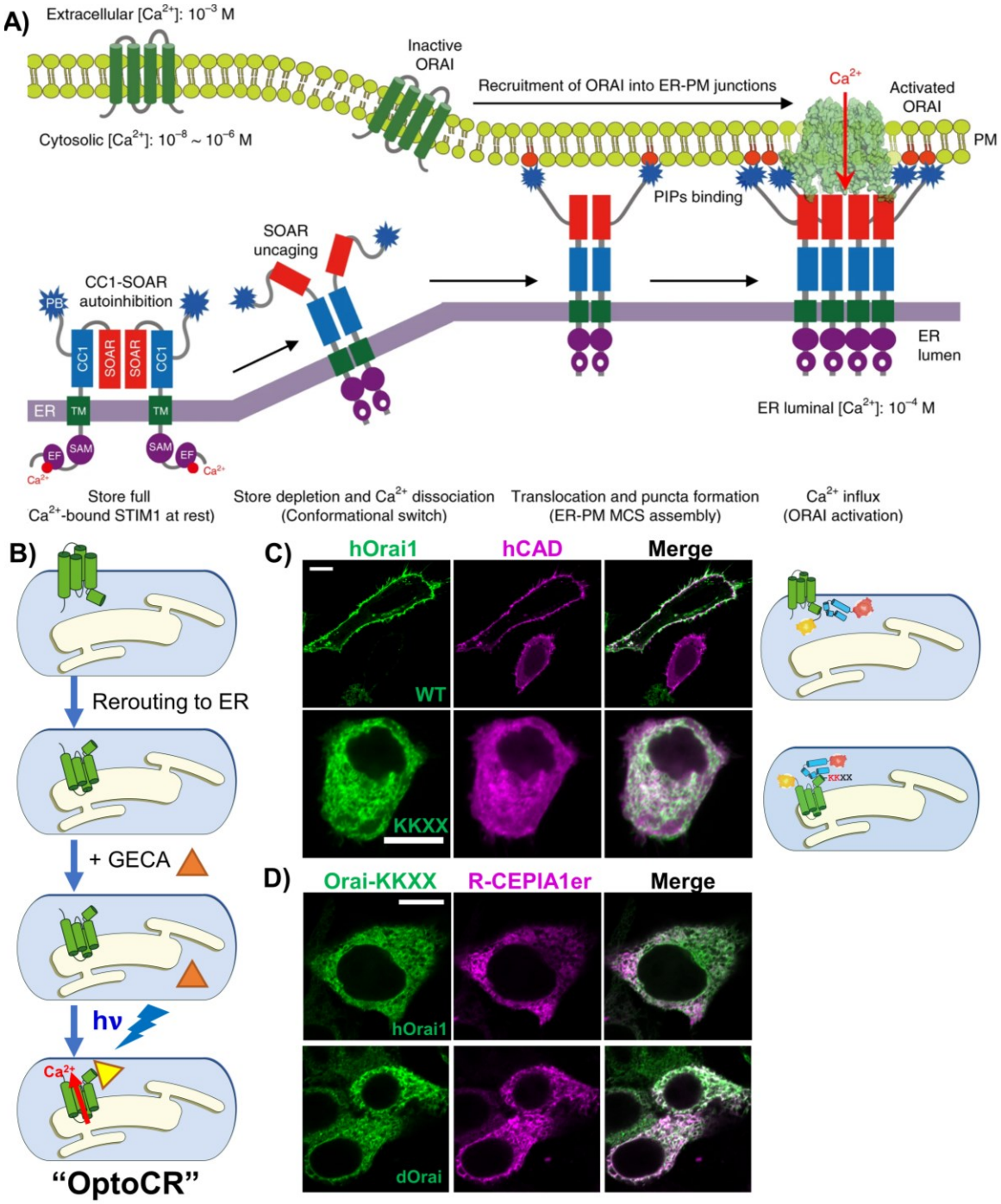

Figure 2.1 C-terminal KKXX motif reroutes Orai channel to ER.

A| Schematics of the molecular mechanism of store depletion-induced STIM1-Orai1 interaction. Adapted under the terms of the Creative Commons CC BY license from Springer Nature, Nature Communications, Guolin Ma et al *Optogenetic engineering to probe the molecular*

choreography of STIM1-mediated cell signaling[187], Copyright 2020. **B|** Schematics of strategy for Orai-based optogenetic control of ER Ca^{2+} release, "OptoCR". The light-blue area represents the cytosol. The light-yellow area represents the ER. The four green cylinders represent an Orai monomer. The orange triangle represents GECA at resting state. The yellow triangle represents GECA at the activated state. The red arrow represents the direction of Ca^{2+} flux mediated by OptoCR. **C|** Localisation of human CAD (hCAD) domain and human Orai1 (hOrai1) pseudocoloured in magenta and green, respectively, with (lower panel) and without (upper panel) KKXX motif observed by confocal microscopy. **D|** Localisation of hOrai1 (upper panel) and fruit fly Orai (dOrai, lower panel) pseudocoloured in green with ER localization marker R-CEPIA1er pseudocoloured in magenta observed by confocal microscopy. In the merged panels, white colour indicates co-localisation. Scale bar, 10 μm.

The Ca^{2+} release-activated Ca^{2+} (CRAC) channel is one of the most successful non-opsin architectures for the engineering of optogenetic Ca^{2+} channels[181]. The activity of the native CRAC channel is mediated by the PM-resident hexameric Orai channel and ER-resident Ca^{2+} sensor stromal interaction molecule 1 (STIM1) proteins (**Figure 2.1A**). Decrease of $[Ca^{2+}]_{ER}$ triggers a series of conformational changes of STIM1 proteins, including an initial oligomerization and the subsequent release of the autoinhibited STIM-Orai activating region (SOAR, also termed CRAC activation domain, CAD). The activated STIM1 oligomers are then recruited to the ER-PM junctions where the exposed SOAR/CAD domain binds to the pore-forming Orai hexamers and activate the channel (**Figure 2.1A**)[188]. Studies have demonstrated that the store-independent binding between Orai and the cytosolic fragment of STIM (STIM1ct) is sufficient to activate CRAC channels and mediate Ca^{2+} influx[189,190].

Based on this elegant mechanism, two major types of designs of genetically encoded Ca^{2+} channel actuators (GECA) have been reported to activate Orai channels (**Figure** 2.3**G**). The first type are designs based on CRY2-STIM1ct fusion proteins, represented by OptoSTIM[191], monSTIM[192], and the photo-crosslinking STIM1ct[187]

variants. The light-induced homo-oligomerization of CRY2 recapitulates the store-dependent homo-oligomerization of STIM1 luminal region, which leads to the subsequent activation of STIM1ct and association with Orai channel. The second type are designs of LOV2-SOAR/CAD fusion proteins, represented by LOVS1K[193], OptoCRAC[194], and BACCS2[195]. In the dark, the tight docking between the Jα helix and the rest of LOV2 domain sterically hinders SOAR/CAD from interacting with Orai channel, mimicking the autoinhibition of STIM1-CC1 domain. The steric hinderance is released upon light stimulation as a result of the unfolding of the Jα helix, facilitating the SOAR/CAD-Orai interaction. Direct optogenetic control of CC1-SOAR autoinhibition has been reported using circularly permutated LOV2 (cpLOV2)[196]. More recently, LOCa3, a novel design of inserting LOV2 into the intracellular loop of hOrai1-H171D-P245T variant, was reported. In this design, the Orai channel is gated by a direct light-induced conformational change, in a binding-independent and store-independent manner[197].

In this work, we describe a functional, ER-localised Orai channel, Orai-KKXX, and a new Orai-KKXX-based GECA for optogenetic ER Ca^{2+} release (**Figure 2.1B**). We demonstrate that the combination of the human homologue-based hOrai1-KKXX and OptoCRAC, and the combination of the fruit fly counterpart dOrai-KKXX and dmBACCS2, can elicit moderate optogenetic Ca^{2+} release from ER to cytosol. This Orai-based optogenetic control of ER Ca^{2+} release, named “OptoCR”, can further induce Ca^{2+}-induced Ca^{2+} release (CICR) via RyR2. We also demonstrate the possibilities of using the binary OptoCR system for spatial control of ER Ca^{2+} release. We envision that OptoCR will serve as a useful tool for studying a wide range of physiological processes associated with ER Ca^{2+} release with good spatiotemporal control.

2.2 Methods and materials

2.2.1 Constructs

Molecular cloning to create the plasmids was performed by PCR using Q5® High-Fidelity DNA Polymerase (New England Biolabs) or CloneAmp™ HiFi PCR Premix (Takara Bio USA), backbone linearization with the appropriate FastDigest restriction enzymes (Thermo Scientific), backbone 5' phosphate removal with FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific), and Gibson Assembly using an in-house prepared mix according to the originally published recipe[198], unless otherwise specified. All mammalian expression vectors used the pCDNA3.1+ backbone, with the open reading frame of N-terminal fluorescent proteins (FP) tags cloned into the *Nhe*I-*Hind*III sites and all C-terminal FP tags cloned into the *EcoR*I-*Xho*I sites. Site-directed mutagenesis was performed using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's instructions.

For Orai-based constructs, hOrai1 gene was subcloned from Orai1-YFP (Addgene 19756, a gift from Anjana Rao). dOrai was subcloned from dmBACCS2-IRES-dOrai-IRES-mCherry (Addgene 72894, a gift from Takao Nakata). The N-terminal transmembrane tag consisted of hSTIM1 signal peptide-YFP (truncated from Addgene 18857, a gift from Tobias Meyer) and transmembrane domain of mSTIM1 (209 – 310) followed by the 36 amino acid linker employed in the concatenated Orai1 hexamer construct[199]. The C-terminal transmembrane tag consisted of the mRyR2 S5 and S6 helices (4765 – 4866), consistent with the ER-anchoring domain of $ChRGR_{ER}$[185]. The C-terminal ER-retention signal KKXX (KKLQ-stop) was introduced by site-directed mutagenesis of cpVenus-

hOrai1-S5S6 and cpVenus-dOrai-S5S6 immediately after the hOrai1 or dOrai ORF. hOrai1 H134S and ANSGA mutations were introduced by site-directed mutagenesis.

For constructs encoding GECAs, OptoSTIM1 was subcloned from pCMV-OptoSTIM1 (Addgene 70159, a gift from Won Do Heo). monSTIM1 (Cry2-E281A-A9-STIM1ct) was subcloned from pCMV-monSTIM1 (Addgene 161795, a gift from W.D.H.). Truncations of monSTIM1 were created by introducing premature stop codons to cpVenus-monSTIM1 or mCerulean3-monSTIM1 by site-directed mutagenesis. OptoCRAC was subcloned from Opto-CRAC version 1 (Addgene 101245, a gift from Yubin Zhou). dBACCS2 was subcloned from dmBACCS2-IRES-dOrai-IRES-mCherry. pCMV-OptoFGFR1 was a gift from W.D.H. (Addgene 59776). mCherry-CAD was a gift from Michael Cahalan (Addgene 73566). The following localization motifs were used to tether the GECAs to the outer surface of mitochondria (TOM20 1-34, MVGRNSAIAAGVCGALFIGYCIYFDRKRRSDPNF), ER (CytERM or P450 1-29, MDPVVVLGLCLSCLLLLSLWKQSYGGGKL) and inner surface of PM (Lyn PM-targeting motif, GCIKSKGKDSA).

For Ca^{2+} imaging, pGP-CMV-NES-jRGECO1a was a gift from Douglas Kim (Addgene 61563). pCMV-R-CEPIA1er was a gift from Masamitsu Iino (Addgene 58216). pGP-CMV-jGCaMP7s was a gift from Douglas Kim and the GENIE Project (Addgene 104463).

Constructs containing hOrai3-KKXX and LOCa3 were created by Y. Wang and colleagues at Beijing Normal University.

2.2.2 **RNA interference**

Small interfering RNAs (siRNAs) targeting hOrai1 (rGrUrC rCrUrC rUrArA rGrArG rArArU rArArG rCrArU rUrUC C and rGrGrA rArArU rGrCrU rUrArU rUrCrU rCrUrU rArGrA rGrGrA rCrArG) and siRNA targeting hOrai3 (rCrArC rUrGrA rArUrU rUrGrG rArUrG rCrArC rCrUrU rGrUT T and rArArA rCrArA rGrGrU rGrCrA rUrCrC rArArA rUrUrC rArGrU rGrUrC) were ordered and synthesized by Integrated DNA Technologies (IDT). These siRNA targeted the 3' UTR of the endogenous transcripts to knockdown the expression of endogenous Orai1 and Orai3 proteins, but not the exogenously expressed hOrai1-KKXX. TEX 615 Transfection Control DsiRNA (IDT) was used as negative control as well as for transfection efficiency evaluation. siRNA oligomers were resuspended, aliquoted, diluted and stored according to manufacturer's recommendations.

2.2.3 **Cell culture and transfection**

HeLa cells and HEK-293 FT cells (Thermo Fisher) were cultured in complete DMEM (HyClone high-glucose DMEM with sodium pyruvate and L-glutamine, supplemented with 10% FBS and 1% antibiotic-antimycotic) in a moist 5% CO_2 atmosphere at 37 °C. For transient transfection, cells were seeded at ~60% confluency onto 22 mm imaging coverslips placed in 6-well plates 24 h before transfection. To co-transfect the Ca^{2+} sensor, Orai channel and optogenetic switch, 0.5 μg plasmid encoding the Ca^{2+} sensor (jRGECO1a, jGCaMP7s, or R-CEPIA1er), 1 μg plasmid encoding Orai channel (hOrai1, hOrai1-KKXX or dOrai-KKXX) and 1.5 μg plasmid encoding the GECA (various versions of OptoCRAC and dmBACCS2) were mixed with 6 μL TurboFect transfection reagent (Thermo Fisher) in 400 μL serum-free DMEM/F12 and incubate for

10 min before adding to the cell culture. For experiments with RNA interference, 100 pmol of siRNA and an extra 5 μL TurboFect transfection reagent was included. Medium was replaced to fresh complete DMEM/F12 4-6 h after transfection. Cells were imaged 72 h after transfection.

Flp-In T-REx HEK-293 RyR2-T4158P stable cell line was a kind gift from S.R.W. Chen of the University of Calgary[200]. Flp-In T-REx HEK-293 RyR2-T4158P stable cell was cultured in complete DMEM supplemented with 50 μg/mL hygromycin (Invitrogen) in a moist 5% CO_2 atmosphere at 37 °C. 1 μg/mL doxycycline was added when the medium was replaced with complete DMEM/F12 to induce the expression of RyR2-T4158P.

HEK-293 Orai1/2/3 triple knockout cell line was generated by Y. Wang and colleagues[201]. Experiments involving this cell line were performed by Y. Wang and colleagues.

2.2.4 Ca^{2+} imaging with optogenetic stimulation

Cells growing on the coverslips were bathed in modified Locke buffer (154 mM NaCl, 5.6 mM KCl, 1 mM $MgCl_2$, 5.6 mM D-glucose) supplemented with 20 mM HEPES and mounted in Attofluor™ Cell Chamber (Thermo Fisher Scientific, Cat. #A7816). Other bath solutions were supplemented with either 500 μM EGTA or 1 mM Ca^{2+}. Rapid change of bath solutions during the imaging were performed in a remove-and-add manner using a homemade solution remover[202]. Cells were imaged on a Nikon Eclipse Ti-E epifluorescence microscope equipped with a 75 W Nikon xenon lamp focused on the back aperture of a 60× objective lens (NA 1.49 oil, Nikon). Images of all channels were acquired

every 5 s by a Photometrics QuantEM 512SC EM-CCD camera at a gain value of 200. To avoid the photoactivation artefacts, jRGECO1a/R-CEPIA1er signal was acquired first in each cycle with a TRITC filter cube, followed by the blue light stimulation with a FITC filter cube (1 mW mm^{-2}) for 100 msec. For the Ca^{2+} imaging using jGCaMP7s, the excitation light (0.5 mW mm^{-2} for 50 msec in every 5 sec cycle) also serves as the stimulation for the GECA. For experiments with RNA interference, the transfection efficiency of siRNA was evaluated by the TEX 615 fluorescence of the negative control under fluorescence microscope. > 90% of cells transfected with negative control siRNA exhibited punctate TEX 615 fluorescence. NIS-Elements AR package software was used for automatic instrument control, data recording and measurement. Data were further analyzed using a custom R script (available at https://github.com/shucez/Microscopy-scripts) where $\Delta F/F_0$ was calculated to normalize the fluorescence change. The analysed data were plotted in GraphPad Prism software. Unless otherwise stated, all plots represent the mean ± standard error mean (s.e.m.) of cells pooled from three independent biological replications with different batches of cell culture transfected separately on different days.

2.2.5 **Confocal microscopy**

At 24 hours after transfection, the coverslips with attached cells were transferred into an Attofluor™ Cell Chamber (Thermo Fisher Scientific, Cat. #A7816) with Hank's balanced salt solution (HBSS) supplemented with 20 mM HEPES. Cells were imaged with a LSM700 laser scanning confocal microscope on a Zeiss AXIO Observer Z1 equipped with a Plan-Apochromat 63× objective lens (NA 1.40 oil, Zeiss). Laser power of 1.80% was used in all channels to minimize photobleaching. YFP and cpVenus were excited by a 488 nm diode laser and the emitted fluorescence below 578 nm was recorded by a

photomultiplier tube (PMT) detector at a gain of 650. mCherry and R-CEPIA1er were excited by a 555 nm diode laser and the emitted fluorescence above 578 nm was recorded by the PMT detector at a gain of 800. Zen black software (Zeiss) was used for automatic instrument control and data recording. For experiments with more than 1 channel, samples with single colour fluorescence were used to obtain the bleed-through coefficients. Bleed-through correction was performed on a pixel-to-pixel basis using a custom Python script to minimize artefact in co-localization experiments. Processed images are pseudo-coloured using Fiji/ImageJ software. Unless otherwise stated, all confocal microscopy pictures are representative of the most typical morphology out of at least three independent biological replications with different batches of cell culture transfected separately on different days.

2.2.6 **FRET imaging**

Cells growing on the coverslips were bathed in modified Locke buffer (154 mM NaCl, 5.6 mM KCl, 1 mM $MgCl_2$, 5.6 mM D-glucose) supplemented with 20 mM and mounted in Attofluor™ Cell Chamber (Thermo Fisher Scientific, Cat. #A7816). Cells were imaged on a Nikon Eclipse Ti-E epifluorescence microscope equipped with a 75 W Nikon xenon lamp focused on the back aperture of a 60× objective lens (NA 1.49 oil, Nikon). A filter wheel was installed between the xenon lamp and the microscope, harbouring a 436/20 filter for exciting mCerulean3/CFP and a 500/20 filter for exciting cpVenus/YFP. Donor channel was acquired with the excitation light passing through the 436/20 filter on the filter wheel and emission light reflected by a cube harbouring 455 LP dichroic mirror and 480/40 emission filter. Acceptor channel was acquired with the excitation light passing through the 500/20 filter on the filter wheel and emission light reflected by a cube

harbouring 515 LP dichroic mirror and 545/30 emission filter. FRET channel was acquired with the excitation light passing through the 436/20 filter on the filter wheel and emission light reflected by a cube harbouring 515 LP dichroic mirror and 545/30 emission filter. All channels were acquired every 10 s by a Photometrics QuantEM 512SC EM-CCD camera at a gain value of 200. NIS-Elements AR package software was used for automatic instrument control, data recording and measurement. HeLa cells transfected with mCerulean3 or cpVenus alone were used to measure the bleed-through coefficients, and HeLa cells transfected with mCerulean-cpVenus fusion protein was used to measure the system-dependent coefficient G by photobleaching[203]. A custom R script was used to calculate the E_{app} for 3-channel-corrected FRET[203]. The analysed data were plotted in GraphPad Prism software. Unless otherwise stated, all plots represent the mean ± standard error mean (s.e.m.) of cells pooled from three independent biological replications with different batches of cell culture transfected separately on different days.

2.3 Results

2.3.1 A C-terminal dilysine motif reroutes functional Orai channels to the ER membrane

Orai1 channels normally localise on the PM and mediate Ca^{2+} influx into the cytosol. To harness the CRAC-based GECAs to conduct ER Ca^{2+} release, it is crucial to reroute the Orai channels to the ER membrane while preserving the proper topology and function. Three strategies were attempted to introduce an ER-retention motif to hOrai1: 1)

an N-terminal signal peptide and type I transmembrane helix from STIM1 (SP-S1TM), as reported for the FIRE system[204]; 2) a C-terminal transmembrane domain of S5 and S6 helices from RyR2, as reported for the $ChRGR_{ER}$[185] design; and 3) a C-terminal KKXX motif. The dilysine motif (KKXX and KXKXX) on the cytoplasmic C-terminus of transmembrane proteins is an ER-retention motif. The KKXX motif is recognised and coated by the Golgi-to-ER retrieving COPI complex[205,206]. When expressed in HeLa cells, SP-S1TM-hOrai1 exhibited an ER whorl-like structure (**Figure** 2.2**C** upper panel), indicating that the N-terminal transmembrane helix tagging interferes with proper folding and assembly of the Orai channel[207,208]. hOrai1 tagged with S5S6 helices exhibited a partial punctate ER structure (**Figure** 2.2**C** middle panel) while hOrai1-KKXX exhibited the expected pattern of ER localisation, without apparent whorl-like or punctate structure (**Figure** 2.2**C** lower panel).

The Orai channel is gated by the binding between the Orai1 C-terminal binding domain (CBD) and the STIM1 SOAR/CAD domain. SOAR/CAD can be overexpressed as a soluble protein in the absence of autoinhibitory CC1 domain or Orai channels (**Figure** 2.2**A**). The soluble SOAR/CAD can be readily recruited by the overexpressed PM-localised Orai channels[209]. Since the binding between Orai and SOAR/CAD domain is a required step for Orai channel activation and gating the Ca^{2+} permeability, we over expressed mCherry-CAD with wildtype (WT) hOrai1 or the tagged variants of hOrai1. Indeed, the WT PM-localised hOrai1 effectively recruited CAD to the PM (**Figure** 2.1**C** upper panel). hOrai1-KKXX co-localised with the ER morphology marker R-CEPIA1er (**Figure** 2.1**D** upper panel) and also recapitulated the recruitment CAD (**Figure** 2.1**C** lower panel). Similarly, the fruit fly Orai homologue (dOrai) also became ER-localised when its C-terminus was tagged with KKXX (**Figure** 2.1**D** lower panel). SP-S1TM-hOrai1 was able

to recruit partially recruit CAD to the regions where it forms the whorl-like structure (**Figure** 2.2**D** upper panel), while the hOrai-S5S6 almost completely abolished the CAD recruitment (**Figure** 2.2**D** middle panel). These results prompted us to consider the C-terminal KKXX motif as the most promising candidate for rerouting a functional Orai channel to the ER membrane, whereas the channel with N- or C-terminal tagging either did not have expected localisation or lost the ability to recruit SOAR/CAD, indicating issues with folding or channel function.

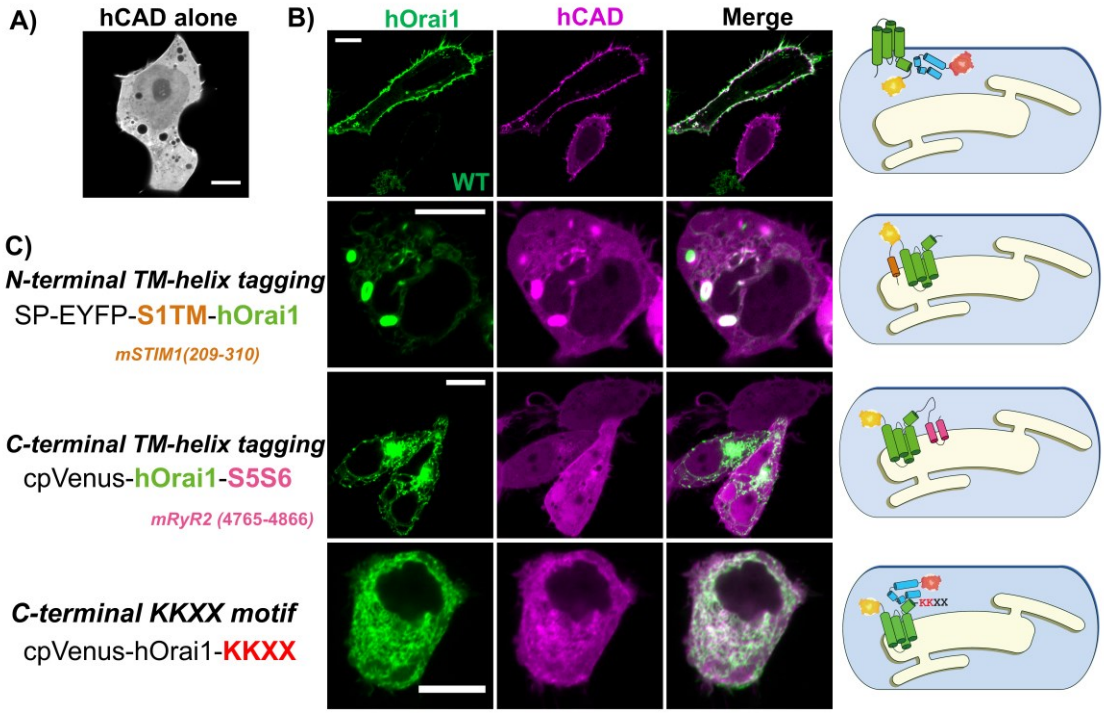

Figure 2.2 Screening ER-localisation tags for functional Orai channel rerouting.

A| Overexpressed human CAD domain alone shows cytosolic distribution in HeLa cells, observed by confocal microscopy. **B|** Localisation of overexpressed PM-localised WT hOrai1 pseudocoloured in green and hCAD pseudocoloured in magenta. The right panel is a schematic of hCAD domain by the hOrai1 channel, where the hOrai1 monomer (green cylinders) is fused with a cpVenus (represented by a FP coloured in yellow) at the C-terminus and the hCAD domain (blue cylinders) is tagged with mCherry (FP coloured in red) at the C-terminus. **C|** Localisation of hOrai1 with N-terminal transmembrane helix (upper panel), C-terminal

transmembrane helices (middle panel), and C-terminal KKXX motif (lower panel, identical to **Figure 2.1C** lower panel) pseudocoloured in green and hCAD pseudocoloured in magenta. In the merged panels, white colour indicates co-localisation. Scale bar, 10 μm. The red helix or helices represent(s) the transmembrane domain fused to the N- or C-terminus of hOrai1.

2.3.2 OptoCRAC and hOrai1-KKXX mediate light-induced Ca^{2+} release (OptoCR).

A few successful GECAs have been developed to activate Orai channels. These GECAs mimic the STIM1 activation by recapitulating either the oligomerization process (e.g., OptoSTIM1 and its improved version monSTIM1, **Figure 2.3G**) or the release of autoinhibition (OptoCRAC for hOrai1 and dmBACCS2 for dOrai, **Figure 2.3G**). In HeLa cells, we confirmed that these GECAs could activate the overexpressed PM-localised hOrai1 (for monSTIM1 and OptoCRAC) or dOrai (for dmBACCS2) and mediate similar level of Ca^{2+} influx (**Figure 2.3A**). Compared to LOV2-based GECAs, monSTIM1 exhibited slower off-kinetics and poorer reversibility.

Based on the CAD recruitment by hOrai1-KKXX, we hypothesized that the KKXX-tagged channels were largely functional and could be activated by the GECAs in a similar fashion as the PM-localised Orai channels. To test this hypothesis, we carried out a Ca^{2+} imaging assay to identify a potent GECA that could activate the Orai-KKXX in response to light stimulation and induce Ca^{2+} release from the ER. For this assay, the cells were transfected with jRGECO1a[210], a red fluorescent Ca^{2+} indicator that allows continuous monitoring of cytosolic $[Ca^{2+}]_{cyt}$ increase without activating LOV2 or CRY2 domains. During the Ca^{2+} imaging, cells were kept in Ca^{2+}-free buffer supplemented with 500 μM EGTA to chelate any possible trace amount of extracellular Ca^{2+} to eliminate any Ca^{2+} influx. Two cycles of blue light stimulation consisting of a 5 min ON phase and 5 min OFF

phase were administered after acquiring the baseline. To validate the assay protocol, OptoFGFR was included as a positive control that induced robust and irreversible Ca^{2+} release (**Figure 2.3B**). For hOrai1-KKXX, OptoCRAC elicited ~20% $\Delta F/F_0$ due to an increase in $[Ca^{2+}]_{cyt}$, a substantially larger response than monSTIM1 (**Figure 2.3CEF**). For dOrai, a modest but notable tendency of Ca^{2+} release could be observed with dmBACCS2 overexpression (**Figure 2.3D**). Even though the response was not statistically significant (**Figure 2.3EF**), the dOrai-KKXX and dmBACCS2 combination was kept in the following experiments. Overall, the combination of hOrai1-KKXX and OptoCRAC yielded the largest Orai-based light-induced Ca^{2+} release, and was hereby named “OptoCR”.

Red fluorescent FP chromophores are known to exhibit enhanced fluorescence when exposed to blue light, a phenomenon known as a “photoactivation artefact”[211,212]. To minimize the effect of photoactivation artefact on Ca^{2+} imaging experiments, we optimised the imaging protocol by acquiring the red channel prior to the blue light illumination at each time point. We reasoned that the off-kinetic of $[Ca^{2+}]_{cyt}$ is much slower than that of the photoactivation. Applying the blue light stimulation after red channel acquisition allows the chromophore to relax in the dark for ~ 4.5 sec. The new protocol resulted in a more stable baseline in the negative control, indicating effective suppression of the photoactivation artefact, while maintaining similar Ca^{2+} response (**Figure 2.3H**). To further differentiate the true responses from photoactivation artefacts, the hOrai1-KKXX and OptoCRAC-mediated increase was repeated using a green fluorescent Ca^{2+} sensor jGCaMP7s (**Figure 2.3I**), suggesting that a true OptoCR-dependent response was achieved by this system.

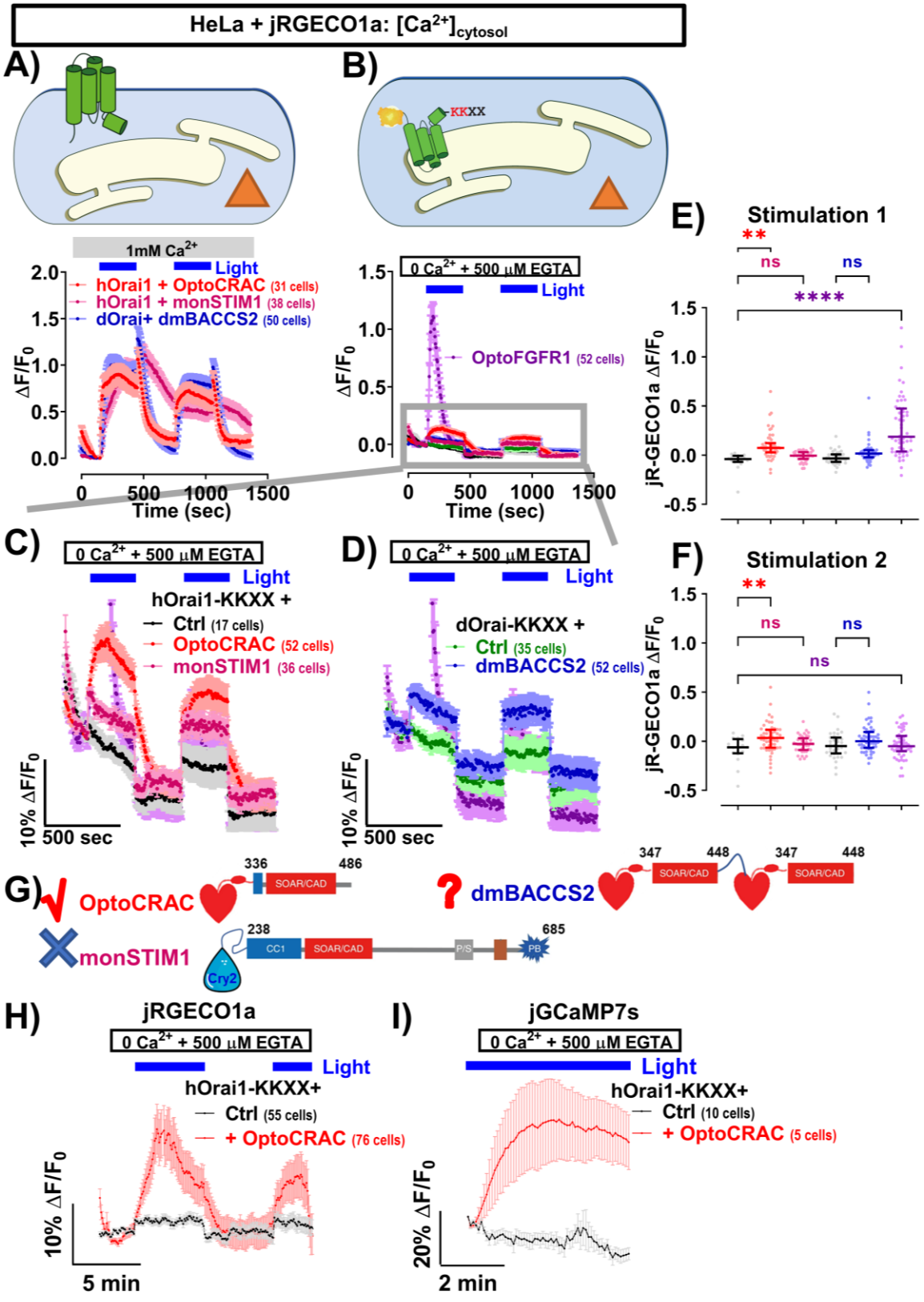
HeLa + jRGECO1a: [Ca2+]cytosol
A)
B)
KKXX
1mM Ca2+
Light
hOrai1 + OptoCRAC (31 cells)
hOrai1 + monSTIM1 (38 cells)
dOrai+ dmBACCS2 (50 cells)
ΔF/F0
Time (sec)
0 Ca2+ + 500 μM EGTA
Light
OptoFGFR1 (52 cells)
C)
hOrai1-KKXX +
Ctrl (17 cells)
OptoCRAC (52 cells)
monSTIM1 (36 cells)
10% ΔF/F0
500 sec
D)
dOrai-KKXX +
Ctrl (35 cells)
dmBACCS2 (52 cells)
E)
Stimulation 1
jR-GECO1a ΔF/F0
**
ns
****
F)
Stimulation 2
G)
OptoCRAC
monSTIM1
dmBACCS2
SOAR/CAD
CC1
P/S
PB
Cry2
336
486
238
685
347
448
H)
jRGECO1a
hOrai1-KKXX+
Ctrl (55 cells)
+ OptoCRAC (76 cells)
5 min
I)
jGCaMP7s
Ctrl (10 cells)
+ OptoCRAC (5 cells)
20% ΔF/F0
2 min

Figure 2.3 Screening for Orai-KKXX activating optogenetic switches.

