UNIVERSITY OF CALIFORNIA
RIVERSIDE

Large Stable Single Clone Combinatorial Libraries in Mammalian Cells ---- Platform Development and Applications

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemical and Environmental Engineering

by

Zening Wang

December 2021

Dissertation Committee:
Dr. Xin Ge, Chairperson
Dr. Ashok Mulchandani
Dr. Rong Hai

The Dissertation of Zening Wang is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

Seven years ago, when I embarked on the journey to the United States, I did not expect that it would take me four years at UW-Madison ending up with a master's (not Ph.D) degree due to a mismatch of research interest. Despite this, I would not say those years were spent in vain. On the contrary, the research experiences as well as the intensive and high-quality courses there have built a solid foundation for me to maneuver in a new research field, resulting in a Ph.D. degree in a shorter period of time at UC, Riverside. Reflecting on these years, there are many people to whom I owe my gratitude for helping me persist in what I am truly passionate about and achieve my goals.

The most prominent person that I want to thank is my advisor, Professor Xin Ge, who welcomed me enthusiastically to join his lab working in the field of immune engineering that I found so fascinating. I still remember the first time that I met with Prof. Ge when I attended the open house back in March 2018. At the dinner party, Prof. Ge shared with me that he benefited a lot from Prof. George Georgiou's guidance during his postdoctoral training at UT, Austin. "My PI treated me so well that I decided to do the same for my students." ---- Prof. Ge's words still echo in my mind with his tremendous and continuous support to me till this day. I would like to thank him for his trust in me to work on important projects. Significant projects often come with great challenges. Prof. Ge has been very approachable guiding me to overcome the difficulties regardless of his hectic schedule. He always recognizes my efforts by even pushing me not to work so hard, which is so reassuring that I can clear my thoughts to adjust my mindset and solve the problems. Throughout the years, Prof. Ge has cultivated my ability of conducting

research efficiently as well as delivering the research findings vividly as manifested by this dissertation. He has broadened my horizons with his insights and support in various ways, helping me grow as a scientist with connections and visions in this field.

Next, I would like to thank Professor Ashok Mulchandani and Professor Rong Hai who have been my committee members since my advancement to candidacy. Their advice and insightful comments have given me unique perspectives on my research and enabled me to improve my dissertation towards perfection. I also appreciate their generous compliments throughout the defense, which encouraged me to work harder in the future.

I am also deeply indebted to Dr. Chuan Chen who used to be a postdoc in our lab. When I initially joined the lab, I was having zero experience in mammalian cell experiments. It was Chuan who taught me all the technique without any reservation. I could not express my gratitude enough to him. I would also like to thank Minhyo Kang who was the master student in our lab for her numerous hours of work to advance my research. Without Chuan and Minhyo, my research would not have been able to be completed nearly as quickly. I wish them all the best and a bright future ahead of them.

I would like to thank my lab members Dr. Kibaek Lee for his help in surface plasmon resonance experiments and Dr. Hyunjun Choe for sharing with me his technique which is very useful in my research. I would like to thank Anqi Zhou for her generous help on the lab supplies when we were in a shortage. I would like to thank the purchasing agent Deanna Vorhees in our department and staffs in corporation yard for their excellent work on logistics in these unprecedent times.

Due to the pandemic, the dissertation defense could be held virtually, making it possible for my parents to attend and witness my achievements from the mainland of China. I could not hold my tears when the picture of me in childhood with my parents in their early 40s showed up on my last slide. There has been a saying that parents are the wall between us and death. Just knowing their presence in the world, I feel so fearless to pursue what I want, not to mention with their unconditional love and support. Thank you, mom and dad, for giving birth to me, raising me to such a healthy and happy person and everything. I hope I could see you in person soon and spend more time taking care of you in the future. I would also like to thank my dear boyfriend Bowei Wang for his company, sharing the joy and tears together at our ups and downs during the last four years. Thank you for making my life even more colorful! May more years to come!

Finally, this research could not be possible without the funding from NIH and UCR.

Thank you all!

To my mom and dad

ABSTRACT OF THE DISSERTATION

Large Single Clone Combinatorial Libraries in Mammalian Cells
---- Platform Development and Applications

by

Zening Wang

Doctor of Philosophy, Graduate Program in Chemical and Environmental Engineering
University of California, Riverside, December 2021
Dr. Xin Ge, Chairperson

Engineering human-related complex proteins has great potential in development of effective therapeutics. Paramount to any directed evolution of such biologics is the construction of large libraries in mammalian cells that provide a native expression environment and allow the direct screening / selection of tens of millions of protein variants for desired properties. The common mammalian cell library construction platforms, such as episomal vectors, viral vectors and transposons, have the issues including low transgene stability, random integration and tug of war between the large library size and single variant per cell. Recently, targeted integration with nucleases has been developed whereas the integration efficiency is low, and off-target effects could result in multiple variants in one cell.

To address these limitations, we developed a mammalian cell library construction platform based on recombinase-mediated cassette exchange (RMCE) which directs the

integration of the transgenes into a single genomic locus, thereby rapidly achieving stable expression and transcriptional normalization. To improve the intrinsically low RMCE efficiency, we developed two complemental approaches: (1) simian virus 40 (SV40) large T-mediated transgene replication in Expi293F™ cells, and (2) nocodazole-mediated CHO cell mitosis phase arrest which not only presents two copies of targeted genomic locus but also facilitates their access due to temporary nuclear membrane disintegration.

Applying these novel approaches, mammalian cell libraries of up to 10 million clones were constructed, and library diversities were validated by deep sequencing and statistical analysis. Furthermore, the utility of the system was illustrated by isolation of GFP variants with enhanced fluorescence, and antibody Fc variants with significantly improved affinities towards specific Fc gamma receptors (FcγRs) important for therapeutic antibody effector functions. Collectively, we developed platform technologies for combinatorial library construction in mammalian cells satisficing all desired features including large diversity, single copy, genomic integration, and defined locus. With contribution to the field of protein engineering, our work will facilitate the discovery of biologics and eventually lead to more and effective therapeutic strategies.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Importance of Mammalian Cell Library

Directed evolution is a common method used in protein engineering that mimics the process of natural selection in a test tube to identify proteins with desired biological functions. The first step of directed evolution is to generate a diverse library of variants followed by screening/selection. Up to date, the construction of libraries in phage[1,2], bacteria[3,4] and yeast[5,6] has enabled the evolution of numerous proteins with new or improved properties, such as reporter proteins, industrial enzymes, antibody fragments. However, engineering proteins with complex structures, such as full-length IgG, T cell receptors or human therapeutic enzymes, is in dire need of mammalian cell systems which harness the expression and secretory apparatus favorable for the production of human proteins properly folded with human-like glycosylation patterns and other post-translational modifications[7].

1.2 Current Mammalian Cell Library Construction Platforms

During the last two decades, several approaches for mammalian cell library construction have been developed, including transient transfection, viral vectors, transposons, nucleases and recombinases. A summary of their characteristics is shown in **Table 1.1**. Initially, transient transfection has been used for library construction in mammalian cells[8–10]. Although relatively large library size could be achieved with this method, the introduced plasmids by transient transfection remain episomal, resulting in

loss of overall DNA within the course of a few days and consequently inefficient selection. The field thus shifted rapidly to genome-integrated library generation possessing high stability of exogenous DNA in cells. Here, we discuss with an emphasis on the methods for stable mammalian cell library construction.

Table 1.1 Summary of current mammalian cell library construction methods and characteristics

Methods	Transgene stability	Copy number per cell	Transcription variation	Transfection/Integration efficiency	Cargo capacity
Transient transfection/ episomal vectors	Low	>> 1	High	High	High
Viral vectors (e.g., Retrovirus, lentivirus)	High	≥ 1	High	High	High
Transposons (e.g., *sleeping beauty*, *piggyBac*)	High	≥ 1	High	High	High
Nucleases (e.g., CRISPR/Cas9, TAELNs)	High	= 1 or 2	Low	Low	Low
Recombinases (e.g., Flp, Cre, Bxb1)	High	= 1	Low	Low	High

1.2.1 Random integration

Viral infection[11] (e.g., retroviruses or lentiviruses) and transposons[12] (e.g., *sleeping beauty* or *piggyBac* transposase) are the two main approaches for random integration. Both have large cargo capacity (8 kb reported for retrovirus[13] and 10 kb for *PiggyBac* transposases[14,15]) without apparent loss of integration efficiency. One aspect of transposons superior to viral-based systems is the simplicity of library preparation. The latter requires a multitude of steps including transfection of packaging cells, generation of viral particle intermediates, and finally infection of library host cells with viral particles[15]. Although both approaches could achieve high integration efficiency, multiple genes encoding different protein variants could be introduced into each cell, impeding the identification of positive clones. This is even more disastrous for antibody engineering, which could cause erroneous combinations of heavy and light chains. To obviate this, the DNA utilized for transfection or viral particles needs to be titrated carefully, thereby enabling an average of one protein variant per transfected cell. However, the library size is compromised at the meantime. In addition, given that the integration within the genome is random, transcription level of different clones could be varied based on the transcriptional activity of the integration locus.

1.2.2 Targeted integration

The fast development of genome editing has created the opportunity for targeted integration which avoids the undesired outcomes associated with random integration by inserting DNA into predetermined genomic locations. This includes nuclease-based platform, represented by CRISPR/Cas9 and TALE nucleases (TALENs), and site-specific

recombinase (SSR)-based system (e.g., Flp, Cre and Bxb1). Since their integration efficiencies are generally low, there is a compelling need to boost it for large mammalian cell library construction.

1.2.2.1 Nuclease-based methods

Targeted genome integration of foreign DNA mediated by nucleases can be mechanistically divided into two steps: (1) the creation of double-strand break (DSB) by nucleases at target site, and (2) the repair of the induced DSB with homology directed recombination (HDR) in the presence of donor transgenes carrying variants[16]. Efforts to enhance the HDR efficiency have been made to increase the genetic diversity and library size.

Mason et al[17] first integrated antibody gene to a hybridoma cell line and utilized degenerated single-stranded oligonucleotides (ssODNs) as donor template at individual complementarity determining regions (CDRs). The authors reported a 35-fold improvement of the Cas9-mediated integration efficiency by optimizing transfection parameters, inhibiting the DNA repair regulator protein and including phophorothioate bonds to ssODNs. Nevertheless, this strategy has its limitation in small cargo size (~ 120 nucleotides). Besides, construction of the library from 10^7 transfected cell resulted in a diversity of 1.47×10^5, equivalent to a 1.47% efficiency.

With a different set-up where dual promoter, full-length IgG-formatted antibody genes were introduced by nucleases to pre-determined site AAVS on genome, the group of McCafferty[18] reported an efficiency of 2.7%-5.1% using TALEN by Maxcyte

electroporation. Notably, this system requires long homology arms (800 base pairs) at left and right sides of the antibody gene and may have high background of non-integrated clones due to the dual promoter present at the upstream of the light and heavy chain genes. Moreover, this method has the issues of off-target and integration to both alleles on the genomes, resulting in a possibility of more than one copy per cell. Eventually, a mammalian display library size of 1.08×10^7 was constructed from transfection of 1.35×10^9 HEK293 suspension cells, equivalent to an efficiency of 0.8%.

1.2.2.2 Recombinase-based methods

Another popular targeted integration method relies on recombinases including the bacteriophage P1 Cre-*loxp* and yeast's Flp-*FRT* that belong to the tyrosine recombinase family, as well as the Int (Bxb1)-*att* system of phage ϕC31 that belongs to the serine family recombinase family[19–22]. In these systems, the site-specific recombinases catalyze the unidirectional integration between two short recognition sites (~30-40 bp), one on the plasmid that carries the gene of interest (GOI) and the other one, identical or similar, located in a chromosome (**Fig. 1.1A**). Unlike nuclease, this whole process has yet to find exposed to the endogenous cellular DNA repair machinery, leading to more defined and precise integration outcomes[16,20]. It also has no limitations to the cargo size. In fact, a recent study showed single-copy genomic insertion of 27 kb synthetic gene circuit in HEK293T cells[23]. To optimize this important system, several protein engineering campaigns have been carried out to improve activity[24,25] stability[26] and codon usage[27,28] of the recombinases and to alter their cognization site specificity[29–31]as well. For instance, since Flp recombinase originates from yeast which grows at 30 °C, it shows thermo-

instability in mammalian cell and thus low recombinase activity. Via directed evolution, Buchholz et al[24] identified thermo-stable Flpe (P2S/L33S/Y108N/S294P) with a 4-fold increase in recombination efficiency. Recently, an improved recombination rate has been achieved using an alternative serine-based recombinase Bxb1, yet only a library of 20,000 clones yielded[32].

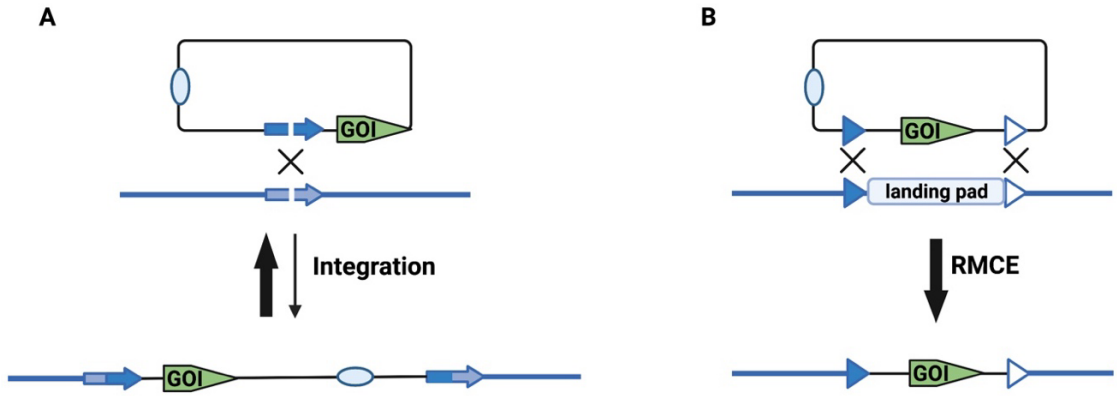


Figure 1.1 Comparison between **(A)** site-specific recombination and **(B)** RMCE to deliver a gene of interest (GOI) into a desired genome location.

One drawback of such integration reaction is that the entire vector plasmid including unrelated elements gets inserted into the genome. Moreover, tyrosine recombinases are reversible[20], leading to more efficient repeated cycles of recombination and excision in the presence of more substrates, i.e., two identical tandem targe sites. To achieve a neat gene insertion, a dual recombinase-mediated cassette exchange (RMCE) reaction was developed with the GOI flanked by a pair of incompatible recognition sites[21,33]. The same set of recognition sites is pre-integrated to the genome to create a landing pad where the exchange plasmid borne GOI could settle leaving behind undesired plasmid sequences.

Moreover, the reversibility of this cassette exchange reaction is low because the compatible combination sites are located on different DNA molecules[21]. Anderson et al[33] improved RMCE efficiency by optimizing Flp/Cre expression plasmids. The authors found Flp and Cre recombinases expressed as Flp-2A-Cre or Flp-IRES-Cre showed higher RMCE efficiency than expressed individually. However, a decent size of SSR-based mammalian cell library has barely been reported compared to nuclease-based systems.

Zhou et al[34] described the presentation of full-length antibodies on the surface of CHO cells by commercial Flp-In technology without diversity reported. Following investigations by other groups have proved its deficiency in large library construction[18,32]. Chen et al[35] performed sequential rounds of RMCE to obtain single landing pad on the CHO cell genome, displayed antibody single-chain variable fragment (scFc) on the cell surface and coupled with somatic hyper mutation for affinity maturation. Nevertheless, affinity improved binders were not isolated until 5 founds of selection, indicating an incompetent system.

1.3 Research Objectives and Thesis Structure

A desired mammalian cell library should encompass large diversity of transgenes with single copy integrated to a defined genomic locus. Assessing the current mammalian cell library construction platforms, only RMCE-based systems could meet those requirements with only one exception ---- the library size. Therefore, the comprehensive

objective of my study is to improve the RMCE efficiency and ultimately construct large mammalian cell library with the ideal characteristics.

RMCE-dependent genomic integration is basically an enzymatic reaction using both GOI and a landing pad as the substrates. Considering substantial efforts have been made to enhance the recombinases themselves, this study focuses on other aspects aiming to improve RMCE efficiency. Based on fundamental principles of chemical engineering, "*the higher substrate and enzyme concentrations, the higher the reaction rate*" and "*the longer reaction time, the more product*", we hypothesize that the RMCE efficiency could be improved by (1) a high copy number of GOI and more recombinases, (2) more available landing pads, and (3) an easier and longer access for GOI and recombinases to the landing pad on the genome. The upcoming chapters present two complementary approaches to fulfill these aspects, followed by large mammalian cell library construction, library size validation by deep sequencing, and successful isolation of affinity improved binders.

Chapter 2 exhibits that simian virus 40 (SV40) large T mediated DNA replication facilitated RMCE by maintaining a high copy number of GOI in transfected cells.

Chapter 3 demonstrates that cell cycle arrest at mitosis phase enhanced RMCE efficiency presumably by removing the nucleus barrier for the genome access, temporarily doubling the amount of landing pad, and pausing chromatin condensation.

Chapter 4 presents the construction of varied size of libraries via cell cycle arrested RMCE (aRMCE) and their diversity validation by deep sequencing and bioinformatics.

Chapter 5 explores an application of the developed aRMCE mammalian cell platform for Fc engineering.

Chapter 6 summarizes overall achievements of the thesis, outlines further improvement strategies, and envisions future applications.

1.4 References

1. Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science* 228, 1315–1317 (1985).

2. Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. Making Antibodies by Phage Display Technology. *Annu Rev Immunol* 12, 433–455 (1994).

3. FRANCISCO, J. A. & GEORGIOU, G. The Expression of Recombinant Proteins on the External Surface of Escherichia coli. *Ann Ny Acad Sci* 745, 372–382 (1994).

4. Mazor, Y., Blarcom, T. V., Iverson, B. L. & Georgiou, G. Isolation of full-length IgG antibodies from combinatorial libraries expressed in Escherichia coli. *Methods Mol Biology Clifton N J* 525, 217–39, xiv (2009).

5. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* 15, 553–557 (1997).

6. Feldhaus, M. J. *et al.* Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. *Nat Biotechnol* 21, 163–170 (2003).

7. Valldorf, B. *et al.* Antibody display technologies: selecting the cream of the crop. *Biol Chem* 0, 000010151520200377 (2021).

8. Higuchi, K. *et al.* Cell display library for gene cloning of variable regions of human antibodies to hepatitis B surface antigen. *J Immunol Methods* 202, 193–204 (1997).

9. Ho, M., Nagata, S. & Pastan, I. Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. *Proc National Acad Sci* 103, 9637–9642 (2006).

10. Bowers, P. M. *et al.* Coupling mammalian cell surface display with somatic hypermutation for the discovery and maturation of human antibodies. *Proc National Acad Sci* 108, 20455–20460 (2011).

11. Kotterman, M. A., Chalberg, T. W. & Schaffer, D. V. Viral Vectors for Gene Therapy: Translational and Clinical Outlook. *Annu Rev Biomed Eng* 17, 63–89 (2015).

12. Hickman, A. B., Chandler, M. & Dyda, F. Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit Rev Biochem Mol* 45, 50–69 (2010).

13. Walther, W. & Stein, U. Viral vectors for gene transfer: a review of their use in the treatment of human diseases. *Drugs* 60, 249–71 (2000).

14. Ding, S. *et al.* Efficient Transposition of the piggyBac (PB) Transposon in Mammalian Cells and Mice. *Cell* 122, 473–483 (2005).

15. Waldmeier, L. *et al.* Transpo-mAb display: Transposition-mediated B cell display and functional screening of full-length IgG antibody libraries. *Mabs* 8, 726–740 (2016).

16. 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).

17. 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).

18. Parthiban, K. *et al.* A comprehensive search of functional sequence space using large mammalian display libraries created by gene editing. *Mabs* 11, 884–898 (2019).

19. Groth, A. C., Olivares, E. C., Thyagarajan, B. & Calos, M. P. A phage integrase directs efficient site-specific integration in human cells. *Proc National Acad Sci* 97, 5995–6000 (2000).

20. Grindley, N. D. F., Whiteson, K. L. & Rice, P. A. Mechanisms of Site-Specific Recombination*. *Annu Rev Biochem* 75, 567–605 (2006).

21. Voziyanova, E. *et al.* Efficient Flp-Int HK022 dual RMCE in mammalian cells. *Nucleic Acids Res* 41, e125–e125 (2013).

22. Voziyanova, E., Anderson, R. P. & Voziyanov, Y. Site-Specific Recombinases, Methods and Protocols. *Methods Mol Biology* 1642, 53–67 (2017).

23. Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res* 42, 13440–13451 (2014).

24. Buchholz, F., Angrand, P.-O. & Stewart, A. F. Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* 16, 657–662 (1998).

25. Voziyanova, E., Anderson, R. P., Shah, R., Li, F. & Voziyanov, Y. Efficient Genome Manipulation by Variants of Site-Specific Recombinases R and TD. *J Mol Biol* 428, 990–1003 (2016).

26. Buchholz, F., Ringrose, L., Angrand, P.-O., Rossi, F. & Stewart, A. F. Different Thermostabilities of FLP and Cre Recombinases: Implications for Applied Site-Specific Recombination. *Nucleic Acids Res* 24, 4256–4262 (1996).

27. Shimshek, D. R. *et al.* Codon-improved Cre recombinase (iCre) expression in the mouse. *Genesis* 32, 19–26 (2002).

28. Raymond, C. S. & Soriano, P. High-Efficiency FLP and ΦC31 Site-Specific Recombination in Mammalian Cells. *Plos One* 2, e162 (2007).

29. Buchholz, F. & Stewart, A. F. Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat Biotechnol* 19, 1047–1052 (2001).

30. Shah, R., Li, F., Voziyanova, E. & Voziyanov, Y. Target-specific variants of Flp recombinase mediate genome engineering reactions in mammalian cells. *Febs J* 282, 3323–3333 (2015).

31. Bolusani, S. *et al.* Evolution of variants of yeast site-specific recombinase Flp that utilize native genomic sequences as recombination target sites. *Nucleic Acids Res* 34, 5259–5269 (2006).

32. 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, gkx183- (2017).

33. Anderson, R. P., Voziyanova, E. & Voziyanov, Y. Flp and Cre expressed from Flp–2A–Cre and Flp–IRES–Cre transcription units mediate the highest level of dual recombinase-mediated cassette exchange. *Nucleic Acids Res* 40, e62–e62 (2012).

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

35. Chen, C., Li, N., Zhao, Y. & Hang, H. Coupling recombinase-mediated cassette exchange with somatic hypermutation for antibody affinity maturation in CHO cells. *Biotechnol Bioeng* 113, 39–51 (2016).

Chapter 2: SV40 Replication Facilitated RMCE Efficiency Improvement

Abstract

Built on the observation that simian virus 40 large T antigen (TAg) indues replication of plasmids bearing a SV40 origin within mammalian cells, this Chapter develops a TAg facilitated recombinase-mediated cassette exchange (RMCE) approach, T-RMCE, with improved efficiency of genomic integration by maintaining a high copy number of the exchange plasmid per cell. In RMCE competent polyclonal 293F cells, expression of TAg increased the RMCE efficiency by 3.4-fold. 293F clones of single landing pad were then generated by two sequential dual RMCE events, of which 15 single cell clones showed RMCE efficiency of 5.5-10.6% in the presence of SV40 TAg. Particularly, for clone D2, the expression of SV40 TAg was confirmed by Western blot, and quantifications of the exchange plasmid by qPCR suggested that the presence of TAg significantly increased the relative copy numbers of the exchange plasmid, e.g., 10.6-fold at day 2, 19.8-fold at day 3, and 7.5-fold at day 4 after transfection. Accordingly, the optimal cultivation time of 4 days was identified for T-RMCE. Finally, a mammalian cell green fluorescent protein (GFP) library size of 1×10^7 was yielded and GFP variants with enhanced fluorescence were isolated after 2 rounds of fluorescence activated cell sorting. The T-RMCE platform technology developed here can be readily applied for protein engineering in mammalian cells.

2.1 Introduction

As a member of *Polymaviridae* family, simian virus 40 (SV40) carries a circular ~5kb double-stranded DNA genome, encoding three structural virion proteins (VP1-3) and two nonstructural proteins called small T antigen (t antigen) and large T antigen (T antigen, or TAg). The 708 aa length multi-domain large T antigen belongs to the early coding unit of SV40, and performs numerous functions including viral DNA replication, virion assembly, and transcriptional regulations[1]. SV40 viral DNA replication is initiated by binding of T antigen to its origin of replication to form a double-hexametric structure allowing conjunction with the cellular replication apparatus[2–4]. In the presence of SV40 TAg, transfected plasmids bearing a SV40 origin can be replicated and maintained at a high copy number[5–7]. Notably, the widely used 293T cell line was generated by stable transfection of SV40 TAg in HEK 293 to enhance transgene expression from vectors carrying the SV40 region of replication[8].

It is reasoned that the higher the cellular titer of the exchange plasmid, the higher the efficiency of recombinase-mediated cassette exchange (RMCE). In this Chapter, we test the hypothesis that expression of SV40 TAg can maintain a high copy number of the exchange plasmid carrying a SV40 origin and thereby improve the RMCE efficiency (**Fig 2.1**). Instead of using adherent cell line 293T, we tested this hypothesis using 293F suspension cells for convenient handling and high scalability. We termed the developed TAg facilitated RMCE approach as T-RMCE.

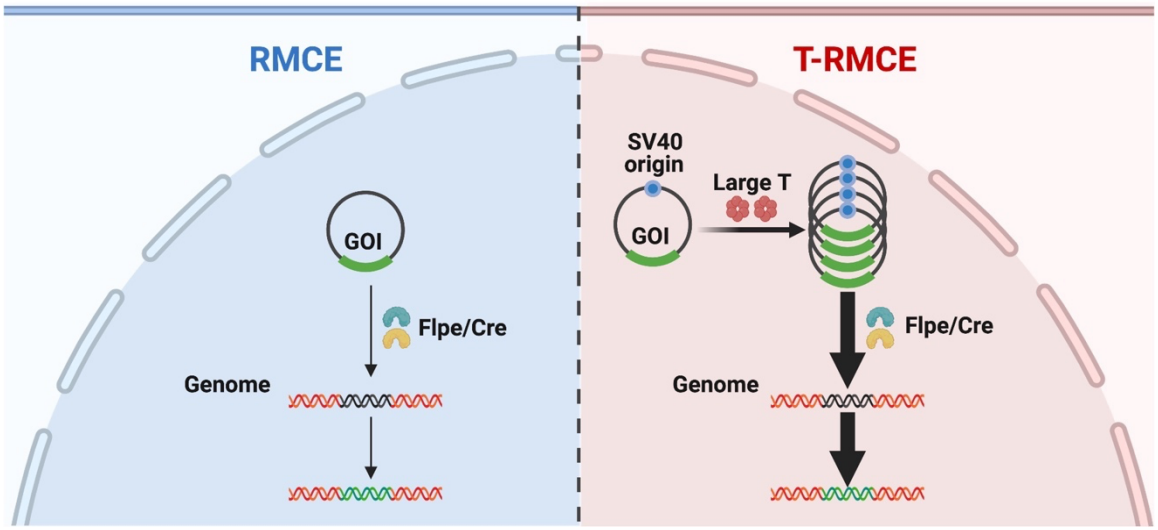


Figure 2.1 Proposed mechanism of the RMCE efficiency enhancement by SV40 Large T. In RMCE (**left**), the exchange plasmid carrying the gene of interest (GOI) was integrated to the landing pad on the genome by the recombinases Flpe and Cre. The RMCE efficiency is low due to the limited number of GOI. In T-RMCE (**right**), the exchange plasmid bearing SV40 origin can be replicated by Large T, providing more GOI and thereby increasing the RMCE efficiency.

2.2 Materials and Methods

2.2.1 Plasmid construction

To reduce the size of plasmids used in RMCE, pcDNA3.1$^{(+)}$ were digested with PvuII, and the obtained 3.3 kb fragment was gel-purified and self-ligated to give pcDNA, in which the hygromycin resistant gene cassette and f1 ori were removed.

For exchange plasmids, the 136 bp double-strand DNA encoding SV40 origin was amplified using pcDNA3.1$^{(+)}$ as the template and cloned into pcDNA using BglII and BamHI to give the plasmid pcDNA-SV40ori, in which the CMV promoter was removed. The gene cassette of FRT-iRFP-loxP was chemically synthesized (IDT) and cloned into

pcDNA-SV40ori with BglII resulted in the exchange plasmid pSV-iRFP. pSV-Puro and pSV-EGFP were prepared similarly.

For the landing pad plasmid encoding hygromycin and EGFP, ppHyFGL, the DNA fragment encoding PGK promoter was chemically synthesized (IDT) and cloned into pcDNA3.1/Hygro(+) at NsiI and XmaI sites to generate pcDNA3.1/PGK-Hygro, in which the SV40 promoter and ori in front of hygromycin resistant gene is replaced with a PGK promoter. The DNA fragment encoding FRT-EGFP-loxP was chemically synthesized (IDT) and cloned into pcDNA3.1/PGK-Hygro at HindIII and XhoI sites to give ppHyFGL. For the landing pad plasmid encoding puromycin, ppPuro was constructed similarly. In brief, the DNA fragments encoding PGK promoter and FRT-Puro-loxp were fused by overlapping PCR then inserted into pcDNA at BglII and BamHI sites to give ppPuro.

For large T expression, genomic DNA of HEK293T cells was extracted with Wizard Genomic DNA Purification Kit (Promega, Cat#A1120) following manufacturer's manual. From the genomic DNA, SV40 TAg gene (Gene ID 29031019) was amplified by PCR and cloned into pcDNA at NheI and XbaI sites to give pLargeT.

For recombinase expression, pcDNA was modified by insertions of a chimeric intron[9] and a Woodchuck hepatitis virus posttranscriptional regulatory element[10] (WPRE) at NheI/KpnI and XhoI/XbaI sites respectively. Obtained pCIW was digested with KpnI/XhoI for cloning Flpe(F70L)[11,12] and Cre recombinases separated by the T2A self-cleaving peptide to generate pF2AC.

GFP mutations were generated by error-prone PCR using the plasmid carrying a GFP clone Q80R gene as the template. The reaction mixture was 10 mM Tris·HCl (pH 8.3), 50 mM KCl, 6.5 mM $MgCl_2$, 1 mM dCTP, 1 mM dTTP, 0.2 mM dATP, 0.2 mM dGTP, 0.3 µM 5′ primer, 0.3 µM 3′ primer, 25.6 pM template DNA pSV-EGFP, 0.5 mM $MnCl_2$, 0.05 U/µl Taq DNA polymerase. The thermal cycling program was 95 °C for 2 min; 25 cycles of 95 °C for 15 s, 56 °C for 30 s, 68 °C for 1 min; final extension at 68 °C for 5 min and storage at 4 °C. The mutated GFP fragments were cloned into pSV-Puro at HindIII and XhoI sites. The ligation mixture was transformed into *E. coli* SS320 competent cells. Library plasmid DNA pSV-GFPm was then miniprepped and ready for transfection.

All restriction enzymes used were purchased from NEB.

2.2.2 Cell culture and transfection

Expi293F™ cells (Thermo Fisher Scientific) were maintained in Expi293 expression media (Gibco, Cat#A1435101) supplemented with 40 U/ml penicillin, and 40 µg/ml streptomycin (Gibco, Cat#15140122) in 125 ml flasks at 37 °C under 8% CO_2 in humidified atmosphere with orbital shaking at 135 rpm. Two days before transfection, cells were maintained in Expi293 expression medium without antibiotics, and 30 min before the transfection, cells were diluted with fresh medium at the density of 10^6 cells per ml.

To generate RMCE ready 293F-puro polyclonal cells, 2×10^7 Expi293F cells were transfected with 10 µg landing pad plasmid ppPuro as well as 30 µg PEI in 10 ml

Expi293 media and incubated for 3 days. Cells were then cultured in fresh media supplemented with 5 μg/ml puromycin (Gibco, Cat#A1113803) for 7 days for selection and subsequently maintained in media with 1 μg/ml puromycin. A similar transfection of Expi293 cells with ppHyGFL was conducted to generate 293F cells carrying the landing pad encoding EGFP, 293F-EGFP, which were selected and maintained with 250 and 50 μg/ml hygromycin respectively. For isolation of 293F-puro single clones, at day 2 after 293F-iRFP cells were transfected with pSV-Puro, cells were seeded into 96-well plates at 200 cells per well and selected in DMEM supplemented with 10% FBS and 5 μg/ml puromycin. After 7-day cultivation, 293F-Puro single cell clones were identified and expanded for further experiments.

In a typical RMCE transfection, RMCE ready 293F cells were seeded in a 14 ml culture tube (VWR Cat#60818-623) at a density of 1×10^6/ml in 1.8 ml Expi293 media. Cell were transfected by adding a mixture of 1 μg exchange plasmid, 1 μg pF2AC and 6 μg PEI MAX 40K (Polysciences, Cat#24765-1) prepared in 0.2 ml Expi293 media, and culture in CO_2 incubator with orbital shaking at 180 rpm for 24 h. For large T facilitated RMCE, 2×10^6 cells were transfected with a mixture of 0.5 μg exchange plasmid, 0.5 μg pLargeT, 1 μg pF2A C and 6 μg PEI. Controls were performed without pLargeT or pF2AC but replaced with same amount of pcDNA. For GFP library transfection, 1×10^8 293F-Puro D2 cells were transfected with 25 μg pSV-GFPm, 25 μg pLargeT, 50 μg pF2AC and 400 μg PEI in one 500 ml flask. Similarly, for qPCR and western blot analysis, 5×10^6 293F-Puro D2 cells were transfected in 125 ml flasks with 2.5 μg pSV-EGFP, 2.5 μg pLargeT or pcDNA and 15 μg PEI.

2.2.3 Flow cytometry

Cell analysis and sorting was performed on a S3e cell sorter (Bio-Rad), equipped with 488/640 nm dual lasers and multiple bandpass/longpass filters. EGFP/GFP mutants and iRFP were excited at 488 nm and 640 nm and detected with 526/48 nm filter (FL1 channel) and 700 nm/LP filter (FL4 channel), respectively. In general, 50,000 events were collected for analysis; FACS for GFP variants screening was performed in purity mode at 2000 – 3000 events/s and the top 0.1-0.3% of the GFP positive cells were collected.

2.2.4 qPCR

5×10^6 293F-Puro D2 cells were incubated with transfection mixtures for 24 h and washed 3 times with 10 ml PBS to remove extracellular plasmids. Cells were seeded in 5 ml fresh media in a new flask and cultured for another 24 h. Then 1×10^6 cells were withdrawn, and the remaining cells were cultured in 5 ml fresh media for 24 h. Cell sampling and passaging were repeated for three more days, to obtain cell samples of day 1 to 4 post transfection. The collected cells were washed with 1 ml PBS for 8 times and then underwent genomic DNA extraction with Wizard Genomic DNA Purification Kit.

The qPCR was performed using iQ5 SYBR Green SuperMix (Bio-Rad Cat#170-8880 with the following primers designed by Primer Premier 5 (Premier Biosoft International): GFP-FWD 5'-TGGCCGACAAGCAGAAGAAC-3' and GFP-REV 5'-ACCATGTGATCGCGCTTCTC-3'; Actin-FWD 5'-AGCGAGCATCCCCCAAAGTT-3' and Actin-REV 5'-GGGCACGAAGGCTCATCATT-3'. 25 μl reaction mixtures were

prepared in low profile 96 well PCR plates (Bio-Rad Cat#MLL9601) containing 100 ng template DNA, 100 nM each primer, and 12.5 µl iQ5 SYBR Green SuperMix. Both water and the template DNA from the cells without transfection were used as controls. Actin was used for normalization. Standard curves were generated using four of 10-fold serial dilutions of template DNA at day 1 post transfection with pSV-EGFP and pLargeT. The reaction proceeded on CFX Connect (Bio-Rad) with following conditions: 95 °C for 10 min, then 40 cycles of 95 °C for 15 seconds and 60 °C for 1 min; a final dissociation step was always performed to obtain the melting curves (thermal profile) of the amplicons obtained in the reactions. All qPCR experiments were performed in triplicates except duplicates for standard curves. qPCR data analysis was performed with Pfaffl method[13]. Briefly, the relative EGFP copy number was calculated as $\frac{(E_{Actin}+1)^{Ct_{Actin}}}{(E_{EGFP}+1)^{Ct_{EGFP}}}$ in which E is qPCR efficiency derived from the standard curve.

2.2.5 Western blot

1×10^6 293F-Puro D2 cells transfected with pSV-EGFP and pLargeT or pcDNA were collected and washed with 1 ml PBS once, then resuspended in 100 µl 2 × reducing Lammeli buffer and boiled at 99 °C for 10 min. After centrifugation at 17000 rpm for 5 min, 10 µl of the supernatant sample was separated by 12% SDS-PAGE and transferred to a PVDF membrane. The membrane was blocked with 5% skim milk in PBS overnight at 4 °C followed by incubation with anti-SV40 T Ag antibody HRP (1:2000, Santa Cruz Biotechnology Cat# sc-147 HRP) at room temperature for 2 h. Immunoreactivity was detected using chemiluminescence (Clarity™ Western ECL substrate Bio-Rad Cat#170-

5060). The amount of total protein was assessed by re-probing with anti-GADPH-HRP antibody (1:3000, Abcam Cat# ab105428).