A| Schematic (upper panel) and cytosolic Ca^{2+} response (lower panel) of GECAs that activate PM-localised hOrai1 or dOrai. Blue bars in the plot represents the period when the blue light stimulation is on. **B|** Schematic (upper panel) and cytosolic Ca^{2+} response (lower panel) of GECAs that activate ER-localised hOrai1-KKXX or dOrai-KKXX. OptoFGFR1 is included as the positive control. **C|** Cytosolic Ca^{2+} response to photostimulation of OptoCRAC and monSTIM in the presence of hOrai1-KKXX. The data is identical as B) with rescaled axis for clarity. **D|** Cytosolic Ca^{2+} response to photostimulation of dmBACCS2 in the presence of dOrai-KKXX. The data is identical as B) with rescaled axis for clarity. B,C,D) The fluctuation of control groups (black and green) in response to blue light stimulation is a result of photoactivation artefact of the indicator jRGECO1a. **E,F|** Comparison of the cytosolic Ca^{2+} response across the GECAs during **E|**the first photostimulation or **F|** the second photostimulation. Each dot represents the average $\Delta F/F_0$ of a cell during the photostimulation subtracted by the average of baseline. The bars represent median and the interquartile range. Comparison was performed using one-way ANOVA followed by Šídák's multiple comparisons test. ns, $p > 0.05$ (not significant); *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$. **G|**Schematic of the primary topology of the optogenetic switches. Red heart, LOV2 domain. Other domains are indicated. **H|** Cytosolic Ca^{2+} response to photostimulation of OptoCRAC and hOrai1-KKXX, after applying correction for the photoactivation artefact. **I|** Cytosolic Ca^{2+} response to photostimulation of OptoCRAC and hOrai1-KKXX free of photoactivation artefact, measured by jGCaMP7s.

2.3.3 Endogenous Orai channels compete with hOrai1-KKXX for OptoCRAC.

Although the PM-localised hOrai1 and the ER-localised hOrai1-KKXX can both be activated by OptoCRAC, OptoCR elicited much smaller increase of $[Ca^{2+}]_{cyt}$ with hOrai1-KKXX than Ca^{2+} influx with WT hOrai1. We reasoned that the endogenous Orai channels localised on the PM might compete with the exogenously expressed hOrai1-KKXX for activated OptoCRAC. In human cells, hOrai3 is the Orai paralogue with the highest affinity to bind SOAR and forms heteromeric channels with the other paralogues[213,214]. To test the hypothesis, we knocked down hOrai1 and hOrai3 using siRNA targeting the 3' UTR of their endogenous transcripts. Since the exogenous hOrai1-KKXX did not share the same sequence as the endogenous transcripts, the siRNAs were only able to knock down the expression of PM-localised Orai channels, but not the ER-localised ones. Compared to

the negative controls, the Orai1 and Orai3 knock down cells exhibited larger OptoCR (**Figure** 2.4**A**) responses mediated by hOrai1-KKXX and OptoCRAC and with faster kinetics (**Figure** 2.4**DE, Table** 2.1). The knock down also enhanced the coupling between hOrai1-KKXX and OptoCRAC, as indicated by the increased apparent FRET efficiency (E_{app}) of the two (**Figure** 2.4**B**). These results suggested that the knock down of endogenous Orai1 and Orai3 reduced the competition, allowing more OptoCRAC to interact with the ER-localised hOrai1-KKXX. On the other hand, knock down of endogenous Orai1 and Orai3 did not affect the OptoCR mediated by dOrai-KKXX and dmBACCS2 of the fruit fly-origin, consistent with the previously reported orthogonality of human and fruit fly machineries of CRAC[195].

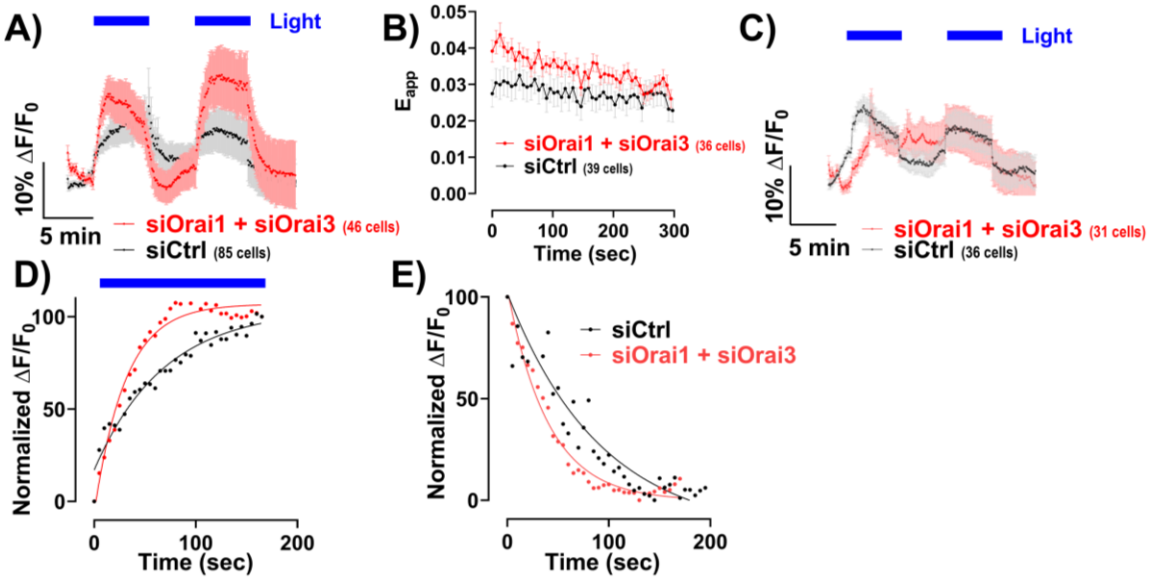

Figure 2.4 Endogenous Orai channel competes for OptoCRAC.

A| Cytosolic Ca^{2+} response of OptoCR mediated by hOrai1-KKXX and OptoCRAC with (red) or without (black) human Orai1 and Orai3 knockdown. **B|** Intermolecular FRET efficiency between mCerulean3-hOrai1-KKXX and cpVenus-OptoCRAC with or without Orai1 and Orai3 knockdown. **C|** Cytosolic Ca^{2+} response of OptoCR mediated by dOrai1 and dmBACCS2 with or without Orai1 and Orai3 knockdown. A,C) The fluctuation of control group (black) in response to blue light

stimulation is a result of photoactivation artefact of the indicator jRGECO1a. **D|** ON-kinetics of OptoCR mediated by hOrai1-KKXX and OptoCRAC with or without Orai1 and Orai3 knockdown, fitted with one-phase exponential association. **E|** OFF-kinetics of OptoCR mediated by hOrai1-KKXX and OptoCRAC with or without Orai1 and Orai3 knockdown, fitted with one-component exponential decay.

Table 2.1 Kinetics parameters for OptoCR mediated by hOrai1-KKXX and OptoCRAC with or without Orai1 and Orai3 knockdown.

Condition	τ_{on}/sec	τ_{off}/sec
siCtrl	68.3	92.07
siO1O3	30.62	40.2

2.3.4 CRY2-based GECA pre-activates hOrai1-KKXX and partially depletes ER Ca^{2+} store.

One intriguing observation of the OptoCR system is the preference of monSTIM1 for PM-localised Orai channels over ER-localised hOrai1-KKXX. monSTIM1 can activate endogenously and exogenously expressed PM-localised hOrai1 channels, but not the ER-localised channels. monSTIM1 is an engineered GECA created by fusing a light-sensitive variant of CRY2 (CRY2-A9-E281A) to the N-terminus of STIM1ct (hSTIM1 238-685)[192]. In addition to the SOAR/CAD region (344–442) that all STIM1-derived GECAs have in common, the STIM1ct fragment also includes the CC1 inhibitory helix (233-343) preceding SOAR/CAD and the C-terminal cytoplasmic tail after SOAR (**Figure** 2.5**A** upper right). Two regions within the cytoplasmic tail are functionally important: 1) the S/TxIP motif (TRIP, 642–645), responsible for constantly tracking the plus-end of microtubule tips via

interactions with EB1; and 2) the polybasic domain (PB, 671–685) that interacts with the negatively charged phosphoinositide (PIPs) on the inner leaflet of PM. Although Orai1 can be sufficiently gated by the interaction between Orai1 and SOAR/CAD, it has been demonstrated that the PB-PIP interaction can greatly facilitate the physical engagement of STIM1-Orai1 binding[215]. We reasoned that the function of TRIP and PB may also contribute to the preferred association between STIM1ct and PM-localised Orai1 over ER-localised Orai1-KKXX.

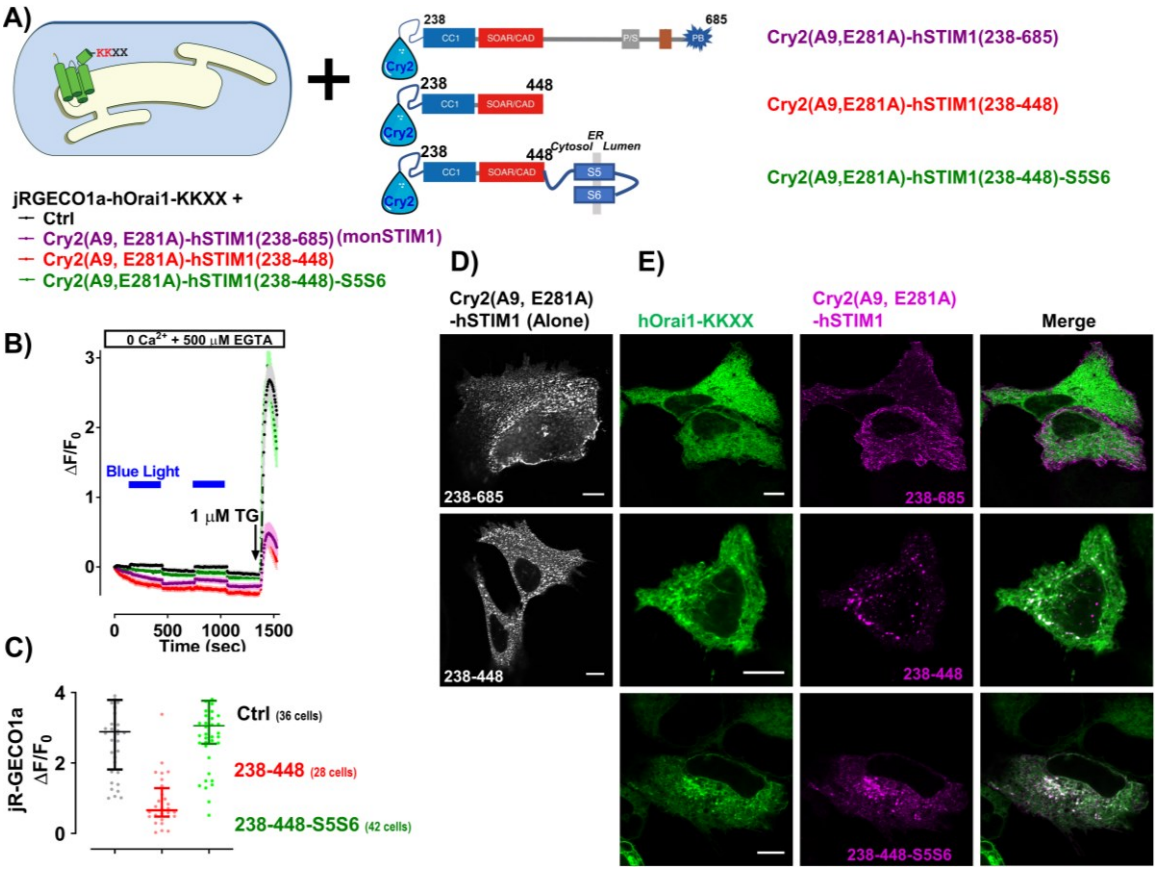

Figure 2.5 Variants of Cry2-hSTIM1 optogenetic switches for hOrai1-KKXX.

A| Schematic of hOrai1-KKXX with the variants of Cry2-based optogenetic switches. **B|** Cytosolic Ca^{2+} response to photostimulation of the Cry2-based optogenetic switch variants and to thapsigargin (TG) treatment. The fluctuation of control group (black) in response to blue light

stimulation is a result of photoactivation artefact of the indicator jRGECO1a. **C|** Quantification of cytosolic Ca^{2+} increase following TG treatment. The fluorescence increase followed by TG treatment indicates the amount of releasable ER Ca^{2+}. **D|** Localisation of Cry2(A9, E281A) fused to hSTIM1 238-685 or 238-448 truncation in the absence of overexpressed Orai channel observed by confocal microscopy. **E|** Localisation of hOrai1-KKXX pseudocoloured in green and the Cry2-hSTIM1-based GECA variants pseudocoloured in magenta, observed by confocal microscopy. Scale bar, 10 μm.

To test this hypothesis, we generated a truncated version of monSTIM1 with the cytosolic tail (449-685) removed (CRY2-A9-E281A)-hSTIM1(238-448). When overexpressed alone in HeLa cells, cpVenus-monSTIM1, either with or without the cytosolic tail, effectively formed puncta when activated by 488 nm laser scanning (**Figure** 2.5**D**). When co-expressed with hOrai1-KKXX in HeLa cells, full length cpVenus-monSTIM1 did not colocalise with mCherry-hOrai1-KKXX under 488 nm photoactivation conditions (**Figure** 2.5**E**, upper panel). However, the deletion of the cytosolic tail rescued the colocalization (**Figure** 2.5**E**, middle panel). These results suggested that the C-terminal cytosolic tail of STIM1ct indeed inhibited the association of GECA with ER-localised Orai1-KKXX.

We further investigated whether the rescued binding between monSTIM1 and hOrai1-KKXX could enhance the function of OptoCR. Ca^{2+} imaging using the above mentioned OptoCR protocol suggested that no substantial OptoCR activity was mediated by either version of monSTIM1 (**Figure** 2.5**B**). Surprisingly, when treated with thapsigargin (TG), overexpression of either version of monSTIM1 resulted in a substantially reduced Ca^{2+} release, consistent with a decreased amount of releasable Ca^{2+} in the ER Ca^{2+} store. When the truncated monSTIM1 was tethered by a C-terminal S5-S6 domain to the ER membrane (**Figure** 2.5**A** lower right), the Ca^{2+} release level by TG treatment was rescued.

The S5-S6 domain that tethered SOAR to the ER likely introduced rigid restrictions to SOAR and rendered it non-functional (**Figure** 2.5**BC**). Overall, it seemed likely that the overexpression of monSTIM1 preactivated the ER-localised hOrai1-KKXX and reduced the releasable Ca^{2+} in the ER, independent of blue light stimulation.

2.3.5 **OptoCR had little effect on $[Ca^{2+}]_{ER}$**

Activation of OptoCR elicited a modest increase of $[Ca^{2+}]_{cyt}$. We further investigated whether activation of OptoCR was causing a decrease in the $[Ca^{2+}]_{ER}$. Ca^{2+} imaging using a red fluorescent ER Ca^{2+} sensor R-CEPIA1er[216] was carried out using the OptoCR protocol. Unfortunately, R-CEPIA1er suffered substantial photobleaching during blue light stimulation, even for the negative control where ~ 10% decrease of R-CEPIA1er fluorescence was caused simply by illuminating with blue light (hOrai1-KKXX alone, **Figure** 2.6**A** black trace). The change of $[Ca^{2+}]_{ER}$ was indistinguishable for OptoCRs mediated by hOrai1-KKXX with OptoCRAC or dOrai-KKXX with dmBACCS2, and also indistinguishable from the negative control (**Figure** 2.6**A**). The positive control OptoFGFR1, which induces sustained activation of RTK and elicited robust and the largest increase of $[Ca^{2+}]_{cyt}$ from ER Ca^{2+} release, also did not result in larger extent of ER depletion compared to the Orai-KKXX based OptoCR or negative control (**Figure** 2.6**A**). Ca^{2+} in the ER is well buffered with most Ca^{2+} bound to the buffering proteins[217]. Our results might suggest that Ca^{2+} signalling downstream of RTK activation does not substantially reduce the free $[Ca^{2+}]_{ER}$.

To understand whether sustained activation of OptoCR system could alter the $[Ca^{2+}]_{ER}$ level, we generated hOrai1-KKXX constructs harbouring H134S mutation or L261A-V262N-H264G-K265A quadruple mutations (or ANSGA). H134S and ANSGA are mutations positioned on the non-pore-lining helices of hOrai1 channels[218,219]. These gain-of-function (GOF) mutations each render hOrai1 constitutively active in the absence of SOAR/CAD binding (**Figure** 2.6**B**), while maintaining highly Ca^{2+}-selective. We reasoned that the H134S and ANSGA mutations would make the ER-localised hOrai1-KKXX constitutively permeable to Ca^{2+}. With a constant Ca^{2+} leakage, we expected that the free $[Ca^{2+}]_{ER}$ level would be lower or completely depleted. Surprisingly, ionomycin treatment mobilised the same amount of ER Ca^{2+} in cells expressing hOrai1-H134S-KKXX or hOrai1-ANSGA-KKXX as the control cells (**Figure** 2.6**B**), suggesting the steady-state of ER Ca^{2+} concentration is largely unaltered even with constitutively active hOrai1-KKXX.

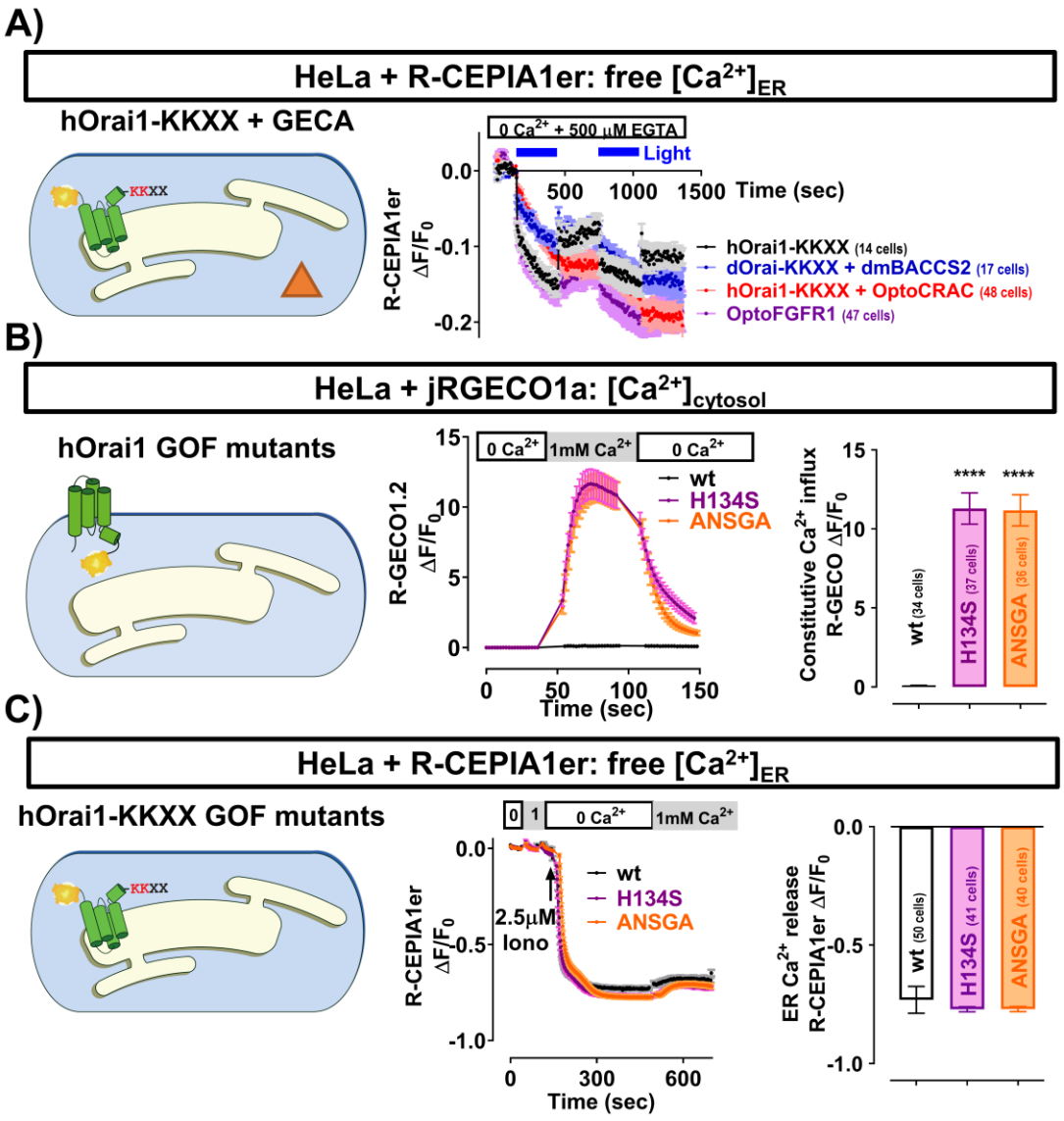

Figure 2.6 Effect of Orai-based Ca^{2+} release on ER free Ca^{2+} concentration.

A| Effect of photo-activated Ca^{2+} release system on free ER Ca^{2+} concentration, measured by R-CEPIA1er. The fluctuation of hOrai1-KKXX only group (black) in response to blue light stimulation is a result of photoactivation artefact of the indicator R-CEPIA1er. **B|** Cytosolic Ca^{2+} response to Ca^{2+} addition in the bath solution in cells overexpressing PM-localised hOrai1-H134S and -ANSGA variants without ER depletion or SOAR/CAD/STIM1 overexpression, measured by jRGECO1a. The increase of $[Ca^{2+}]_{cyt}$ indicates constitutive Ca^{2+} influx mediated by the bespoken variants. **C|** Effect of overexpression of ER-localised hOrai1-H134S-KKXX and hOrai1-ANSGA-KKXX on ER free Ca^{2+} concentration. The fluorescence decrease followed by ionomycin treatment (iono) indicates the amount of ER free Ca^{2+} mobilised by ionomycin.

2.3.6 **Orai3-KKXX and LOCa3-KKXX did not mediate functional OptoCR.**

We envisioned that similar OptoCR strategies could be applied to other Orai homologues in addition to hOrai1-KKXX and dOrai. In human cells, Orai3 associates with SOAR/CAD with higher affinity than Orai1[214]. In HEK-293 cells with Orai1, 2, and 3 triple knocked out (KO), OptoCRAC could activate overexpressed hOrai3 localised to the PM (**Figure** 2.7**A**), suggesting that GECAs can activate hOrai3-formed channels. Accordingly, we suspected that when Orai3 is rerouted to ER, it might mediate a larger OptoCR with OptoCRAC. However, OptoCRAC could not induce ER Ca^{2+} release when co-expressed with hOrai3-KKXX (**Figure** 2.7**B**). This indicated hOrai3-KKXX might not be fully functional when rerouted to the ER. Endogenously or exogenously expressed Orai3 has the unique pharmacological characteristic of being activated by 50 μM 2-aminoethoxydiphenyl borate (2-APB) independent of store depletion[220]. If the ER-localised hOrai3-KKXX was functional, 2-APB should activate hOrai3-KKXX and induce a Ca^{2+} release. Indeed, the response of hOrai3-KKXX to 50 μM 2-APB was barely detectable by Ca^{2+} imaging with jRGECO1a (**Figure** 2.7**C**). The fact that the ER-localised hOrai3-KKXX was neither activated optogenetically by OptoCRAC nor pharmacologically 2-APB indicated that the hOrai3-KKXX was not fully functional on the ER membrane, and that hOrai3 may require the PM environment for its proper functioning.

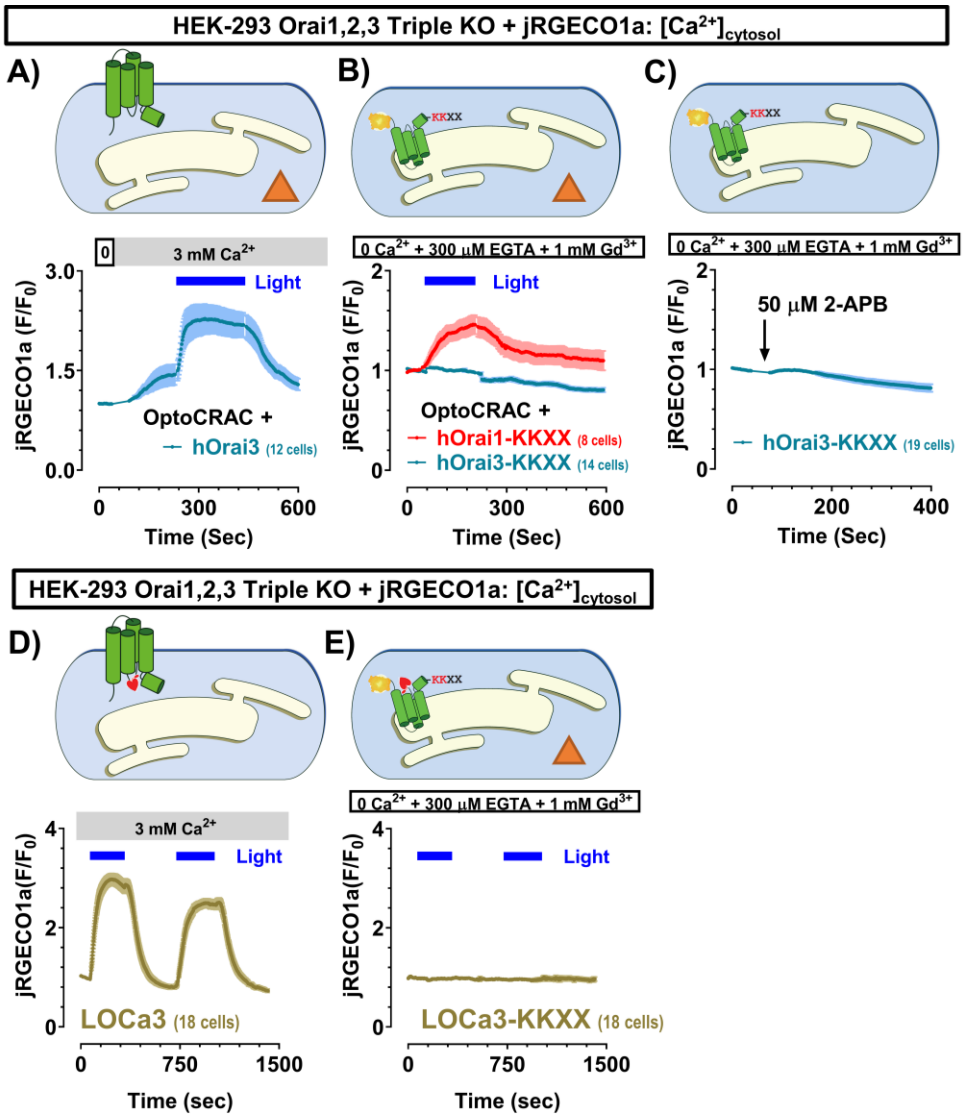

Figure 2.7 KKXX-tagged Orai3 and LOCa3 do not function as Ca^{2+} release channels.

A| Schematic (upper) and cytosolic Ca2+ response (lower) to photostimulation in Orai1,2,3 triple knockout HEK-293 cells overexpressing PM-localised hOrai3 and OptoCRAC. **B|** Schematic (upper) and cytosolic Ca2+ response (lower) to photostimulation in Orai1,2,3 triple knockout HEK-293 cells overexpressing OptoCRAC with ER-localised hOrai1-KKXX (red) or hOrai3-KKXX (turquoise). **C|** Schematic (upper) and cytosolic Ca2+ response (lower) to 2-APB treatment in Orai1,2,3 triple knockout HEK-293 cells overexpressing hOrai3-KKXX. **D,E|** Schematic (upper) and cytosolic Ca2+ response (lower) to photostimulation in Orai1,2,3 triple knockout HEK-293 cells overexpressing **D|** LOCa3 or **E|** ER-localised LOCa3-KKXX. All plots in this figure are generated using the data recorded by Y.W. and colleagues from one replication, representative of three independent biological replications.

LOCa3, a single-component light-activated Orai channel with a LOV2 domain inserted in the intracellular loop of Orai1, has been shown to mediate Ca^{2+} influx when overexpressed on the PM[197]. LOCa3 is directly gated by a blue light-induced conformational change of the LOV2 domain which results in channel opening. We expected the KKXX-tagged LOCa3 would be able to mediate OptoCR. Unfortunately, LOCa3-KKXX did not induce an increase in $[Ca^{2+}]_{cyt}$ with blue light stimulation. It is likely that the ER rerouting caused a loss of functionality.

2.3.7 **Spatial control of OptoCR using organelle tethered GECA**

The OptoCR system consisted of two components: the ER-resident Orai-KKXX and the cytosolic GECA. We envisioned that this binary tool could be used to elicit Ca^{2+} release at ER-organelle and ER-PM junctions. We fused common localisation sequences to the N-terminus of GECAs to tether the GECAs with the outer surface of mitochondria (TOM20 motif), ER (CytEM motif) and inner surface of the PM (Lyn motif). We hoped that the TOM20-tagged GECAs would activate OptoCR at ER-mitochondria junctions (**Figure** 2.8**A**), and that the Lyn-tagged GECAs would activate OptoCR at ER-PM junction (**Figure** 2.8**C**). In HEK-293 Orai1,2,3 triple KO cells, OptoCR using the human homologues (hOrai1-KKXX and OptoCRAC) always yielded larger response than the fruit fly homologues (dOrai-KKXX and dmBACCS2). Accordingly, we focused our efforts on the human homologues.

The human homologues of OptoCR at the ER-PM junction elicited that largest $[Ca^{2+}]_{cyt}$ response, while the response at ER-mitochondria junctions was much milder. To demonstrate that the OptoCR can tolerate the effect of subcellular targeting motifs, we

included a control of CytEM-GECAs fusion proteins and Orai-KKXX channels, of which both components of the binary system were localised on the ER membrane (**Figure** 2.8**B**). The human version of OptoCR tolerated the N-terminal tagging of CytEM, whereas the $[Ca^{2+}]_{cyt}$ response of the fruit fly system was abolished. In conclusion, hOrai1-KKXX and Lyn-OptoCRAC are able to mediate release of ER Ca^{2+} at ER-PM junctions.