2.2.6 GFP variants identification, production and characterization

Genomic DNA was extracted from 1×10^6 of cells from the GFP library post the 2nd round of FACS using Wizard Genomic DNA Purification Kit. The fragments encoding GFP variants were PCR amplified with 1 µg extracted genomic DNA as the template and 2 µl Herculase II fusion DNA polymerase (Agilent, Cat# 600675) in a 100-µl reaction. The forward primer was specific to 5' untranslated region on the integrated landing pad and the reverse primer recognizes the C-terminal of GFP (Fig. 2.3). The thermal cycling program was 98 °C for 30 s; 30 cycles of 98 °C for 10 s, 63 °C for 20 s, 72 °C for 40 s; final extension at 72 °C for 2 min and storage at 4 °C. The PCR products were gel purified, digested with NheI and EcoRI, and cloned into pMoPac[14] vector with no signal peptide for cytoplasmic expression. The expressed GFP variants were purified by Ni-NTA chromatography (HisPur Ni-NTA resin, Thermo Fisher Scientific #88222) according to the manufacture's protocol. The excitation spectra were measured on BioTek Synergy H1 with emission at 550 nm and excitation from 250 to 520 nm; the emission spectra were measured with excitation at 488 nm and emission from 515 to 700 nm. The fluorescence intensity was indicated by RFU at emission 528 nm excited by 488 nm.

2.3 Results

2.3.1 SV40 TAg expression improved RMCE efficiency in polyclonal cells

SV40 TAg could bind with SV40 origin to initiate DNA replication and maintain the plasmids episomal in high copy number. To test the hypothesis that TAg-based plasmid replication will improve the RMCE efficiency, we first generated a polyclonal stable cell line 293F-RMCE which has the RMCE recombination sites on the genome (**Fig. 2.2A, Step 1**). 293F-RMCE cells were then co-transfected with the EGFP exchange plasmid carrying SV40 origin (pSV-EGFP) and TAg expression plasmid (pLargeT), pSV-EGFP and Flpe/Cre expression plasmid (pF2AC) or pSV-EGFP and pLargeT and pF2AC, respectively (**Fig. 2.2A, Step 2**). At day 3 and 5 post transfection, the percentages of the EGFP positive cells were detected by flow cytometry as an indicator of the RMCE efficiency. As shown in **Fig. 2.2B**, the cells transfected with both pLargeT and pF2AC had higher RMCE efficiency (6.78% at day 3 and 3.96% at day 5) than the ones with only pLargeT (0.55% at day 3 and 0.11% at day 5) or the ones with only pF2AC (2.01% at day 3 and 1.43% at day 5). This result provided us confidence to proceed to single landing pad stable cell line generation for further evaluation.

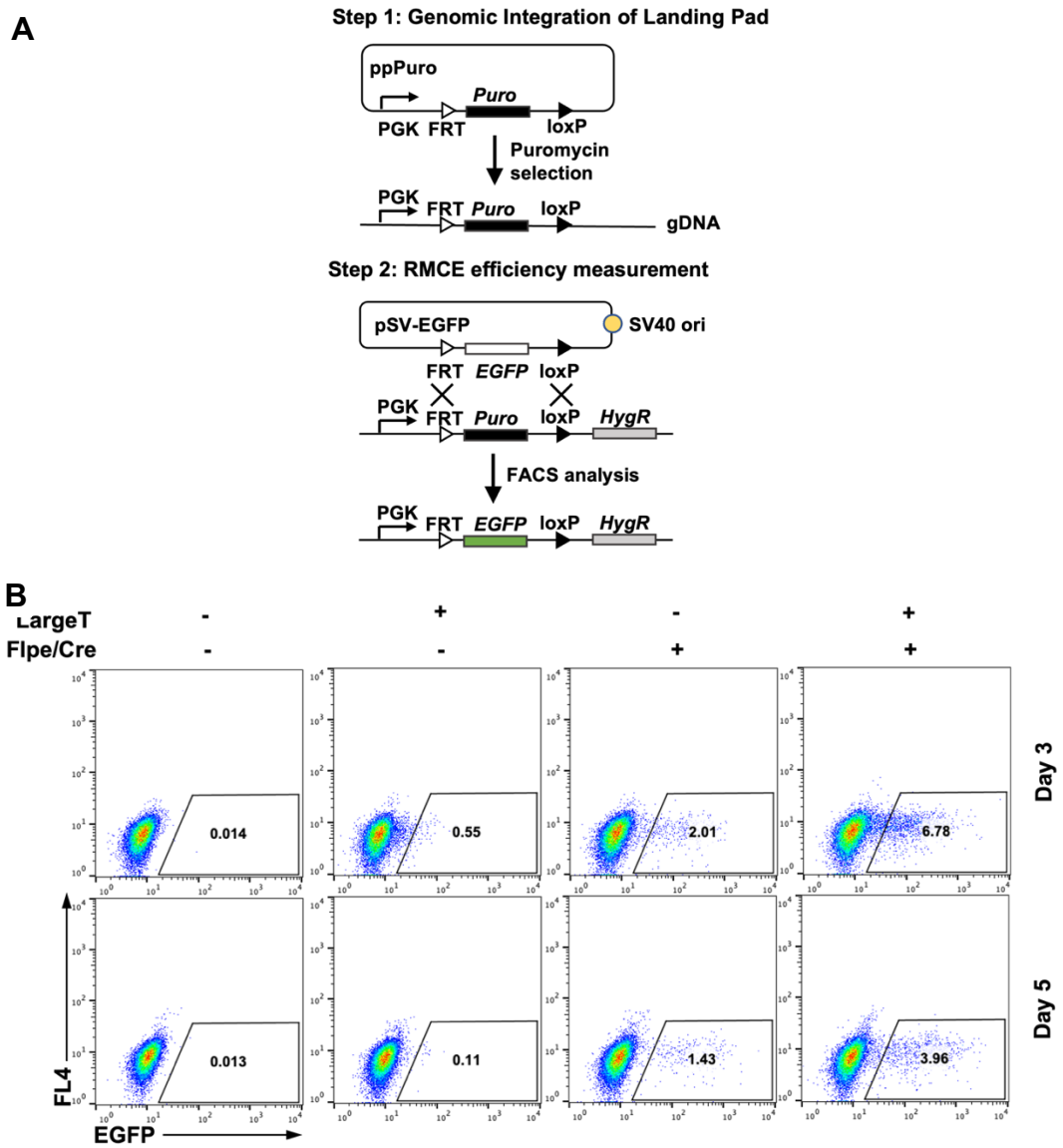


Figure 2.2 Large T mediated GOI replication improved RMCE efficiency in polyclonal cells. **(A)** Workflow of the proof-of-concept experiment. (**Step 1**) Generation of polyclonal cells with RMCE recombination sites on the genome, named 293F-RMCE. (**Step 2**) 293F-RMCE cells were transfected with the EGFP exchange plasmid bearing SV40 origin pSV-EGFP, the Large T expression plasmid pLargeT and the recombinases expression plasmid pF2AC. Large T can bind to the SV40 origin and thereby replicate pSV-EGFP to increase the exchange plasmid number. The RMCE efficiency was indicated by the percentage of the EGFP positive cells which can be determined by flow cytometry analysis. **(B)** Flow cytometry analysis of cells transfected with exchange plasmid pSV-EGFP in the presence or absence of large T and/or Flpe/Cre expression plasmids. EGFP positive cells 3- or 5-days after transfection were detected by flow cytometry.

2.3.2 Generation of single landing pad RMCE cell clones

To generate cell clones with a single landing pad, three sequential steps of transfection and selection / screening were performed[15,16]. In Step 1, Expi293F cells were transfected with a plasmid carrying a CMV promoter, EGFP gene flanked by the FRT and loxP recombination sites (FRT-EGFP-loxP) and a hygromycin resistance marker for selection. Transfected cells were selected with 250 μg/mL hygromycin for 7 days and maintained with 50 μg/mL hygromycin for 8 more days to expand the cells and clear out the free plasmids. Flow cytometry analysis indicated that ~38% of the survived cells were EGFP positive (**Fig. 2.3**, **Step 1**), suggesting that genomic integration was achieved. Although rare, it is possible that more than one landing pads can be integrated per cell. In Step 2, the 293F-EGFP cells obtained in Step 1 were co-transfected with two plasmids, one harboring a promoter-less FRT-iRFP-loxP fragment for exchange and the other encoding an expression cassette for Flpe and Cre recombinases. At day 10 post-transfection, 0.04% cells exhibited iRFP signals, indicating occurrence of RMCE because iRFP can only be produced upon genomic integration into the landing pad which follows a CMV promotor (**Fig. 2.3**, **Step 2**). Approximately 2×10^7 cells were sorted by FACS to collect the RFP^+ only cells. After expansion, flow cytometry confirmed that over 96% of the isolated cells were $iRFP^+$ only. In Step 3, the gnomically integrated iRFP gene was replaced with a puromycin resistance gene. Similarly, a promoter-less exchange vector was used for transfection, and thus the puromycin resistant phenotype can only exhibit upon RMCE occurrence. Following monoclonal culture under puromycin selection, all of the 15 clones we identified were both iRFP and EGFP negative as shown by the flow

cytometer analysis (**Fig 2.3**, **Step 3**). Considering the possibility of simultaneously replacing multiple landing pads in one cell with the same cassette during two successive rounds of RMCE is extremely low, presumably the obtained puromycin-resistant non-fluorescent cells (293F-Puro) should contain a single landing pad.

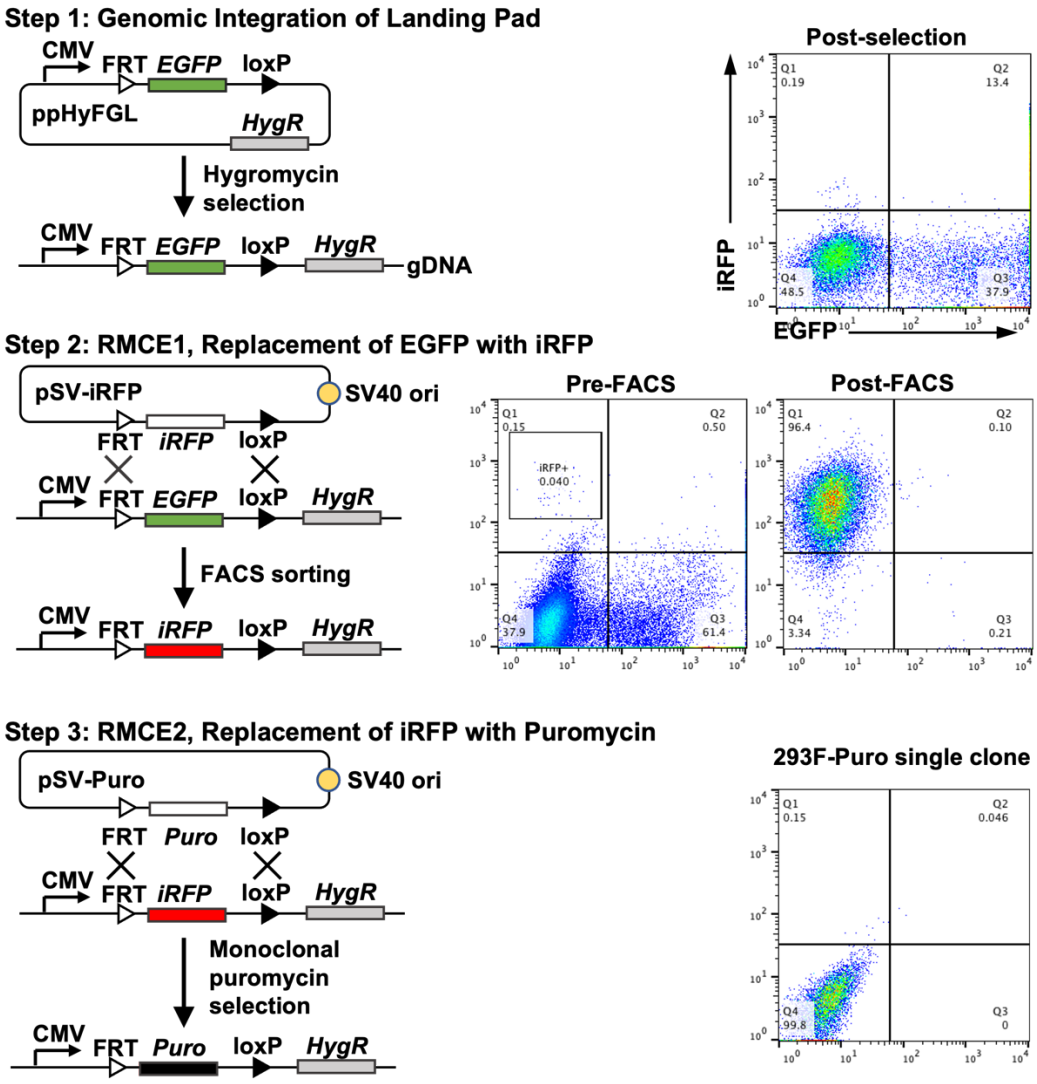


Figure 2.3 Generation of RMCE-competent 293-F cells carrying a single expression landing pad on their genome. (**Step 1**) EGFP expression cassette is randomly integrated into genomic loci under hygromycin selection to give EGFP positive cells (293F-EGFP). (**Step 2**) In RMCE1, the EGFP gene at the landing pad is replaced with an iRFP gene on the exchange plasmid and iRFP^{+} EGFP^{-} cells are selected by FACS to give 293F-iRFP. (**Step 3**) In RMCE2, the iRFP gene at the landing pad is replaced with a puromycin resistance gene on the exchange plasmid to give 293F-Puro. Due to low occurrence of multiple replacements in one cell, two successive rounds of RMCE guarantee that majority of resulting 293F-Puro cells carry a single expression RMCE landing pad. In each step, the empty rectangles on the promoter-less exchange plasmids indicate no expression. In contrast, the filled rectangles on genomes indicate expression enabled when integrated at the downstream of a CMV promotor. Flow cytometry plots show cell populations for each step.

2.3.3 Screening 293F-Puro single cell clones for high RMCE efficiency under TAg expression

To select a candidate who has the highest RMCE efficiency for future large library construction, the 15 single clones were co-transfected with the plasmids for EGFP exchange (SV40 origin included), Flpe/Cre and TAg expression **(Fig. 2.4A)** followed by the flow cytometry analysis of the percentages of EGFP positive cells as an indicator of replacement efficiency. Owing to the promoter-less design of the exchange plasmid, no EGFP positive cells were detected in 293F wild type (WT) cells as the control **(Fig. 2.4B)**, whereas the replacement efficiency under TAg expression for 15 tested 293F-Puro single cell clones ranged from 4.4% to 10.6%, with an average of 6.6% EGFP+ cells **(Fig. 2.4C)**. Clone D2 exhibited the highest efficiency (10.6**%**) among tested clones, and thus was selected for further study.

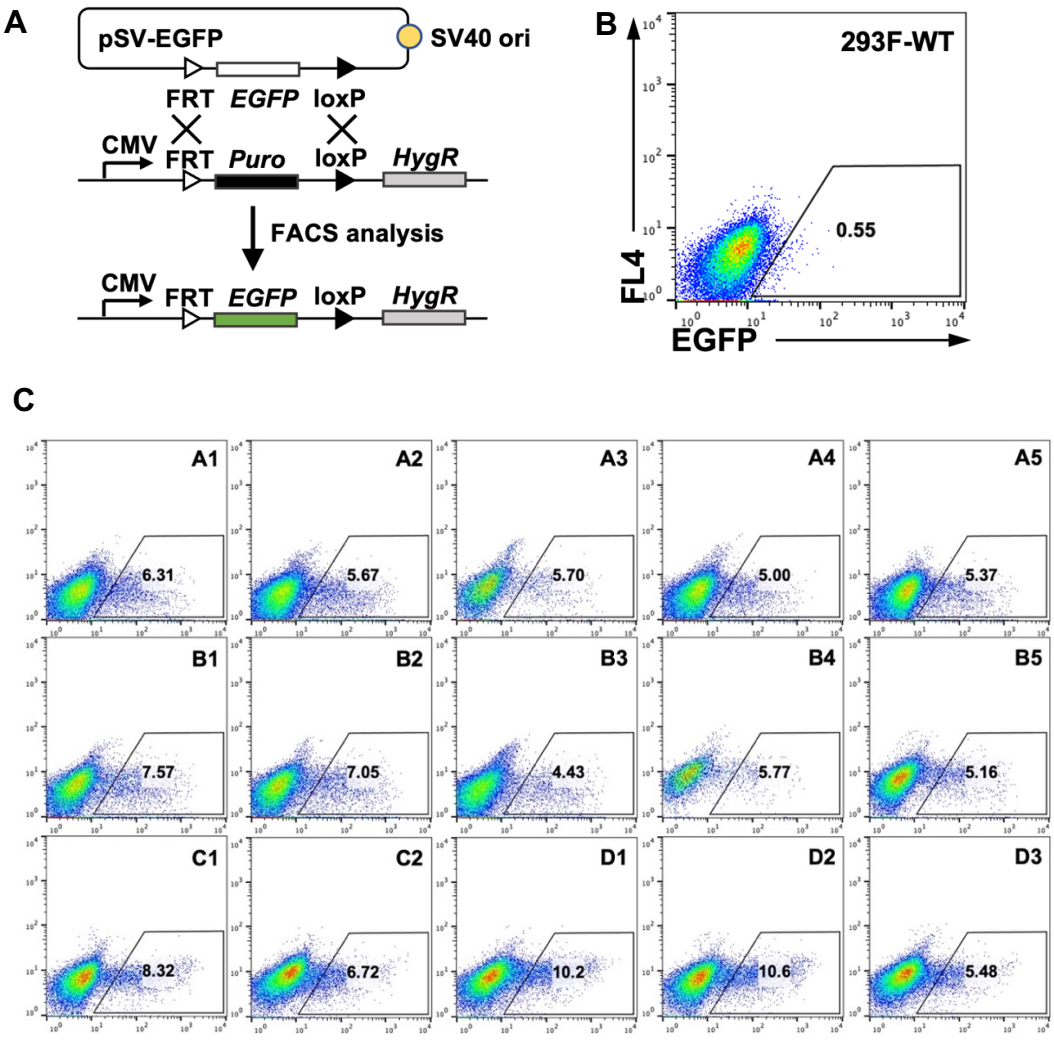
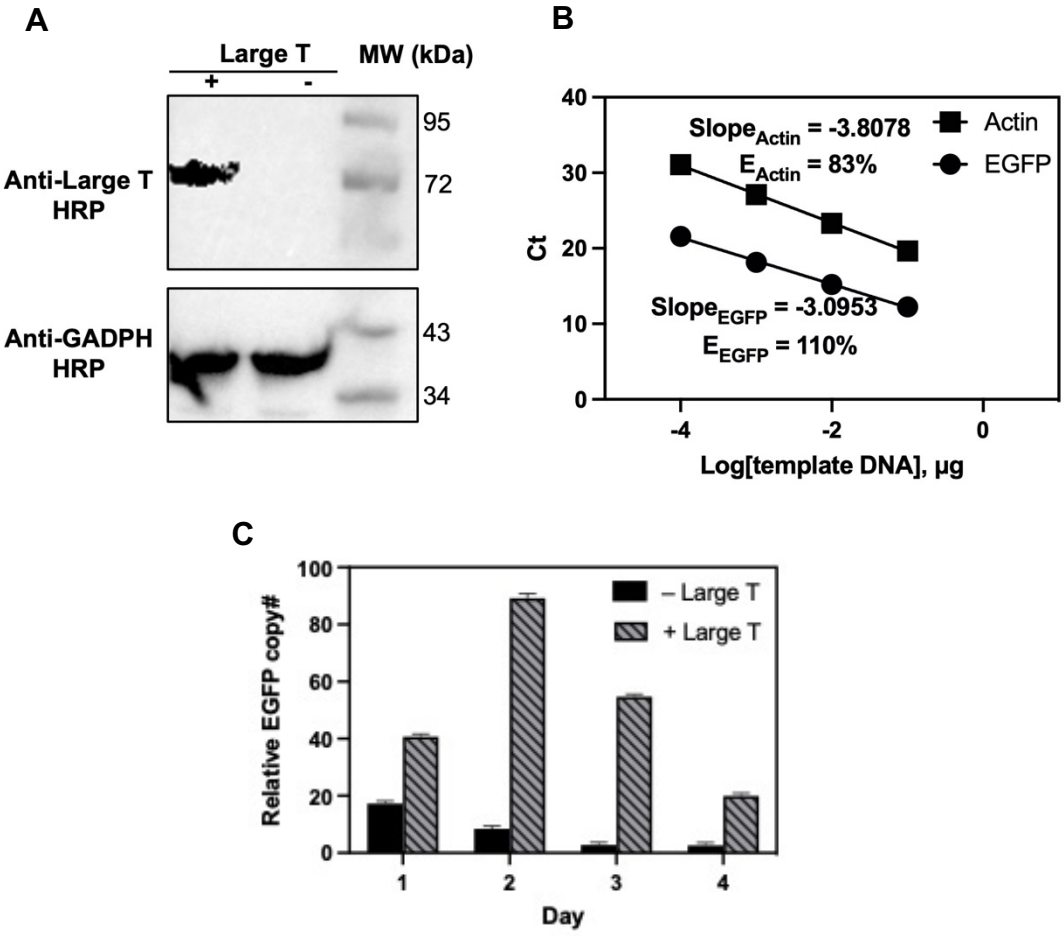


Figure 2.4 Screening for monoclonal cell lines with high efficiency of Large T replication facilitated RMCE (T-RMCE). **(A)** RMCE between the EGFP exchange plasmid carrying SV40 origin (pSV-EGFP) and the landing pad on the genome. EGFP can only be expressed following the successful integration. **(B)** 293-F WT and **(C)** 15 RMCE-competent clones co-transfected with pSV-EGFP and plasmids for large T (pLargeT) and Flpe/Cre expression (pF2AC). GFP positive cells were detected by flow cytometry on Day 3 post transfection.