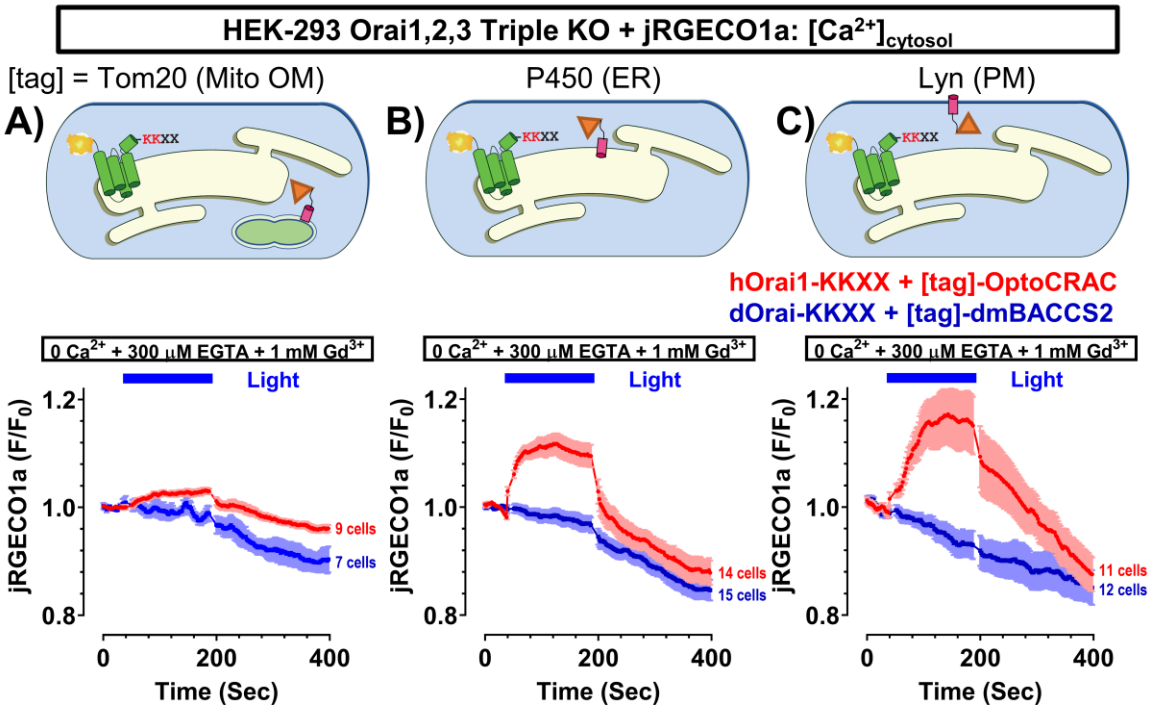

Figure 2.8 Light-activated Ca^{2+} release at organelle junctions.

Schematic (upper) and cytosolic Ca2+ response (lower) to photostimulation in Orai1,2,3 triple knockout HEK-293 cells overexpressing the indicated OptoCR combinations with GECAs tethers to **A|** outer mitochondria membrane using TOM20 motif for Ca^{2+} release at ER-mitochrondria junction, **B|** ER membrane using CytEM motif, or **C|** PM using Lyn motif for Ca^{2+} release at ER-PM junction. All plots in this figure are generated using the data recorded by Y.W. and colleagues from one replication, representative of three independent biological replications.

2.3.8 **OptoCR promotes RyR2-mediated CICR.**

The endogenous ER Ca^{2+} in non-excitable cells is mostly release via IP_3Rs. In addition to IP_3Rs, excitable cells also express RyRs on the ER and the sarcoplasmic reticulum (SR) in muscle cells. In cardiac muscles, RyRs are activated by the local increase of $[Ca^{2+}]_{cyt}$ from L-type Ca^{2+} channel (LTCC, or dihydropyridine receptor, DHPR) localised on the transverse tubules (t-tubules)[221]. RyR channels, in turn, mediates a much larger SR Ca^{2+} release flux into the cytosol. We envisioned that such Ca^{2+}-induced Ca^{2+} release (CICR) process might be initiated by OptoCR. We expressed the hOrai1-KKXX and OptoCRAC in the T-REx 293 cell line stably overexpressing RyR2-T4158P. We found that, in Ca^{2+}-free bath solution, blue light stimulation could induce oscillations of $[Ca^{2+}]_{cyt}$ (**Figure** 2.9**B**), but not in the control group without OptoCRAC (**Figure** 2.9**B**). This suggested that the Ca^{2+} release from OptoCR can initiate the CICR by RyR2-T4158P. It is promising that OptoCR may be applied to excitable cells to achieve light-induced CICR and the associated physiological outcomes.

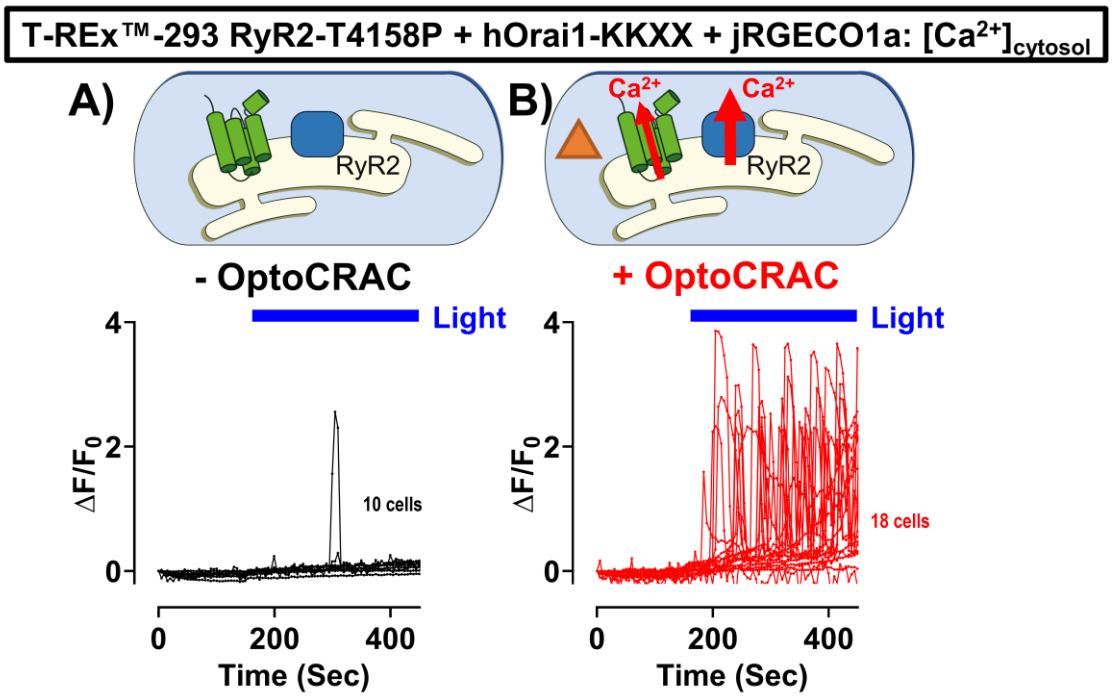

Figure 2.9 Light-activated Ca^{2+}-induced Ca^{2+} release (CICR).

Schematic (upper) and cytosolic Ca2+ response (lower) to photostimulation in T-REx 293 RyR2-T4158P stable cells overexpressing **A|** hOrai1-KKXX, in the absence of OptoCRAC, or **B|** both hOrai1-KKXX and OptoCRAC. Lines represent traces recorded from single cells.

2.4 Discussion

In this work, we demonstrate a novel strategy of light-induced Ca2+ release using CRAC channel-based GECAs. This system is named 'OptoCR' (**Figure 2.1B**). The canonical CRAC channels function as hexamers of Orai proteins[199]. In the hexameric channel, the N-terminal TM1 helices of each subunit forming the Ca^{2+}-permeable pore, with the other helices of each subunit (TM2 – TM4) arranged radially around the pore[222,223]. The TM4 helices are positioned in the outermost layer of the hexamer. In particular, the

C-terminal cytosolic extension helices, TM4ext, have been shown to bind directly with SOAR/CAD[214]. It is reasonable to speculate that rigid steric restriction on either N- or C-terminus of Orai channel may be highly disruptive to the channel function. To reroute the natively PM-localised Orai channel to ER membrane, we demonstrated that adding the C-terminal KKXX motif could result in ER localization while preserving Orai's ability to recruit and bind the SOAR/CAD domain. Fusing the STIM1 transmembrane domain to the N-terminus resulted in whorl-like folding/assembly issues, and fusing RyR2 S5S6 transmembrane domain to the C-terminus abolished the CAD recruitment, possibly due to the disruptive nature of transmembrane helices to Orai1 TM1 and TM4ext. The fact that KKXX-tagged hOrai1 can still recruit and bind the SOAR/CAD domain indicated that the KKXX-mediated ER-rerouting likely have not disrupted the architecture of the channel.

Both the PM-localised hOrai1 and the ER-localised hOrai1-KKXX can be activated by OptoCRAC and mediate Ca^{2+} flux, but OptoCR induced much smaller increases in $[Ca^{2+}]_{cyt}$. Competition from the endogenous Orai1 and Orai3 has been shown to contribute to the reduced response of OptoCR, as Orai1/3 knockdown moderately rescued OptoCR mediated by hOrai1-KKXX and OptoCRAC, but not dOrai and dmBACCS2. GECAs with monSTIM1-like designs preferentially interact with the PM-localised hOrai1 channels over the ER-localised hOrai-KKXX, possibly facilitated by the STIM1-PM and STIM1-MT interactions[189]. Other contributing factors to the reduced response of hOrai1-KKXX may include the absence of negative transmembrane potential and regulating proteins that enhances the store-operated Ca^{2+} entry (SOCE) on the ER membrane. Improving the compatibility of GECA with the ER-localised Orai-KKXX may be a promising direction for optimising the OptoCR system.

Pharmacological treatments typically induce thorough and irreversible Ca^{2+} release with ~ 60% decrease in R-CEPIA1er fluorescence. Optogenetic tools that facilitate light-induced ER Ca^{2+} release is much milder (15 – 20% by OptoFGFR1 including possible ~ 10% decreased by photobleaching). Nonetheless, the very mild OptoCR was sufficient to induce RyR2-mediated CICR and Ca^{2+} oscillation. This indicates that the approaches for inducing moderate Ca^{2+} release might be better positioned to recapitulate the intensity, duration, and reversibility of the ER Ca^{2+} release under physiological conditions. It is desirable that a variety of tools for OptoCR be engineered to recapitulate the different strengths and kinetic characteristics to mobilise ER Ca^{2+}. We have demonstrated that Orai-KKXX channels and the associated GECAs can serve as useful prototypes for inducing subtle $[Ca^{2+}]_{cyt}$ increase. Alternative architectures of Ca^{2+} channels could be explored for other desired characteristics.

OptoCR is normally composed of an ER-resident channel and a soluble GECA. By targeting the GECA to a specific subcellular location, the binary system conveniently facilitates Ca^{2+} release at the specific ER-organelle or ER-PM junctions. Endogenous STIM proteins and Orai channels mediate Ca^{2+} influx at the ER-PM junction, known as store-operated Ca^{2+} entry (SOCE). In this work, we demonstrate that it is also possible to elicit Ca^{2+} release at the ER-PM junction using Lyn-OptoCRAC-activated OptoCR. Subcellularly targeted GECA may be an elegant tool for studying ER-organelle Ca^{2+} communication and may be applicable to other organelles.

The Ca^{2+} release from the current OptoCR design is very mild. Its application may be limited to the processes only associated with transient, mild local Ca^{2+} release. Although OptoCR largely functions as expected, the activity of Orai-KKXX channels may

be compromised due to the non-native context of Orai channels (ER vs. PM). For one, it is believed that the ER potential is maintained at 0 mV despite the movement of ions across the ER membrane[224,225]. The current density of Orai channels at 0 mV is about 1/5 of that at -70 mV[226], the resting potential of PM where the endogenous Orai channels are located. It is therefore promising to engineer a Orai-KKXX variant with larger conductance around 0 mM. For another, the ER retention likely alters the glycosylation of the Orai channel. Specifically, Orai-KKXX is likely to have the mannose-rich N-linked core glycosylation structure at the Asn223 residue, rather than the sialic acid-containing mature structure when trafficked to the PM. Studies have shown that the glycosylation of hOrai1 channels may influence the extent of SOCE in a cell type-specific manner[227]. In this study, the effect of glycosylation was not investigated. The altered glycosylation state of membrane proteins associated with rerouted trafficking may be an important consideration to engineer the future versions of OptoCR.

Chapter 3

3 A genetic platform for mammalian cell-based directed evolution using serine recombinase Bxb1

*Chapter 3 is an original work of Shuce Zhang, unless otherwise stated. S.Z. was responsible for formed the idea, designed the study, established the methodology, constructed all plasmids, established candidate monoclonal cell lines and performed all flow cytometry experiments, except the experiments performed in the LLP cell line (**Figure 3.6B,D**) which were performed by collaborators, Dr. François St-Pierre and colleagues, at Baylor College of Medicine.*

3.1 Introduction

The past few decades have witnessed tremendous success in protein engineering facilitated by directed evolution. In a typical process of directed evolution, an ensemble of mutant peptides or proteins is expressed in a cellular or cell-free system from a DNA or RNA library. *Escherichia coli* is the most used cellular system to express and engineer soluble proteins by virtue of the highly efficient and mutually exclusive nature of plasmid transformation. On the other hand, engineering of membrane proteins and many other proteins that function in the context of mammalian signalling pathways still presents challenges in prokaryotic organisms such as *E. coli*.

One major limitation that hinders library screen in mammalian cells is a lack of a strict maintenance of the "genotype-phenotype" linkage that is necessary for directed evolution in conventional plasmid transfection. When a mixture of different plasmids is introduced to mammalian cells by transfection, the cells tend to take up multiple plasmids and express the genes following the mammalian promoter. Consequently, the transfected cells exhibit either a mixed phenotype of several expressed protein variants, or the phenotype of the dominant gain-of-function or loss-of-function variant. This inherent problem of co-transfection and co-expression complicates the identification and separation of variants in library screening. As a result, screening DNA libraries in mammalian cells typically necessitates miniprepping of plasmids individually followed by separate transfections (for example, engineering of ion channels[197] and voltage sensors[228]), or inefficiently transfecting a highly diluted library (for example, screening libraries of voltage sensors[14,229]) such that each cell gets zero or one plasmid. For these reasons, It is highly desirable to engineer a mammalian cell-based library screening system where only one variant is functionally expressed in each cell.

Site-directed recombination provides promising solutions to this problem. In **Section 1.3**, I reviewed the molecular mechanism and the reported systems for mammalian platforms for library screening using recombinase. In these systems, pooled libraries to be screened are cloned into a plasmid vector harbouring a recognition site for the recombinase. A host cell line is engineered to contain a single copy of landing pad in its genome harbouring a recognition site for the same recombinase. When transfected with the pooled library in bulk, only one molecule becomes integrated into the landing pad locus in each cell. However, the existing systems typically suffer problems such as low

efficiency, reversible integration and excision, and expression of non-integrated plasmid molecules[230], making them less than ideal for rapid screening in directed evolution.

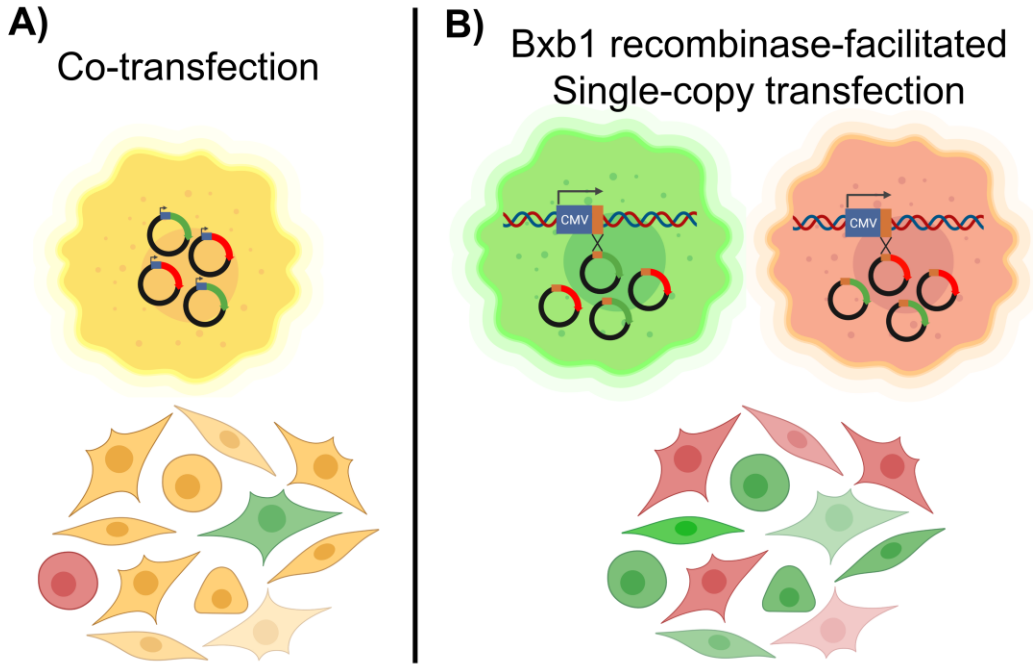

Figure 3.1 Schematic for canonical co-transfection and recombinase-mediated single copy integration.

A| Mammalian cells can take up and express multiple plasmid molecules. As a result, the phenotype reflects the average of the transfected library members. **B|** Using an engineered host cell line harbouring only one copy of the landing pad, a recombinase integrates only one molecule from the transfected library in each cell. Unintegrated molecules lack mammalian promoters and do not express in the host cell. Therefore, each cell expresses only one molecule of the library. All schematics in this chapter are created with BioRender.com

In this study, we designed a genetic platform to transfect mammalian cells in a "one-cell-one-copy" fashion (**Figure** 3.1**B**). We confirmed that Bxb1 mediates more efficient recombination than Flp. We engineered a HEK-293 FT-derived stable cell line

harbouring one locus of the landing pad as the second generation host cell line and designed a dual-function pBAD-derived construct for both mammalian cell transfection and prokaryotic expression. Notably, the dual-function pBAD-derived construct did not exhibit promoter-less expression. We demonstrated that only one plasmid per cell will express under a CMV promoter when integrated into the host genome. This system works more efficiently (5-10%) than traditional methods using diluted libraries. We also showed that the Bxb1-facilitated recombination system can integrate and express a linear DNA donor such as PCR product. We further proposed a third-generation host cell line featuring a RMCE design for better compatibility with linear DNA donors, and an “on-demand” fluorescent reporter for landing pad, recombination, and photoactivation, improving compatibility with various fluorescence-based phenotype readout. We envision that this “one-cell-one-copy” mammalian cell platform will serve as an efficient solution for engineering a wide range of optogenetic tools.

3.2 Methods and materials

3.2.1 Constructs

To create the first-generation (Flp) system donor plasmid (pBAD-*FRT*-EGFP-SPA-HO1), a sequence encoding *FRT*-EGFP was subcloned into pBAD-smURFP-RBS-HO1 plasmid (Addgene #80341) using the *BamH*I and *EcoR*I sites. A start codon was placed in frame with the EGFP sequence after the *BamH*I site for proper prokaryotic expression. The EGFP sequence is also in frame with the *FRT* landing pad (p*FRT*/lacZeo) in the host Flp-In™ 293 T-REx cell line for proper recombination and expression. A *Xho*I site was introduced after the 6×His tag using the QuikChange Lightning Site-Directed Mutagenesis

Kit according to the manufacturer's protocol (Agilent). A synthetic polyadenylation signal was added in front of this XhoI by a subcloning procedure.

To create the second-generation system landing pad plasmid (pLJM1-*FRT*-*attP*-mCherry), a gBlock (Integrated DNA Technologies) encoding a Kozak sequence, *FRT*, Bxb1 *attP*, and mCherry was cloned into pLJM1-EGFP vector (Addgene #19319) using *Nhe*I and *EcoR*I sites. To create the second-generation system donor plasmids (pBAD-*FRT*-*attB*-EGFP/mCerulean3/miRFP680/smURFP), a gBlock (Integrated DNA Technologies) encoding *FRT* and Bxb1 *attB* was cloned into the first-generation donor vector together with a PCR product of either EGFP, mCerulean3, miRFP680 or smURFP and synthetic polyadenylation sequence, using the *BamH*I and *EcoR*I sites.

To create the third-generation system landing pad plasmid (pLJM1-*attP*-EBFP-*attP**), the insert sequence encoding a Kozak sequence, wild type Bxb1 *attP*, EBFP and a mutant Bxb1 *attP* (designated as *attP**) was synthesized on two gBlocks with an overlapping sequence of 56 bp and subcloned into pLJM1-EGFP vector using NheI and *EcoR*I sites. To create the third-generation system donor plasmid (pBAD-*attB*-smURFP-*attB**), the insert sequence encoding a wild type Bxb1 *attB*, smURFP, a 6×His tag, a herpes simplex virus thymidine kinase polyadenylation signal and a mutant Bxb1 *attB* (designated as *attB**) was synthesized on two gBlocks with an overlapping sequence of 109 bp and subcloned into the second-generation donor plasmid pBAD-*FRT*-*attB*-mCerulean3 vector using *BamH*I and *Xho*I sites. All cloning steps were performed using an in-house prepared Gibson Assembly mix according to the originally published recipe[198] unless otherwise specified. All constructs were amplified and purified using the Thermo Fisher GeneJET Plasmid Miniprep or Maxiprep kit following transformation of the Thermo

Fisher ElectroMAX electrocompetent cells (for regular constructs) or NEB® Stable chemically competent cells (for lentiviral constructs). Cloning results were confirmed by Sanger Sequencing (MBSU, University of Alberta). pCAG-NLS-HA-Bxb1 (Addgene plasmid #51271) and pOG44 (Thermo Fisher) were used to express Bxb1 and Flp recombinases in the cell lines, respectively.

3.2.2 **Cell culture and stable cell line generation**

HEK-293 FT cells (Thermo Fisher) and Flp-In™ T-Rex 293 cells (Thermo Fisher) are cultured in complete DMEM (HyClone high-glucose DMEM with sodium pyruvate and L-glutamine, supplemented with 10% FBS and1% antibiotic-antimycotic) in a moist 5% CO_2 atmosphere at 37 °C. To generate the monoclonal stable cell lines harbouring the landing pad by lentivirus transduction, the packaging HEK-293 FT cells were seeded at ~70% confluency in a T25 flask 24 hrs before transfection. 3239 ng pLJM1 landing pad transfer plasmid, 2674 ng pCMV-dR8.2 dvpr (Addgene #8455) and 724 ng pCMV-VSV-G (Addgene #8454) was mixed with 13.2 µL TurboFect transfection reagent (Thermo Fisher) in 400 µL serum-free DMEM/F12 and incubate for 10 min before adding to the cell culture. Medium was replaced to fresh complete DMEM/F12 (DMEM/F12 Nutrient Mixture with L-glutamine and HEPES supplemented with 10% FBS and 1% antibiotic-antimycotic) 6 hrs after transfection. Supernatant was harvested 48 hrs after transfection and passed through a 40 µm filter to remove possible cell debris. The packaging cells were cultured in complete DMEM/F12 for another 24 hrs. HEK-293 FT cells to be transduced were seeded in 6-well plates to achieve ~80% confluency 24 hrs after seeding. Freshly filtered supernatant was added to this cell culture at a ratio of 1:3 to 1:5. 24 hrs after infection, medium was replaced to fresh complete DMEM/F12. 2 µg/mL puromycin dihydrochloride

(Gibco) was added to the medium and changed every 2-3 days until minimal cell death was observed. The polyclonal cells that survived the puromycin selection were split in two halves for cell sorting and cryopreservation, respectively. mCherry-positive single cells were sorted to 96-well plates in a 1 cell/well fashion. 7 days after sorting, these 96-well plates were imaged on an ImageXpress high-content analysis system (Molecular Devices) and inspected for monoclonality and mCherry expression. The monoclonal stable cell lines were cultured in complete DMEM or complete DMEM/F12 supplemented 2 μg/mL puromycin dihydrochloride, cryopreserved within three generations of passages and were subjected to functional and genomic tests to identify the clones harbouring single copy of landing pad.

3.2.3 **Flow cytometry**

Briefly, cells were seeded at ~60% confluency in 6-well plates 24 hrs before transfection. For recombination assay in HEK-293 FT cells, 1 μg landing pad plasmid, 1 μg Bxb1 plasmid and 2 μg donor DNA (plasmid or PCR product) are mixed with 8 μL TurboFect transfection reagent (Thermo Fisher) in 400 μL serum-free DMEM/F12 and incubate for 10 min before adding to the cell culture. For recombination assay in landing pad-expressing stable cells, 2 μg Bxb1 plasmid and 2 μg donor DNA (plasmid or PCR product) are mixed with 8 μL TurboFect transfection reagent (Thermo Fisher) in 400 μL serum-free DMEM/F12 and incubate for 10 min before adding to the cell culture. Medium was replaced to fresh complete DMEM/F12 4 hrs after transfection. Cells were analysed and/or sorted 72 hrs after transfection. To objectively measure the viability, cells were trypsinised followed by neutralisation by the old medium that contains detached/dead cells. Cells are then stained with 0.05% LIVE/DEAD™ Fixable Near IR (780) Viability dye

(Invitrogen) in Ca^{2+}/Mg^{2+}-free HBSS containing 250 μM EGTA for 30 min. Cells are resuspended in Ca^{2+}/Mg^{2+}-free HBSS containing 250 μM EGTA for flow cytometry analysis and sorting.

For the first generation (Flp) system, flow cytometry assays were performed on a BD Accuri C6 flow analyser with the associated software. EGFP fluorescence was excited by a 488 nm laser and the emitted light was collected after passing through a 533/30 filter. mCherry fluorescence was excited by a 488 nm laser and the emitted light was collected after passing through a 670 LP filter. For the second and third generation (Bxb1) systems, flow cytometry assays were performed on a BD FACSMelody cell sorter equipped with a 100 μm sorting nozzle. mCerulean3 fluorescence was excited by a 405 nm laser and the emitted light was collected after reflected by a 448/45 mirror. EGFP fluorescence was excited by a 488 nm laser and the emitted light was collected after passing through a 527/32 filter. mCherry fluorescence was excited by a 488 nm laser and the emitted light was collected after passing through a 700/54 filter. miRFP680 fluorescence was excited by a 640 nm laser and the emitted light was collected after passing through a 660/10 filter. Near IR (780) viability staining (Invitrogen) was excited by a 640 nm laser and the emitted light was collected after passing through a 783/56 filter. Compensation was set up with the cells expressing or stained with a single fluorophore. Cells were first gated using FSC-A vs SSC-A, SSC-H vs SSC-W, and FSC-H vs FSC-W to identify single cells. The viability staining-negative population was the gated to analyse and quantify the live cells. Gates for EGFP-positive, mCerulean3-positive, and miRFP680 cells were set up with the mCherry-expressing live cell negative control. Data collection and sorting set up were performed in BD FACSChorus software. Further analysis and plotting were performed on FlowJo v10.8 software. Unless otherwise stated, all flow cytometry plots are generated

from one out of at least three independent biological replications with different batches of cell culture transfected separately on different days.

3.2.4 **Quantification of landing pad copy number**

Stable cell lines and control HEK-293 FT cells were growing in 6-well plates to ~1×10^6 cells/well. Cells were detached using trypsin, neutralized in DMEM, and washed in HBSS before genome DNA extraction using QIAamp DNA Blood Mini Kit (Qiagen). Landing copy numbers was measured by qPCR using an assay targeting lentivirus vector[231] (primer 1: TCTCGACGCAGGACTCG, primer 2: TACTGACGCTCTCGCACC, probe: /6-FAM/ATCTCTCTCCTTCTAGCCTC/6-TAMRA/) and an assay targeting mCherry[232] (primer 1: GAGGCTGAAGCTGAAGGAC, primer 2: GATGGTGTAGTCCTCGTTGTG, probe: /6-FAM/CCAACTTGATGTTGACGTTGTAGGCG/6-TAMRA/). An assay targeting Poly(rC)-binding protein 2 (PCBP2) locus[233] (primer 1: TTGTGTCTCCAGTCTGCTTG, primer 2: AGGTGGTGGTGGTGGTA, probe: /6-FAM/CCCTCTCCTGGCTCTAAATGTTGTGT/6-TAMRA/) was used as an internal reference. It has been reported that cell lines derived from HEK-293 cells consistently contain three copies of PCBP2 gene[234]. Synthetic gBlocks of the amplicons of the abovementioned assays were serially diluted and used as the standards for qPCR reactions. Molar concentrations relative to that of the internal reference were used to calculate the copy number.

A genome walking analysis using TaKaRa Lenti-X™ Integration Site Analysis Kit was performed to locate the insertion sites of landing pads in the genome. Briefly, genome DNA extracted described above was digested using DraI, SspI and HpaI restriction

enzymes, respectively. Digested genome fragments were ligated with GenomeWalker adapters followed by a nested PCR with a pair of primers biding to 3' UTR and the genome walker adapter. PCR reaction was analysed by agarose gel electrophoresis. All non-smearing bands were excised, recovered using GeneJET Gel Extraction Kit (Thermo Fisher) followed by Sanger sequencing (MBSU, University of Alberta) using a 3' LRT-binding primer (GCTCCTCTGGTTTCCCTTTCGCTTTCAA). Genome locus of the insertion site was identified by querying the Sanger sequencing results on the NCBI blast server (https://blast.ncbi.nlm.nih.gov/) and manually inspected using the UCSC genome browser (https://genome.ucsc.edu/).

3.3 Results

3.3.1 First Generation platform cell line: the commercial Flp-FRT recombination-based system is inefficient

As discussed in **Section 1.3.1**, the Flp-In™ T-REx 293 cell line harbours a single copy of a landing pad containing the *FRT* sequence following an SV40 promoter. A donor plasmid can be incorporated at the *FRT* site and expressed under the SV40 promoter. Although we envisioned possible problems with this system (for example, reversibility, low efficiency, low expression level of SV40 promoter), since the Flp-In™ T-REx 293 cell line is commercially available and presents as a potentially ready-to-use solution, we decided to design our first-generation library screening system based on the landing pad of this cell line. Accordingly, a donor plasmid based on pBAD-HO1 vector was constructed for

this purpose, with a *FRT* inserted between the start codon and the GOI while maintaining the in-frame translation of the latter. pBAD is an arabinose-inducible expression vector in *E. coli* and is the most commonly used vector in the Campbell lab. Notably, pBAD vector does not contain any promoter or other elements known to function in mammalian cells. Upon integration, the GOI of the integrated copy will be placed downstream of the SV40 promoter (**Figure** 3.2**A**).

To test whether this system is able to express the GOI on the donor plasmid efficiently, we first chose EGFP as the GOI, reasoning that EGFP is usually well folded and easily expressed in mammalian cells, and that the percentage of EGFP-positive population could represent the efficiency of genomic integration. We co-transfected the Flp-In™ T-REx 293 cells with the pBAD-*FRT*-EGFP donor plasmid and pOG44 (CAG-Flp) according to the manufacturer's protocol (**Figure** 3.2**B**, third panel). We also used a donor plasmid-only transfected sample to test for possible integration-independent expression (**Figure** 3.2**B**, second panel), and an un-transfected sample as the negative control (**Figure** 3.2**B**, first panel). A sample transfected with pcDNA-EGFP plasmid served as a positive control, where EGFP will express from a CMV promoter from the plasmid. However, we were unable to detect any increase in the percentage of GFP-positive population for the co-transfection group (**Figure** 3.2**B**, third panel). Other studies have reported an integration efficiency of ~0.015% measured by enumerating the colonies surviving hygromycin selection[109]. This efficiency is comparable with the percentage of EGFP-positive cells in the negative control sample (**Figure** 3.2**B**, first panel). It is not surprising that an increase due to the "true" integration was not detected in this experiment. We concluded that it was not practical to use Flp-mediated recombination to screen the libraries in Flp-In™ T-REx 293 cells without antibiotic selection.

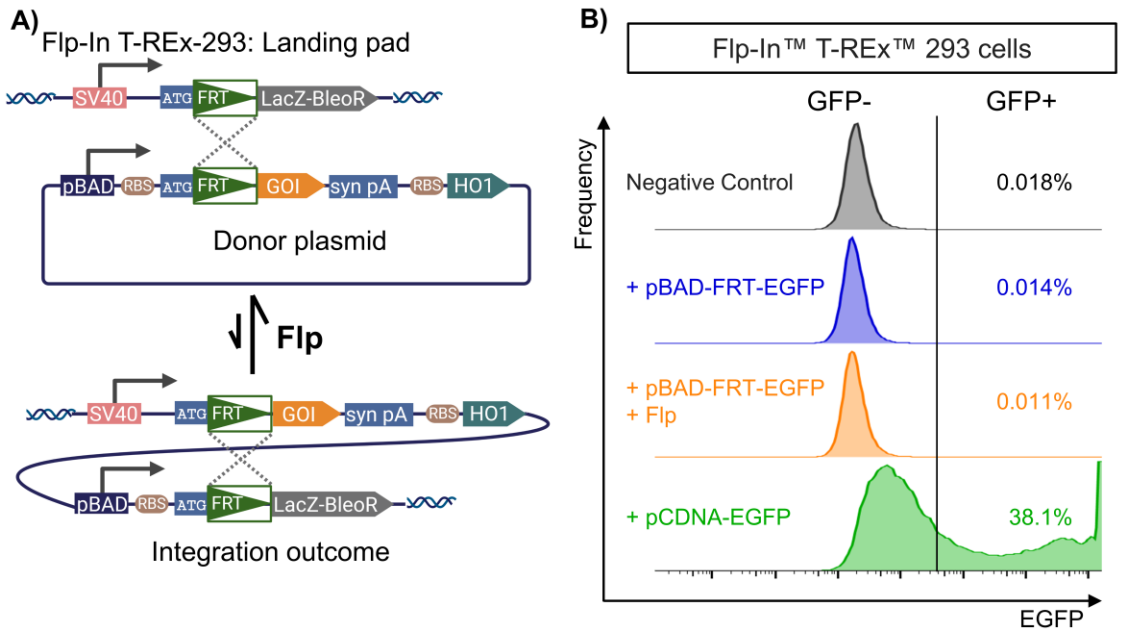

Figure 3.2 First-generation system using Flp-*FRT* recombination.

A| Schematic of Flp-mediated recombination between the *FRT* sites of the SV40 locus of Flp-In™ T-REx 293 cell line and a pBAD-based donor vector. **B|** Recombination efficiency is measured as the percentage of GFP-positive cells 72 hrs post-transfection by flow cytometry. Negative control, untransfected Flp-In™ T-Rex-293 cell; pBAD-*FRT*-EGFP, donor plasmid without mammalian promoter; Flp, pOG44 plasmid encoding Flp under CAG promoter.

3.3.2 Bxb1 mediates more efficient recombination than Flp in transiently transfected cells

We hypothesized that the low efficiency of the first generation system was due to the reversibility of the tyrosine recombinase Flp, which favours the intramolecular recombination (removal) over intermolecular (integration). If the reversibility is the limiting factor, then using an irreversible recombination system such as Bxb1 would greatly improve the recombination efficiency. To test this hypothesis, we designed a landing pad vector carrying both *FRT* and *attP* sequences downstream of a CMV promoter, allowing

for direct comparison of Bxb1- and Flp-mediated recombination using identical landing pad plasmid and donor plasmid. An ORF encoding mCherry was placed downstream of the *FRT* and *attP* sites, in frame with the start codon. We also designed a donor vector with *FRT* and *attB* sequences. The newly designed landing pad and donor vector would be compatible with both Flp and Bxb1 recombinases. That is, recombination with either recombinase will produce an in-frame expression cassette of the GOI (EGFP in this experiment) encoded on the donor plasmid (**Figure** 3.3**A**). In other words, the EGFP-expressing population represents the cells that underwent successful *attP-attB* recombination. Such recombination, in theory, should lead to a promoter-less mCherry locus and stop its expression in mammalian cells. However we found that, in practice, most cells that were EGFP-positive remained mCherry-positive, possibly due to the long lifetime of mCherry protein in the cytosol, and the possibility that not all transiently transfected landing pad plasmid molecules have reacted with donor plasmid. Therefore, the mCherry-positive population roughly represents the cells that took up at least one copy of the landing pad plasmid molecules. Using this system, it was possible to obtain a well-controlled comparison of the recombination efficiencies by co-transfecting the landing pad and donor plasmid with either recombinase (**Figure** 3.3**C**). Indeed, in transiently transfected HEK-293 FT cells, Bxb1 mediated recombination was found to be 3× more efficient recombination than Flp (**Figure** 3.3**C, F**). It should also be noted that the recombination efficiency of Flp improved to 4.10% of all cells and 11.3% in landing pad-expressing cells. This remarkable improvement might be a result of the presence of a stronger promoter (CMV vs. SV40) and the presence of multiple copies of landing pad in the transient transfection system.

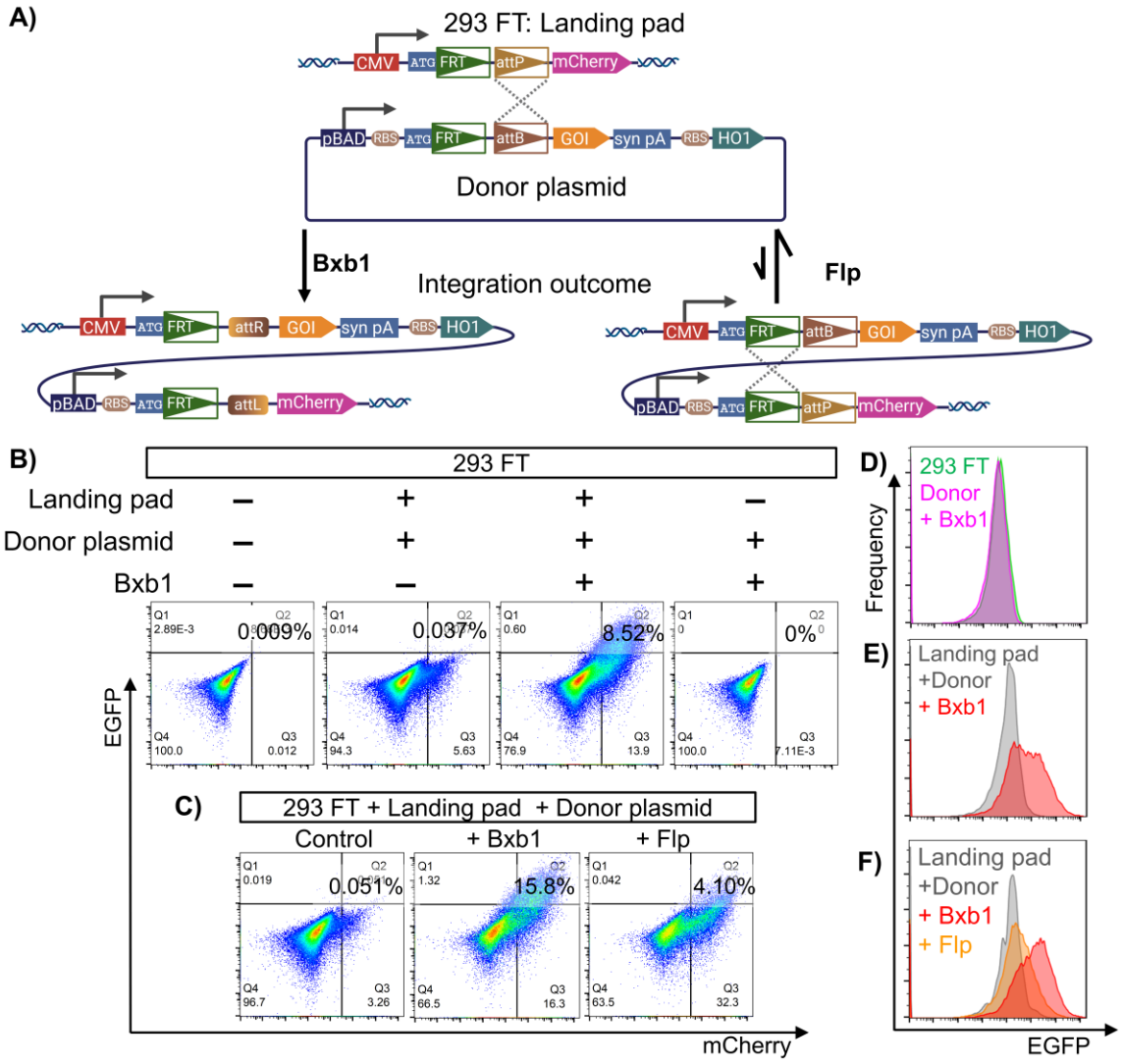

Figure 3.3 Bxb1-mediated recombination is more efficient than Flp in transiently transfected HEK-293 FT cells.

A| Schematic of the landing pad and the donor plasmid compatible with both Flp and Bxb1. **B, D-E|** Fluorescence representing expression of the landing pad (mCherry) and recombination (EGFP) was measured by flow cytometry in HEK-293FT cells transiently transfected with some or all of the components of the system. Histogram of EGFP fluorescence in all single cells is shown in D) and in mCherry-positive cells shown in E). **C, F|** Recombination efficiencies of Flp and Bxb1 were compared by co-transfecting 293-FT cells with the landing pad and donor plasmids with plasmids encoding either Flp or Bxb1 recombinase.

We also aimed to demonstrate that the expression of GOI is the result of specific recombination mediated by Bxb1. Indeed, minimal EGFP fluorescence was detected in cells transfected with only the landing pad and donor plasmid, but not Bxb1 (**Figure** 3.3**B, E**). EGFP was not detected in cells transfected with only Bxb1 and the donor plasmid, but not with the landing pad plasmid (**Figure** 3.3**B, D**). This result indicated that the Bxb1-mediated recombination was specific to the *attP* located on the landing pad, but not the pseudo-*attP* sites in the native genome. It also suggested that the expression of GOI is depended on the CMV promoter located on the landing pad, and notably, that the un-integrated free donor plasmid molecules did not contribute to the expression. In previously reported Bxb1 system, 30 - 40% of landing pad cells transfected with Matreyek lab's[116–118] donor plasmids exhibited transient, recombination-independent GOI expression (**Figure** 3.6**B**). These donor constructs have the Kozak sequence and start codon downstream of *attP*/*attB* sites. This promoter-less expression typically lasted for ~7 days before disappearing (**Figure** 3.6**B**), effectively behaving like transiently transfected pcDNA constructs. In our pBAD-based donor plasmid, Kozak sequence is not present. Instead, a prokaryotic RBS is placed 9 nt upstream of the start codon following the pBAD promoter in order to facilitate seamless *E. coli* expression (**Figure** 3.6**C**). The Kozak sequence is present only in the landing pad plasmid upstream of the *FRT-attP* sites, but not in the donor plasmid. This design not only eliminated the recombination-independent expression issue of Matreyek's design (**Figure** 3.3**B, Figure** 3.6**A**), but also facilitated the dual application of *E. coli* expression and mammalian Bxb1 recombination using the same vector, without the need of subcloning. We envisioned that these two major improvements would expedite the turnaround time of each round of directed evolution, resulting in a time of ~10 days / round.

3.3.3 Engineering monoclonal stable cell line with a single copy landing pad as the second-generation platform cell line

In the aforementioned transient transfection system, we have demonstrated that the newly designed landing pad and donor plasmid system can achieve efficient and specific Bxb1-mediated recombination. We further aimed to achieve single-copy incorporation and expression. For this purpose, it is crucial to establish a monoclonal stable cell line that harbours just one single copy of the landing pad. Briefly, HEK-293 FT cells were transduced with lentivirus carrying the landing pad sequence, selected for puromycin resistance, sorted by mCherry fluorescence and expanded into monoclonal stable cell lines. Lentivirus transduction can produce cells that vary in their insertion sites and copy numbers. We therefore kept a number of single cell clones and profiled for their landing pad copy numbers, recombination efficiency, and growth rate (**Figure** 3.4**A**).

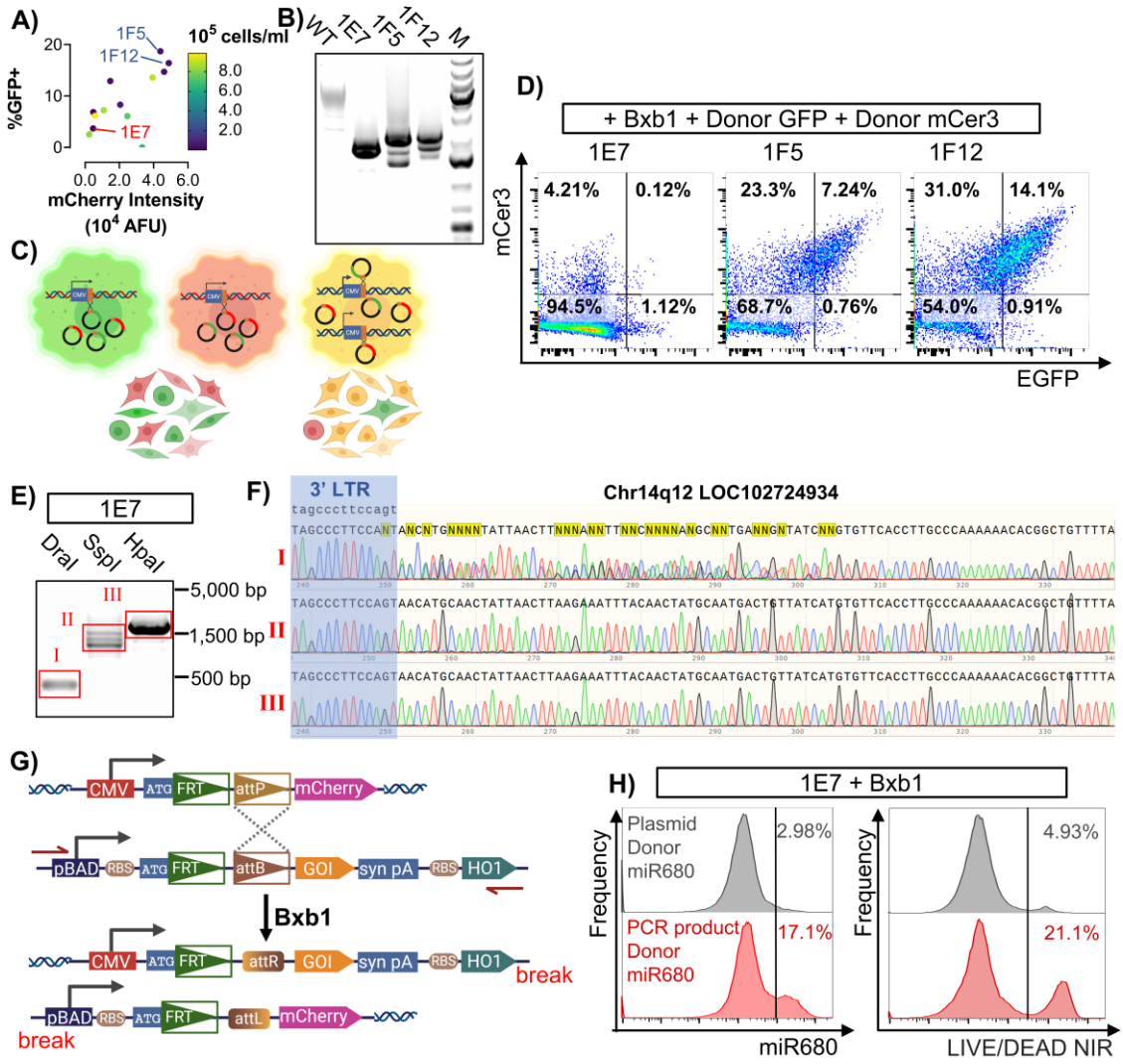

Figure 3.4 Establishment of the second-generation stable host cell line for Bxb1-mediated recombination.

A| Quantification of median mCherry fluorescence intensity, EGFP+ population percentage and cell density by flow cytometry, indicating landing pad expression, recombination efficiency and cell proliferation rate, respectively. **B|** Gel electrophoresis diagram of nested PCR using a lentiviral specific primer and an adapter primer of the GenomeWalker libraries of 4 cell lines with HpaI digestion. **C|** Schematics of the two-colour mutual exclusion assay. **D|** Quantification for the two-colour mutual exclusion assay. The presence of EGFP+ mCer3+ population indicates the presence of multiple copies of landing pad in the stable cell line. **E|** Gel electrophoresis diagram of nested PCR of the GenomeWalker libraries of 1E7 stable cell line constructed using three restriction enzymes. **F|** Sanger sequencing results of the indicated bands in E) using a lentiviral specific primer. **G|** Schematic of recombination using linear PCR product as donor DNA. **H|** Quantification of recombination efficiency (left) and viability (right) of 1E7 stable cell line using either circular plasmid DNA or PCR product as donor.

To ensure that the stable cell line contains only a single copy of landing pad, we first tested whether the cells functionally express only one variant of the donor plasmid. For this purpose, we employed a two-colour mutual exclusion assay[116]. We transfected each strain with Bxb1 and a mixture of two donor plasmids encoding different fluorescent proteins, mCerulean3 and EGFP. We envisioned that, for strains of cells with single copy of landing pad, each cell is only able to incorporate one plasmid molecule. Since Bxb1 recombination is irreversible, this cell will express the colour of fluorescence of the incorporated FP. Therefore, the cells that have undergone successful recombination will exhibit either cyan (from mCerulean3) or green fluorescence (from EGFP), but not both. In other words, the two colours will appear mutually exclusive to each other (**Figure** 3.4**C** left panel). In contrast, in strains with multiple copies of landing pad, a substantial population of double-positive cells will be expected to present (**Figure** 3.4**C** right panel). The 1E7 clone exhibited mutual exclusion between co-transfected donor plasmids encoding EGFP and mCerulean3 (**Figure** 3.4**D**). In contrast, 1F5 and 1F12 clones showed substantial double-positive population and were believed to contain multiple copies of functional landing pad, unsuitable for serving as the platform cell line. This assay also allowed us to quantify the recombination efficiency of these cell strains. For 1E7, the overall incorporation efficiency was calculated as the sum of percentages of mCerulean3-positive and EGFP-positive populations. This efficiency (~5%) was comparable or slightly higher than that observed in Matreyek's system[116], possibly due to a stronger promoter (CMV vs. Ptet).

To further confirm that the 1E7 clone contains single copy of landing pad, we performed a genome walking analysis of the genomic DNA. Briefly, genomic DNA was extracted from monolayer of cell culture and digested with restriction enzymes (DraI, SspI,

and HpaI) to generate a library of shorter gDNA fragments with blunt ends, which were subsequently ligated with GenomeWalker adapters using T4 ligase. The GenomeWalker provided a universal primer binding site on both ends of all fragments in the library. A nested PCR was performed using this library as the template to amplify the fragments containing binding sites for primers targeting lentivirus 3' UTR. The IE7 strain, which was shown to express only one variant in the two-colour mutual exclusion assay, yielded a single band in the amplification of DraI- and HpaI-digested gDNA libraries (**Figure** 3.4**E**). In the SspI-digested library, three bands of similar sizes (~1,500 bp) could be observed from the agarose gel electrophoresis (**Figure** 3.4**E**). Nonetheless, a clean, unambiguous Sanger sequencing reading was obtained when these bands were pooled and purified (**Figure** 3.4**F**). The sequencing result is consistent with the bands obtained from the other two libraries. A BLAST search of the sequence suggested that the insertion site is located in the chromosome 14 q12 region (**Figure** 3.4**F**). The 1F5 and 1F12 clones were included as controls. These two clones exhibited multiple distinct bands of nested PCR product using the HpaI-digested gDNA library (**Figure** 3.4**B**), consistent with the functional non-exclusive phenotype in the two-colour mutual exclusion assay. Overall, the 1E7 clone was identified, functionally and genetically, to harbour a single copy of landing pad and considered suitable to serve as the platform cell line for mammalian cell-based directed evolution.

3.3.4 Engineering the next-generation platform cell line for linear donor DNA

Bxb1 recombinase has been shown to tolerate both linear and circular DNA as its substrate in mammalian cells[235]. Libraries in the form of linear DNA can be conveniently prepared by a variety of techniques, including EP-PCR and synthetic combinatorial libraries. Being able to screen linear DNA libraries would represent an attractive option that avoids the extra steps of subcloning. Having established that 1E7 clone as the platform cell line for mammalian cell-based directed evolution by Bxb1-mediated DNA recombination, we further set out to explore the possibility of using linear DNA as the donor for the recombination. To prepare linear DNA donor, we performed PCR to amplify the GOI from the donor plasmid encoding miRFP680 using a pair of primers that bind to pBAD promoter and HO1, respectively (**Figure** 3.4**G**). The PCR amplicon was chosen such that it contains the *attB* site that is necessary for Bxb1-mediated recombination, and the synthetic poly(A) signal for stabilising the mRNA in the host cell (**Figure** 3.4**G**). We also used circular donor plasmid encoding miRFP680 as the control. We first gated the live cells that are negative for the viability stain from the single cells. In the live cell population, the recombination efficiency was quantified by the percentage of miRFP680-positive cells. Circular plasmid donor yielded a recombination efficiency of 2.98%, while using PCR product as the donor DNA increased the efficiency to 17.1% (**Figure** 3.4**H**), a 474% improvement from circular DNA donor. However, using PCR product as the donor DNA also caused substantially increased cell death (21.1% vs. 4.93%, **Figure** 3.4**H**). This might be caused by the structural changes introduced to the chromosome. After the recombination between an *attB* site on a linear DNA molecule and a genomic *attP* site, the chromosome becomes fragmented into two parts (**Figure** 3.4**G**), one of which does

not have a centromere. While both ends of the break already exist on the linear donor DNA, such structural change effectively introduces a DSB into the chromosome that might elicit DNA damage repair responses and lead to subsequent cell death.

To facilitate the incorporation of linear DNA donor, we set out to design a third generation of platform cell line. The new design features a RMCE type recombination (**Figure** 3.5**AB**), where two *attP* sites were placed in the landing pad. Correspondingly, two *attB* sites were placed on the donor vector flanking the GOI, which undergo recombination with the two landing-pad *attP* sites, respectively (**Figure** 3.5**AB**). These two pairs of *attP*/B sites can both be recognised by the Bxb1 recombinase, only differing in the central dinucleotide (GT vs. GA) that renders them functionally orthogonal[235]. Effectively, the region flanked by the *attP* sites on the landing pad is exchanged with the region flanked by the *attB* sites (**Figure** 3.5**AB**). Such RMCE design does not introduce permanent breaks in the genome, regardless of the topology of donor DNA (circular or linear).

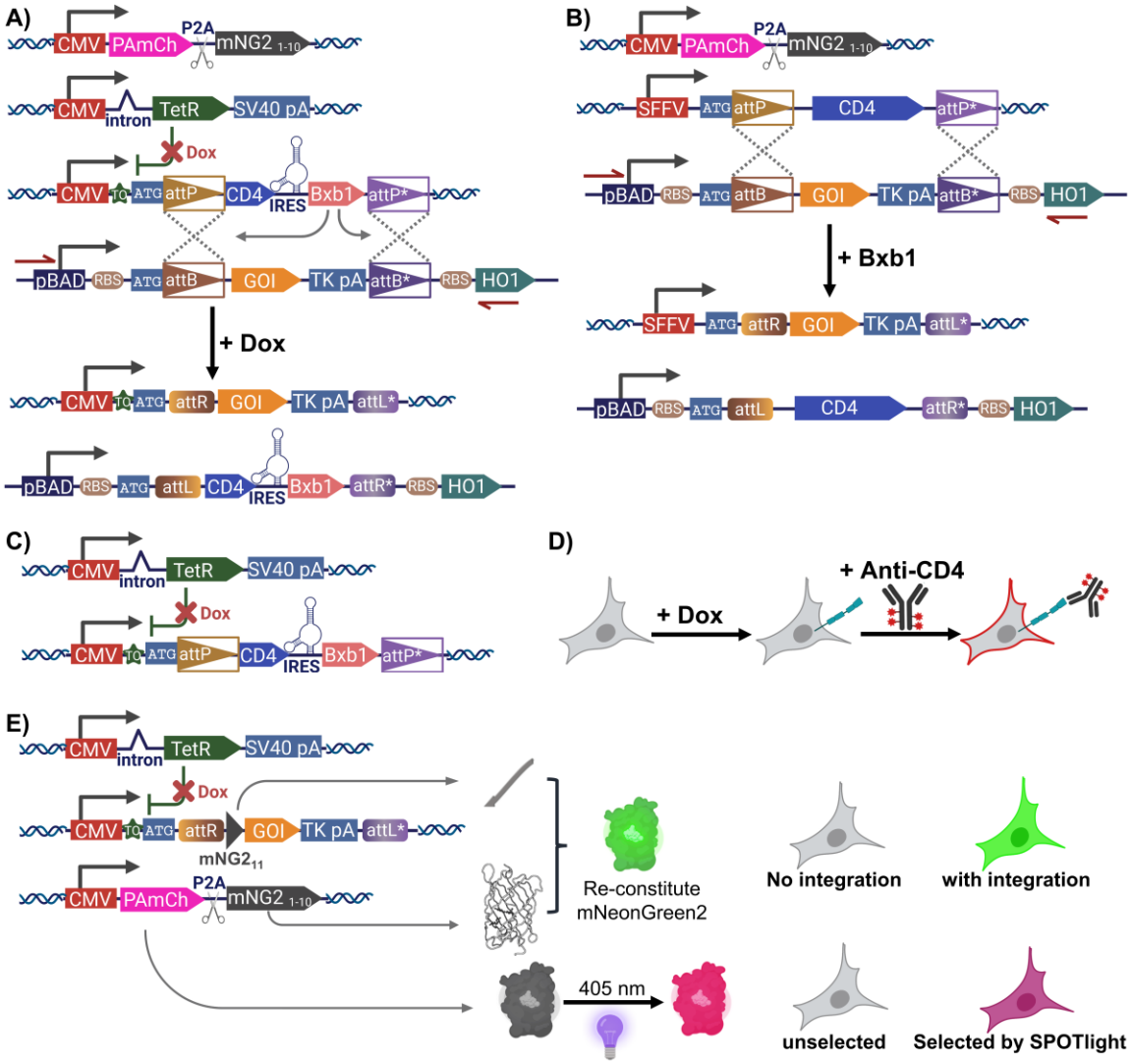

Figure 3.5 Design of the third-generation platform cell line for linear donor DNA by RMCE.

A| Schematic of the third-generation host cell line, version 1. TetR from the Flp-In™ T-REx facilitates the doxycycline-inducible expression of the landing pad cassette under CMV promoter and $TetO_2$ sequence. The landing pad encodes human CD4, followed by IRES sequence and a Bxb1 ORF, flanked by the Bxb1 *attP* and *attP**. Donor DNA represents a PCR product of the donor plasmid encoding the GOI and a TK poly(A), flanked by *attB* and *attB**. Directional integration is achieved by RMCE between the landing pad and the donor DNA by inducing Bxb1 expression with doxycycline. A separate expression cassette introduced by lentiviral transduction encodes PAmCherry, P2A, and $mNG2_{1-10}$ under a CMV promoter. "*" designates the functionally orthogonal *attP*/B mutants. **B|** Schematic of the third-generation host cell line, version 2. Version 2 and version 1 shares the same donor DNA and RMCE mechanism, only differing in its landing pad. The landing pad of version 2 uses a strong

constitutive SFFV promoter and does not encode the Bxb1 recombinase. A separate plasmid needs to be co-transfected with the donor DNA. **C|** Schematic for the doxycycline-inducible landing pad for version 1 host cell. **D|** Schematic for the optional hCD4 staining with tuneable colour for version 1 host cell. **E|** Schematic for the optional fluorescence reporter facilitated by the feature cassette. PAmCherry can be activated by SPOTlight protocol to allow microscopy-base cell tagging. mNeonGreen2$_{1-10}$ can be used with donor GOIs with a mNeonGreen2$_{11}$ to reconstitute mNeonGreen fluorescence as a recombination indicator.

Another consideration in designing the third-generation platform was to make more fluorescent channels available for the GOI or assay readout. This required us to reduce or eliminate as many fluorophores occupied by the landing pad. In the 1E7 clone, the landing pad carried an mCherry ORF as a placeholder (**Figure** 3.3**A**). This fluorophore was only useful during the sorting and screening of the single cell clones. Since the mCherry protein has a considerably long lifetime in mammalian cells, its fluorescence may be present for several days after its expression has been inactivated by the recombination, consistent with previously reported slow decay of inactivated genes by landing pad recombination[117]. This feature makes the 1E7 clone unsuitable for evolving any GOI with a red fluorescence readout. In the design of the third generation system, the placeholder gene is replaced by the ORF of hCD4. CD4 is a cell surface marker for certain types of T cells. The expression of CD4 can be conveniently assayed by staining the cell surface with fluorophore-conjugated anti-hCD4 antibody (**Figure** 3.5**D**). This staining allows us to select a colour that is compatible with any other fluorophores that are present in the system when sorting single cells to establish the platform cell line. During the directed evolution workflow, the optional CD4 staining is compatible with live cells in that it does not require fixation or permeabilization of the cell (**Figure** 3.5**CD**). A stronger promoter (SFFV) was

used to drive a Kozak-*attP*-hCD4-*attP** cassette (**Figure** 3.5**B**), where *attP** indicates an orthogonal mutant of *attP* that differs by the central dinucleotide[235].

It was previously reported that for a tetOn CHO cell line, transiently transfected Bxb1 can mediate removal of genomic DNA sequence flanked by *attB* and *attP* sites at an efficiency of 26.4%[236]. Since *attB* and *attP* sites were placed in cis, it was most likely that the recombination efficiency was limited by the expression of transfected plasmid encoding Bxb1. In a another study, expressing Bxb1 from the genomic landing pad under a TetOn promoter can improve the recombination efficiency by 2-4 fold compared to Bxb1 expressed from a co-transfected plasmid[117]. Inspired by these observations, we envisioned that expressing Bxb1 from a genomic locus will produce more efficient recombination than co-transfecting a Bxb1-encoding plasmid. Therefore, we also designed a version that expresses Bxb1 from the genomic landing pad place holder for the third generation of platform cell lines (**Figure** 3.5**A**). In this design, a Kozak-*attP*-hCD4-IRES-Bxb1-*attP** cassette is expressed from a CMV promoter controlled by $TetO_2$ sequence. In Flp-In™ T-REx 293 cells where the tetracycline repressor protein (TetR) is stably expressed, the binding of TetR to the $TetO_2$ inhibits the transcription initiation. Following the addition of tetracycline/doxycycline, TetR dissociates from $TetO_2$, relieving the transcription inhibition. This design ensures that the Bxb1 is only expressed on demand in response to doxycycline induction, rather than expressed constitutively (**Figure** 3.5**AC**).

In the form of delivery, we decided to use pcDNA5/*FRT* vector for engineering the third-generation platform cell line with Flp-In™ T-REx 293 cell line. The Flp-In™ T-REx 293 cell line contains a single *FRT* site (**Figure** 3.2**A**). The Flp-mediated recombination

will produce stably transfected cells harbouring only a single copy of landing pad locus. In the second generation, the landing pad was delivered by lentivirus transduction, which typically produce cells with several random insertions in the genome[237] and requires tedious genotyping and phenotyping of a fairly large number of candidate clones. We envisioned that the usage of Flp-In™ T-REx 293 cell line with pcDNA5/*FRT* vector would greatly reduce the workload of clone identification.

In addition to the landing pad and donor construct, we designed an extra feature into the construct by including the genes for photo-activatable mCherry (PAmCherry) protein and the 1-10th strands of split-mNeonGreen2 ($mNG2_{1-10}$), separated by a P2A self-cleaving peptide (**Figure** 3.5**E**). PAmCherry is nonfluorescent until photoactivation with violet light, which can be employed to tag the cells of interest using a microscopy-based phenotype screening assay, the SPOTlight, technique[238] (**Figure** 3.5**E**). The $mNG2_{1-10}$ is a nonfluorescent part of mNeonGreen that can reconstitute a fluorescent mNeonGreen with the presence of the 11th strand of split-mNeonGreen2 ($mNG2_{11}$). By tagging the GOI on the donor plasmid with $mNG2_{11}$, the cells undergoing successful recombination will contain fluorescent reconstituted mNeonGreen proteins, which serves as a reporter for recombination (**Figure** 3.5**E**). Overall, the constitutively expressed nonfluorescent PAmCherry and mNeonGreen, like the hCD4, provide options for “on-demand” fluorescence readout while avoiding permanently occupying fluorescent channels. The construct that encodes PAmCherry and $mNG2_{1-10}$ is delivered by lentiviral transduction. Currently, work is in progress to create the stable cell lines for the third-generation design.