2.3.4 TAg expression increased the exchange plasmid copy number

To test whether TAg can maintain the exchange plasmid number in high copy number, western blot was employed to detect the TAg expression and qPCR was

conducted on the 293F-Puro D2 cells from day 1 to day 4 post co-transfection with the EGFP exchange plasmid carrying SV40 origin and the TAg expression plasmid or empty plasmid as control. The Flp/Cre expression plasmid was not transfected to avoid underestimation of the exchange plasmid copy number due to the integration. As the result shown in **Fig 2.5A**, TAg expression can still be detected by western blot at day 4 post transfection. To obtain the qPCR efficiency, the qPCR standard curves were generated using four of 10-fold serial dilutions of template DNA at day 1 post transfection with pSV-EGFP and pLargeT (**Fig 2.5B**). The EGFP and actin qPCR efficiency was derived as 110% and 83%, respectively, which were thereby used to calculate the EGFP copy number normalized by actin. The negative controls, i.e., qPCR reactions with the template DNA extracted from the cells without transfection or without template DNA, showed similar EGFP Ct value around 35, resulting in an EGFP copy number of 3.8×10^{-6} normalized to the actin copy number. In contrast, for the cells transfected with the exchange plasmid carrying the *EGFP* gene, at day 1 post transfection, the EGFP copy number in the cells with TAg expression was 40.68, over 2-fold more than that without (**Fig 2.5CD**). The difference became even larger at day 2 as the EGFP copy number in the cells with TAg expression increased by over 2-fold compared to that at day 1while that without TAg kept decreasing over the 4-day cultivation (**Fig 2.5CD**). This is strong evidence showing the correlation between the TAg expression and the increased number of the exchange plasmid carrying SV40 origin, thus the improved RMCE efficiency.



D

	- Large T			+ Large T			Fold of copy # increase
	Ct-GFP	Ct-actin	Relative EGFP copy #	Ct-GFP	Ct-actin	Relative EGFP copy #	
Day1	13.52±0.26	21.34±0.09	17.33±0.87	12.46±0.14	21.45±0.08	40.68±0.94	2.4±1.1
Day2	15.26±0.08	22.30±0.11	8.45±1.01	12.04±0.00	22.23±0.82	89.18±1.64	10.6±1.6
Day3	15.51±0.15	20.76±0.04	2.76±0.91	11.67±0.21	20.97±0.06	54.69±0.89	19.8±1.0
Day4	15.13±0.07	20.21±0.09	2.65±1.00	12.78±0.15	20.67±0.25	19.33±1.04	7.5±1.0

Figure 2.5 Large T expression and exchange plasmid quantification in 293F-Puro D2 clone. (**A**) Western blot of large T expression. Whole cell samples transfected with pLargeT were subjected to analysis. GADPH was used as the loading control. (**B**) qPCR standard curves. The qPCR efficiency E was calculated as $(10^{-slope}-1) \times 100\%$. (**C**) Relative amount of EGFP gene determined by qPCR. 293F-Puro D2 cells were co-transfected with exchange plasmid pSV-EGFP and large T expression plasmid pLargeT (+ Large T) or empty plasmid pcDNA (- Large T) without Flpe/Cre. Actin was used for normalization. **(D)** Cycle threshold (Ct) values of qPCR. The relative copy numbers of GFP, normalized with the copy number of actin, was calculated as $\frac{(E_{Actin}+1)^{Ct_{Actin}}}{(E_{EGFP}+1)^{Ct_{EGFP}}}$.

2.3.5 Impact of cultivation time on RMCE efficiency

As we observed the exchange plasmid number still maintained in a high level at day 4, we argued that a longer culture time after transfection could further improve the RMCE efficiency. To test this hypothesis, 24 h after the transfection with the plasmids for EGFP exchange, TAg and Flp/Cre expression, cells were cultured for 2-5 days and measured for percentages of EGFP+ cells by flow cytometry. Results indicated that the RMCE efficiency plateaued at day 4 (12.2%) and slightly reduced on day 5 (**Fig 2.6**), likely due to a faster growth rate of unaffected cells by transfection reagents as well as the expression of recombinases or TAg. The results also guided us to harvest cells at day 4 post transfection for library construction.

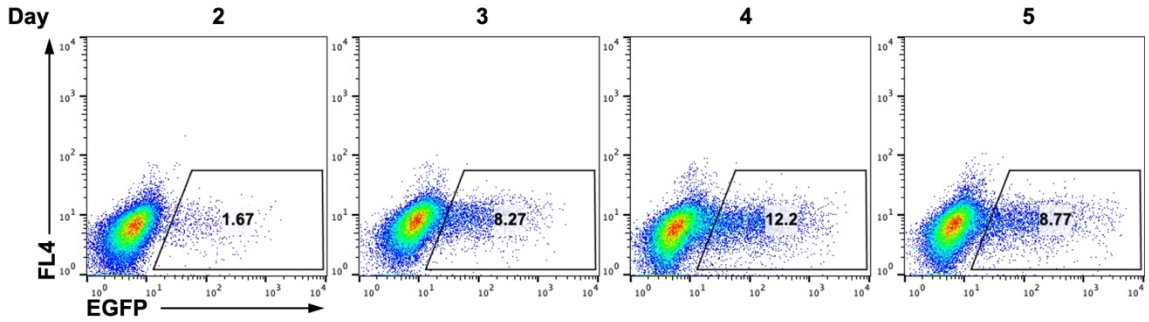


Figure 2.6 Effect of post-transfection cultivation time on RMCE efficiency. 293F-Puro D2 cells were co-transfected with pSV-EGFP, pLargeT and pF2AC. Cells were sampled 2-5 days after transfection and analyzed by flow cytometry for EGFP positive populations.

2.3.6 Isolation of GFP variants with enhanced green fluorescence

A *E. coli* library of 1×10^8 GFP variants was generated by error-prone PCR with wild type GFP as template and inserted into pSV-GFP plasmid (**Fig 2.7A**). The wild type GFP

in fact contains an innocuous Q80R mutation that is present in most cDNA constructs derived from the original sequence due to a PCR error[17,18]. 1×10^8 293F-Puro cells were co-transfected with the GFP library exchange plasmid and the plasmids for Flpe/Cre and TAg expression. At day 4 post transfection, the top 0.21% of the GFP+ cells were sorted out by FACS as round 1. After 2 rounds of FACS, there was a significant rightward shift of the cell population, indicating a successful enrichment for the cells with enhanced green fluorescence (**Fig 2.7B**). Therefore, the genomic DNA from the GFP library post round 2 were extracted and the GFP variant gene was amplified by PCR and cloned into *E. coli* expression vector pMopac for identification, production and characterization. In total, 16 colonies were randomly picked for sanger sequencing, leading to four GFP variants identified (**Fig 2.8A**). The emission spectra at 550 nm showed all GFP variants have maximum excitation at 488 nm except the clone #6 (I167V/Q80R) have two excitation peaks at 399 nm and 488 nm, similar as WT (**Fig 2.8B**). Clone #1 was the most abundant (10/16) with only one mutation S65T and #2 also has one mutation F64L. Excited by the wavelength 488 nm, both clones showed over 10-fold improvement on the fluorescence intensity at emission wavelength 526 nm. In fact, those two mutations exist in the well-established EGFP amino acid sequence whereas they shared a similar intensity at 526 nm. Clone #8 (T62S/S65T/P192L) have novel mutations and showed moderate improvement compared to WT.

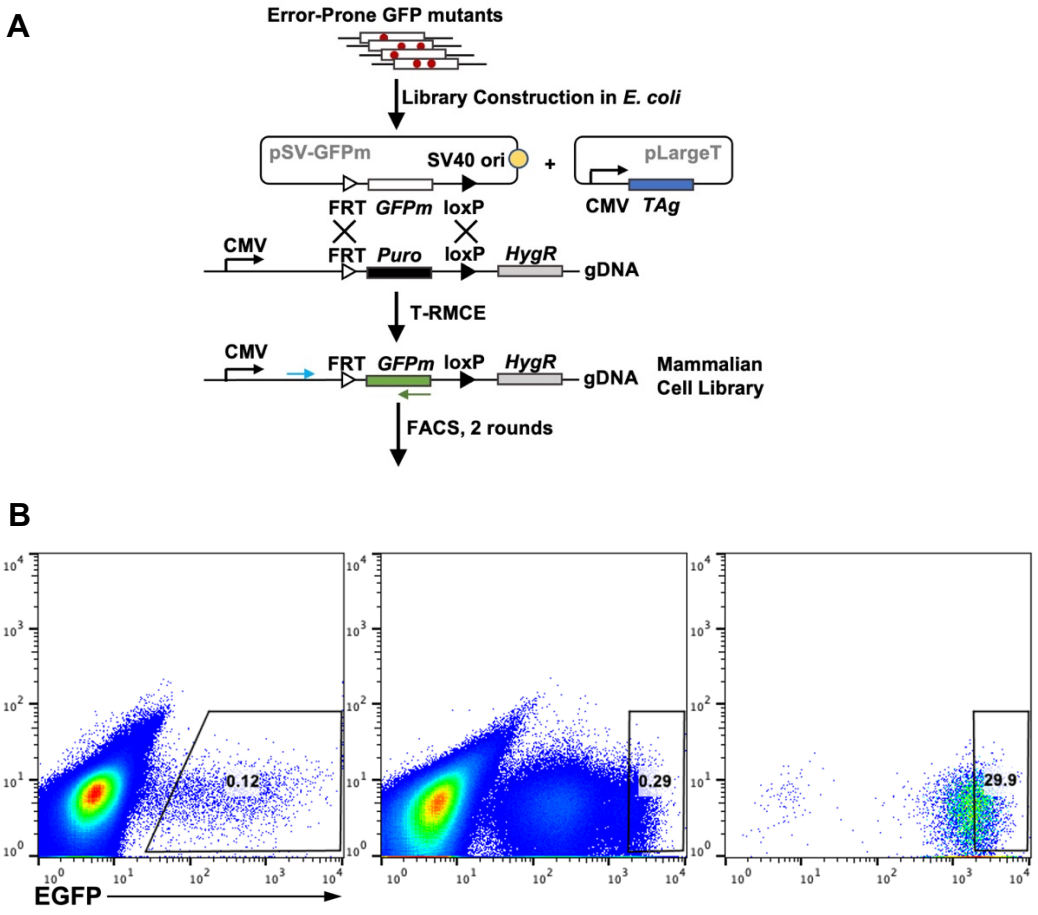


Figure 2.7 Construction and screening GFP library in mammalian cells. (**A**) Workflow of the GFP library construction in mammalian cell via T-RMCE. The blue and green arrows indicate the forward and reverse primers, respectively, for GFP variants identification by PCR. (**B**) Screening of the GFP library by two rounds of FACS.

A

Clone	Occurrence	Mutations	RFU*
Q80R	-	Q80R	2,067
1	10	S65T, Q80R	25,172
3	3	F64L, Q80R	26,849
6	2	I167V, Q80R	3,106
8	1	T62S, S65T, Q80R, P192L	12,050
EGFP	-	F64L, S65T, H231L	25,097

*RFU was measured at 528 nm excited by 488 nm.

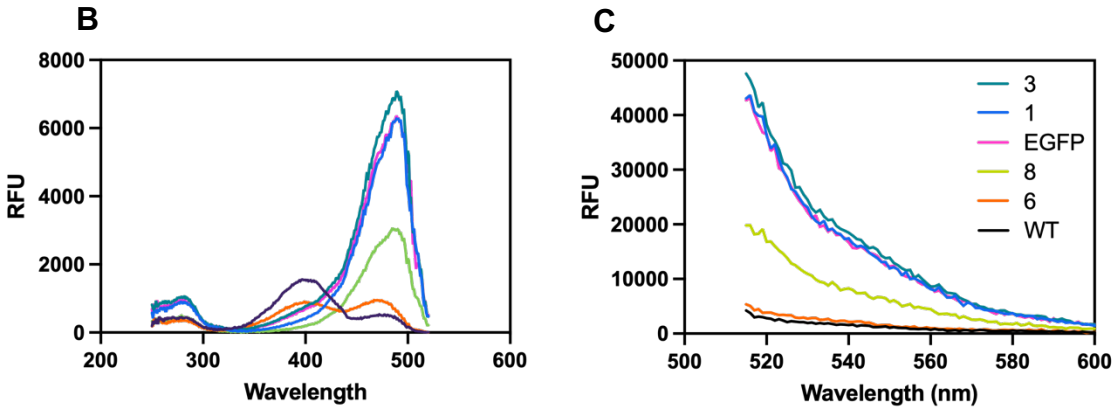


Figure 2.8 GFP variants identification and characterization. (**A**) Summary of the Identified GFP variants compared to the parental GFP Q80R and EGFP **(B**) Excitation spectra with emission at 550 nm. (**C**) Emission spectra with excitation at 488 nm. The parental GFP and EGFP were used as controls. All were measured at 16 nM.

2.4 Discussion and Conclusions

Here we report a simple and robust system to enhance RMCE efficiency in 293F cells based on the well-established SV40 large T facilitated plasmid replication (T-RMCE). We generated single clones of 293F cells with presumably single landing pad on the genome by two sequential rounds of RMCE based on the extremely low possibility of multiple RMCE events taking place in one cell. Chuan et al[15] utilized the same method to generate the single landing pad cell lines in a previous study, which was confirmed by Southern blot analysis. We believe the future mammalian cell library constructed with this cell line will be single copy per cell.

The RMCE efficiency facilitated by TAg in clone D2 could reach 10% at day 3 post transfection. qPCR and western blot analysis confirmed that the exchange plasmid carrying SV40 can be replicated and maintained at a high-level copy number even at day 4 post transfection under the expression of TAg. We further found 4-day cultivation after transfection was optimal for library construction.

Based on the TAg facilitated RMCE efficiency 10%, we presumably constructed a size of 10^7 mammalian cell GFP library from 10^8 transfected cells. After 2 rounds of FACS selection, green fluorescence enhanced proteins were isolated with 2 of the mutations were reported showing the highest signals and the other novel variants showing moderate signal. The results indicated a successful GFP evolution.

Our TAg facilitated RMCE system is established in suspension cell with great scalability, allowing to create libraries of up to 2×10^8 clones from 2×10^9 293F suspension cells in 1 L media. Moreover, it is feasible to electroporate up to 10^{11} cells in a 30-min cycle which could yield a library of 10^{10} clones. Considering the library size is comparable to the capacity of phage library albeit with a significant "overhead", we envision our approach could not be limited to affinity maturation or protein optimization, but rather the discovery of novel target proteins with complex structures.

2.5 References

1. Sullivan, C. S. & Pipas, J. M. T Antigens of Simian Virus 40: Molecular Chaperones for Viral Replication and Tumorigenesis. *Microbiol Mol Biol R* 66, 179–202 (2002).

2. Fanning, E. & Zhao, K. SV40 DNA replication: From the A gene to a nanomachine. *Virology* 384, 352–359 (2009).

3. McKay, R. & DiMaio, D. Binding of an SV40 T antigen-related protein to the DNA of SV40 regulatory mutants. *Nature* 289, 810–813 (1981).

4. Hassell, J. A. & Brinton, B. T. SV40 and Polyomavirus DNA Replication. in vol. 31 (1996).

5. Heinzel, S. S., Krysan, P. J., Calos, M. P. & DuBridge, R. B. Use of simian virus 40 replication to amplify Epstein-Barr virus shuttle vectors in human cells. *J Virol* 62, 3738–46 (1988).

6. Chittenden, T., Frey, A. & Levine, A. J. Regulated replication of an episomal simian virus 40 origin plasmid in COS7 cells. *J Virol* 65, 5944–51 (1991).

7. Mahon, M. J. Vectors bicistronically linking a gene of interest to the SV40 large T antigen in combination with the SV40 origin of replication enhance transient protein expression and luciferase reporter activity. *Biotechniques* 51, 119–126 (2011).

8. 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).

9. 326. Insertion of an Internal Intron Increases Transgene Expression Levels from a Bicistronic Lentivirus Vector. *Mol Ther* 15, S123 (2007).

10. Pañeda, A. *et al.* 1079. Effect of WPRE on Transgene Expression Using Different Promoters in the Context of Hydrodynamically Delivered Plasmid Vectors. *Mol Ther* 13, S413–S414 (2006).

11. Buchholz, F., Angrand, P.-O. & Stewart, A. F. Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* 16, 657–662 (1998).

12. Buchholz, F., Ringrose, L., Angrand, P.-O., Rossi, F. & Stewart, A. F. Different Thermostabilities of FLP and Cre Recombinases: Implications for Applied Site-Specific Recombination. *Nucleic Acids Res* 24, 4256–4262 (1996).

13. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res* 29, e45–e45 (2001).

14. Hayhurst, A. *et al.* Isolation and expression of recombinant antibody fragments to the biological warfare pathogen Brucella melitensis. *J Immunol Methods* 276, 185–196 (2003).

15. Chen, C., Li, N., Zhao, Y. & Hang, H. Coupling recombinase-mediated cassette exchange with somatic hypermutation for antibody affinity maturation in CHO cells. *Biotechnol Bioeng* 113, 39–51 (2016).

16. Qiao, J., Oumard, A., Wegloehner, W. & Bode, J. Novel Tag-and-Exchange (RMCE) Strategies Generate Master Cell Clones with Predictable and Stable Transgene Expression Properties. *J Mol Biol* 390, 579–594 (2009).

17. Zimmer, M. Green Fluorescent Protein (GFP): Applications, Structure, and Related Photophysical Behavior. *Chem Rev* 102, 759–782 (2002).

18. Tsien, R. Y. THE GREEN FLUORESCENT PROTEIN. *Biochemistry-us* 67, 509–544 (1998).

Chapter 3: Improved RMCE Efficiency by Cell Cycle Arrest

Abstract

Cell cycle progression influences the RMCE event in many aspects including endocytosis, intracellular trafficking, nuclear entry as well as chromatid structure and number. In this Chapter, we constructed RMCE-competent CHO cell line carrying a single landing pad on its genome and evaluated the effect of cell cycle arrest on the RMCE efficiency using different chemical inhibitors. Results showed that the RMCE efficiency was increased significantly up to 10% with nocodazole arresting the cells at the beginning of the mitosis phase. We further optimized the drug concentration, treatment duration as well as the cultivation time post transfection, and established the aRMCE approach which is applicable to different CHO cell clones and gene of interest independent. Overall, the aRMCE approach makes it feasible to construct large combinatorial libraries in mammalian cells.

3.1 Introduction

RMCE requires the presence of the associated recombinases and the gene of interest (GOI) at the landing pad site on chromosome usually in cell nucleus. Transfection is frequently conducted with cationic polymers or liposome forming complex with an recombinase expression plasmid and an exchange plasmid carrying GOI. Following cellular uptake, the exchange plasmid must traffic across the cytoplasm and the nuclear membrane to encounter the landing pad. For recombinases, multiple shuttles between cytoplasm and nucleus are required, as transcription and translation occur in nucleus and

ribosome respectively and the produced recombinases need to translocate back to the nucleus for RMCE. Limited by these transportation steps, efficiency of RMCE is usually low.

The state of cell cycle is known to modulate cellular uptake, trafficking in the cytoplasm and nuclear entry of transfected DNA[1,2]. For mammalian cells, cell cycle is composed of interphase (G_1, S, and G_2 phases) and the mitotic (M) phase[3]. Cell grows normally in G_1 phase and DNA is synthesized during S phase. In G_2 phase, cell makes proteins and organelles in preparation for cell division. During mitosis, sister chromatids separate into daughter cells to finish cell division. Notably, in interphase or non-dividing (G_0) phase, the nuclear envelop is largely impermeable, acting as a barrier for macromolecules. However, during mitosis, the nuclear envelop breaks down, allowing cytoplasmic plasmids and proteins near the nucleus access to the nuclear space[2].

Previous studies indicated that manipulation of cell cycle progression can significantly affect gene delivery and protein expression level. Brunner[4] et al found that transgene expression was 30- to over 500- fold higher when transfection with polycation or liposome was performed during S or G_2 phase than G_1 phase, likely due to enhanced nuclear entry[4]. However, this cell-cycle dependency was not observed with viral transfection systems which have a highly efficient nuclear entry machinery[5–7]. In addition, cell cycle-dependent internalization of the DNA-liposome complexes can impact transfection efficiency. Mannisto[8] et al showed that the rate of endocytosis increased gradually during the G_1, S and G_2 phases, but drops drastically at the beginning of mitosis. Transgene expression level is also linked with plasmid compaction state

which is affected by chromatid structure at different cell cycle phases. Marenzi[9] et al. found when transfected in G_0 or G_1 phase, plasmid is packaged to a compact form due to the condensed chromatids, leading to a minimal expression of the reporter gene. In contrast, the high expression level was seen with transfection in S phase.

As proposed in **Chapter 1**, the RMCE efficiency can be improved by more available exchange plasmids, recombinases and an easier and longer access to more landing pads. We hypothesize that these beneficial events can be achieved via manipulation of the cell cycle progression (**Fig. 3.1**). More specifically, S or G2 phase can provide more plasmids due to high endocytic capacity and facilitate landing pad accessibility due to chromatin relaxation. M phase can double the amount of landing pads, remove the nuclear barrier, enhance recombinase expression, and provide more exchange plasmids at the recombination site.

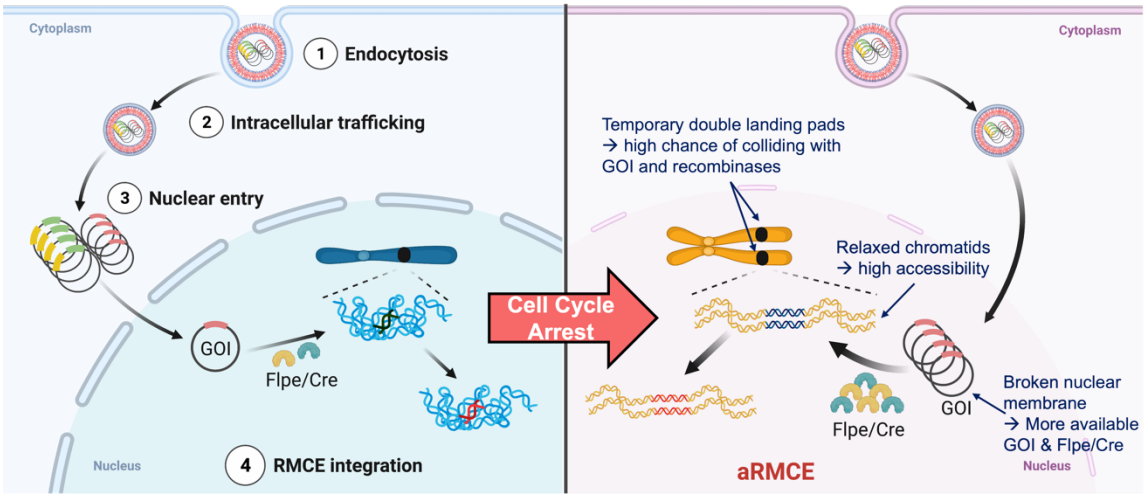


Figure 3.1 Hypothesis of the RMCE efficiency improvement by cell cycle arrest.

In this Chapter, we constructed RMCE-competent CHO cell line carrying a single landing pad on its genome and evaluated the effect of cell cycle arrest on the RMCE efficiency. Results showed that the RMCE efficiency was improved significantly when the cells were arrested at M phase by nocodazole and moderately at S phase by hydroxyurea. With the optimized drug concentration and culturing duration, a feasible approach for construction of large combinatorial libraries in mammalian cells was established.

3.2 Materials and Methods

3.2.1 Plasmid construction

The DNA fragment encoding FRT-EGFP-loxP was chemically synthesized (IDT) and cloned into pcDNA3.1/Hygro$^{(+)}$ with HindIII and XhoI restriction sites to give the landing pad plasmid pcHyFGL. To reduce the size of vectors used in RMCE, pCDNA3.1$^{(+)}$ were digested with PvuII, and the obtained 3.3 kb fragment was gel-purified and self-ligated to give pcDNA, in which the hygromycin resistant gene cassette and f1 ori were removed. The gene cassette of FRT-iRFP-loxP was synthesized and cloned into pcDNA with BglII/BamHI sites for CMV promoter removal and resulted in pEx-iRFP. Other exchange plasmids were prepared similarly, for puromycin resistance gene (pEx-Puro), EGFP (pEx-EGFP), human anti-MMP9 IgG1 L13^{10} heavy chain (pEx-HC), and full length human IgG1 L13 (pEx-IgG). The C-terminal of heavy chain was fused via a $(GSGSS)_2$ linker with transmembrane domain (TM) from platelet derived

growth factor receptor beta (PDGFRβ). For recombinase expression, the gene encoding Flpe(F70L)[11,12] and Cre recombinases separated by the T2A self-cleaving peptide (F2AC) was assembled by overlap PCR. pcDNA was modified by insertions of a chimeric intron[13] and a Woodchuck hepatitis virus posttranscriptional regulatory element[14] (WPRE) at NheI/KpnI and XhoI/XbaI sites respectively. Obtained pCIW was digested with KpnI/XhoI followed by ligation with F2AC to generate pF2AC. The simian virus 40 (SV40) nuclear localization signal (NSL) sequence was introduced by primers during PCR at the upstream of the genes encoding Flpe and Cre. All restriction enzymes used were purchased from NEB.

3.2.2 Cell culture and transfection

CHO-K1 (ATCC, Cat# CCL-61) were cultured in Ham's F-12K (Kaighn's) medium (Gibco, Cat#21127022) supplemented with 10% fetal bovine serum (FBS, Gibco, Cat# 10437028), 50 U/ml penicillin, and 50 µg/ml streptomycin (Gibco, Cat#15140122) at 37 °C, in 5% CO_2 humidified atmosphere. Hygromycin (Gibco, Cat#10687010) at 350 µg/ml and 50 µg/ml and puromycin (Gibco, Cat#A1113803) at 5 µg/ml and 1 µg/ml were used for selection and maintaining cell cultures respectively. In a typical RMCE transfection, CHO cells were seeded in 6-well plates (Falcon, Cat#353046) at a density of 2×10^5 cells per well in 2 ml F-12K media. After cultivation for 24 h, the supernatants were replaced with 1.8 ml fresh F-12K media, and cells were transfected by adding transfection mixtures, which were prepared in 0.2 ml Expi293 media (Gibco, Cat#A1435101) containing 1 µg exchange plasmid, 1 µg pF2AC and 8 µg PEI MAX 40K (Polysciences, Cat#24765-1). Following incubation for 24 h, cells were detached

with 0.25% trypsin-EDTA (Gibco, Cat#25200056), and cultured in a 10 cm tissue culture dish (Falcon, Cat#353003) with 10 ml fresh F-12K media until flow cytometry analysis. To generate landing pad CHO-EGFP cells, 1×10^6 CHO-K1 cells were seeded into a 10 cm plate and transfected with 3 μg pcHyFGL and 12 μg PEI. One day after the transfection, cells were changed with fresh media and selected with 350 μg/ml hygromycin for 7 days, then maintained with 50 μg/mL hygromycin. To obtain single clones of CHO-Puro, cells were diluted and seeded into 96-well plates at a density of 0.5 cell/well. After 7-day culture, single cell clones were identified and expanded for further study.

3.2.3 Chemical-mediated cell cycle arrest RMCE (aRMCE)

The following chemicals and concentrations were evaluated: 3 μM, 6 μM and 12 μM aphidicolin (Sigma-Aldrich, Cat #A0781); 0.5%, 1% and 2% DMSO (Fisher Scientific Cat#BP231-100); 1mM, 2 mM and 4 mM hydroxyurea (Acros Organics, Cat#AAA1083103); 20 μM, 40 μM and 80 μM lovastatin (Acros Organics, Cat#AC458830050); 100 μM, 200 μM, 400 μM mimosine (Fisher Scientific Cat#50-194-8071); 325 nM, 650 nM 1.3 μM nocodazole (Fisher Scientific, Cat#AC358240100); and 2.5 mM, 5 mM and 10 mM thymidine (Acros Organics, Cat#AC226740050). CHO-Puro clones were seeded at 2.5×10^5 cells per well in 6-well plates and cultured for 24 h to achieve ~5×10^5 cells per well. Culture supernatants were replaced with 1.8 mL fresh F-12K media containing the chemicals at their pre-determined concentrations. After a brief incubation of 30 minutes, 0.2 mL transfection mixtures were added and cells were cultured for 24 h. Cells were then detached with 0.25% trypsin-EDTA (Gibco,

Cat#25200056), and split into 2 wells in 6-well plates. For 24-h drug treatment groups, cells were cultured in fresh media and analyzed by flow cytometry. For 48-h drug treatment groups, cells were incubated with a second dose of drugs in fresh media for 24 h, then washed with PBS to remove the drugs and grown in fresh media until flow cytometry analysis.

3.2.4 Flow Cytometry

Cell analysis and sorting was performed on a S3e cell sorter (Bio-Rad), equipped with 488/640 nm dual lasers and multiple bandpass / longpass filters. EGFP and iRFP were excited at 488 nm and 640 nm and detected with 526/48 (FL1 channel) and 700/LP (FL4 channel) filters, respectively. Cells displaying human mAb L13 was stained with 1 μg/ml mouse anti-human IgG (Fab)-PE (Invitrogen Cat#MA1-10377) for excitation at 488 nm and detection with 593/40 filter (FL2 channel). In general, 50,000 events were collected for analysis; at least 5×10^6 cells were sorted at 2000 – 3000 events/s.

3.3 Results

3.3.1 Generation of RMCE ready single cell clones

To generate cell clones with a single landing pad, three sequential steps of transfection and selection / screening were performed[15,16]. In Step1, CHO-K1 cells were transfected with a plasmid carrying a CMV promoter, EGFP gene flanked by the FRT and loxP recombination sites (FRT-EGFP-loxP) and a hygromycin resistance marker for selection. Transfected cells were selected with 350 μg/mL hygromycin for 7 days and

maintained with 50 μg/mL hygromycin for 8 more days to expand the cells and clear out the free plasmids. Flow cytometry analysis indicated that 43% of the survived cells were EGFP positive (**Fig. 3.2**, Step 1), suggesting that genomic integration was achieved. Although rare, it is possible that more than one landing pads can be integrated per cell. In Step 2, the CHO-EGFP cells obtained in Step1 were co-transfected with two plasmids, one harboring a promoter-less FRT-iRFP-loxP fragment for exchange and the other encoding an expression cassette for Flpe and Cre recombinases. At 48-h post-transfection, ~ 1% cells exhibited iRFP signals, indicating occurrence of RMCE because iRFP can only be produced upon genomic integration into the landing pad which follows a CMV promotor (**Fig. 3.2**, Step 2). Approximately 5×10^6 cells were sorted by FACS to collect the top 0.44% RFP^+ cells. After expansion, flow cytometry confirmed that 89% of the isolated cells were $iRFP^+$ and $EGFP^-$, and the top 50% $iRFP^+$ cells were then re-sorted by FACS. In Step 3, the gnomically integrated iRFP gene was replaced with a puromycin resistance gene. Similarly, a promoter-less exchange vector was used for transfection, and thus the puromycin resistant phenotype can only exhibit upon RMCE occurrence. Following puromycin selection, 90% of the survived cells were both iRFP and EGFP negative as shown by the flow cytometer analysis (**Fig 3.2**, Step 3). Considering the possibility of simultaneously replacing multiple landing pads in one cell with the same cassette during two successive rounds of RMCE is extremely low, presumably the obtained puromycin-resistant non-fluorescent cells (CHO-Puro) should contain a single landing pad.

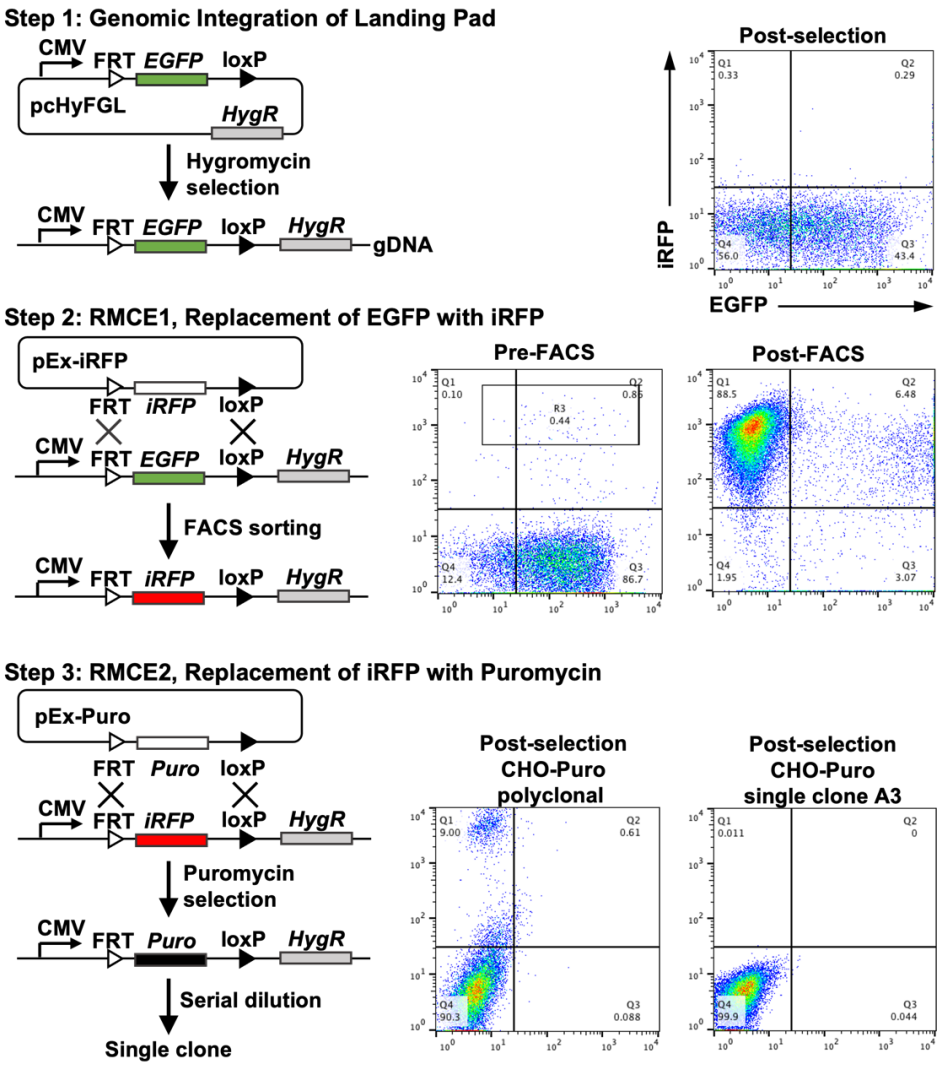


Figure 3.2 Generation of RMCE-competent CHO cells carrying a single expression landing pad on their genome. **(Step 1)** EGFP expression cassette is randomly integrated into genomic loci under hygromycin selection. The resulting EGFP positive cells (CHO-EGFP) mostly have one and possibly have two or more landing pads on their genomes. **(Step 2)** In RMCE1, the EGFP gene at the landing pad is replaced with an iRFP gene on the exchange plasmid and iRFP^{+} EGFP^{-} cells are selected by FACS to give CHO-iRFP. **(Step 3)** In RMCE2, the iRFP gene at the landing pad is replaced with a puromycin resistance gene on the exchange plasmid to give CHO-Puro. Due to low occurrence of multiple replacements in one cell, two successive rounds of RMCE guarantee that majority of resulting CHO-Puro cells carry a single expression RMCE landing pad. In each step, the empty rectangles on the promoter-less exchange plasmids indicate no expression. In contrast, the filled rectangles on genomes indicate expression enabled when integrated at the downstream of a CMV promotor. Flow cytometry plots show cell populations for each step.