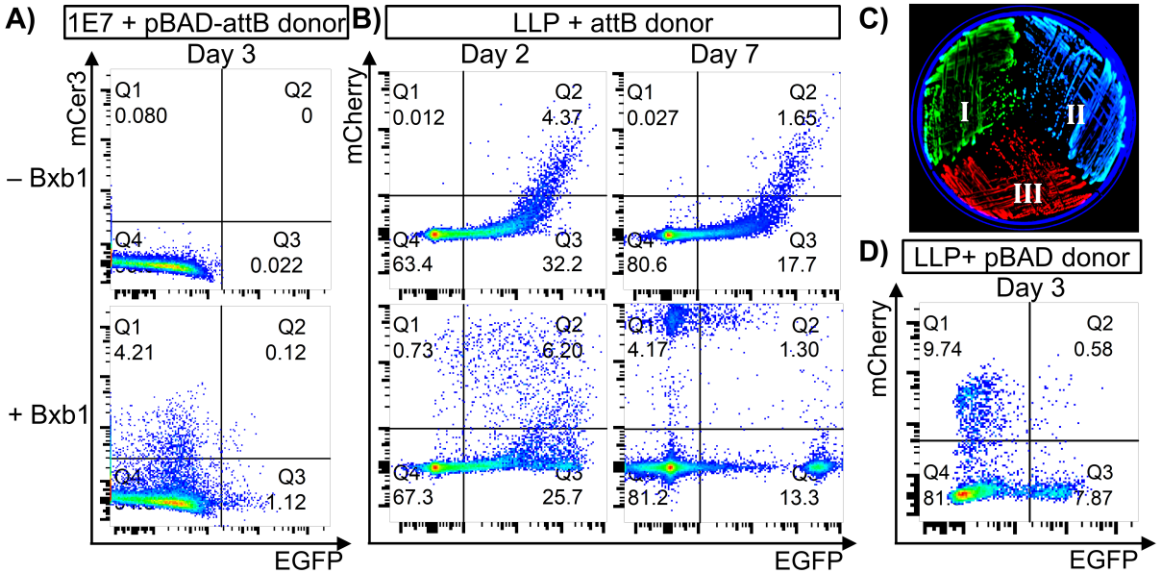

Figure 3.6 Elimination of recombination-independent expression.

A| 1E7 stable cells transfected with a mixture of pBAD-*attB*-EGFP and pBAD-*attB*-mCerulean3 plasmids, in the absence (upper) or presence (lower) of Bxb1 co-transfection. Flow cytometry analysis performed 3 days after transfection. **B|** LLP stable cells[117] transfected with a mixture of promoter-less *attB*-EGFP and *attB*-mCerulean3 plasmids, in the absence (upper) or presence (lower) of Bxb1 co-transfection. Flow cytometry analysis performed 2 days (left) or 7 days (right) after transfection. **C|** An LB-agar plate streaked with *E. coli* DH10B colonies transformed with pBAD-*attB*-EGFP (I), pBAD-*attB*-mCerulean3 (II), and pBAD-*attB*-mCherry (III). **D|** LLP stable cells[117] co-transfected with Bxb1 and a mixture of pBAD-*attB*-EGFP-*attB** and pBAD-*attB*-mCerulean3-*attB** plasmids. Flow cytometry analysis performed 2 days after transfection. Panels B and D are generated using the data recorded by F.S.-P. and colleagues from one replication, representative of three independent biological replications.

3.4 Discussion

Mammalian cell-based directed evolution represents a promising solution to engineer proteins that require mammalian cellular context for their proper function. The lack of mutual exclusiveness in mammalian cell transfection is one of the major limitations for realizing this goal. Previous efforts in mammalian cell-based library screening relied on

either the highly diluted inefficient transfection or purification of plasmids of different variants in parallel.

In this work, we demonstrated that library screening can potentially be achieved by introducing a GOI into specific genomic landing pad using site-specific recombinases (**Figure** 3.3**B-F**). Serine recombinase Bxb1 is highly specific and more efficient than the commercially available Flp-based host cell system. We also demonstrated that engineering a host cell line with a single copy of landing pad is crucial for achieving mutual exclusiveness of transfection (**Figure** 3.4**B-F**). 5-10% integration efficiency can be achieved with circular donor plasmid in an optimised system, and even higher efficiency may be expected by expressing Bxb1 from a genomic locus rather than co-transfection, and by using linear PCR product as donor DNA rather than circular plasmid (**Figure** 3.4**D,H**). Overall, the single-copy landing pad host cell line, represented by the second-generation design 1E7 clone, has achieved site-specific clonal integration of donor DNA in a “transformation-style” “one-cell-one-copy” fashion. This approach also represents a more efficient and economical solution compared to diluted transfection or parallel plasmid purification.

With a traditional *E. coli*-based directed evolution strategy, it is usually necessary to generate variant libraries in the form of circular plasmid DNA to facilitate transformation. Subcloning libraries into circular plasmid vectors not only introduces extra steps in the workflow, it is also common that the parental plasmid carryover contamination or growth biases might be introduced during subcloning and compromise the library diversity. In mammalian cell-based systems, the circular topology or elements such as replication origin and ampicillin resistance gene, is no longer required for transfection and genome

incorporation. Linear DNA libraries can be conveniently prepared by a variety of methods, such as EP-PCR, gblock, synthetic combinatorial libraries, etc. Using linear DNA library as recombination donor represents a promising and convenient solution for mammalian cell-based directed evolution.

With our second-generation platform cell line 1E7 clone, we demonstrated that linear DNA donors such as PCR product, is capable of genome incorporation and protein expression. However, incorporation of linear DNA donor seems to cause increased cell death in the second-generation design. One possible explanation is that the second-generation design uses a single *attP-attB* recombination, and recombination with a linear DNA may generate breakage on the chromosome (**Figure** 3.4**GH**). To address this issue, we used the RMCE design – two functionally orthogonal Bxb1 *attP* sites on the landing pad and two correspondingly orthogonal *attB* sites before and after the GOI on the donor DNA (**Figure** 3.5**AB**). RMCE can incorporate the linear DNA fragments by replacing the regions of DNA flanked by attachment sites, without disrupting the global chromosomal structure or introducing permanent DNA breaks.

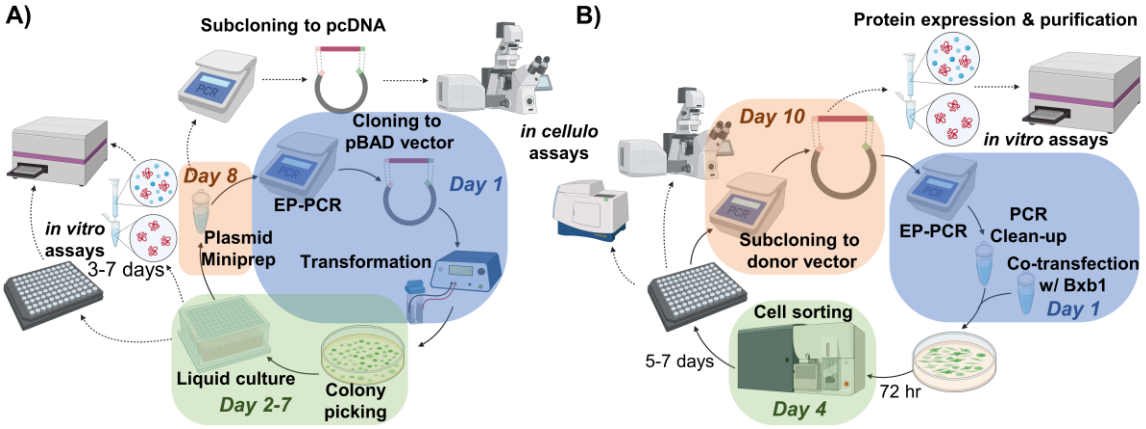

Figure 3.7 Schematic of iterative workflows for *E. coli*- and mammalian cell-based directed evolution.

A| Schematic for *E. coli*-based iterative directed evolution workflow. Assuming that mutant libraries are generated by EP-PCR, and that the colony picking, liquid culture and *in vitro* assays are performed for ~ 5 batches during each round of evolution. **B|** Schematic for mammalian cell-based iterative directed evolution workflow. Assuming donor DNA uses linear libraries generated by EP-PCR.

We envision that the mammalian cell-based library screening will facilitate an iterative directed evolution workflow (**Figure** 3.7**B**) with comparable throughput turnaround time compared to the traditional *E. coli*-based workflow (**Figure** 3.7**A**). A typical transfection with 1 × 10^6 cells with a recombination of 5% yields 50,000 cells with successful recombination, comparable to 50 agar plates of *E. coli* colonies with 1,000 each. That is, mammalian cell-based library screening using Bxb1-mediated recombination can assay a considerably diverse library with a throughput comparable to an extensive colony screening. During the *E. coli*-based directed evolution, the time-limiting factor is usually the picking, culturing, and testing the colonies. Testing enough numbers of colonies would typically take a few days. In the mammalian cell-based workflow, identification of the winning variant is performed by fluorescence-activated cell

sorting (FACS) in a rapid and automated fashion. Instead of testing the variants, expansion of single cells into a small clone becomes the time-limiting step. Overall, the turnaround time required to complete each round of evolution is expected to be comparable or slightly longer for the mammalian cell-based workflow. It is promising to further accelerate this process by employing robust single cell-level assaying techniques.

Chapter 4

4 Understanding the dissociation of PhoCl

Chapter 4 is a collaborative work and has been published as part of the article Lu, X., Wen, Y., Zhang, S., Zhang, W., Chen, Y., Shen, Y., Lemieux, M.J. and Campbell, R.E., (2021). Photocleavable proteins that undergo fast and efficient dissociation. Chemical Science, 12(28), 9658-9672. S.Z. was responsible for designing, implementing, analysing, and writing computational simulation experiments. Y.C. assisted with the computational simulation. X.L. assembled all constructs, screened new variants, performed protein characterizations and cell-based experiments as well as the manuscript composition. Y.W. was responsible for protein crystallization, X-ray data collection and structure determination. W.Z. conceived the idea of light-induced cell death experiment and performed the initial test. Y.S. contributed to manuscript edits. M.J.L. and R.E.C. supervised and acquired funding for this work.

4.1 Introduction

As a burgeoning range of biological techniques, optogenetics, which involves the use of light and genetically encodable light-sensitive proteins, enables unprecedented levels of control of numerous biological processes ranging from cellular activities to animal behaviours[239–241]. Genetically encodable light-sensitive proteins are a key component of optogenetic actuators that facilitates the light-controlled physiological processes[242]. Photocleavable protein (PhoCl) is the sole member of a distinct class of optogenetic actuators and uniquely enables irreversible optogenetic activation via a mechanism that requires cleavage of a covalent bond.

Although a wide range of photocleavable groups have been investigated[243–245], PhoCl is the first reported fully genetically encodable photocleavable moiety. The first generation PhoCl was engineered from a circularly permuted (cp) green-to-red photoconvertible fluorescent protein (FP) that, when illuminated with ~400 nm violet light, undergoes a main chain-breaking β-elimination at the green chromophore[246]. The bond cleavage produces a large N-terminal empty barrel fragment and a small C-terminal peptide fragment which spontaneously dissociate (**Figure** 4.1**AB**). Compared to the other photoreceptor proteins used in optogenetics, PhoCl has some unique beneficial properties, such as robust heterologous expression, self-sufficient chromophore formation[247], and a uniquely irreversible photocleavage mechanism[246,248,249]. Due to this property, PhoCl1 can be used as a relatively simple yet versatile tool for covalent caging of a protein of interest that can be subsequently activated irreversibly by light. Although PhoCl1 has been applied in a growing number of cell physiology and biomaterials applications[250–254], a substantial drawback has been the relatively long half-time ($t_{1/2}$) of fragment dissociation ($t_{1/2}$ ~ 500 s)[250].

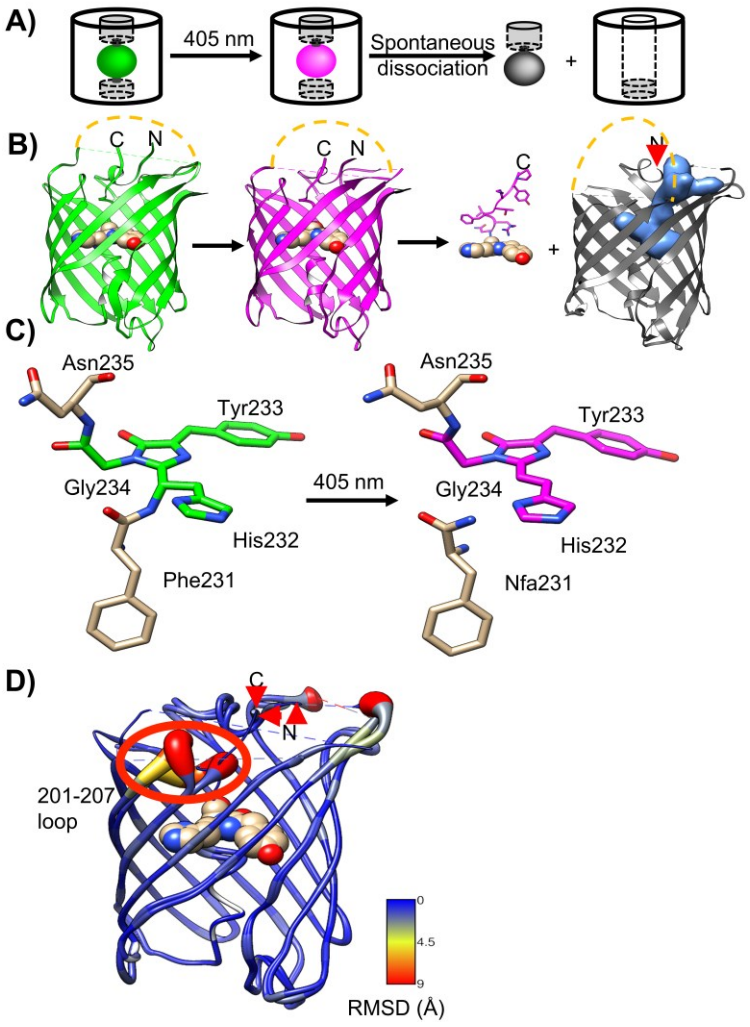

Figure 4.1 Overview of PhoCl1 structure and function.

A| Schematic of PhoCl photoconversion and dissociation. **B|** Representation of PhoCl1 structures. Green state of PhoCl1: green structure (PDB ID: 7DMX), the red state: magenta structure (PDB ID: 7DNA), and the cleaved PhoCl1 empty barrel: grey structure (PDB ID: 7DNB). The cp linkers that connect the original N- and C- termini were not determined experimentally and are schematically represented in orange dash lines. Water-filled cavity is represented as a blue space-filling volume. Dissociated peptide fragment is shown as sticks for the peptide portion and spheres for the chromophore portion. **C|** The experimentally determined structure of the PhoCl1 chromophore in the green and red states. **D|** Structure alignment of the PhoCl1 red state and empty barrel colored by root-mean-square deviation (RMSD) values. The RMSD values are indicated by scale bar and coil thickness (thicker coil indicates higher RMSD).

To increase the general utility and applicability of PhoCl, we undertook the development of second generation PhoCl variants (PhoCl2) with improved rate and efficiency of dissociation. To achieve this goal, it was crucial to obtain structural and mechanistic insights of the dissociation of PhoCl to facilitate rational design and targeted mutagenesis during directed evolution. We first solved the X-ray crystal structures of the green state, the red state, and the cleaved empty barrel of PhoCl1. Using these structures as a starting point, molecular dynamic (MD) simulations were performed to reveal the details of the dissociation mechanism. We identified a possible path for the dissociation of the C-terminal peptide and further identified the 201-207 loop on the N-terminal barrel that undergoes major conformational changes during the dissociation. Informed by these structural and mechanistic insights, my colleagues engineered the PhoCl2c variant with higher dissociation contrast ratio and the PhoCl2f variant with faster dissociation rate. Compared to the original PhoCl1, these PhoCl2 variants exhibit improved dissociation both as purified proteins and in cell-based experiments.

4.2 **Methods**

X. Lu and Y. Wen performed the expression, purification, and structure determination of the PhoCl1 protein. PhoCl1 crystal structures were deposited in protein data bank (PDB) and made available in the green state (before photoconversion, PDB ID: 7DMX), red state (after photoconversion, PDB ID: 7DNA) and empty barrel (after dissociation, PDB ID: 7DNB).

To prepare the starting structure for MD, the red state chromophore in the protonated state, designated as “RCP”, was generated in Avogadro (version 1.2.0)[255] , and parameterized using R.E.D. Server Development 2.0[256]. The circular cp linker (residues 147-172), which was not resolved in the crystal structure of the photo-converted PhoCl1, was modelled in Rosetta (version 2018.33.60351) using the kinematic closure (KIC) with fragment protocol[257]. Briefly, 10,000 models were generated and the structure with the largest negative score was selected to represent the starting conformation of the cp linker. The experimental structure of the PhoCl1 red state structure, edited to include the parameterized RCP chromophore and the cp linker, was subsequently solvated in LEaP (AmberTools19) with explicit water (TIP3P) and 1 sodium counterion in an octagon periodic boundary condition with each side no closer than 30 Å away from the protein.

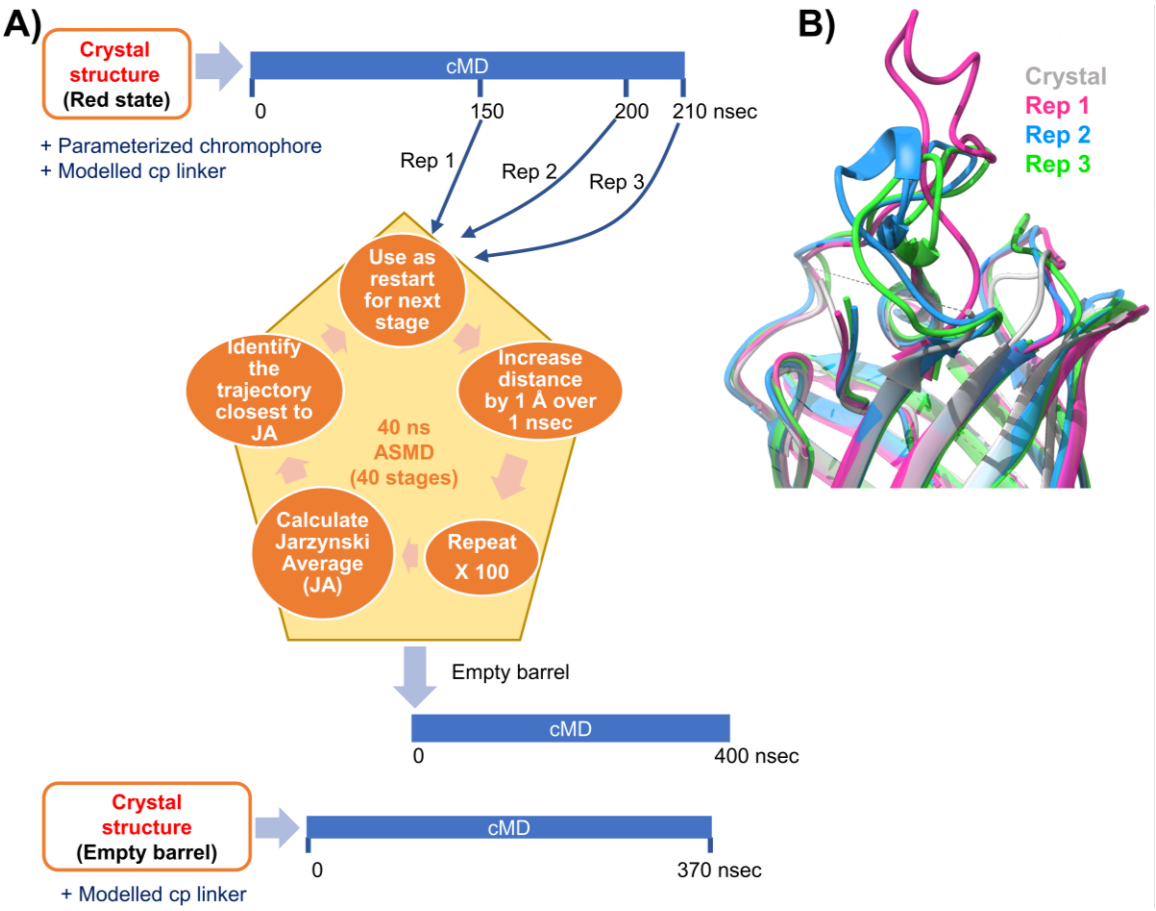

Figure 4.2 Workflow of molecular dynamic simulations.

A| Schematic of the workflow of molecular dynamic simulations. Rep, replication. **B|** Structure alignment of the crystal structure of the PhoCl1 red state and the simulated initial stages of 3 ASMD replications with different conformations of the cp linker. The crystal structure of the PhoCl1 red state is shown in silver, Rep1 is shown in magenta, Rep2 is shown in blue and Rep3 is shown in green.

All MD simulations are performed with Amber18 with GPU parallelization on Cedar cluster, Compute Canada. The system was minimized, heated to 303.15 K and equilibrated with decreasing restraints for the protein. An unconstrained production simulation (cMD) was carried out at 303.15 K for 210 ns under isothermal–isobaric (NPT)

ensemble. The restart file at the end of the cMD served as the initial coordinate of the adaptive steered MD (ASMD). The distance between the centres-of-mass (COMs) for the barrel (residues 1-231) and the C-terminal peptide (residues 232-242) were subjected to a harmonic constraint (rk2 = 18) increasing by 40 Å over 40 ns, without specifying the direction of dissociation. Such process was divided into 40 stages, where the distance between the COMs increased by 1 Å over 1 ns in each stage with 100 replications. The average of potential of mean force (PMF) was calculated by Jarzynski Equality[258] after all replicates had completed. The replicate closest to the average was chosen to represent the trajectory of that stage whose restart file served as the initial coordinate of the next stage. To investigate whether the conformation of the cp linker affects the dissociation pathway, the same ASMD simulation was repeated twice restarting from 150 ns and 200 ns of the cMD, respectively. An additional 400 ns unconstrained cMD was performed starting from the dissociated barrel structure derived from ASMD. An unconstrained cMD was also carried out for 370 ns using the empty barrel structure which was prepared using the same protocol as the cMD for the red state crystal structure. The prolonged simulation time with these two cMD allowed sufficient sampling of the loop conformation before and after the transition. The trajectories of the chosen replicate from each stage were concatenated and aligned with the program CPPTRAJ (AmberTools19)[259]. Free energy along the dissociation process was calculated using molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) method quasi-harmonic entropy calculation[260] with 20 snapshots from each stage. Root-mean-square deviation (RMSD) analysis and movie preparation were performed with VMD (version 1.9.4a31)[261] software. The scripts used for MD simulations are available for download at https://github.com/shucez/PhoCl-MD.

4.3 Results

4.3.1 Changes in crystal structure associated with photoconversion and dissociation.

To gain a better understanding of the structural changes associated with photoconversion and peptide dissociation, and to aid the protein engineering efforts for next generation of PhoCl, my colleagues X. Lu and Y. Wen determined the X-ray crystal structures of the PhoCl1 green state (2.1 Å resolution, PDB ID: 7DMX), red state (2.3 Å resolution, PDB ID: 7DNA), and the empty barrel fragment (2.8 Å resolution, PDB ID: 7DNB) (**Figure** 4.1**B**).

As expected on the basis of the known photoconversion mechanism[246,248,249], the red state of PhoCl1 was cleaved at the amide bond connecting Phe231 to His232 (**Figure** 4.1**C**). Other than the changes in the vicinity of Phe231 and the chromophore, the overall structures of the green and red states are essentially identical. The structure of the PhoCl1 empty barrel revealed substantial conformational changes relative to the red state, as expected given the complete absence of the C-terminal fragment. **Figure** 4.1**D** visualizes the conformational changes between PhoCl1 red state and the empty barrel by colouring the cartoon coil alignment using the RMSD values between the two structures and adjusting the coil diameter such that a thicker coil indicates higher RMSD. This representation revealed that the most substantial, and observable, conformational changes occur in the 201-207 loop. Specifically, in the empty barrel structure, the 201-207 loop region is “folded in” towards the center of the barrel, relative to its position in the red state structure. We reasoned that the conformational change of 201 -207 loop likely served

to accommodate the absence of C-terminal fragment. In addition, we also speculated that the 201-207 loop may play a role in the dissociation of the C-terminal fragment.

4.3.2 Simulating PhoCl1 dissociation by adaptive steered molecular dynamics (ASMD)

To test this hypothesis and gain further insight into the dissociation process of PhoCl1 at the molecular level, we devised an adaptive steered molecular dynamics (ASMD) scheme (**Figure** 4.2**A**) to investigate the dissociation pathway. The cp linker, which joins the original N- and C-termini of mMaple, was not solved in the crystal structures. However, the cp linker is positioned on the same side of the barrel as the C-terminal fragment and is likely to interact with the fragment during its dissociation. To minimize the bias from the starting conformation of the cp linker on the dissociation pathway, we obtained three structures with distinct conformations of the modelled cp linker (**Figure** 4.2**B**) that later served as the initial coordinates of the ASMD. These three starting structures were taken from a 210 ns unconstrained canonical molecular dynamics (cMD) trajectory, during which no spontaneous dissociation occurred and most of the movements take place at the cp linker while the rest of the protein remains relatively still, including the 201-207 loop. During the ASMD, the distance between the centres-of-mass of the N-terminal barrel and the C-terminal peptide were increased by 1 Å over 1 ns for each of 40 iterative stages, to give a total final displacement of 40 Å. Each stage was repeated 100 times, from which an averaged potential mean of force was calculated and the replicate closest to the average was taken as the starting point for the next stage. We simulated

the dissociation through 40 such stages. Notably, although this sampling is termed as "steered" MD, the "steering" simply refers to the constraint imposed on the distance between the centres-of-mass of the two fragments; the route or direction of dissociation was never specified and was only the result of the MD simulation (**Figure** 4.2**A**). In all 3 distinct ASMD replications, the C-terminal peptide dissociated through the space between the flexible cp linker (residue 147-172) and the 201-207 loop* (**Figure** 4.3**A,B**).

* also see the supplementary movie from the publication[262]:
https://www.rsc.org/suppdata/d1/sc/d1sc01059j/d1sc01059j1.mp4

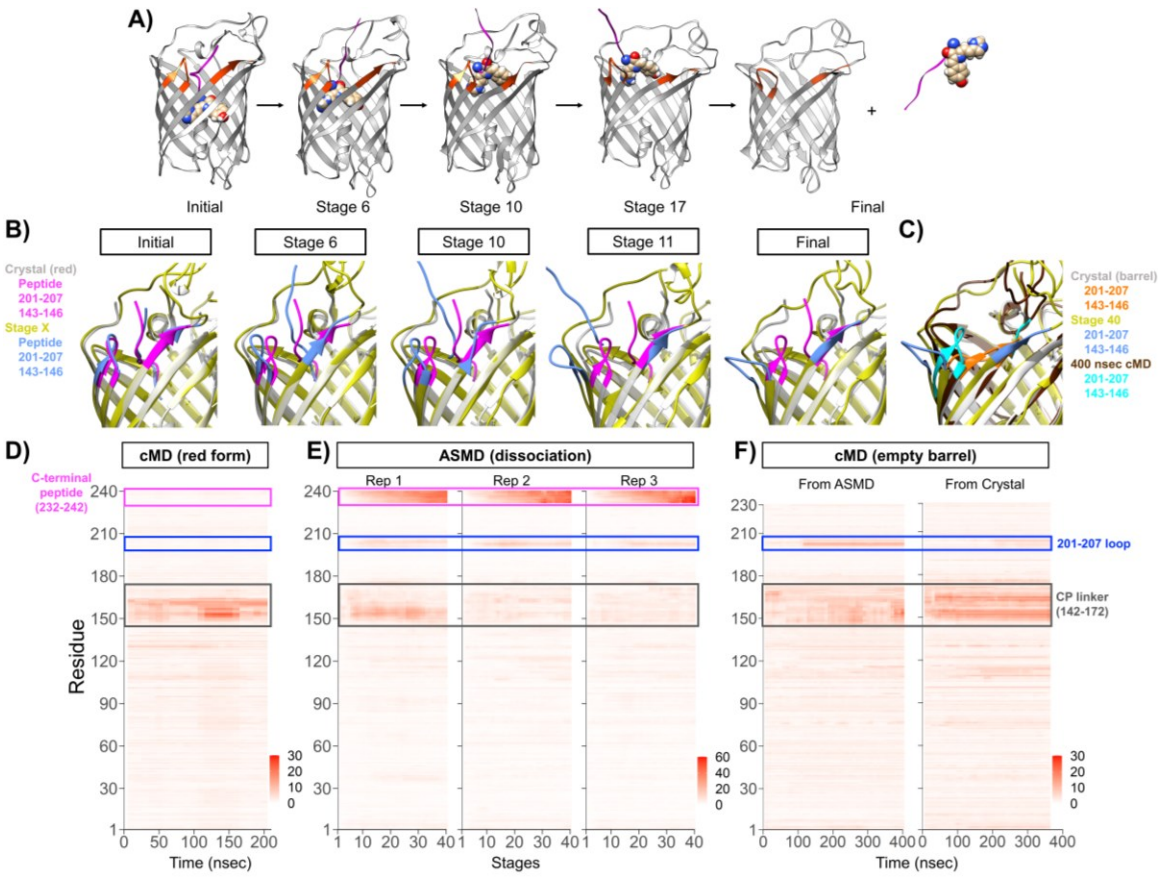

Figure 4.3 Simulating the dissociation process of PhoCl1.

A| Representation of the simulated dissociation process. Photo-converted PhoCl1 is shown in grey, dissociated peptide fragment is shown as magenta ribbon for the peptide portion and spheres for the chromophore. Residues within and near the 201-207 loop are highlighted in orange. **B|** Structure alignment of the crystal structure of the PhoCl1 red state and the simulated structure from different stages of ASMD replication 3 (Rep 3). Crystal structure of the PhoCl1 red state is shown in silver, with residues within and near the 201-207 loop highlighted in magenta. Simulated structures from different stages are shown in yellow, with residues within and near the 201-207 loop highlighted in blue. **C|** Structure alignment of the crystal structure of the PhoCl1 empty barrel, the simulated structure from final stage of ASMD Rep3, and the structure of the empty barrel after 400 ns cMD succeeding the ASMD. Crystal structure of the PhoCl1 empty barrel is shown in silver, with residues within and near the 201-207 loop highlighted in orange. ASMD final stage structure is shown in yellow, with residues within and near the 201-207 loop highlighted in blue. The structure at 400 ns from the cMD following the ASMD is shown in brown, with residues within and near the 201-207 loop highlighted in cyan. **D|** Heatmaps of the residue-wise average RMSD of 10 ns windows during the cMD from the red state crystal structure. The RMSD values were colored according to the scale bar. **E|** Heatmaps of the residue-wise average RMSD of each stage for the 3 replications

of ASMD. **F|** Heatmaps of the residue-wise average RMSD of 10 ns windows during the cMD from (left) the barrel structure derived from the ASMD simulation and (right) the crystal structure of the empty barrel. The RMSD values were calculated with each of their first frame as reference and colored according to the scale bar. The C-terminal peptide is enclosed in a magenta box, the 201-207 loop is enclosed in a blue box and the cp linker is enclosed in a grey box. Rep, replication.

Analysis of the RMSD values for all residues over time, with the initial stage as the reference, revealed that the C-terminal peptide showed the highest RMSD due to its physical displacement (**Figure** 4.3**D**). The region with the second highest RMSD was the 201-207 loop, consistent with conformational changes observed in the experimental crystal structures. Compared to the crystal structure of the empty barrel, inspection of the dynamic conformational changes during stages of dissociation reveals that the 201-207 loop "flips" out of the barrel rather than "folds in" towards the barrel, presumably to minimize steric interactions with the C-terminal peptide as it dissociates (**Figure** 4.3**C**). During the additional 400 ns of unconstrained cMD following dissociation by AMSD, this loop repositioned to a conformation similar to the one observed in the crystal structure of the red state (**Figure** 4.3**C,F**, **Figure** 4.4**I-K**). The repositioning of the loop 201-207 was also observed in another 370 ns unconstrained cMD starting from the crystal structure of the empty barrel[†] (**Figure** 4.3**F**). It is possible that crystal packing interactions help to stabilize this relatively flexible loop in the "folded in" conformation observed in the empty barrel structure. Dynamic conformational changes of the 201-207 loop were not observed

[†] also see the supplementary movie from the publication[262]:
https://www.rsc.org/suppdata/d1/sc/d1sc01059j/d1sc01059j2.mp4

during 210 ns of unrestrained cMD of the red state (**Figure** 4.3**D**). Overall, these results demonstrate that the 201-207 loop undergoes a dissociation-dependent conformational change, rather than being intrinsically flexible, providing support for the conclusion that the 201-207 loop would be the key region of interest in our subsequent engineering.

X. Lu continued to perform directed evolution of PhoCl1 through a combination of targeted and random mutagenesis, and discovered four key mutations found within or near 201-207 loop that had a substantial and favorable effect on the PhoCl dissociation rate: Lys202Leu and Ile207Ser in the loop and the combination of Val143Met and Ala144Val on the adjacent β-strand 11.

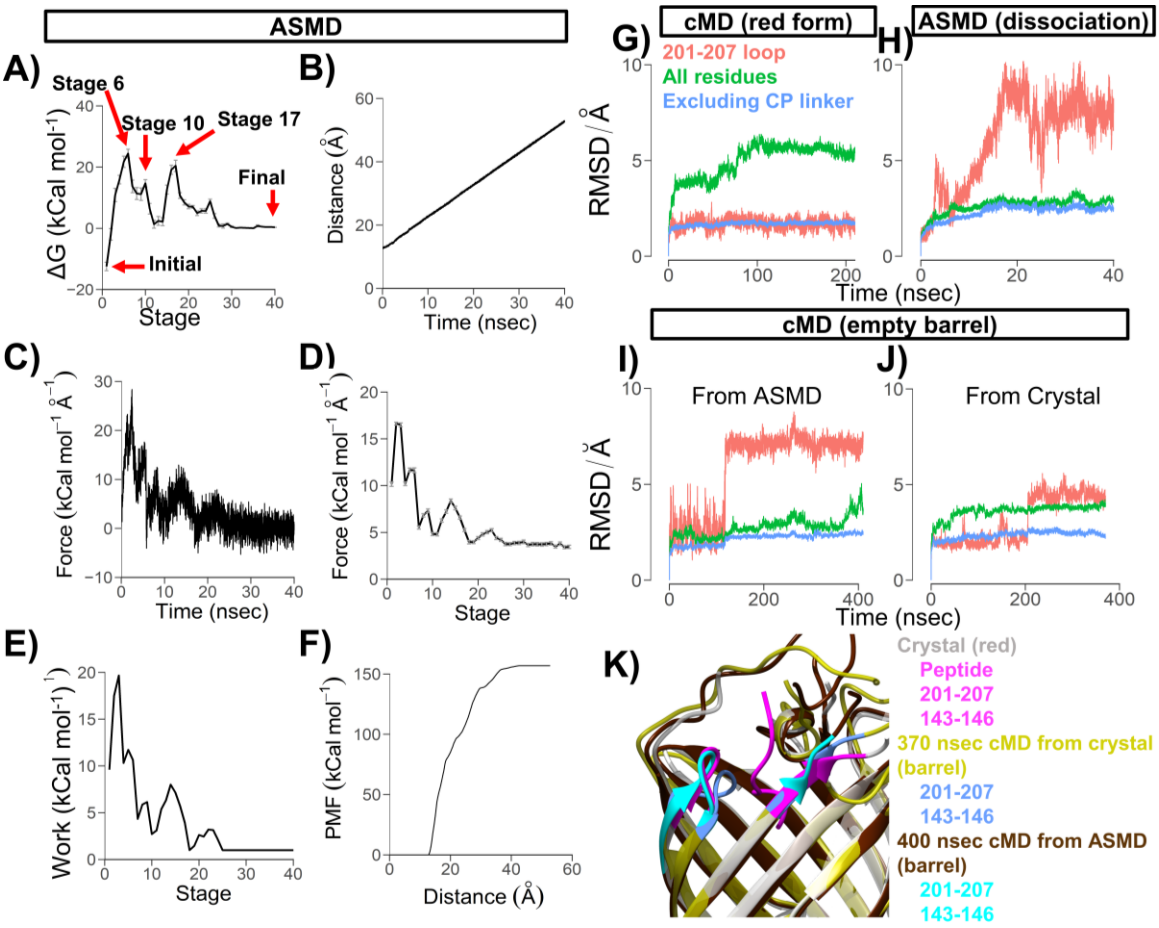

Figure 4.4 Additional information on the molecular dynamic simulation of the dissociation process.

A| Gibbs free energy of activation over stages from Rep3 with the final stage as reference (ΔG = 0 kCal/mol). Values are means ± SEM (n = 20 snapshots per stage). The stages represented in **Figure** 4.3**AB** are indicated by red arrows. **B|** The distance between the centres-of-mass of the C-terminal peptide and the N-terminal barrel increased by 40 Å over 40 ns imposed by a harmonic restraint. **C|** The force of the harmonic restraint over time to maintain a constant speed of movement. The curve is smoothened by averaging the adjacent 10 snapshots. **D|** The average force of the harmonic restraint in each stage by averaging all snapshots within a stage, and **E|** the work done by such force in each stage. **F|** The cumulative potential of mean force (PMF) throughout the ASMD. **G-J|** The RMSD over time for **G|** the cMD with the red state crystal structure, **H|** the ASMD Rep 3, **I|** the empty barrel following ASMD Rep3, and **J|** the empty barrel starting from the empty barrel crystal structure. Green, RMSD including all residues; blue, all residues other than cp linker; pink; the 201-207 loop. **K|** A dissociation-dependent conformational change of the 201-207 loop takes place during ASMD followed by its repositioning during the cMD that follows the ASMD.

4.4 Discussion

To summarize, we have reported the X-ray crystal structures of PhoCl1 and MD simulations of the dissociation process which contribute to an improved understanding of the molecular mechanisms of the dissociation pathway. The slow kinetics of PhoCl1 dissociation indicates a high energy barrier during the dissociation process, which is unlikely to be sampled by unconstrained cMD. Indeed, the 210 ns cMD simulation starting with PhoCl1 in the red state did not show spontaneous dissociation. In this work, we employed ASMD for enhanced sampling of the dissociation pathway. Our ASMD scheme divided the 40 Å total displacement into 40 stages with each stage increasing 1 Å over 1 ns. Each stage was simulated in 100 replicates, and the replicate closest to the PMF calculated by Jarzynski Equality represented the outcome of this stage and served as the start point for all replicates in the next stage. This ASMD scheme effectively samples the possible dissociation pathways, without relying on the human assumptions of the dissociation route. Firstly, all three ASMD simulations using distinct cp linker conformations as starting coordinates showed that the C-terminal fragment exited the barrel following the same route – the space between the flexible cp linker (residue 147-172) and the 201-207 loop. Secondly, the 201-207 loop exhibited the largest RMSD during ASMD, other than C-terminal fragment and the cp linker. This “flipping out” movement of the 201-207 loop only occurred during ASMD, but not in the cMD of PhoCl1 in the red state, indicating that the 201-207 loop exhibited specific, dissociation-related conformational change. Thirdly and experimentally, many mutations associated with improved kinetics of dissociation were mapped on or near the 201-207 loop on the tertiary structure. The results above indicated that the exit route observed in the ASMD was likely

to be the *bona fide* dissociation pathway of PhoCl1. Interestingly, both cMD simulations of the empty barrel, whether following ASMD with a “flipped”-out conformation, or using the crystal structure with a “folded-in” conformation, exhibited a reposition of the 201-207 loop to a conformation similar as observed in the red state. This may suggest that the “folded-in” conformation was adopted in the empty barrel to satisfy certain interactions required for crystal packing.

The structural and mechanistic insights provided by MD simulations enabled my colleagues to engineer the second generation of PhoCl variants (PhoCl2) with improved dissociation efficiency (PhoCl2c) and kinetics (PhoCl2f). The mechanisms by which the key favorable mutations are exerting their observed influence on peptide dissociation are unclear. Two possible mechanisms are: 1) Decreasing the energy barrier for peptide dissociation leading to faster dissociation kinetics; and 2) Decreasing stabilization of intact PhoCl, or increasing stabilization of the empty barrel, which would likely have little effect on the kinetics, but may lead to a greater overall efficiency of dissociation. To further engineer the next generation of PhoCl, based on the first proposed mechanism, mutations may be introduced to facilitate peptide dissociation by reducing steric hindrance and/or increasing the conformational flexibility of the barrel in the vicinity of the dissociable peptide; based on the second proposed mechanism, mutations may be introduced to destabilize the red state and shift the thermodynamic equilibrium for dissociation further towards the dissociated peptide and empty barrel state, leading to the more efficient dissociation. We anticipate further computational studies to model the newly discovered PhoCl variants, simulate with more comprehensive sampling, and provide more detailed mechanism for the PhoCl dissociation.

Chapter 5

5 Conclusion and future directions

5.1 Summary of the thesis

Optogenetic tools — genetically encoded proteins that enable optical visualization or manipulation of physiological states — have revolutionized all areas of cell biology. Optogenetic tools have attracted wide attention of protein engineering that optimises the photoreceptive domains for manipulation in mammalian context. So far, directed evolution of optogenetic tools still heavily relied on prokaryotic or lower eukaryotic systems, such as *E. coli* and yeast. These systems may lack the mammalian signalling pathways and may have very different mechanisms for protein folding, trafficking, and post-translational modification. Despite the success with the *E. coli* system in evolving GECIs[212,263–265] and other fluorescent sensors[266–268] at Campbell lab, these non-mammalian systems have remained unsatisfactory and inappropriate for other optogenetic tools, such as GEVI[269–271], GECA[187,197,272], and other tools that are membrane proteins or requires mammalian signalling pathways for proper function.

In **Chapter 1**, I reviewed the methods in mammalian cell-based library screening systems. Libraries that are generated outside of living cells can be delivered via nucleic acid transfection, viral transduction, and recombinase-mediated genome integration. These libraries can be prepared and delivered either individually (arrayed library) or as a mixture of library members (pooled library). Since mammalian cells do not natively have an *E. coli*-like clonal expression mechanism, pooled library needs to be delivered either

as a highly diluted pool of plasmid, or using a site-directed recombinase to integrate one single copy of DNA into a genomic landing pad locus. On the other hand, variants can be generated within the mammalian cells (*in situ* mutagenesis), which uses similar mechanisms of antibody diversification in lymphocytes and proves useful for engineering humanised antibody. *In situ* mutagenesis also facilitates virus-mediated continuous evolution in mammalian cells when the phenotype is coupled with the virus proliferation.

In **Chapter 2**, I described my original work of developing an optogenetic tool for light-induced ER Ca^{2+} release, named OptoCR. OptoCR is a binary system consisting of 1) an Orai channel (hOrai1 or dOrai) tagged with a C-terminal dilysine motif (KKXX) that reroutes the channel to the ER-membrane, and 2) a GECA (OptoCRAC or dmBACCS2) that activates the ER-localised Orai-KKXX. We demonstrated that OptoCR releases ER Ca^{2+} in response to blue light stimulation, resulting in reversible $[Ca^{2+}]_{cyt}$ increase. The effect of $[Ca^{2+}]_{ER}$ was not detectible, possibly because of the mild outflux of ER Ca^{2+} and the buffering of $[Ca^{2+}]_{ER}$. Nonetheless, in cells overexpressing RyR2-T4158P and OptoCR, blue light stimulation induced oscillation of $[Ca^{2+}]_{cyt}$, indicating that the Ca2+ release from OptoCR could activate RyR-T4158P and induce CICR, which plays an important role in the excitability of myocytes and other excitable cells. On the other hand, the GECA component can be tagged with a localisation motif for spatial targeting to specific subcellular locations. For example, hOrai1-KKXX and Lyn-OptoCRAC can induce ER Ca^{2+} release at the ER-PM junction. So far, OptoCR is the first non-opsin, second messenger-independent system for optogenetic ER Ca^{2+} release. We anticipate that OptoCR will prove a useful and physiological tool in calcium signalling research as an alternative to traditional pharmacological treatments.

In **Chapter 3**, I described the development of a novel mammalian cell-based library screening system using Bxb1 recombinase-mediated genomic integration for directed evolution. This system is composed of 1) an engineered mammalian cell line harbouring a single landing pad locus in the genome, and 2) a donor plasmid that serves as the vector for variant delivery. Using a transient transfection system in HEK-293 FT cells, we demonstrated that the donor plasmid did not drive the expression of GOI *per se*, and that the expression of GOI required the presence of all three components: the landing pad, the donor plasmid, and the expression of Bxb1 recombinase. We engineered a stable cell line by lentivirus transduction and identified the 1E7 strain to contain a single landing pad locus in its genome. When transfected with pooled donor plasmid molecules encoding different FPs, these FPs were mutually exclusively expressed in each cell, with overall transfection efficiency ~ 5%. In this work, the donor plasmid vector was engineered from the arabinose-inducible *E. coli* expression vector pBAD and could be used directly for prokaryotic expression without subcloning. Furthermore, the pBAD based vector was shown to be compatible with Matreyek lab's LLP landing pad cell line and eliminated the recombination-independent transient expression. Overall, we provided an improved solution specifically designed for protein engineers that enabled efficient recombination-dependent genome integration and clonal expression. We envision that this system will serve as a useful tool for protein engineering and enable fast iterative directed evolution workflow in mammalian cells.

In **Chapter 4**, I described my computational simulation of the dissociation of N- and C-terminal fragments of photoconverted PhoCl1 protein. Starting with the crystal structures of PhoCl1 solved by my colleagues, we employed ASMD as the enhanced sampling methods to simulate the dissociation pathway with only an increasing distance

constraint, but not the direction constraint. All three replications of ASMD simulation identified a common route of dissociation where the dissociable peptide exited the N-terminal barrel through the space between the flexible cp linker and the 201-207 loop. In addition, we identified a dissociation-dependent “flipping-out” movement of the 201-207 loop, which otherwise adopted a relatively stable conformation before and after the dissociation, similar to the conformation observed in the crystal structure of the red state. Consistent with the MD-derived structural and mechanistic insights of PhoCl dissociation, directed evolution also identified mutations mapped within or in vicinity to the 201-207 loop that are associated with improved dissociation kinetics and/or kinetics.

5.2 Future directions

The OptoCR system in **Chapter 2** demonstrated the feasibility of re-targeting PM-localised Ca^{2+} channel to ER membrane for optogenetic ER Ca^{2+} release. The OptoCR system is also by far the first optogenetic ER Ca^{2+} release without using opsin or second messenger. Nonetheless, there is plenty of room for improvement for OptoCR: The most prominent issue of the OptoCR system is that weak ER Ca^{2+} release. OptoCR system that can mediate stronger ER Ca^{2+} release should be developed, with which the strength could be conveniently modulated by light stimulation intensity or frequency. To enhance the strength of OptoCR, several possible strategies may be pursued: 1) Other Ca^{2+} channel with larger unit conductance may be tested using similar rerouting strategy. 2) Orai channel may be engineered to improve the folding when rerouted to ER (e.g., for Orai3 and LOCa3), and improve the current at the membrane potential similar to ER membrane

(~ 0 mV), and to investigate the possible effect of the altered glycosylation state. 3) Novel light gating mechanisms for Orai and other channels may be explored. In particular, the recently reported cpLOV2[196] domain that facilitates N-terminal conformational modulation may be useful for gating the Orai channels, whose first transmembrane helix (TM1) makes up the channel pore. On the other hand, in addition to the experiments performed in HEK-293 and HeLa cells, we anticipate exciting applications of the OptoCR system in other physiological models, especially in excitable cells where CICR may be induced by OptoCR. We also anticipate that OptoCR may prove useful for studies on Ca^{2+} signalling at ER-organelle and ER-PM junctions.

The Bxb1-based mammalian cell library screening system described in **Chapter 3** achieved single cell-level variant separation and clonal expression. The design of pBAD-based donor plasmid not only eliminated the recombination-independent transient expression in previously reported LLP cell line from Matreyek lab, but also provided a convenient shuttle for seamless mammalian and *E. coli* expression without the need of subcloning. Due to the single *attP* × *attB* recombination design, the circular plasmid topology is required for post-recombination genome stability. One immediate improvement, as discussed in **Section 3.3.4**, is to adopt the RMCE design where two orthogonal pairs of *attP* × *attB* recombination can swap in a GOI cassette from donor DNA of both linear and circular topology. The dual-recombination RMCE design will enable screening plus the option of a recombination marker using the split mNeonGreen2 system or the option of optical tagging using PAmCherry for SPOTlight system. Nonetheless, the Bxb1 clonal expression system. inevitably requires handling of single cells. Developing robust and affordable sequence recovery methods of the cells of interest, either via single-cell sequencing, or long-read sequencing of pooled cells, would also be highly beneficial

to mammalian cell-based screening. Overall, we anticipate a mammalian cell-based library screening will facilitate an iterative directed evolution workflow (**Figure** 3.7**B**) with comparable throughput turnaround time compared to the traditional *E. coli*-based workflow.

In **Chapter 4**, the molecular simulations identified a possible dissociation pathway through which the C-terminal dissociable peptide exits the N-terminal barrel. To sample the dissociation, ASMD was used to artificially increase the distance between the centres-of-mass of the N- and C-terminal fragments. Among 100 replicates in each stage, only the replicate closest to the average PMF was used as the starting coordinate for the next stage. Therefore, ASMD inevitably under-sampled the coordinate and conformation landscape and may be produce artifact. Nonetheless, the structural and mechanistic insights provided by ASMD, especially the dissociation-dependent conformational change, have led to successful engineering of PhoCl2 variants where the emerging mutations are mapped within or in proximity to the 201-207 loop. To understand the dissociation more thoroughly and to pinpoint specific interactions that contributed to the energy barriers, it would be highly beneficial to model the PhoCl2 variants, simulate their dissociation and benchmark the experimental data. For this goal, other enhanced sampling methods, such as umbrella sampling or replica exchange. We anticipate that PhoCl2 variants will prove useful in an increased range of optogenetic applications due to their faster and more efficient photo-induced cleavage and dissociation.

5.3 Other research projects during my PhD studies

5.3.1 Development of a SARS-CoV-2 main protease sensor

Section 5.3.1 was an original work of Shuce Zhang. S.Z. was responsible for formed the idea, designed the study, established the methodology, constructed all plasmids, and performed all live cell imaging experiments. Yoojin Choi assisted with E. coli expression and in vitro assays under the direct supervision of S.Z.

COVID-19, the pandemic disease caused by a novel coronavirus, has infected millions of people and caused over a hundred thousand deaths worldwide. A critical event of the viral infection and replication cycle is the cleavage of a large immature polyprotein into many smaller functional proteins by the proteases encoded in the virus genome. These proteases are essential for the replication and propagation of the virus, making it an attractive target for antiviral therapies. Indeed, inhibition of the main protease (M^{pro} or $3CL^{pro}$) has been a successful strategy in treating both COVID-19 (e.g., Oral administration of nirmatrelvir/ritonavir, i.e., Paxlovid®[273]) and feline coronavirus diseasese[274]. To identify small molecules as potent protease inhibitors against coronavirus, it is essential to have a sensitive and robust assay that is easy to implement and scalable for high-throughput screening. Currently, most assays for screening coronavirus protease inhibitors use purified protease and chemically synthetic probes and are performed in vitro, which, however, can be very artificial due to the lack of cellular relevance. We have hereby designed a cell-based protease assay system for functionally testing inhibition of protease activity.

At the core of our cell-based assay system will be a fluorescent protein-based biosensor for the protease. Briefly, the biosensor contains two fluorescent proteins, mCerulean3 and cpVenus, connected by a linker harbouring the substrate sequence of M^{pro}. An active protease will cleave the substrate sequence connecting the FRET pair, resulting in a lower FRET efficiency (**Figure 1 AB**). Indeed, a substantial decrease was observed with a fluorescence microscope when the reporter is co-expressed with M^{pro} (**Figure 1 CD**). Such a FRET reporter is not only compatible with microscopy, but also suitable for quantitative measurement by flow cytometry. Distinct distributions of CFP-normalized FRET were observed for fully active (wild type), catalytically dead (C145A) and pharmacologically inhibited (GC373 and GC376) M^{pro} (**Figure 1 EF**), suggesting a partial inhibition of M^{pro} by GC373 and GC376 at 20 μM.

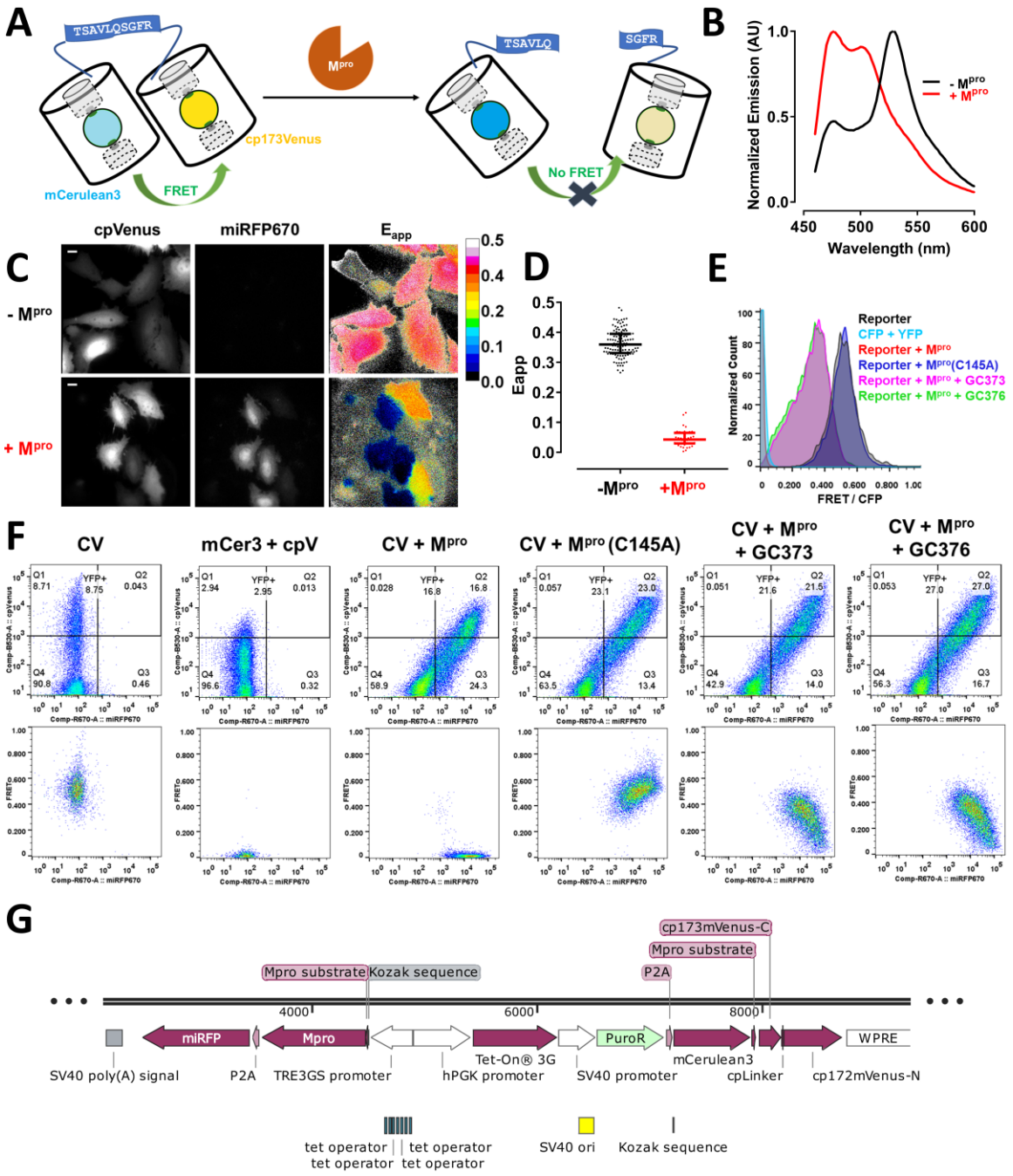

Figure 5.1 Genetically encoded FRET reporter for coronavirus Mpro.

A| Schematic of the mechanism of the FRET reporter. **B|** Fluorescence emission spectrum before (-M^{pro}) and after (+M^{pro}) treating with the protease, excited at 430nm. **C|** Transiently expression of FRET reporter with (upper panel) or without (lower panel) co-transfection with Mpro-P2A-miRFP670. The cpVenus channel denotes the expression of the reporter, and the

miRFP670 channel that of M^{pro}. E_{app}, 3-channel corrected apparent FRET efficiency. **D|** Comparison of reporter E_{app} in the presence and absence of Mpro. **E|** CFP-normalized FRET signal measured by flow cytometry. **F|** Scatter plots of flow cytometer experiments. Upper panel, cpVenus versus miRFP670 fluorescence. Lower panel, ratio of FRET / CFP versus miRFP670 fluorescence. **G|** TetOne lentiviral construct with constitutively expression of reporter and inducible expression of M^{pro} for generating stable cell lines for drug screening assays.

We further generated a tri-cistronic lentiviral construct with constitutively expression of reporter and doxycycline-inducible expression of M^{pro}. This construct allows us to conveniently engineer mammalian cell lines from different backgrounds to stably express this system. Compounds from large chemical libraries can be applied to the cells (+ doxycycline) and the effect of protease inhibition can be determined by examining the change in the fluorescence emission profile. In this way, the inhibitors can be assayed in living cells under conditions that are more physiologically relevant than tests with purified proteins. The cell line can be maintained and expanded rapidly to adapt to a large scale of screening experiment, without needing to purify the protease or synthesize peptide probes.

5.3.2 **Emission ratiometric imaging of REX-GECO1**

Section 5.3.2 is a collaborative work and has been published as part of the article Tang, L., Zhang, S., Zhao, Y., Rozanov, N.D., Zhu, L., Wu, J., Campbell, R.E. and Fang, C., (2021). Switching between ultrafast pathways enables a green-red emission ratiometric fluorescent-protein-based Ca^{2+} biosensor. International journal of molecular sciences, 22(1), 445. S.Z. was responsible for performing, analysing, and writing of all live cell imaging experiments. L.T., N.D.R., L.Z. and J.W. performed the ultra-fast spectroscopy studies. C.F. and R.E.C supervised and acquired funding for this work.

Ratiometric indicators with long emission wavelengths are highly preferred in modern bioimaging and life sciences. My collaborators, Dr. Chong Fang and colleagues at Oregon State University elucidated the working mechanism of a standalone red fluorescent protein (FP)-based Ca^{2+} biosensor, REX-GECO1[275], using a series of spectroscopic and computational methods. Upon 480 nm photoexcitation, the Ca^{2+}-free biosensor chromophore becomes trapped in an excited dark state. Binding with Ca^{2+} switches the route to ultrafast excited-state proton transfer through a short hydrogen bond to an adjacent Glu80 residue, which is key for the biosensor's functionality. Inspired by the 2D-fluorescence map, we demonstrated the application of REX-GECO1 for Ca^{2+} imaging for the first time with a red/green emission ratio change ($\Delta R/R_0$) of ~300%, in the ionomycin-treated human HeLa cells outperforming many FRET- and single FP-based indicators. These spectroscopy-driven discoveries enable targeted design for the next-generation biosensors with larger dynamic range and longer emission wavelengths.

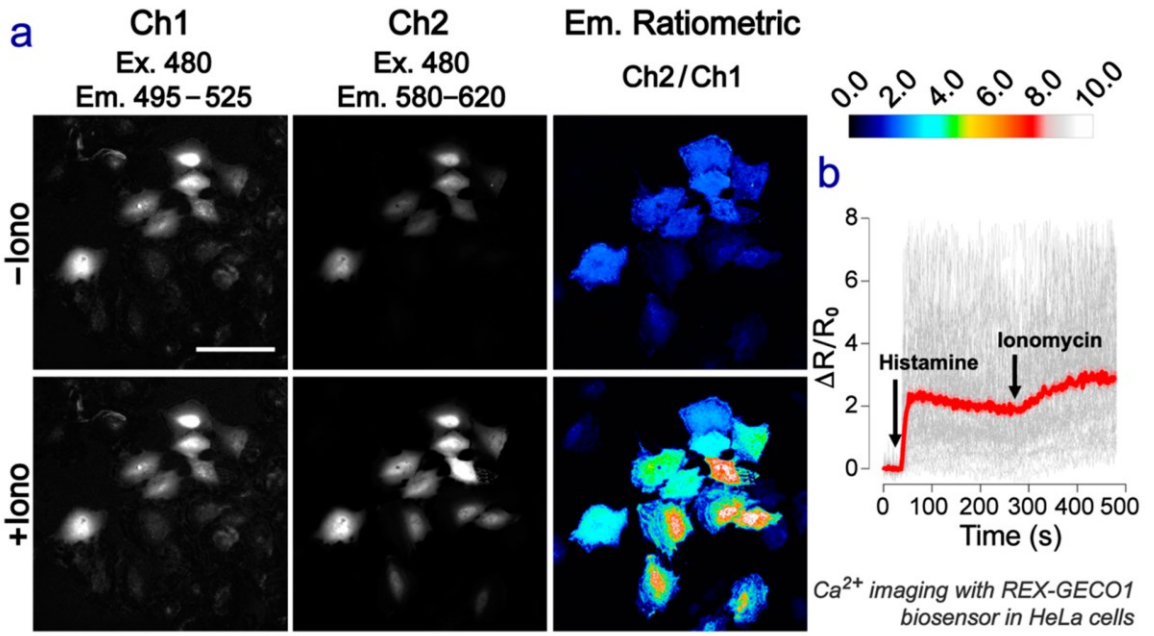

Figure 5.2 Emission-ratiometric imaging of the REX-GECO1 biosensor.

A| Confocal images of HeLa cells expressing REX-GECO1 before (low Ca^{2+}) and after (high Ca^{2+}) ionomycin treatment. Scale bar = 100 µm. **B|** Time-lapse Ca^{2+} imaging shown by the red/green fluorescence intensity ratio change ($\Delta R/R_0$) after 480 nm excitation. Gray lines: single-cell traces. Red line with pink shade: mean ± s.e.m.

To demonstrate the biosensor capability, we expressed REX-GECO1 (K_d for Ca^{2+} of 240 nM) in cultured human cervical cancer cells (HeLa) and imaged the histamine- and ionomycin-induced increases in Ca^{2+} concentration in the cytosol. As expected on the basis of our spectroscopic measurements with purified proteins, excitation at ~480 nm resulted in green and red emission in the Ca^{2+}-free state, which becomes dominantly red in the Ca^{2+}-bound state‡ (**Figure** 5.2**a**). After background subtraction, the average $\Delta R/R_0$

‡ Also see Video S1 from the original publication[265].

is ~300% with sufficient signal-to-noise ratios in both green and red channels (70 cells in total, *n*=3), where R is the red/green emission ratio (**Figure** 5.2**b**, the red plateau reaching ~3 against the vertical axis at later time). The dynamic range of REX-GECO1 as an emission-ratiometric indicator (**Figure** 5.2) outperforms many FRET-based indicators[276,277] and commonly used intensiometric indicators such as jRGECO1a[210] and K-GECO1[278]. Such dynamic range also outperforms that derived from the blue shift of the emission peak (F_{590} / F_{610}, $\Delta R/R_0$ of 34.4%) when excited at 550 nm.