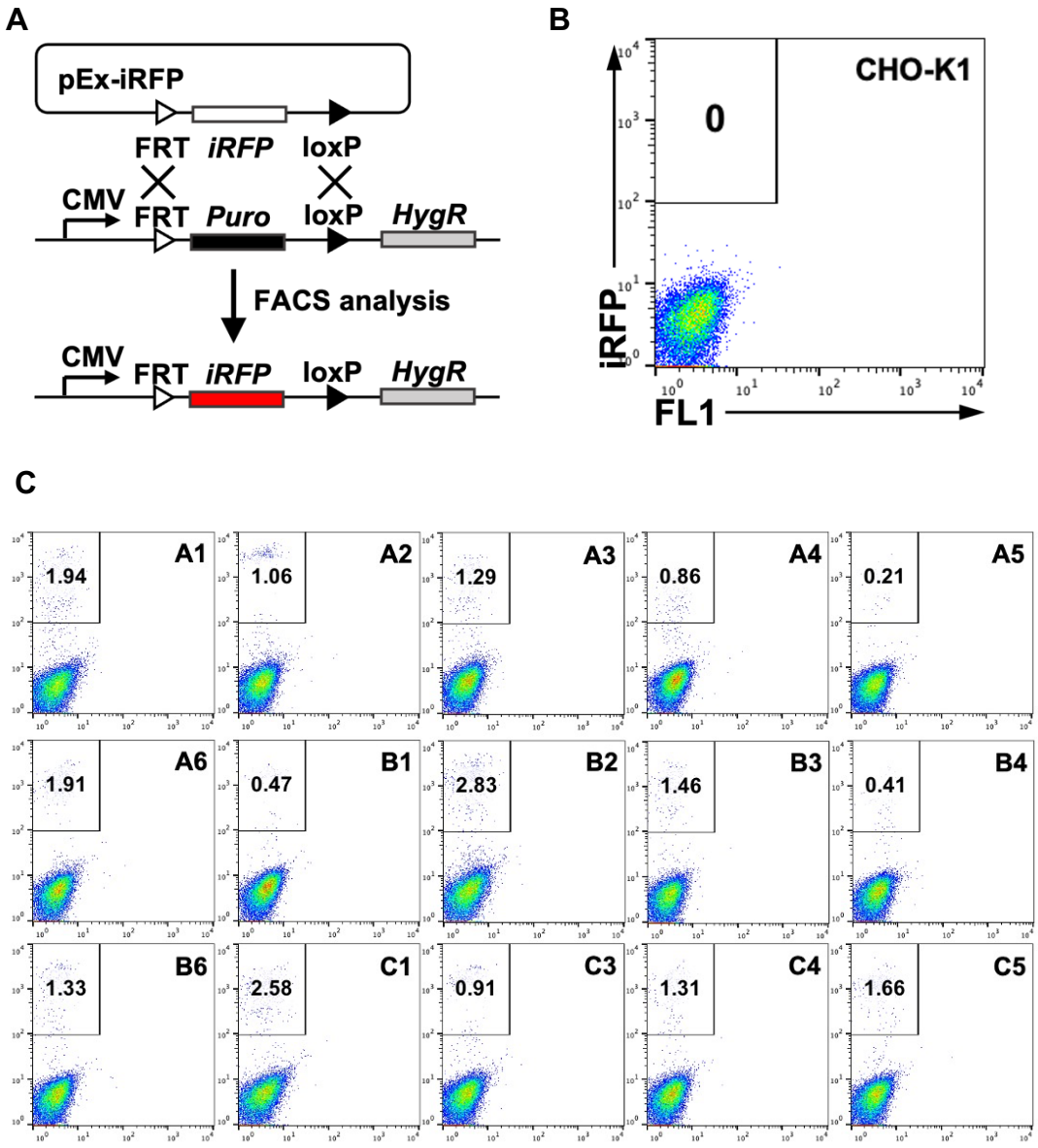


Figure 3.3 Screen monoclonal cell lines for high RMCE efficiency. **(A)** RMCE between the iRFP exchange plasmid pEx-iRFP and the landing pad on the genome. iRFP can only express following the successful integration. **(B)** CHO-K1 on day 2 after transfection with pEx-iRFP. **(C)** 15 RMCE-competent clones on day 2 after transfection with pEX-iRFP. All with the expression of Flpe and Cre.

After serial dilution and expansion of the obtained CHO-Puro cells, 15 single clones of normal cell morphology were identified for RMCE efficiency tests. Following the co-transfection of CHO-Puro single clone cells with plasmids for iRFP exchange and Flpe-Cre expression (**Fig. 3.3A**), the percentages of iRFP positive cells were determined by flow cytometry as an indicator of replacement efficiency. Whereas no iRFP positive cells were detected in CHO-K1 cells as the control (**Fig. 3.3B**), owing to the promoter-less design of the exchange plasmid, the replacement efficiency for 15 tested CHO-Puro single cell clones ranged from 0.5% to 2.8%, with an average of 1.35% $iRFP^+$ cells (**Fig. 3.3C**). Clone B2 exhibited the highest efficiency among tested clones, and thus was selected for further study. In addition, co-transfection of B2 cells with both EGFP and iRFP exchange plasmids (1:1 molar) were conducted, and at 10 days post-transfection, only iRFP or EGFP single positive but not double positive cells were detected, validating that B2 clone carries one landing pad (**Fig 3.4**).

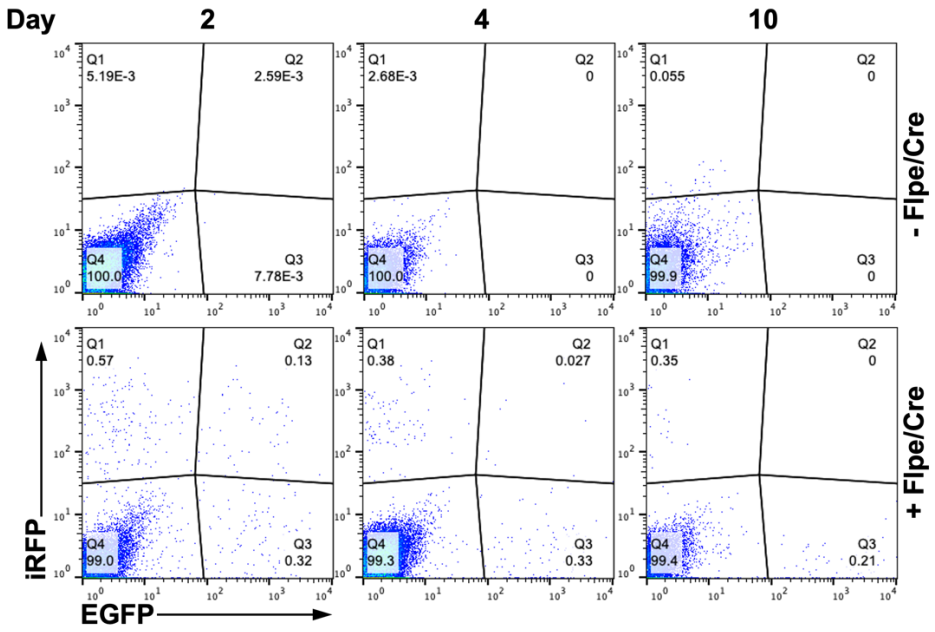


Figure 3.4 Co-transfection of EGFP and iRFP exchange plasmids (1:1) to CHO-Puro clone B2 with or without Flpe and Cre expression. Cells were analyzed at day 2, 4, and 10 post transfection.

3.3.2 Cell cycle arrest improved the RMCE efficiency

To test the hypothesis that chemical-mediated cell cycle arrest can improve RMCE efficiency, we studied seven chemicals which can temporarily hold cell cycle progression at various phases: *e.g.*, DMSO and lovastatin block cells at early G1 phase; mimosine arrests cells at late G1 phase; aphidicolin, hydroxyurea and thymidine inhibit DNA replication at early S phase; and nocodazole inhibits microtubule formation at G2/M phase (**Fig 3.5A**). In the presence of each chemical at their predetermined concentrations, cells of CHO-Puro clone B2 were co-transfected with iRFP exchange and Flpe-Cre expression plasmids. After 24 h incubation, cells were grown in fresh media for 2 days

and tested by flow cytometry. The percentages of $iRFP^+$ cells were detected to study the impacts of these chemicals on RMCE efficiency. Results indicated that compared to the basal efficiency of 1.5 ± 0.3% without drug treatment, DMSO (efficiency = 1.5 ± 0.2%), lovastatin (1.7 ± 0.2%), aphidicolin (0.9 ± 0.2%) and thymidine (1.1 ± 0.2%), which arrest cells in G1 or early S cycle before DNA replication, did not significantly change the RMCE efficiency (**Fig 3.5B**). In contracts, cells treated with 200 μM mimosine (3.6 ± 0.2%), 2 mM hydroxyurea (3.8 ± 1.3%) and 650 nM nocodazole (12.9 ± 2.3%) exhibited 2.4-, 2.5-, and 8.6-folds improvements on RMCE efficiency (**Fig 3.5BC**). Because mimosine negatively impacted on cell viability (data not shown), it was not further considered. For hydroxyurea and nocodazole, two other concentrations were tested next. Results indicated a positive correlation between hydroxyurea concentrations and RMCE efficiency, *i.e.*, 2.5 ± 0.8% with 1 mM hydroxyurea, and 5.3 ± 1.0% with 4 mM hydroxyurea (**Fig 3.6AB**). Nonetheless, higher concentrations of hydroxyurea were not tested due to the significant cell death at 4 mM hydroxyurea. This dosage-dependency was also observed for 325 nM (8.8 ± 1.6%) and 650 nM nocodazole, but the efficiency plateaued without further improvement with 1.3 μM nocodazole (**Fig 3.6C**).

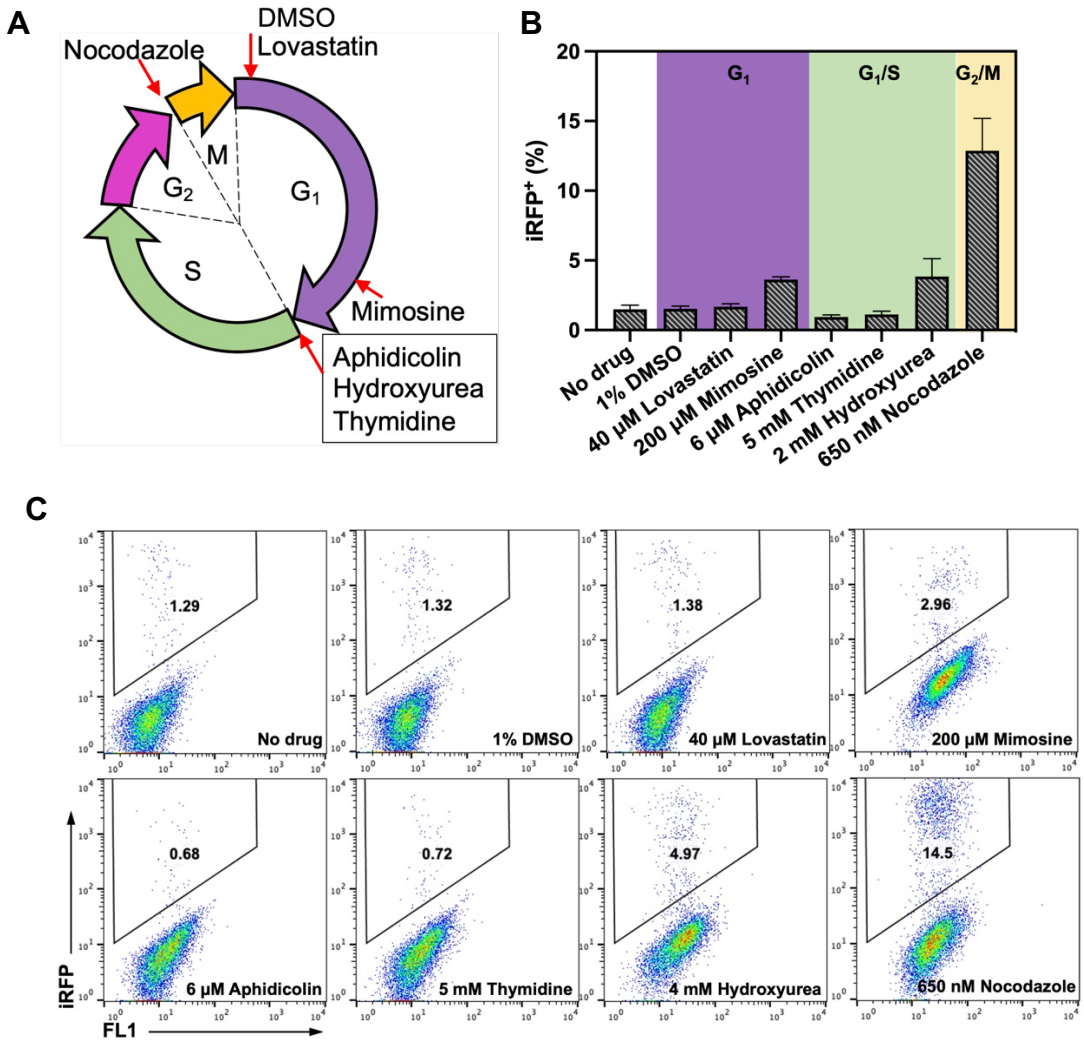


Figure 3.5 Drug-mediated cell cycle arrest and integration efficiency of aRMCE. **(A)** Chemical inhibitors arrest cell cycle at specific phases: DMSO and lovastatin block at early G_1; mimosine arrests cells at late G_1; aphidicolin, hydroxyurea and thymidine inhibit DNA replication at early S phase; nocodazole inhibits microtubule formation at G_2/M phase. **(B)** The effect of chemical inhibitor treatments on aRMCE efficiency. CHO-Puro B2 cells were transfected with pEx-iRFP and pF2AC under 24-h chemical treatment. Flow cytometry analysis was conducted on day 2 post transfection to detect $iRFP^+$ cells. (n = 3) **(C)** Flow cytometry plots of representative trials of aRMCE.

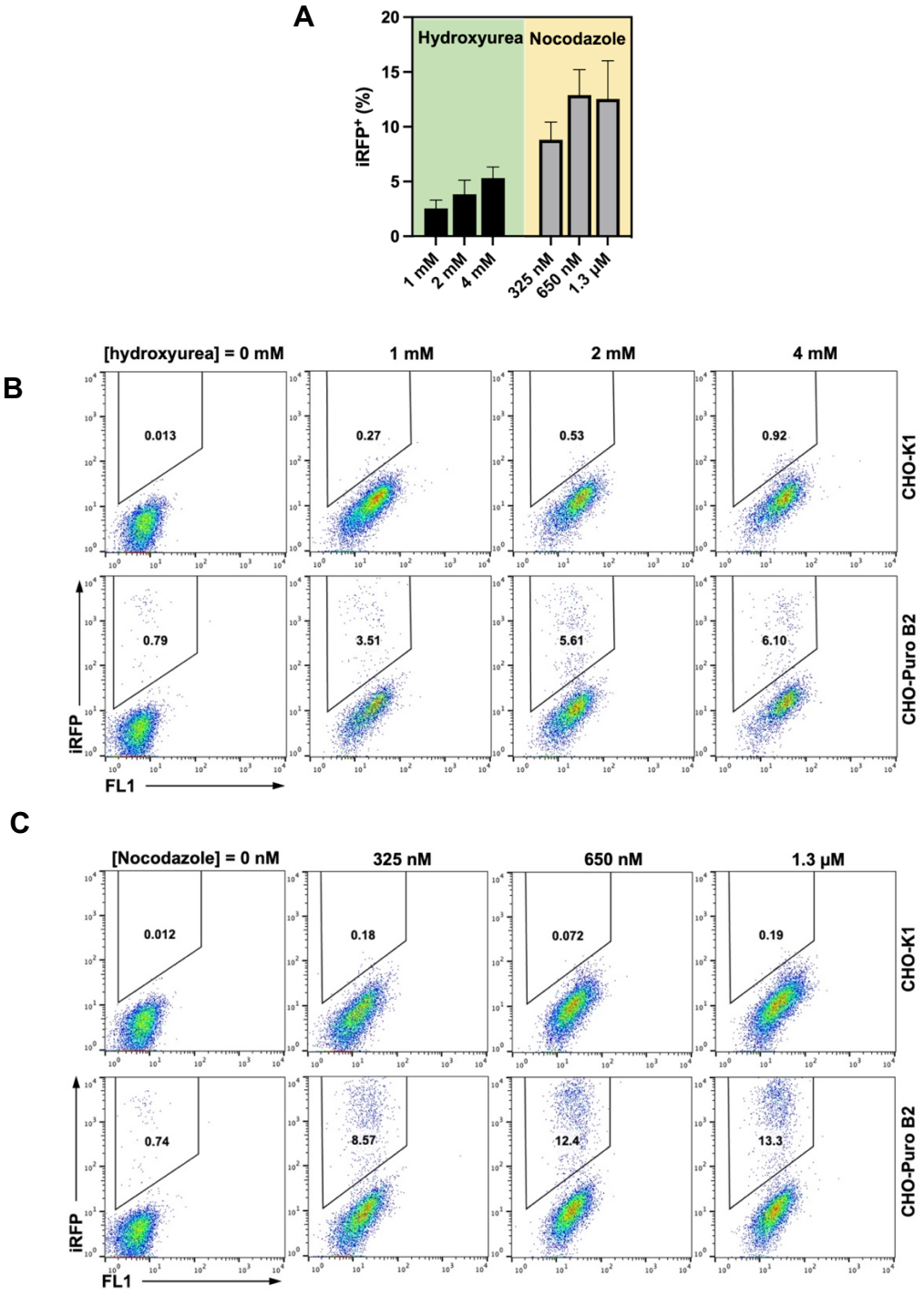


Figure 3.6 The effect of inhibitor concentrations on aRMCE efficiency. **(A)** Percentages of iRFP positive cells; **(B)** The effect of 1, 2, 4 mM hydroxyurea; **(C)** The effect of 325, 650, 1300 nM nocodazole. Cells were transfected with pEx-iRFP under 24 h drug treated and analyzed by flow cytometry on day 2 post transfection.

3.3.3 Extended drug treatment duration caused low cell viability

We next questioned whether a prolonged period of drug treatment can further improve the RMCE efficiency. After 24-h transfection in the presence of cell cycle arrest drugs, the culture supernatant was replaced with fresh media to remove the transfection reagent while still containing the same drugs. After another 24 h incubation with the drugs, cells were cultured in fresh media until flow cytometry analysis. Results indicated that for these cells treated for 48 h with 1 mM hydroxyurea, no significant difference on RMCE efficiency was observed (**Fig 3.7AB**), whereas 48-h cell cycle arrest with 650 nM nocodazole achieved 20.0 ± 0.7% RMCE efficiency, significantly higher than the 24-h treatment group (**Fig 3.7AC**). It was also found that 48 h cell cycle arrest with nocodazole caused a substantial drop in cell viability. However, 24 h treatment with 650 nM nocodazole did not significantly compromise cell viability, and thus these conditions were used in following experiments.