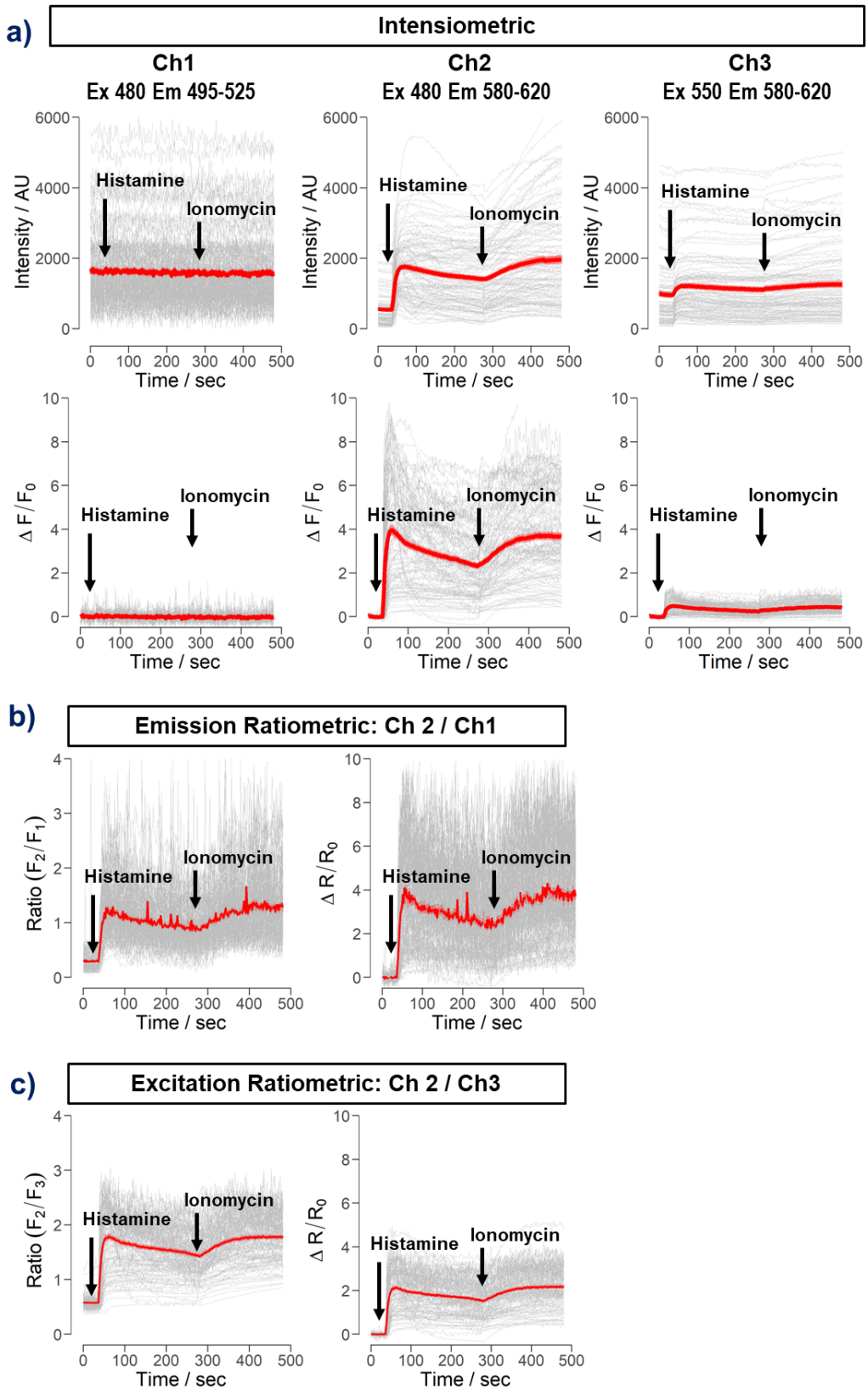
a)
Intensiometric
Ch1
Ex 480 Em 495-525
Ch2
Ex 480 Em 580-620
Ch3
Ex 550 Em 580-620
Intensity / AU
Histamine
Ionomycin
Time / sec
Δ F/F0
b)
Emission Ratiometric: Ch 2 / Ch1
Ratio (F2/F1)
Δ R/R0
c)
Excitation Ratiometric: Ch 2 / Ch3
Ratio (F2/F3)

Figure 5.3 Raw experimental data and change in fluorescence signal at each individual imaging channel.

A| The intensiometric signal intensities (upper panels) and the associated fluorescence intensity changes ($\Delta F/F_0$) are displayed for Ch1, Ch2, and Ch3, respectively. Traces for individual cells (gray) lead to mean value for all the ROIs (red) and the standard error of the mean (s.e.m., pink shade around the red curve). **B|** Fluorescence signal intensity ratios (left) and the red/green emission ratio changes (right) for the emission-ratiometric method, and **C|** Fluorescence signal intensity ratios (left) and the red/green emission ratio changes (right) excitation-ratiometric method demonstrate the ratiometric sensing capability of REX-GECO1 in HeLa cells.

Although the Ca^{2+}-dependent ratio-fold change is currently lower than GEM-GECO1 (emission ratio)[263] and REX-GECO1 (excitation ratio)[275], additional engineering and optimization can likely produce REX-GECO1 variants with substantially improved dual-colour ratiometric emission, an intrinsic property of the CRO as delineated (*vide supra*). In contrast to another emission-ratiometric biosensor GEM-GECO1[263], wherein ESPT is inhibited upon Ca^{2+} binding that leads to the colour change from green to blue[279], the ESPT pathway is enabled with Ca^{2+} binding in REX-GECO1. More importantly, the achieved fluorescence is ~100 nm red-shifted relative to GEM-GECO1. However, the REX-GECO1 red to green emission ratiometric method suffers from higher noise relative to the excitation ratiometric method[275] due to the intrinsically dim emission at 510 nm when excited at 480 nm.

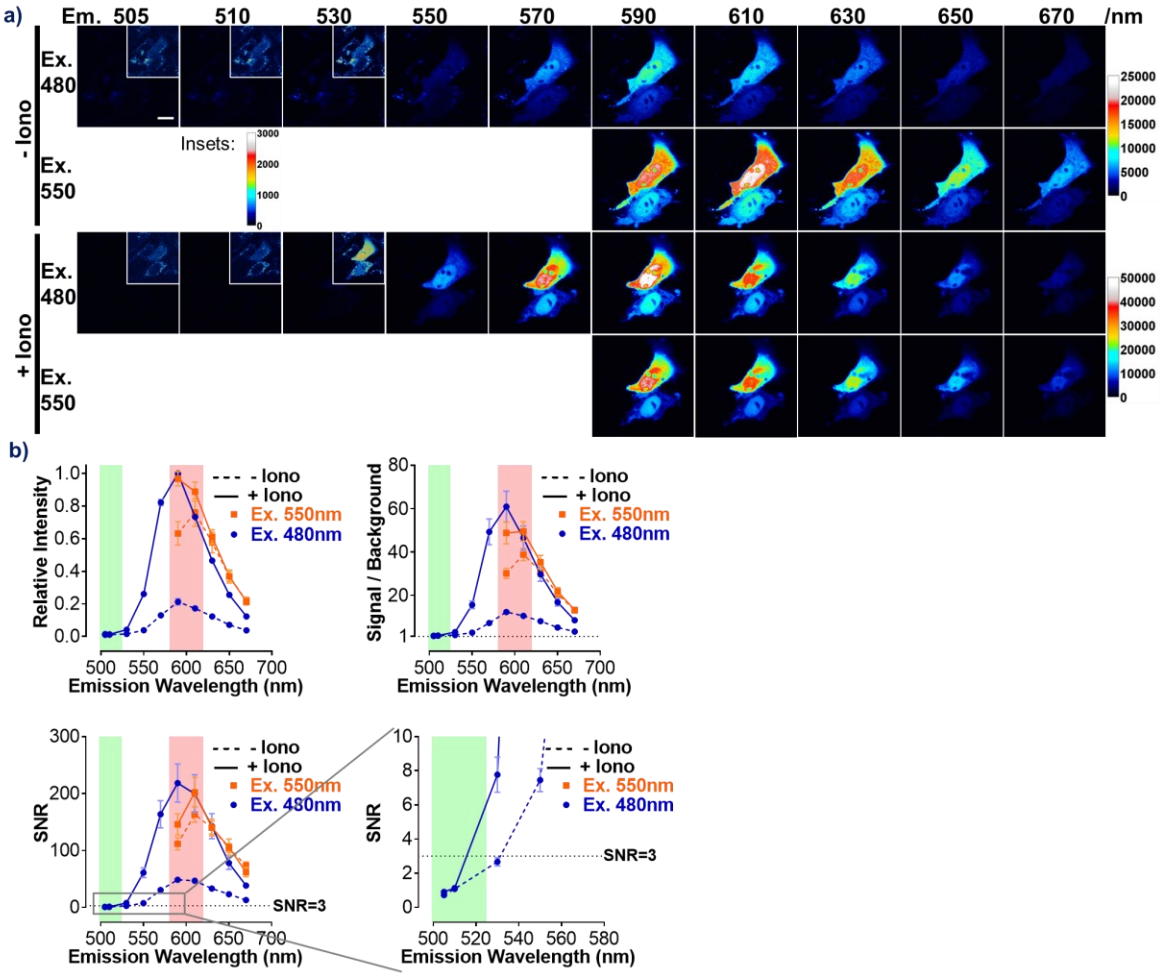

Figure 5.4 Spectral imaging of REX-GECO1.

A| Manual λ-scan performed on HeLa cells expressing REX-GECO1 before (upper two rows) and after (lower two rows) the histamine and ionomycin (Iono) treatment with 480 or 550 nm excitation. Scale bar = 10 µm. In the green channels (Ex. 480 nm, Em. 505, 510, and 530 nm, see the insets of three upper left cells), the autofluorescence is comparable to the biosensor signal. The collection window was shifted across the spectrum with a 20 nm increment (except the first one at 505 nm). **B|** Analysis of manual λ-scan imaging data (with mean filtering) on HeLa cells expressing REX-GECO1 before (dashed curves) and after (solid curves) ionomycin treatment with 480 nm (blue) or 550 nm (orange) excitation. A total of 26 ROIs from 6 cells were recorded, and the data are represented as mean ± s.e.m. The selected wavelength ranges of the green (495–525 nm) and red (580–620 nm) emission channels are highlighted by the light green and red shades, respectively. Note that in the enlarged plot (lower right panel), weak emission signal in the green channel satisfies the common SNR>3 criteria.

For the first time, the green/red dual-colour Ca^{2+} sensing was achieved in human cells with a single FP-based biosensor, which offers a new design strategy for Ca^{2+} biosensors. Moving forward, REX-GECO1 can be engineered to further improve its dynamic range as an emission-ratiometric reporter. This could be achieved by directed evolution with strategic mutations[280–282] to enhance red emission from the Ca^{2+}-bound state, while reducing the CRO cavity volume with more hydrophobic residue sidechains to trap the Ca^{2+}-free biosensor in an ESPT-incapable, brighter green state.

5.3.3 **Characterising the pH dependence of lactate indicator eLACCO1.1**

Section 5.3.3 is an collaborative work and has been published as part of the article Nasu, Y., Murphy-Royal, C., Wen, Y., Haidey, J.N., Molina, R.S., Aggarwal, A., Zhang, S., Kamijo, Y., Paquet, M.E., Podgorski, K. and Drobizhev, M. (2021). A genetically encoded fluorescent biosensor for extracellular L-lactate. Nature communications, 12(1), 1-12. S.Z. was responsible for performing and analysing the in-situ pH titration experiment (Supplementary Figure 11) and expressed the protein for crystallography. Y.N. developed eLACCO1.1 and performed in vitro characterization. R.S.M. and M.D. measured one-photon absorbance spectra and two-photon excitation spectra. C.M.R. and J.H. performed acute brain slice imaging. Y.W. determined the crystal structure of eLACCO1. A.A. performed stopped-flow experiment. A.A. and K.P. performed the imaging of primary neurons and data analysis. Y.K. performed screening of leader sequence. M.-E.P produced AAV. M.J.L., K.P., J.S.B., G.R.G. and R.E.C. supervised research. Y.N. and R.E.C. wrote the manuscript.

L-Lactate has been increasingly considered to play important roles as an energy currency in a variety of mammalian tissues. My colleagues Y. Nasu and collaborators developed a green fluorescent genetically encoded biosensor for extracellular L-lactate, named eLACCO1.1. Green fluorescent sensors are known to be sensitive over a range of pH close to the physiological pH, which may suffer from pH-dependent imaging artefacts. In this case, the ligand L-lactate also contains an acidic carboxylic group that may affect the fluorescence of eLACCO1.1 by altering the pH, in addition to the canonical binding to the sensor. Therefore, it is important to characterise and understand the pH dependence of the eLACCO1.1 fluorescence on the surface of living mammalian cells. To achieve this

goal, I performed *in situ* pH titration of eLACCO1.1 using total internal reflection fluorescence (TIRF) microscopy.

HeLa cells seeded onto coverslips were co-transfected with pDisplay-pHuji (Addgene plasmid #61556) and pAMEXT-eLACCO1.1 or deLACCO. Forty-eight hours after transfection, the coverslips were transferred into Attofluor™ Cell Chamber with HBSS supplemented with 20 mM HEPES (Gibco, Cat. #15630130) and 10 mM 2-deoxy-D-glucose (Sigma-Aldrich Cat. #D8375-1G) at pH 7.05. Other bath solutions were supplemented with 10 mM L-lactate (Sigma-Aldrich Cat. #71718-10G) and subsequently adjusted to their respective pH values. Rapid change of bath solutions during the image was performed in a remove-and-add manner using a homemade solution remover previously described. Cells are imaged on a Nikon Eclipse Ti-E epifluorescence microscope equipped with a 488 nm argon laser and a 543 nm He-Ne laser focused on the back aperture of a 60X oil TIRF objective lens (NA 1.49, Nikon). TIRF setup was achieved by a TI-TIRF-E Motorized Illuminator Unit (Nikon) to reduce the contribution of fluorophores that are not localized on the plasma membrane. Images were acquired every 10 sec by a Photometrics QuantEM 512SC EM-CCD camera at a gain value of 500. To avoid the photoactivation artefacts, pHuji signal was acquired first in each cycle with the 543 nm laser with a TRITC filter cube followed by the acquisition of eLACCO1.1 signal with 488 nm laser and a FITC filter cube. NIS-Elements AR package software was used for automatic instrument control, data recording and measurement. Data was further analyzed and normalized to the intensity at pH 7.99 using a custom R script (available at https://github.com/shucez/eLACCO_manuscript_TIRF_deltaF_F0) and plotted in GraphPad Prism software.

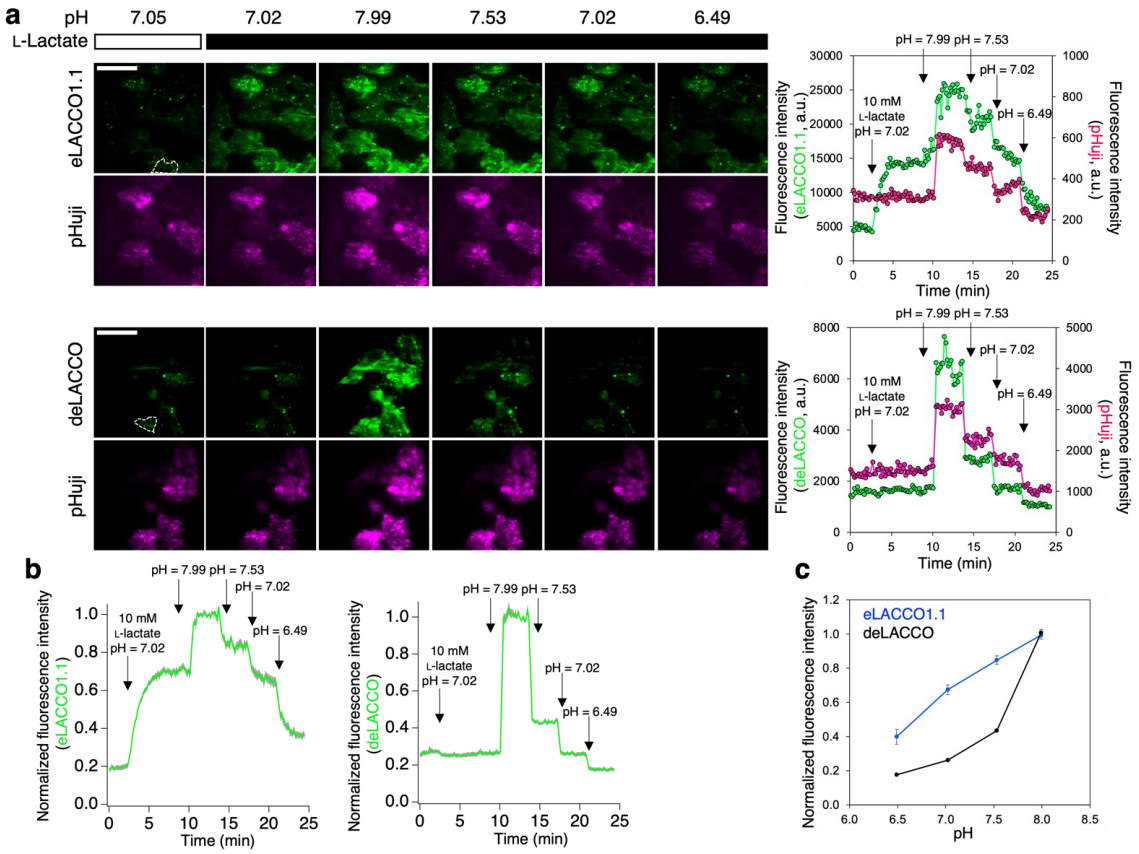

Figure 5.5 pH titration of eLACCO1.1 on the surface of live HeLa cells.

A| Representative images of HeLa cells expressing eLACCO1.1 or deLACCO and exposed to different pH values. A genetically encoded red fluorescent pH indicator pHuji was simultaneously expressed on the cell surface. Representative traces of fluorescence intensity (right) correspond to dashed white lines on the images. Scale bars, 50 µm. **B|** The time course of the normalized fluorescence intensity of eLACCO1.1 or deLACCO. *n* = 29 and 23 cells for eLACCO1.1 and deLACCO, respectively (mean ± s.e.m.). **C|** pH titration curves of eLACCO1.1 and deLACCO in the presence of 10 mM L-lactate. *n* = 20 data points at each pH from (b) (mean ± s.d.).

The TIRF imaging allowed recording of fluorescence intensity only in close proximity to the cell surface, greatly reducing the interference of intracellular fluorescence.

The *in situ* pH titration exhibited the expected pH dependence, consistent with the pH titration of purified protein[§].

[§] See Figure 2e, Supplementary Figure 4d and Supplementary Figure 7 of the original publication[267].

5.3.4 **Automating high-throughput plasmid purification**

Section 5.3.4 is a collaborative work with Eric Y. Fan-Lou.

As discussed in **Section 1.1.3.1**, screening of arrayed plasmid libraries requires purifying the plasmid encoding each library member separately. Automating the otherwise laborious procedure can great enhance the capacity to perform library screening. In this work, we implemented automatic magnetic beads-based plasmid purification using King Fisher Duo instrument. We performed systematic comparisons of the commercially available products and evaluated the feasibility and user experience in this process. In summary, the automated plasmid purification yielded comparable quantity and quality as the manual column-based counterparts. Magnetic beads-based purification allowed substantially less hands-on bench time and more walk-off time.

This work was supported by a research service contract with Galenvs Sciences Inc., Montreal, QC Canada. Due to the confidentiality articles in the contract, details of the work will not be disclosed in the thesis. E. Y. Fan-Lou contributed to the project.

References

1. Voskarides, K. Directed evolution. The legacy of a Nobel Prize. *J. Mol. Evol.* **89**, 189–191 (2021).

2. Fus-Kujawa, A. *et al.* An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. *Front Bioeng Biotechnol* **9**, 701031 (2021).

3. Boussif, O. *et al.* A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7297–7301 (1995).

4. Kumar, P., Nagarajan, A. & Uchil, P. D. Electroporation. *Cold Spring Harb. Protoc.* **2019**, db.top096271 (2019).

5. Kim, T. K. & Eberwine, J. H. Mammalian cell transfection: the present and the future. *Anal. Bioanal. Chem.* **397**, 3173–3178 (2010).

6. DuBridge, R. B. *et al.* Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Mol. Cell. Biol.* **7**, 379–387 (1987).

7. Murphy, A. J. M., Kung, A. L., Swirski, R. A. & Schimke, R. T. cDNA expression cloning in human cells using the p λ DR2 episomal vector system. *Methods* **4**, 111 – 131 (1992).

8. Kunaparaju, R., Liao, M. & Sunstrom, N.-A. Epi-CHO, an episomal expression system for recombinant protein production in CHO cells. *Biotechnol. Bioeng.* **91**, 670–677 (2005).

9. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766 (2014).

10. Corbett, K. S. *et al.* SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* **586**, 567–571 (2020).

11. Castanotto, D. & Rossi, J. J. Construction and transfection of PCR products expressing siRNAs or shRNAs in mammalian cells. *Methods Mol. Biol.* **252**, 509–514 (2004).

12. Liu, Z. *et al.* Sustained deep-tissue voltage recording using a fast indicator evolved for two-photon microscopy. *Cell* **0**, (2022).

13. Villette, V. *et al.* Ultrafast Two-Photon Imaging of a High-Gain Voltage Indicator in Awake Behaving Mice. *Cell* **179**, 1590-1608.e23 (2019).

14. Piatkevich, K. D. *et al.* A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat. Chem. Biol.* **14**, 352–360 (2018).

15. Sun, F. *et al.* A genetically encoded fluorescent sensor enables rapid and specific detection of dopamine in flies, fish, and mice. *Cell* **174**, 481-496.e19 (2018).

16. Sun, F. *et al.* Next-generation GRAB sensors for monitoring dopaminergic activity in vivo. *Nat. Methods* **17**, 1156–1166 (2020).

17. Dong, C. *et al.* Psychedelic-inspired drug discovery using an engineered biosensor. *Cell* **184**, 2779-2792.e18 (2021).

18. Wan, J. *et al.* A genetically encoded sensor for measuring serotonin dynamics. *Nat. Neurosci.* **24**, 746–752 (2021).

19. Feng, J. *et al.* A genetically encoded fluorescent sensor for rapid and specific in vivo detection of norepinephrine. *Neuron* **102**, 745-761.e8 (2019).

20. Jing, M. *et al.* A genetically encoded fluorescent acetylcholine indicator for in vitro and in vivo studies. *Nat. Biotechnol.* **36**, 726–737 (2018).

21. Latorraca, N. R., Venkatakrishnan, A. J. & Dror, R. O. GPCR dynamics: Structures in motion. *Chem. Rev.* **117**, 139–155 (2017).

22. Lu, X. *et al.* Detecting rapid pan-cortical voltage dynamics in vivo with a brighter and faster voltage indicator. *bioRxiv* (2022) doi:10.1101/2022.08.29.505018.

23. Zhang, D.-Y., Lau, C.-P. & Li, G.-R. Human Kir2.1 channel carries a transient outward potassium current with inward rectification. *Pflugers Arch.* **457**, 1275–1285 (2009).

24. Tseng, W. C., Haselton, F. R. & Giorgio, T. D. Transfection by cationic liposomes using simultaneous single cell measurements of plasmid delivery and transgene expression. *J. Biol. Chem.* **272**, 25641–25647 (1997).

25. Cohen, R. N., van der Aa, M. A. E. M., Macaraeg, N., Lee, A. P. & Szoka, F. C., Jr. Quantification of plasmid DNA copies in the nucleus after lipoplex and polyplex transfection. *J. Control. Release* **135**, 166–174 (2009).

26. Vodicka, M. A. Determinants for lentiviral infection of non-dividing cells. *Somatic Cell Genet.* **26**, 35–49 (2001).

27. Elegheert, J. *et al.* Lentiviral transduction of mammalian cells for fast, scalable and high-level production of soluble and membrane proteins. *Nat. Protoc.* **13**, 2991–3017 (2018).

28. Finkelshtein, D., Werman, A., Novick, D., Barak, S. & Rubinstein, M. LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7306–7311 (2013).

29. Doms, R. W. & Trono, D. The plasma membrane as a combat zone in the HIV battlefield. *Genes Dev.* **14**, 2677–2688 (2000).

30. Yee, J. K., Friedmann, T. & Burns, J. C. Generation of high-titer pseudotyped retroviral vectors with very broad host range. *Methods Cell Biol.* **43 Pt A**, 99–112 (1994).

31. Joglekar, A. V. & Sandoval, S. Pseudotyped Lentiviral vectors: One vector, many guises. *Hum. Gene Ther. Methods* **28**, 291–301 (2017).

32. Crawford, K. H. D. *et al.* Protocol and reagents for pseudotyping Lentiviral particles with SARS-CoV-2 Spike protein for neutralization assays. *Viruses* **12**, (2020).

33. Cribbs, A. P., Kennedy, A., Gregory, B. & Brennan, F. M. Simplified production and concentration of lentiviral vectors to achieve high transduction in primary human T cells. *BMC Biotechnol.* **13**, 98 (2013).

34. Ricks, D. M., Kutner, R., Zhang, X.-Y., Welsh, D. A. & Reiser, J. Optimized lentiviral transduction of mouse bone marrow-derived mesenchymal stem cells. *Stem Cells Dev.* **17**, 441–450 (2008).

35. Kutner, R. H., Zhang, X.-Y. & Reiser, J. Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors. *Nat. Protoc.* **4**, 495–505 (2009).

36. Wardill, T. J. *et al.* A neuron-based screening platform for optimizing genetically-encoded calcium indicators. *PLoS One* **8**, e77728 (2013).

37. Qudsia, S. *et al.* A novel lentiviral scFv display library for rapid optimization and selection of high affinity antibodies. *Biochem. Biophys. Res. Commun.* **499**, 71–77 (2018).

38. Alonso-Camino, V. *et al.* CARbodies: Human antibodies against cell surface tumor antigens selected from repertoires displayed on T cell chimeric antigen receptors. *Mol. Ther. Nucleic Acids* **2**, e93 (2013).

39. Boutin, S. *et al.* Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors. *Hum. Gene Ther.* **21**, 704–712 (2010).

40. Pupo, A. *et al.* AAV Vectors: The Rubik's Cube of Human Gene Therapy. *Mol. Ther.* (2022) doi:10.1016/j.ymthe.2022.09.015.

41. Samulski, R. J., Chang, L. S. & Shenk, T. A recombinant plasmid from which an infectious adeno-associated virus genome can be excised in vitro and its use to study viral replication. *J. Virol.* **61**, 3096–3101 (1987).

42. Xiao, X., Li, J. & Samulski, R. J. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J. Virol.* **72**, 2224–2232 (1998).

43. Burger, C. *et al.* Recombinant AAV viral vectors pseudotyped with viral capsids from serotypes 1, 2, and 5 display differential efficiency and cell tropism after delivery to different regions of the central nervous system. *Mol. Ther.* **10**, 302–317 (2004).

44. Kwon, I. & Schaffer, D. V. Designer gene delivery vectors: molecular engineering and evolution of adeno-associated viral vectors for enhanced gene transfer. *Pharm. Res.* **25**, 489–499 (2008).

45. Barnes, L. F., Draper, B. E., Chen, Y.-T., Powers, T. W. & Jarrold, M. F. Quantitative analysis of genome packaging in recombinant AAV vectors by charge detection mass spectrometry. *Mol Ther Methods Clin Dev* **23**, 87–97 (2021).

46. Kotin, R. M. *et al.* Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 2211–2215 (1990).

47. Duan, D. *et al.* Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term Episomal persistence in muscle tissue. *J. Virol.* **73**, 861–861 (1999).

48. Nakai, H. *et al.* AAV serotype 2 vectors preferentially integrate into active genes in mice. *Nat. Genet.* **34**, 297–302 (2003).

49. Roos, J. *et al.* STIM1, an essential and conserved component of store-operated Ca2+ channel function. *J. Cell Biol.* **169**, 435–445 (2005).

50. Liou, J. *et al.* STIM is a Ca2+ sensor essential for Ca2+-store-depletion-triggered Ca2+ influx. *Curr. Biol.* **15**, 1235–1241 (2005).

51. Feske, S. *et al.* A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* **441**, 179–185 (2006).

52. Vig, M. *et al.* CRACM1 is a plasma membrane protein essential for store-operated Ca2+ entry. *Science* **312**, 1220–1223 (2006).

53. Bock, C. *et al.* High-content CRISPR screening. *Nature Reviews Methods Primers* **2**, 1–23 (2022).

54. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

55. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).

56. Yan, X. *et al.* High-content imaging-based pooled CRISPR screens in mammalian cells. *J. Cell Biol.* **220**, (2021).

57. Kanfer, G. *et al.* Image-based pooled whole-genome CRISPRi screening for subcellular phenotypes. *J. Cell Biol.* **220**, (2021).

58. Binan, L. *et al.* Live single-cell laser tag. *Nat. Commun.* **7**, 11636 (2016).

59. Adamson, B. *et al.* A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867-1882.e21 (2016).

60. Jetzt, A. E. *et al.* High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* **74**, 1234–1240 (2000).

61. Hill, A. J. *et al.* On the design of CRISPR-based single-cell molecular screens. *Nat. Methods* **15**, 271–274 (2018).

62. Nikolaitchik, O. A. *et al.* Dimeric RNA recognition regulates HIV-1 genome packaging. *PLoS Pathog.* **9**, e1003249 (2013).

63. Pradhan, S., Varsani, A., Leff, C., Swanson, C. J. & Hariadi, R. F. Viral aggregation: The knowns and unknowns. *Viruses* **14**, 438 (2022).

64. Chan, R. *et al.* Retroviruses human immunodeficiency virus and murine leukemia virus are enriched in phosphoinositides. *J. Virol.* **82**, 11228–11238 (2008).

65. Hammarstedt, M., Wallengren, K., Pedersen, K. W., Roos, N. & Garoff, H. Minimal exclusion of plasma membrane proteins during retroviral envelope formation. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7527–7532 (2000).

66. Wang, Y., Wang, Z. & Li, G. Engineered retroviruses map ligand-receptor interactions. *Nature methods* vol. 19 408–410 (2022).

67. Wang, Y. *et al.* Directed Evolution: Methodologies and Applications. *Chem. Rev.* **121**, 12384–12444 (2021).

68. Taube, R. *et al.* Lentivirus display: stable expression of human antibodies on the surface of human cells and virus particles. *PLoS One* **3**, e3181 (2008).

69. Granieri, L., Baret, J.-C., Griffiths, A. D. & Merten, C. A. High-throughput screening of enzymes by retroviral display using droplet-based microfluidics. *Chem. Biol.* **17**, 229–235 (2010).

70. Dobson, C. S. *et al.* Antigen identification and high-throughput interaction mapping by reprogramming viral entry. *Nat. Methods* **19**, 449–460 (2022).

71. Urban, J. H. & Merten, C. A. Retroviral display in gene therapy, protein engineering, and vaccine development. *ACS Chem. Biol.* **6**, 61–74 (2011).

72. Zhou, C. & Shen, W. D. Mammalian Cell Surface Display of Full Length IgG. in *Antibody Engineering: Methods and Protocols, Second Edition* (ed. Chames, P.) 293–302 (Humana Press, 2012).

73. Zhou, C., Jacobsen, F. W., Cai, L., Chen, Q. & Shen, W. D. Development of a novel mammalian cell surface antibody display platform. *MAbs* **2**, 508–518 (2010).

74. Meinke, G., Bohm, A., Hauber, J., Pisabarro, M. T. & Buchholz, F. Cre Recombinase and Other Tyrosine Recombinases. *Chem. Rev.* **116**, 12785–12820 (2016).

75. Merrick, C. A., Zhao, J. & Rosser, S. J. Serine Integrases: Advancing Synthetic Biology. *ACS Synth. Biol.* **7**, 299–310 (2018).

76. Zhang, M., Yang, C., Tasan, I. & Zhao, H. Expanding the Potential of Mammalian Genome Engineering via Targeted DNA Integration. *ACS Synth. Biol.* **10**, 429–446 (2021).

77. Grindley, N. D. F., Whiteson, K. L. & Rice, P. A. Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* **75**, 567–605 (2006).

78. Mclellan, M. A., Rosenthal, N. A. & Pinto, A. R. Cre- loxP-mediated recombination: general principles and experimental considerations. *Curr Protoc Mouse Biol* **7**, 1–12 (2017).

79. Weinberg, B. H. *et al.* High-performance chemical- and light-inducible recombinases in mammalian cells and mice. *Nat. Commun.* **10**, 4845 (2019).

80. Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA cloning using in vitro site-specific recombination. *Genome Res.* **10**, 1788–1795 (2000).

81. Freuler, F., Stettler, T., Meyerhofer, M., Leder, L. & Mayr, L. M. Development of a novel Gateway-based vector system for efficient, multiparallel protein expression in Escherichia coli. *Protein Expr. Purif.* **59**, 232–241 (2008).

82. Giuraniuc, C. V., MacPherson, M. & Saka, Y. Gateway vectors for efficient artificial gene assembly in vitro and expression in yeast Saccharomyces cerevisiae. *PLoS One* **8**, e64419 (2013).

83. Weinberg, B. H. *et al.* Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* **35**, 453–462 (2017).

84. Friedland, A. E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).

85. Yehl, K. & Lu, T. Scaling computation and memory in living cells. *Curr. Opin. Biomed. Eng.* **4**, 143–151 (2017).

86. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**, 448–452 (2013).

87. Guiziou, S., Mayonove, P. & Bonnet, J. Hierarchical composition of reliable recombinase logic devices. *Nat. Commun.* **10**, 456 (2019).

88. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science* **353**, aad8559 (2016).

89. Broach, J. R. & Hicks, J. B. Replication and recombination functions associated with the yeast plasmid, 2 mu circle. *Cell* **21**, 501–508 (1980).

90. Dutcher, S. K. & Hartwell, L. H. The role of S. cerevisiae cell division cycle genes in nuclear fusion. *Genetics* **100**, 175–184 (1982).

91. Reynolds, A. E., Murray, A. W. & Szostak, J. W. Roles of the 2 microns gene products in stable maintenance of the 2 microns plasmid of Saccharomyces cerevisiae. *Mol. Cell. Biol.* **7**, 3566–3573 (1987).

92. Sternberg, N. & Hamilton, D. Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *J. Mol. Biol.* **150**, 467–486 (1981).

93. Segev, N. & Cohen, G. Control of circularization of bacteriophage P1 DNA in Escherichia coli. *Virology* **114**, 333–342 (1981).

94. Sternberg, N., Hamilton, D., Austin, S., Yarmolinsky, M. & Hoess, R. Site-specific recombination and its role in the life cycle of bacteriophage P1. *Cold Spring Harb. Symp. Quant. Biol.* **45 Pt 1**, 297–309 (1981).

95. Ghosh, K., Lau, C.-K., Gupta, K. & Van Duyne, G. D. Preferential synapsis of loxP sites drives ordered strand exchange in Cre-loxP site-specific recombination. *Nat. Chem. Biol.* **1**, 275–282 (2005).

96. Voziyanov, Y., Lee, J., Whang, I., Lee, J. & Jayaram, M. Analyses of the first chemical step in Flp site-specific recombination: Synapsis may not be a pre-requisite for strand cleavage. *J. Mol. Biol.* **256**, 720–735 (1996).

97. Stark, W. M., Sherratt, D. J. & Boocock, M. R. Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. *Cell* **58**, 779–790 (1989).

98. Guo, F., Gopaul, D. N. & van Duyne, G. D. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40–46 (1997).

99. Reed, R. R. & Grindley, N. D. Transposon-mediated site-specific recombination in vitro: DNA cleavage and protein-DNA linkage at the recombination site. *Cell* **25**, 721–728 (1981).

100. Li, W. *et al.* Structure of a synaptic gammadelta resolvase tetramer covalently linked to two cleaved DNAs. *Science* **309**, 1210–1215 (2005).

101. Reed, R. R. & Moser, C. D. Resolvase-mediated recombination intermediates contain a serine residue covalently linked to DNA. *Cold Spring Harb. Symp. Quant. Biol.* **49**, 245–249 (1984).

102. Smith, M. C. M. *et al.* Sandmeyer SB Phage-encoded serine integrases and other large serine recombinases. *Mobile DNA III* 253–272 (2015).

103. Pokhilko, A. *et al.* The mechanism of ϕC31 integrase directionality: experimental analysis and computational modelling. *Nucleic Acids Res.* **44**, 7360–7372 (2016).

104. Khaleel, T., Younger, E., McEwan, A. R., Varghese, A. S. & Smith, M. C. M. A phage protein that binds ϕ C31 integrase to switch its directionality. *Mol. Microbiol.* **80**, 1450 – 1463 (2011).

105. Fan, H.-F., Hsieh, T.-S., Ma, C.-H. & Jayaram, M. Single-molecule analysis of ϕC31 integrase-mediated site-specific recombination by tethered particle motion. *Nucleic Acids Res.* **44**, 10804–10823 (2016).

106. Chi, X., Zheng, Q., Jiang, R., Chen-Tsai, R. Y. & Kong, L.-J. A system for site-specific integration of transgenes in mammalian cells. *PLoS One* **14**, e0219842 (2019).

107. Olivares, E. C. *et al.* Site-specific genomic integration produces therapeutic Factor IX levels in mice. *Nat. Biotechnol.* **20**, 1124–1128 (2002).

108. O'Gorman, S., Fox, D. T. & Wahl, G. M. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**, 1351–1355 (1991).

109. Wissink, E. M., Fogarty, E. A. & Grimson, A. High-throughput discovery of post-transcriptional cis-regulatory elements. *BMC Genomics* **17**, 177 (2016).

110. Thyagarajan, B. *et al.* Creation of engineered human embryonic stem cell lines using phiC31 integrase. *Stem Cells* **26**, 119–126 (2008).

111. Lieu, P. T. *et al.* Generation of site-specific retargeting platform cell lines for drug discovery using phiC31 and R4 integrases. *J. Biomol. Screen.* **14**, 1207–1215 (2009).

112. Russell, J. P., Chang, D. W., Tretiakova, A. & Padidam, M. Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells. *Biotechniques* **40**, 460, 462, 464 (2006).

113. Xu, Z. *et al.* Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol.* **13**, 87 (2013).

114. Chao, G., Travis, C. & Church, G. Measurement of large serine integrase enzymatic characteristics in HEK293 cells reveals variability and influence on downstream reporter expression. *FEBS J.* **288**, 6410–6427 (2021).

115. Mulholland, C. B. *et al.* A modular open platform for systematic functional studies under physiological conditions. *Nucleic Acids Res.* **43**, e112 (2015).

116. Matreyek, K. A., Stephany, J. J. & Fowler, D. M. A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res.* **45**, e102 (2017).

117. Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res.* **48**, e1 (2020).

118. Shukla, N., Roelle, S. M., Suzart, V. G., Bruchez, A. M. & Matreyek, K. A. Mutants of human ACE2 differentially promote SARS-CoV and SARS-CoV-2 spike mediated infection. *PLoS Pathog.* **17**, e1009715 (2021).

119. Inniss, M. C. *et al.* A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO Cells. *Biotechnol. Bioeng.* **114**, 1837–1846 (2017).

120. Xiong, K. *et al.* An optimized genome-wide, virus-free CRISPR screen for mammalian cells. *Cell Rep Methods* **1**, (2021).

121. Briney, B., Inderbitzin, A., Joyce, C. & Burton, D. R. Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* **566**, 393–397 (2019).

122. Bashford-Rogers, R. J. M., Smith, K. G. C. & Thomas, D. C. Antibody repertoire analysis in polygenic autoimmune diseases. *Immunology* **155**, 3–17 (2018).

123. Jung, D., Giallourakis, C., Mostoslavsky, R. & Alt, F. W. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol.* **24**, 541–570 (2006).

124. Schatz, D. G. & Swanson, P. C. V(D)J recombination: mechanisms of initiation. *Annu. Rev. Genet.* **45**, 167–202 (2011).

125. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988).

126. Ru, H. *et al.* Molecular mechanism of V(D)J recombination from synaptic RAG1-RAG2 complex structures. *Cell* **163**, 1807 (2015).

127. Chi, X., Li, Y. & Qiu, X. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology* **160**, 233–247 (2020).

128. Di Noia, J. M. & Neuberger, M. S. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007).

129. Peled, J. U. *et al.* The biochemistry of somatic hypermutation. *Annu. Rev. Immunol.* **26**, 481–511 (2008).

130. Rogozin, I. B. & Diaz, M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J. Immunol.* **172**, 3382–3384 (2004).

131. Yeap, L.-S. & Meng, F.-L. Cis- and trans-factors affecting AID targeting and mutagenic outcomes in antibody diversification. *Adv. Immunol.* **141**, 51–103 (2019).

132. Sale, J. E. & Neuberger, M. S. TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity* **9**, 859–869 (1998).

133. Cumbers, S. J. *et al.* Generation and iterative affinity maturation of antibodies in vitro using hypermutating B-cell lines. *Nat. Biotechnol.* **20**, 1129–1134 (2002).

134. Wang, L., Jackson, W. C., Steinbach, P. A. & Tsien, R. Y. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16745–16749 (2004).

135. Majors, B. S., Chiang, G. G., Pederson, N. E. & Betenbaugh, M. J. Directed evolution of mammalian anti-apoptosis proteins by somatic hypermutation. *Protein Eng. Des. Sel.* **25**, 27–38 (2012).

136. Martin, A. & Scharff, M. D. Somatic hypermutation of the AID transgene in B and non-B cells. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12304–12308 (2002).

137. Chen, S. *et al.* Affinity maturation of anti-TNF-alpha scFv with somatic hypermutation in non-B cells. *Protein Cell* **3**, 460–469 (2012).

138. Al-Qaisi, T. S., Su, Y.-C. & Roffler, S. R. Transient AID expression for in situ mutagenesis with improved cellular fitness. *Sci. Rep.* **8**, (2018).

139. Chen, H. *et al.* Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. *Nat. Biotechnol.* **38**, 165–168 (2020).

140. Thiel, V., Herold, J., Schelle, B. & Siddell, S. G. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.* **82**, 1273–1281 (2001).

141. Kielczewska, A. *et al.* Development of a potent high-affinity human therapeutic antibody via novel application of recombination signal sequence-based affinity maturation. *J. Biol. Chem.* **298**, 101533 (2022).

142. Gallo, M., Kang, J. S. & Pigott, C. R. Sequence diversity generation in immunoglobulins. *US Patent* (2013).

143. Gallo, M. L., Kang, J. S., Pigott, C. R., Chai, F. & Abby, L. I. N. Single variable domain t-cell receptors. *Patent* (2017).

144. Donovan, K. F. *et al.* Creation of novel protein variants with CRISPR/Cas9-mediated Mutagenesis: Turning a screening by-product into a discovery tool. *PLoS One* **12**, e0170445 (2017).

145. Ipsaro, J. J. *et al.* Rapid generation of drug-resistance alleles at endogenous loci using CRISPR-Cas9 indel mutagenesis. *PLoS One* **12**, e0172177 (2017).

146. Hendel, S. J. & Shoulders, M. D. Directed evolution in mammalian cells. *Nat. Methods* **18**, 346–357 (2021).

147. Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).

148. Gutierrez-Triana, J. A. *et al.* Efficient single-copy HDR by 5' modified long dsDNA donors. *Elife* **7**, (2018).

149. Mason, D. M. *et al.* High-throughput antibody engineering in mammalian cells by CRISPR/Cas9-mediated homology-directed mutagenesis. *Nucleic Acids Res.* **46**, 7436–7449 (2018).

150. Erdogan, M., Fabritius, A., Basquin, J. & Griesbeck, O. Targeted In Situ Protein Diversification and Intra-organelle Validation in Mammalian Cells. *Cell Chem Biol* **27**, 610-621.e5 (2020).

151. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

152. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).

153. Kurt, I. C. *et al.* CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2021).

154. Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).

155. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729–aaf8729 (2016).

156. Liu, L. D. *et al.* Intrinsic nucleotide preference of diversifying base editors guides antibody ex vivo affinity maturation. *Cell Rep.* **25**, 884-892.e3 (2018).

157. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).

158. Fijalkowska, I. J. & Schaaper, R. M. Mutants in the Exo I motif of Escherichia coli dnaQ: defective proofreading and inviability due to error catastrophe. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 2856–2861 (1996).

159. Berman, C. M. *et al.* An Adaptable Platform for Directed Evolution in Human Cells. *J. Am. Chem. Soc.* **140**, 18093–18103 (2018).

160. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5547–5551 (1992).

161. Hecht, B., Müller, G. & Hillen, W. Noninducible Tet repressor mutations map from the operator binding motif to the C terminus. *J. Bacteriol.* **175**, 1206–1210 (1993).

162. English, J. G. *et al.* VEGAS as a Platform for Facile Directed Evolution in Mammalian Cells. *Cell* **178**, 748-761.e17 (2019).

163. Strauss, E. G., Rice, C. M. & Strauss, J. H. Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology* **133**, 92–110 (1984).

164. Frolov, I., Hardy, R. & Rice, C. M. Cis-acting RNA elements at the 5’ end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. *RNA* **7**, 1638–1651 (2001).

165. Strauss, J. H. & Strauss, E. G. The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* **58**, 491–562 (1994).

166. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral mutation rates. *J. Virol.* **84**, 9733–9748 (2010).

167. Huston, M. A general hypothesis of species diversity. *Am. Nat.* **113**, 81–101 (1979).

168. Popa, S. C., Inamoto, I., Thuronyi, B. W. & Shin, J. A. Phage-assisted continuous evolution (PACE): A guide focused on evolving protein-DNA interactions. *ACS Omega* **5**, 26957–26966 (2020).

169. Berridge, M. J., Bootman, M. D. & Roderick, H. L. Calcium signalling: dynamics, homeostasis and remodelling. *Nat. Rev. Mol. Cell Biol.* **4**, 517–529 (2003).

170. Berridge, M. J. Calcium signalling remodelling and disease. *Biochem. Soc. Trans.* **40**, 297–309 (2012).

171. Clapham, D. E. Calcium signaling. *Cell* **80**, 259–268 (1995).

172. Berridge, M. J., Lipp, P. & Bootman, M. D. The versatility and universality of calcium signalling. *Nat. Rev. Mol. Cell Biol.* **1**, 11–21 (2000).

173. Berridge, M. J. Calcium signalling, a spatiotemporal phenomenon. in *Calcium - A Matter of Life or Death* 485–502 (Elsevier, 2007).

174. Iino, M. Spatiotemporal dynamics of Ca2+ signaling and its physiological roles. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **86**, 244–256 (2010).

175. Mekahli, D., Bultynck, G., Parys, J. B., De Smedt, H. & Missiaen, L. Endoplasmic-reticulum calcium depletion and disease. *Cold Spring Harb. Perspect. Biol.* **3**, a004317 (2011).

176. Raffaello, A., Mammucari, C., Gherardi, G. & Rizzuto, R. Calcium at the Center of Cell Signaling: Interplay between Endoplasmic Reticulum, Mitochondria, and Lysosomes. *Trends Biochem. Sci.* **41**, 1035–1049 (2016).

177. Zheng, S., Wang, X., Zhao, D., Liu, H. & Hu, Y. Calcium homeostasis and cancer: insights from endoplasmic reticulum-centered organelle communications. *Trends Cell Biol.* **0**, (2022).

178. Wang, Q.-C. *et al.* TMCO1 Is an ER Ca(2+) Load-Activated Ca(2+) Channel. *Cell* **165**, 1454–1466 (2016).

179. Mason, M. J., Garcia-Rodriguez, C. & Grinstein, S. Coupling between intracellular Ca2+ stores and the Ca2+ permeability of the plasma membrane. Comparison of the effects of thapsigargin, 2,5-di-(tert-butyl)-1,4-hydroquinone, and cyclopiazonic acid in rat thymic lymphocytes. *J. Biol. Chem.* **266**, 20856–20862 (1991).

180. Fasolato, C., Hoth, M. & Penner, R. A GTP-dependent step in the activation mechanism of capacitative calcium influx. *J. Biol. Chem.* **268**, 20737–20740 (1993).

181. Tan, P., He, L., Huang, Y. & Zhou, Y. Optophysiology: Illuminating cell physiology with optogenetics. *Physiol. Rev.* **102**, 1263–1325 (2022).

182. Kim, N. *et al.* Spatiotemporal control of fibroblast growth factor receptor signals by blue light. *Chem. Biol.* **21**, 903–912 (2014).

183. Airan, R. D., Thompson, K. R., Fenno, L. E., Bernstein, H. & Deisseroth, K. Temporally precise in vivo control of intracellular signalling. *Nature* **458**, 1025–1029 (2009).

184. Yu, G. *et al.* Optical manipulation of the alpha subunits of heterotrimeric G proteins using photoswitchable dimerization systems. *Sci. Rep.* **6**, 35777 (2016).

185. Asano, T., Igarashi, H., Ishizuka, T. & Yawo, H. Organelle Optogenetics: Direct Manipulation of Intracellular Ca2+ Dynamics by Light. *Front. Neurosci.* **12**, 561 (2018).

186. Saße, P., Jangsangthong, W., Gottschalk, A., Lehnart, S. & Dura, M. Polypeptides for light induced ca2 release and constructs containing the same. *European Patent* (2022).

187. Ma, G. *et al.* Optogenetic engineering to probe the molecular choreography of STIM1-mediated cell signaling. *Nat. Commun.* **11**, 1039 (2020).

188. Yeung, P. S.-W., Yamashita, M. & Prakriya, M. Molecular basis of allosteric Orai1 channel activation by STIM1. *The Journal of Physiology* **598**, 1707–1723 (2020).

189. Yuan, J. P. *et al.* SOAR and the polybasic STIM1 domains gate and regulate Orai channels. *Nat. Cell Biol.* **11**, 337–343 (2009).

190. Park, C. Y. *et al.* STIM1 clusters and activates CRAC channels via direct binding of a cytosolic domain to Orai1. *Cell* **136**, 876–890 (2009).

191. Kyung, T. *et al.* Optogenetic control of endogenous Ca(2+) channels in vivo. *Nat. Biotechnol.* **33**, 1092–1096 (2015).

192. Kim, S. *et al.* Non-invasive optical control of endogenous Ca2+ channels in awake mice. *Nat. Commun.* **11**, 210 (2020).

193. Pham, E., Mills, E. & Truong, K. A synthetic photoactivated protein to generate local or global Ca(2+) signals. *Chem. Biol.* **18**, 880–890 (2011).

194. He, L. *et al.* Near-infrared photoactivatable control of Ca(2+) signaling and optogenetic immunomodulation. *Elife* **4**, e10024 (2015).

195. Ishii, T. *et al.* Light generation of intracellular Ca(2+) signals by a genetically encoded protein BACCS. *Nat. Commun.* **6**, 8021 (2015).

196. He, L. *et al.* Circularly permuted LOV2 as a modular photoswitch for optogenetic engineering. *Nat. Chem. Biol.* **17**, 915–923 (2021).

197. He, L. *et al.* Engineering of a bona fide light-operated calcium channel. *Nat. Commun.* **12**, 164 (2021).

198. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

199. Cai, X. *et al.* The Orai1 Store-operated Calcium Channel Functions as a Hexamer. *The Journal of biological chemistry* vol. 291 25764–25775 Preprint at https://doi.org/10.1074/jbc.M116.758813 (2016).

200. Xiao, Z. *et al.* Enhanced Cytosolic Ca2+ Activation Underlies a Common Defect of Central Domain Cardiac Ryanodine Receptor Mutations Linked to Arrhythmias. *J. Biol. Chem.* **291**, 24528–24537 (2016).

201. Zheng, S. *et al.* Calcium store refilling and STIM activation in STIM- and Orai-deficient cell lines. *Pflugers Arch.* **470**, 1555–1567 (2018).

202. Zhang, S., He, L., Zhou, Y. & Wang, Y. Fluorescence-based ratiometric measurement of CRAC channel activity in STIM-Orai-overexpressing HEK-293 cells. *Methods Mol. Biol.* **1843**, 17–39 (2018).

203. Zal, T. & Gascoigne, N. R. J. Photobleaching-corrected FRET efficiency imaging of live cells. *Biophys. J.* **86**, 3923–3939 (2004).

204. Fahrner, M. *et al.* A coiled-coil clamp controls both conformation and clustering of stromal interaction molecule 1 (STIM1). *J. Biol. Chem.* **289**, 33231–33244 (2014).

205. Jackson, L. P. *et al.* Molecular basis for recognition of dilysine trafficking motifs by COPI. *Dev. Cell* **23**, 1255–1262 (2012).

206. Ma, W. & Goldberg, J. Rules for the recognition of dilysine retrieval motifs by coatomer. *EMBO J.* **32**, 926–937 (2013).

207. Costantini, L. M., Fossati, M., Francolini, M. & Snapp, E. L. Assessing the tendency of fluorescent proteins to oligomerize under physiologic conditions. *Traffic* **13**, 643–649 (2012).

208. Xu, F. *et al.* COPII mitigates ER stress by promoting formation of ER whorls. *Cell Res.* **31**, 141–156 (2021).

209. Zhou, Y. *et al.* STIM1 dimers undergo unimolecular coupling to activate Orai1 channels. *Nat. Commun.* **6**, 8395 (2015).

210. Dana, H. *et al.* Sensitive red protein calcium indicators for imaging neural activity. *Elife* **5**, (2016).

211. Akerboom, J. *et al.* Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Front. Mol. Neurosci.* **6**, 2 (2013).

212. Wu, J. *et al.* Improved orange and red Ca^{2}± indicators and photophysical considerations for optogenetic applications. *ACS Chem. Neurosci.* **4**, 963–972 (2013).

213. Yoast, R. E., Emrich, S. M. & Trebak, M. The anatomy of native CRAC channel(s). *Curr Opin Physiol* **17**, 89–95 (2020).

214. Baraniak, J. H., Jr *et al.* Orai channel C-terminal peptides are key modulators of STIM-Orai coupling and calcium signal generation. *Cell Rep.* **35**, 109322 (2021).

215. Perni, S., Dynes, J. L., Yeromin, A. V., Cahalan, M. D. & Franzini-Armstrong, C. Nanoscale patterning of STIM1 and Orai1 during store-operated Ca2+ entry. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5533-42 (2015).

216. Suzuki, J. *et al.* Imaging intraorganellar Ca2+ at subcellular resolution using CEPIA. *Nat. Commun.* **5**, 4153 (2014).

217. Prins, D. & Michalak, M. Organellar calcium buffers. *Cold Spring Harb. Perspect. Biol.* **3**, a004069–a004069 (2011).

218. Zhou, Y. *et al.* The STIM1-binding site nexus remotely controls Orai1 channel gating. *Nat. Commun.* **7**, 13725 (2016).

219. Yeung, P. S.-W. *et al.* Mapping the functional anatomy of Orai1 transmembrane domains for CRAC channel gating. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5193–E5202 (2018).

220. Motiani, R. K., Abdullaev, I. F. & Trebak, M. A novel native store-operated calcium channel encoded by Orai3. *J. Biol. Chem.* **285**, 19173–19183 (2010).

221. Poláková, E., Zahradníková, A., Jr, Pavelková, J., Zahradník, I. & Zahradníková, A. Local calcium release activation by DHPR calcium channel openings in rat cardiac myocytes. *J. Physiol.* **586**, 3839–3854 (2008).

222. Hou, X., Pedi, L., Diver, M. M. & Long, S. B. Crystal structure of the calcium release-activated calcium channel Orai. *Science* **338**, 1308–1313 (2012).

223. Hou, X., Burstein, S. R. & Long, S. B. Structures reveal opening of the store-operated calcium channel Orai. *Elife* **7**, e36758 (2018).

224. Koshita, M. & Hotta, K. Relationship between membrane potential and calcium ion fluxes in the fragmented sarcoplasmic reticulum. *Jpn. J. Physiol.* **31**, 109–120 (1981).

225. Sanchez, C. *et al.* Tracking the sarcoplasmic reticulum membrane voltage in muscle with a FRET biosensor. *J. Gen. Physiol.* **150**, 1163–1177 (2018).

226. Zhang, X. *et al.* Distinct pharmacological profiles of ORAI1, ORAI2, and ORAI3 channels. *Cell Calcium* **91**, 102281 (2020).

227. Dörr, K. *et al.* Cell type-specific glycosylation of Orai1 modulates store-operated Ca2+ entry. *Sci. Signal.* **9**, ra25 (2016).

228. Platisa, J., Vasan, G., Yang, A. & Pieribone, V. A. Directed Evolution of Key Residues in Fluorescent Protein Inverses the Polarity of Voltage Sensitivity in the Genetically Encoded Indicator ArcLight. *ACS Chem. Neurosci.* **8**, 513–523 (2017).

229. Park, J. *et al.* Screening fluorescent voltage indicators with spontaneously spiking HEK cells. *PLoS One* **8**, e85221 (2013).

230. Matreyek, K. A. *et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018).

231. Malnati, M. S. *et al.* A universal real-time PCR assay for the quantification of group-M HIV-1 proviral load. *Nat. Protoc.* **3**, 1240–1248 (2008).

232. Jacobs, C. L., Badiee, R. K. & Lin, M. Z. StaPLs: versatile genetically encoded modules for engineering drug-inducible proteins. *Nat. Methods* **15**, 523–526 (2018).

233. Christodoulou, I. *et al.* Measurement of lentiviral vector titre and copy number by cross-species duplex quantitative PCR. *Gene Ther.* **23**, 113–118 (2016).

234. Lin, Y.-C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5**, 4767 (2014).

235. Ghosh, P., Kim, A. I. & Hatfull, G. F. The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB. *Mol. Cell* **12**, 1101–1111 (2003).

236. Tomimatsu, K. *et al.* Multiple expression cassette exchange via TP901-1, R4, and Bxb1 integrase systems on a mouse artificial chromosome. *FEBS Open Bio* **7**, 306–317 (2017).

237. Zhou, S. *et al.* A self-inactivating lentiviral vector for SCID-X1 gene therapy that does not activate LMO2 expression in human T cells. *Blood* **116**, 900–908 (2010).

238. Lee, J. *et al.* Versatile phenotype-activated cell sorting. *Sci Adv* **6**, (2020).

239. Tischer, D. & Weiner, O. D. Illuminating cell signalling with optogenetic tools. *Nature Reviews Molecular Cell Biology* **15**, 551–558 (2014).

240. Liu, Q. & Tucker, C. L. Engineering genetically-encoded tools for optogenetic control of protein activity. *Current Opinion in Chemical Biology* **40**, 17–23 (2017).

241. Johnson, H. E. & Toettcher, J. E. Illuminating developmental biology with cellular optogenetics. *Current Opinion in Biotechnology* **52**, 42–48 (2018).

242. Lu, X., Shen, Y. & Campbell, R. E. Engineering Photosensory Modules of Non-Opsin-Based Optogenetic Actuators. *International Journal of Molecular Sciences* **21**, 6522 (2020).

243. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nature Chemical Biology* **3**, 769–772 (2007).

244. Lee, H. M., Larson, D. R. & Lawrence, D. S. Illuminating the chemistry of life: Design, synthesis, and applications of caged and related photoresponsive compounds. *ACS Chemical Biology* **4**, 409–427 (2009).

245. Klán, P. *et al.* Photoremovable protecting groups in chemistry and biology: Reaction mechanisms and efficacy. *Chemical Reviews* **113**, 119–191 (2013).

246. Mizuno, H. *et al.* Photo-induced peptide cleavage in the green-to-red conversion of a fluorescent protein. *Molecular Cell* **12**, 1051–1058 (2003).

247. Davidson, M. W. & Campbell, R. E. Engineered fluorescent proteins: Innovations and applications. *Nature Methods* **6**, 713–717 (2009).

248. Nienhaus, K., Nienhaus, G. U., Wiedenmann, J. & Nar, H. Structural basis for photo-induced protein cleavage and green-to-red conversion of fluorescent protein EosFP. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 9156–9159 (2005).

249. Tsutsui, H. *et al.* The E1 Mechanism in Photo-Induced β-Elimination Reactions for Green-to-Red Conversion of Fluorescent Proteins. *Chemistry and Biology* **16**, 1140–1147 (2009).

250. Zhang, W. *et al.* Optogenetic control with a photocleavable protein, PhoCl. *Nat. Methods* **14**, 391–394 (2017).

251. Shadish, J. A., Strange, A. C. & DeForest, C. A. Genetically Encoded Photocleavable Linkers for Patterned Protein Release from Biomaterials. *J. Am. Chem. Soc.* **141**, 15619–15625 (2019).

252. Endo, M., Iwawaki, T., Yoshimura, H. & Ozawa, T. Photocleavable Cadherin Inhibits Cell-to-Cell Mechanotransduction by Light. *ACS Chem. Biol.* **14**, 2206–2214 (2019).

253. Xiang, D. *et al.* Hydrogels With Tunable Mechanical Properties Based on Photocleavable Proteins. *Front Chem* **8**, 7 (2020).

254. Reed, E. H., Schuster, B. S., Good, M. C. & Hammer, D. A. SPLIT: Stable Protein Coacervation Using a Light Induced Transition. *ACS Synth. Biol.* **9**, 500–507 (2020).

255. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* **4**, 17 (2012).

256. Vanquelef, E. *et al.* R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Research* **39**, 511–517 (2011).

257. Wink, L. H., Baker, D. L., Cole, J. A. & Parrill, A. L. A benchmark study of loop modeling methods applied to G protein-coupled receptors. *Journal of Computer-Aided Molecular Design* **33**, 573–595 (2019).

258. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **78**, 2690–2693 (1997).

259. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation* **9**, 3084–3095 (2013).

260. Miller, B. R. *et al.* MMPBSA.py: An efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation* **8**, 3314–3321 (2012).

261. Humphrey, W., Dalke, A. & Schulten, K. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).

262. Lu, X. *et al.* Photocleavable proteins that undergo fast and efficient dissociation. *Chem. Sci.* **12**, 9658–9672 (2021).

263. Zhao, Y. *et al.* An expanded palette of genetically encoded Ca^{2+} indicators. *Science* **333**, 1888–1891 (2011).

264. Qian, Y., Rancic, V., Wu, J., Ballanyi, K. & Campbell, R. E. A bioluminescent Ca2+ indicator based on a topological variant of GCaMP6s. *Chembiochem* **20**, 516–520 (2019).

265. Tang, L. *et al.* Switching between ultrafast pathways enables a green-red emission ratiometric fluorescent-protein-based Ca2+ biosensor. *Int. J. Mol. Sci.* **22**, 445 (2021).

266. Zhao, Y., Shen, Y., Wen, Y. & Campbell, R. E. High-performance intensiometric direct- and inverse-response genetically encoded biosensors for citrate. *ACS Cent. Sci.* **6**, 1441–1450 (2020).

267. Nasu, Y. *et al.* A genetically encoded fluorescent biosensor for extracellular L-lactate. *Nat. Commun.* **12**, 7058 (2021).

268. Wu, S.-Y. *et al.* A sensitive and specific genetically-encoded potassium ion biosensor for in vivo applications across the tree of life. *PLoS Biol.* **20**, e3001772 (2022).

269. Zou, P. *et al.* Bright and fast multicoloured voltage reporters via electrochromic FRET. *Nat. Commun.* **5**, 4625 (2014).

270. Abdelfattah, A. S., Rancic, V., Rawal, B., Ballanyi, K. & Campbell, R. E. Ratiometric and photoconvertible fluorescent protein-based voltage indicator prototypes. *Chem. Commun. (Camb.)* **52**, 14153–14156 (2016).

271. Abdelfattah, A. S. *et al.* A Bright and Fast Red Fluorescent Protein Voltage Indicator That Reports Neuronal Activity in Organotypic Brain Slices. *J. Neurosci.* **36**, 2458–2472 (2016).

272. Ma, G. *et al.* Optogenetic toolkit for precise control of calcium signaling. *Cell Calcium* **64**, 36–46 (2017).

273. Saravolatz, L. D., Depcinski, S. & Sharma, M. Molnupiravir and nirmatrelvir-ritonavir: Oral COVID antiviral drugs. *Clin. Infect. Dis.* (2022) doi:10.1093/cid/ciac180.

274. Vuong, W. *et al.* Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun.* **11**, 4282 (2020).

275. Wu, J. *et al.* A long Stokes shift red fluorescent Ca2+ indicator protein for two-photon and ratiometric imaging. *Nat. Commun.* **5**, 5262 (2014).

276. Nagai, T. & Miyawaki, A. A high-throughput method for development of FRET-based indicators for proteolysis. *Biochem. Biophys. Res. Commun.* **319**, 72–77 (2004).

277. Chiu, T.-Y., Chen, P.-H., Chang, C.-L. & Yang, D.-M. Live-cell dynamic sensing of Cd(2+) with a FRET-based indicator. *PLoS One* **8**, e65853 (2013).

278. Shen, Y. *et al.* A genetically encoded Ca2+ indicator based on circularly permutated sea anemone red fluorescent protein eqFP578. *BMC Biol.* **16**, 9 (2018).

279. Oscar, B. G. *et al.* Excited-state structural dynamics of a dual-emission calmodulin-green fluorescent protein sensor for calcium ion imaging. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10191–10196 (2014).

280. Shen, Y., Nasu, Y., Shkolnikov, I., Kim, A. & Campbell, R. E. Engineering genetically encoded fluorescent indicators for imaging of neuronal activity: Progress and prospects. *Neurosci. Res.* **152**, 3–14 (2020).

281. Chen, C. *et al.* Ultrafast excited-state proton transfer dynamics in dihalogenated non-fluorescent and fluorescent GFP chromophores. *J. Chem. Phys.* **152**, 021101 (2020).

282. Chen, C. & Fang, C. Devising efficient red-shifting strategies for bioimaging: A generalizable donor-acceptor fluorophore prototype. *Chem. Asian J.* **15**, 1514–1523 (2020).