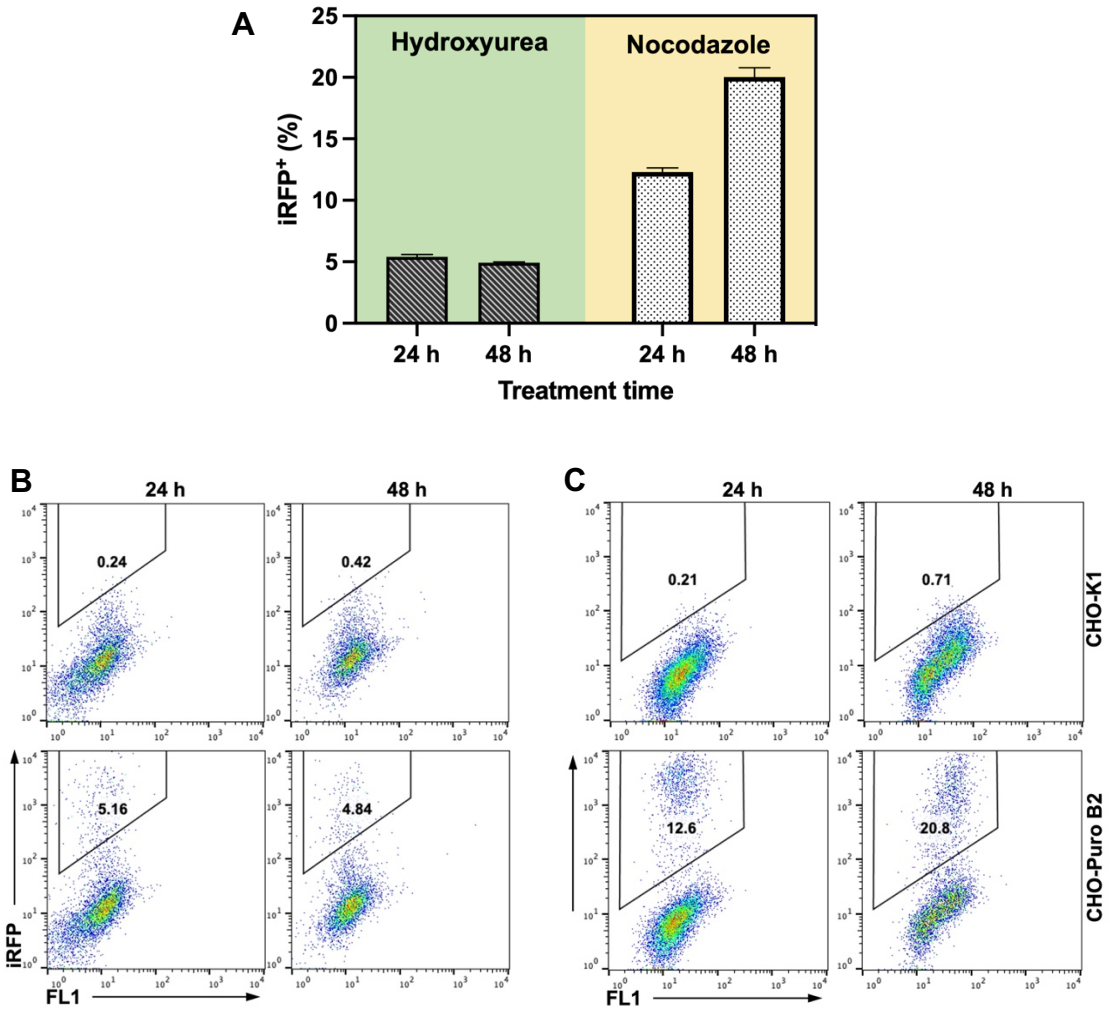


Figure 3.7 The effect of arrest duration on aRMCE efficiency. **(A)** Percentages of iRFP positive cells. (n = 3) **(B)** 4 mM hydroxyurea treatment; **(C)** 650 nM nocodazole treatment. Cells were transfected with pEx-iRFP under either 24 h or 48 h drug treatment and analyzed by flow cytometry on day 2 after transfection.

3.3.4 Prolonged post-transfection cultivation did not improve aRMCE efficiency

Following RMCE mediated recombination, chromosomes must be orderly packaged for their proper segregation into daughter cells. As cell cycle arrest chemicals interfere the cell cycle progression, we argued that a longer culture time after transfection could be

needed for cells to restore their cell cycle and normal growth. To test this hypothesis, after the 24 h chemical-mediated RMCE transfection period, cells were cultured for 2-5 days and measured for percentages of iRFP+ cells by flow cytometry. Results indicated that with 1-4 mM hydroxyurea treatment, the RMCE efficiency maintained in the first three days after transfection and slightly reduced on day 4 (**Fig 3.8A**). Similarly, prolonged post-transfection cultivation did not improve RMCE efficiency for the cells treated with nocodazole (**Fig 3.8B**). In fact, a gradual decrease of iRFP+ cell was observed, presumably due to a faster growth rate of untransfected cells. The results also suggested that combination of increased nocodazole concentration at 1.3 μM with prolonged cultivation did not surpass the efficiency obtained with 650 nM nocodazole. Collectively, 24-h treatment with 650 nM nocodazole and cell harvest at 48 h after transfection was chosen as the condition for cell cycle arrest RMCE (aRMCR).

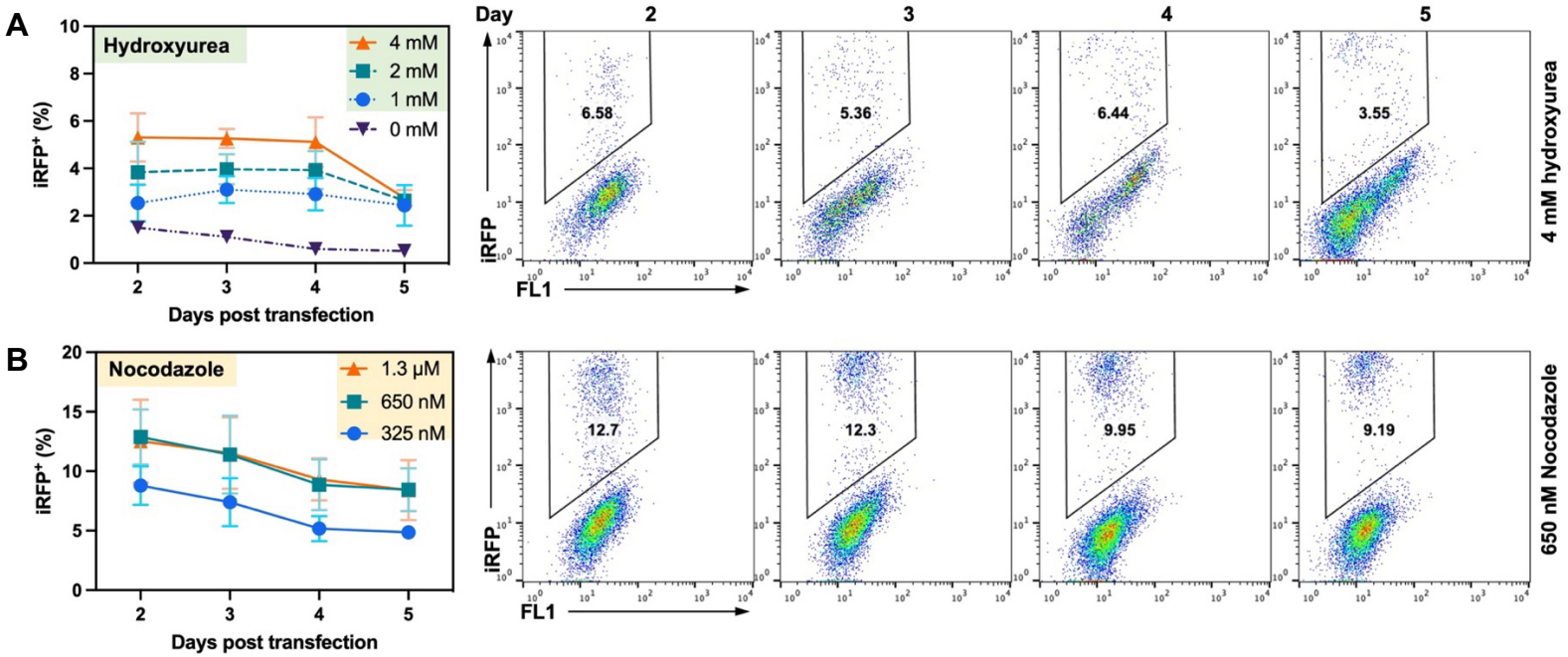


Figure 3.8 Effect of post transfection duration on aRMCE efficiency. Cells were transfected with pEx-iRFP under the treatment with **(A)** 1, 2, 4 mM hydroxyurea or **(B)** 325, 650, 1300 nM nocodazole for 24 h, and analyzed by flow cytometry on day 2, 3, 4, 5 post transfection (n = 3). No chemical treatment as the control.

3.3.5 Nuclear localization signal did not improve aRMCE efficiency

As genomic integration takes place in nucleus, we asked whether the nuclear localization signal (NSL) can further improve the aRMCE efficiency. To facilitate the nuclear entry of the recombinases, simian virus 40 (SV40) nuclear localization signal (NLS) was fused to the N-termini of either Flpe (1×NLS) or both Flpe and Cre (2×NLS) (**Fig 3.9A**). Following co-transfection and cell culture, FACS analysis suggested that adding NSL did not improve aRMCE efficiency when treated with 650 nM nocodazole (**Fig. 3.9BC**). In fact, nocodazole arrests cells at M phase during which the nuclear membrane has disintegrated, explaining why NLS had no impact on the RMCE efficiency[23].

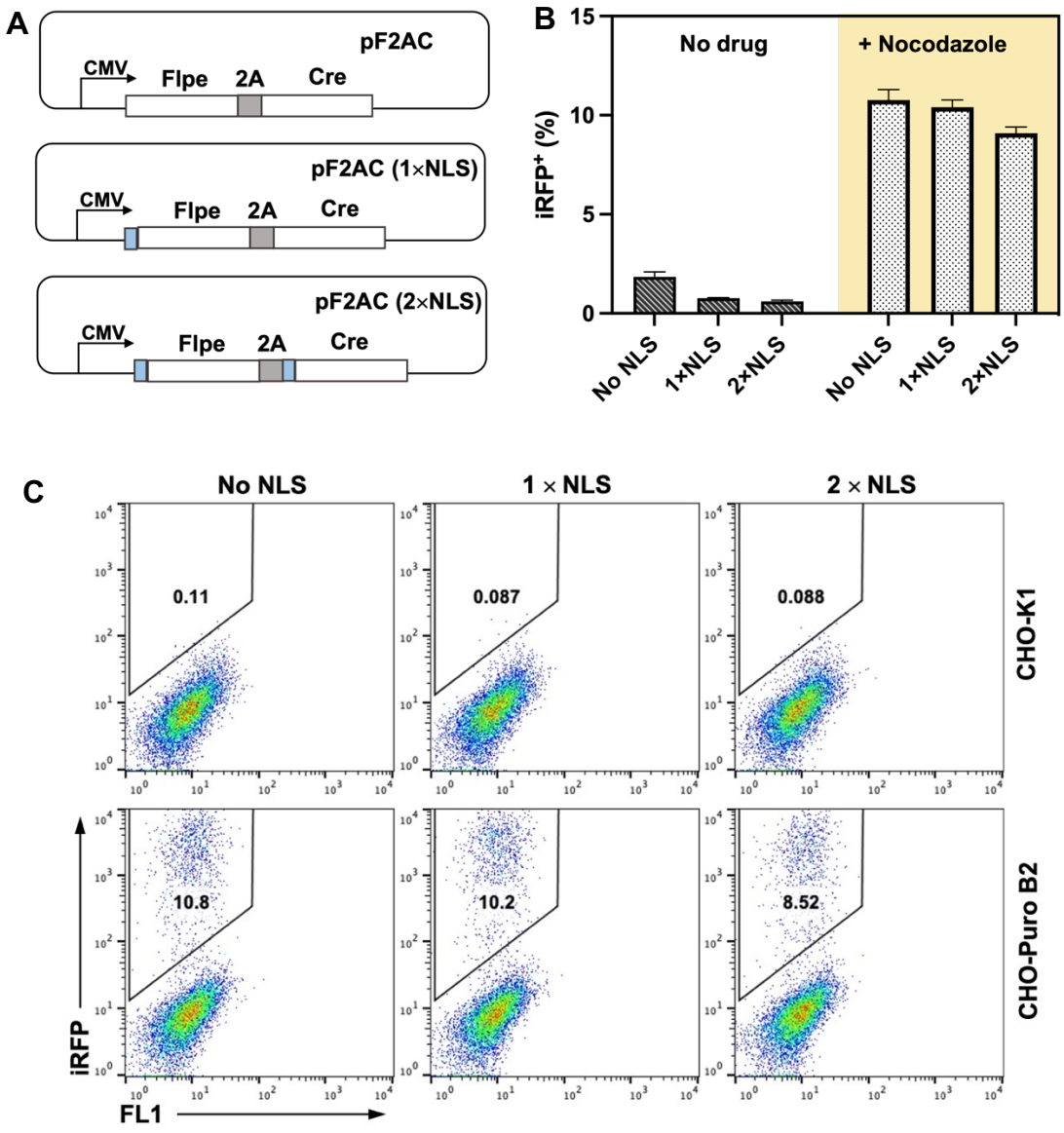


Figure 3.9 The effect of nuclear localization signal (NLS) on aRMCE efficiency. **(A**) The recombinase plasmid pF2AC. Simian virus 40 NLS (blue blocks) was fused to the N-terminus of Flpe (1×NLS), or both N-termini of Flpe and Cre (2×NLS) on pF2AC. **(B)** Cells of CHO-Puro B2 clone were treated with 650 nM nocodazole for 24 h, transfected, and FACS analyzed on day 2 after transfection (n = 3). No NLS or no drug treatment as controls were shown here. **(C)** Flow cytometry plots of the trials with nocodazole treatment. CHO-K1 as control was shown here.

3.3.6 Efficiency improvement by aRMCE was cell clone independent and gene of interest (GOI) independent

To test whether the RMCE efficiency improvement by cell cycle arrest is cell clone dependent, we conducted nocodazole-mediated cell arrest on clones A6 and C5. and found the RMCE efficient increased to ~ 7% after nocodazole treatment (**Fig. 3.10**), while no drug controls gave 1.91% and 1.66% (**Fig. 3.3C**). Moreover, we transfected pEx-EGFP and pEx-IgG to clone B2 with nocodazole arrest, respectively and found 10.8% of cells were EGFP positive (**Fig. 3.11A**) and 9.5% of cells displayed IgG (**Fig. 3.11B**), suggesting the cell cycle arrest method to improve RMCE efficiency is applicable to different replacement genes.

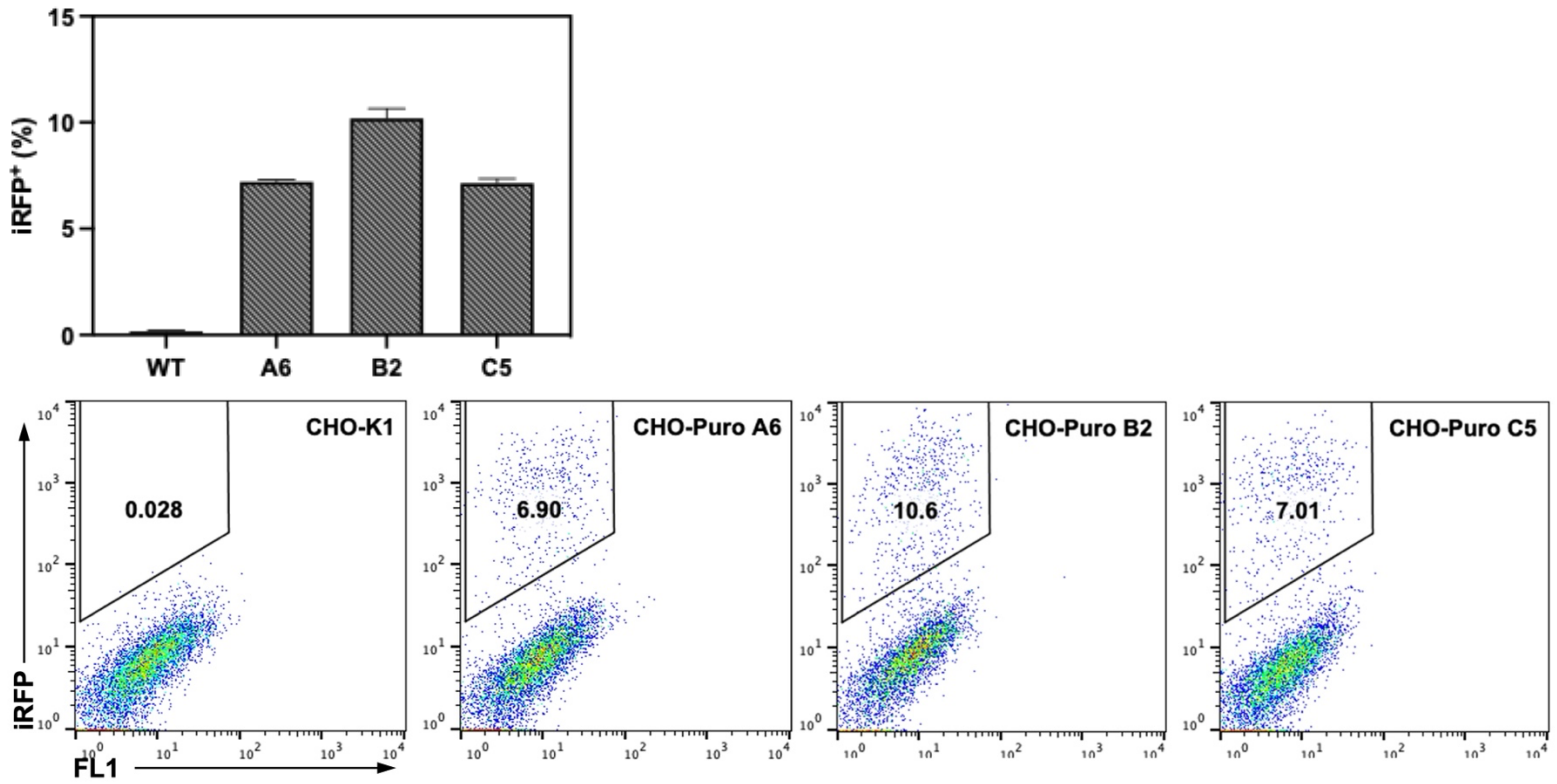


Figure 3.10 aRMCE with three CHO-Puro cell lines. Cells of A6, B2 and C5 clones were treated with 650 nM nocodazole for 24 h, transfected, and FACS analyzed on day 2 after transfection (n = 3). CHO-K1 was used as control.

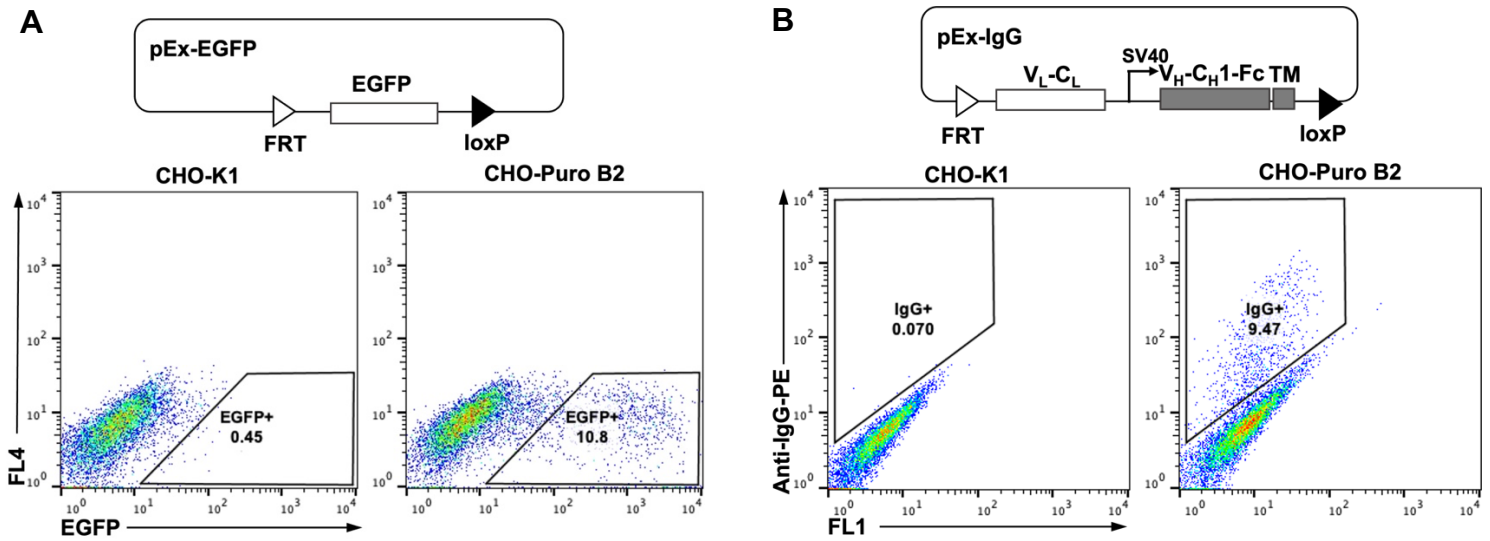


Figure 3.11 aRMCE for **(A)** EGFP and **(B)** human IgG1.CHO-Puro B2 cells were treated with 650 nM nocodazole for 24 h, transfected with the associated exchange plasmids, and analyzed by FACS on day 2 (EGFP) or day 4 (IgG1) post transfection. CHO-K1 was used as control. The exchange plasmids are shown as well.

3.3.7 Validation of single landing pad

The clone B2 cells were co-transfected with EGFP and iRFP exchange plasmids (1:1 ratio), and dual-color flow cytometry indicated that in the presence of Flpe/Cre, nocodazole treatment resulted in 7.7% iRFP+ and 6.4% GFP+ cells (**Fig. 3.12**), with double positive cells at undetectable level at day 26 post-transfection. This result further validated that B2 clone carries one landing pad. Notably, the temporary double positive events present in early days after transfection were presumably caused by either two copies of chromosomes in the cells arrested at mitosis, or the residual mRNA / protein remained in the divided cells. This notion was supported by the observation that the double positive population gradually decreased over the time and eventually disappeared on day 26 while the single positive population fluctuated between 10.5% and 14.0%. It also suggested that the inhibition of division on certain cells was long lasting even after the nocodazole was removed.

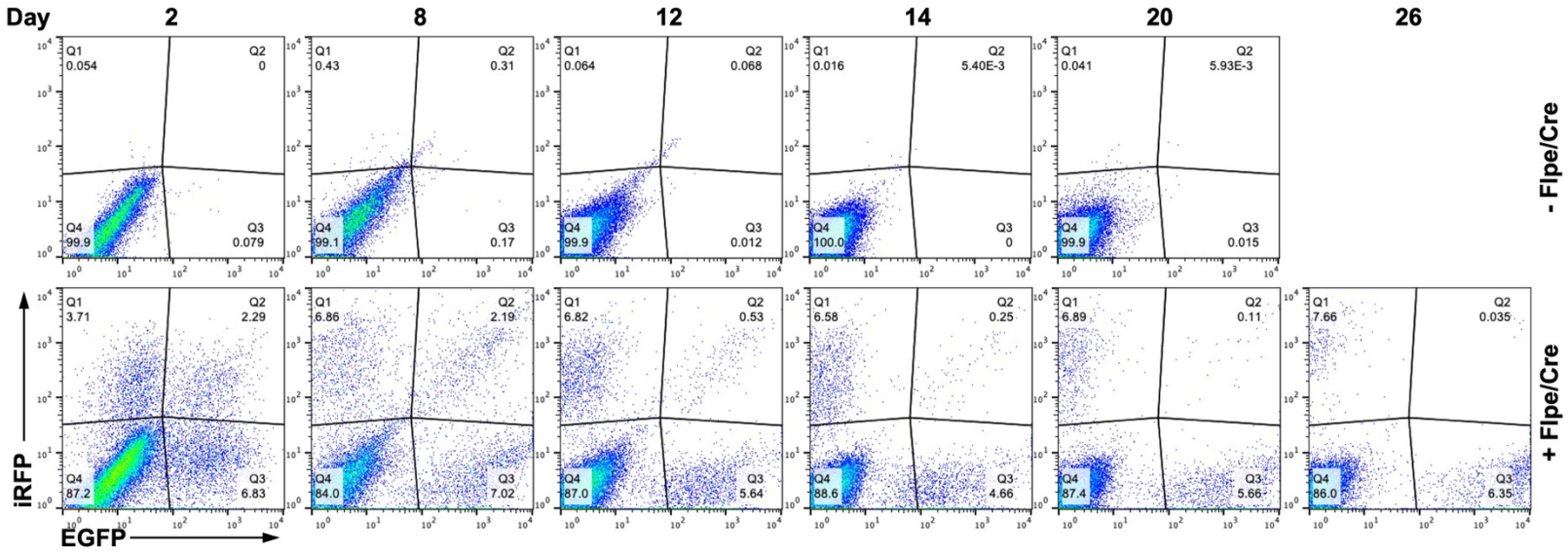


Figure 3.12 aRMCE with co-transfection of EGFP and iRFP exchange plasmids (1:1). Cells of CHO-Puro B2 clone treated with 650 nM nocodazole for 24 h, co-transfected, and FACS analyzed 2-26 days post-transfection. Cells transfected with the vector without Flpe and Cre genes were used as control.

3.4 Discussion and Conclusions

In this Chapter, two sequential rounds of RMCE were used for the generation of CHO cell clones carrying single landing pad on their genome. Co-transfection with the exchange plasmids encoding either *EGFP* or *iRFP* showed no double positive cells, confirming the single landing pad of the constructed RMCE-ready CHO cell lines. In addition, the singularity of the landing pad has been also validated by Southern blot analysis on a RMCE cell line generated by the same approach[16].

Next, the effect of cell cycle progression on the RMCE efficiency was investigated for the first time. We report that the RMCE efficiency on day 2 post transfection was improved by 8.6-fold with 650 nM nocodazole 24 h treatment. Presumably, this significant improvement on RMCE efficiency can attribute to multiple effects of nocodazole-mediated cell arresting at early mitosis: (1) doubling the amount of the landing pad per cell, (2) prolonging available time for the recombination event, (3) disintegrating the barrier – nucleus membrane, (4) facilitating DNA packaging and repairing, and (5) delaying the further condensation of chromatin and thus providing an easier access to the recombination site.

Although aphidicolin, thymidine and hydroxyurea all arrest cells in late G1 or early S phases by preventing DNA synthesis, an improved RMCE efficiency was only observed with hydroxyurea treatment but not with aphidicolin or thymidine. This might be due to their different intrinsic mechanisms: aphidicolin inhibits DNA polymerase[17]; thymidine creates an unbalanced dNTP pool[18,19]; and hydroxyurea inhibits ribonucleotide reductase

that catalyzes the synthesis of deoxyribonucleotides from ribonucleotides[20]. In addition, two 12-h sequential treatments with aphidicolin and thymidine were required for cell arrest at the entry of S phase[21,22], and the one 24-h treatment protocol used in this study may be less effective.

We further optimized the drug concentration, drug treatment time and culturing time after drug treatment. Results showed 24-h treatment with 650 nM nocodazole gave the optimal cell morphology with highest RMCE efficiency at 48 h after transfection, which was thereby chosen as the condition for cell cycle arrest RMCE. This approach, termed aRMCE, has also been proved GOI-independent and could be applied to other cell clones as well.

Fusing NLS to the recombinases did not improve RMCE efficiency in the cells treated with 650 nM nocodazole. One possibility could be that the nuclear membrane has disintegrated in mitosis when NLS had no impact on nuclear entry. However, it is surprising to see that NLS also failed to increase the RMCE efficiency on the cells without nocodazole treatment. Given that an earlier study showed that only 1% - 5% of the plasmids per cell could enter the nucleus after PEI transfection[24], the limiting factor of the RMCE event could more likely to be the exchange plasmid and/or the landing pad available for recombination.

The aRMCE efficiency still has the potential for further improvement. For instance, electroporation could be a good alternative to PEI considering higher gene targeting efficiency has been observed in nuclease-based systems[25,26]. In terms of recombinases, serine-based recombinases such as Bxb1 could be used which has been reported with

high recombination rate and irreversibility[27–29]. Moreover, recombinases could be stably expressed instead of being transfected each time, a strategy to increase the input of the exchange plasmids. It is noteworthy that inducible promoters should be used to avoid the repeated cycles of RMCE or cell toxicity[30].

3.5 References

1. Symens, N. *et al.* Intracellular partitioning of cell organelles and extraneous nanoparticles during mitosis. *Adv Drug Deliver Rev* 64, 78–94 (2012).

2. Bai, H., Lester, G. M. S., Petishnok, L. C. & Dean, D. A. Cytoplasmic transport and nuclear import of plasmid DNA. *Bioscience Rep* 37, (2017).

3. Nurse, P. A Long Twentieth Century of the Cell Cycle and Beyond. *Cell* 100, 71–78 (2000).

4. Brunner, S. *et al.* Cell cycle dependence of gene transfer by lipoplex, polyplex and recombinant adenovirus. *Gene Ther* 7, 401–407 (2000).

5. Fay, N. & Panté, N. Nuclear entry of DNA viruses. *Front Microbiol* 6, 467 (2015).

6. Rijck, J. D., Vandekerckhove, L., Christ, F. & Debyser, Z. Lentiviral nuclear import: a complex interplay between virus and host. *Bioessays* 29, 441–451 (2007).

7. Suzuki, Y. & Craigie, R. The road to chromatin — nuclear entry of retroviruses. *Nat Rev Microbiol* 5, 187–196 (2007).

8. Männistö, M. *et al.* The role of cell cycle on polyplex-mediated gene transfer into a retinal pigment epithelial cell line. *J Gene Medicine* 7, 466–476 (2005).

9. Marenzi, S. *et al.* Efficiency of expression of transfected genes depends on the cell cycle. *Mol Biol Rep* 26, 261–267 (1999).

10. Lopez, T. *et al.* Functional selection of protease inhibitory antibodies. *Proc National Acad Sci* 116, 16314–16319 (2019).

11. Buchholz, F., Angrand, P.-O. & Stewart, A. F. Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* 16, 657–662 (1998).

12. Buchholz, F., Ringrose, L., Angrand, P.-O., Rossi, F. & Stewart, A. F. Different Thermostabilities of FLP and Cre Recombinases: Implications for Applied Site-Specific Recombination. *Nucleic Acids Res* 24, 4256–4262 (1996).

13. 326. Insertion of an Internal Intron Increases Transgene Expression Levels from a Bicistronic Lentivirus Vector. *Mol Ther* 15, S123 (2007).

14. Pañeda, A. *et al.* 1079. Effect of WPRE on Transgene Expression Using Different Promoters in the Context of Hydrodynamically Delivered Plasmid Vectors. *Mol Ther* 13, S413–S414 (2006).

15. Qiao, J., Oumard, A., Wegloehner, W. & Bode, J. Novel Tag-and-Exchange (RMCE) Strategies Generate Master Cell Clones with Predictable and Stable Transgene Expression Properties. *J Mol Biol* 390, 579–594 (2009).

16. Chen, C., Li, N., Zhao, Y. & Hang, H. Coupling recombinase-mediated cassette exchange with somatic hypermutation for antibody affinity maturation in CHO cells. *Biotechnol Bioeng* 113, 39–51 (2016).

17. Baranovskiy, A. G. *et al.* Structural basis for inhibition of DNA replication by aphidicolin. *Nucleic Acids Res* 42, 14013–14021 (2014).

18. Chabes, A. & Stillman, B. Constitutively high dNTP concentration inhibits cell cycle progression and the DNA damage checkpoint in yeast Saccharomyces cerevisiae. *Proc National Acad Sci* 104, 1183–1188 (2007).

19. Hyland, P. L. *et al.* Effect of a dCTP:dTTP pool imbalance on DNA replication fidelity in friend murine erythroleukemia cells. *Environ Mol Mutagen* 36, 87–96 (2000).

20. Xu, Y., Singh, A. & Alter, G. M. Hydroxyurea Induces Cytokinesis Arrest in Cells Expressing a Mutated Sterol-14α-Demethylase in the Ergosterol Biosynthesis Pathway. *Genetics* 204, 959–973 (2016).

21. Jackman, J. & O'Connor, P. M. Methods for Synchronizing Cells at Specific Stages of the Cell Cycle. *Curr Protoc Cell Biology* 00, 8.3.1-8.3.20 (1998).

22. Humphrey, T., Brooks, G. & Harper, J. V. Cell Cycle Control, Mechanisms and Protocols. *Methods Mol Biology Clifton N J* 296, 157–166 (2004).

23. Vaughan, E. E. & Dean, D. A. Intracellular Trafficking of Plasmids during Transfection Is Mediated by Microtubules. *Mol Ther* 13, 422–428 (2006).

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

25. Parthiban, K. *et al.* A comprehensive search of functional sequence space using large mammalian display libraries created by gene editing. *Mabs* 11, 884–898 (2019).

26. 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).

27. 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, 87–87 (2013).

28. Yamaguchi, S. *et al.* A Method for Producing Transgenic Cells Using a Multi-Integrase System on a Human Artificial Chromosome Vector. *Plos One* 6, e17267 (2011).

29. 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, gkx183- (2017).

30. Andrusaite, A. & Milling, S. Should we be more cre-tical? A cautionary tale of recombination. *Immunology* 159, 131–132 (2020).

Chapter 4: Cell Cycle Arrest RMCE Library Construction and Diversity Validation by Deep Sequencing

Abstract

Next generation sequencing has enabled new and advanced ways to assess the diversity of combinatorial libraries. In this Chapter, we applied high-throughput Illumina sequencing to acquire "digital" insights on the diversity and mutation profiles of the antibody fragment crystallizable (Fc) libraries constructed in mammalian cells via aRMCE, of which the mutations were generated by error-prone PCR. We custom-designed our deep sequencing pipeline, including sample preparation, data processing, and bioinformatics analysis to ensure the authenticity of the library diversity. As a result, the unique clone numbers of the libraries derived from 2×10^5 and 1×10^6 transfected cells were conservatively estimated to be 18,678 and 112,406, respectively. Additional library profiling indicated a significant uneven distribution of the clone abundance, an average DNA mutation of 3.1 bp and pronounced substitution preference to AT→GC and AT→TA, which are consistent with the characteristics of the error-prone PCR generated library and the results from Sanger sequencing. Furthermore, we performed statistical analysis of mark-recapture models and estimated the aRMCE efficiency of 3.7% - 5.8%, with the library derived from 1×10^7 transfected cells. Overall, our diversity validation suggested that a library of at least 10^7 Fc variants was generated by transfection of 3×10^8 cells via aRMCE

4.1 Introduction

The ability to generate a large combinatorial library of protein variants is critical for any directed evolution endeavor. Mounting evidence suggested that the chance of discovering protein mutants with desired properties, e.g., specific binding, improved affinity, enzymatic activity, is proportionally corelated with the library diversity. To determine library size, traditionally, transformed or transfected cells are serially diluted, cultured under selection and counted, e.g., resistant colonies grown on agar plates containing the associated antibiotics. In addition, libraries are profiled by conducting Sanger sequencing with a small number of clones, usually dozens. This is less efficient and only provides limited information about the library.

Developed in 2000's, next-generation sequencing (NGS) technologies enabled massive parallel sequencing with throughput of millions or even billions of reads[1–4]. Since their introduction, NGS and associated bioinformatics have profoundly impacted many aspects of life sciences[1,5–9], including immune repertoire engineering. The Quake group first reported high-throughput sequencing of the variable domain of the antibody heavy chain in zebrafish in 2009[10]. Later, the Georgiou group sequenced antibody variable region gene repertoires from bone marrow plasma cells of immunized mice to isolate antigen specific antibodies without the screening step[11]. The same group also developed high-throughput single cell sequencing to determine the paired heavy- and light-chain variable regions in a given B-cell repertoire in human cells[12]. Recently, the Burton group sequenced the circulating B cell population of ten individuals at an unprecedented depth, revealing the commonality of human antibody repertoires despite

exceptional diversity[13]. Furthermore, the Ning Jiang group developed tetramer-associated T-cell receptor (TCR) sequencing to link TCR sequences to their cognate antigens in single cells at high throughput[14].

In **Chapter 3**, we have demonstrated that aRMCE efficiency can reach approximately 10% as determined by the percentage of the positive cell population on flow cytometer. However, the effective diversity of constructed combinatorial libraries cannot be simply extrapolated by flow cytometer, as some variants may carry stop codons, reading-frame shifts, or mutations that jeopardize expression and display of the target protein. Encouraged by technological advancements on NGS, in this Chapter, we applied high-throughput Illumina sequencing which provided "digital" insights on the diversity and mutation profiles of our combinatorial libraries constructed in mammalian cells. More specifically, we chose antibody fragment crystallizable (Fc) as our target owing to its important role in antibody-mediated immune response by recruiting effector cells, such as T cells, natural killer cells, and macrophages. In addition, Fc's biological function is fully dependent on its correct N-glycosylation at N297, and thus an ideal target for mammalian system. Following error-prone PCR, CHO cells of 10^5-10^8 orders of magnitude were transfected via aRMCE with diversified Fc fragments (**Fig. 4.1**). The libraries derived from 2×10^5, 1×10^6 and 1×10^7 transfected cells were subjected to deep sequencing, bioinformatics, and statistical analysis in this Chapter. Moreover, the library derived from 3×10^8 transfected cells will be screened for high affinity binder to FcγRs in **Chapter 5**.

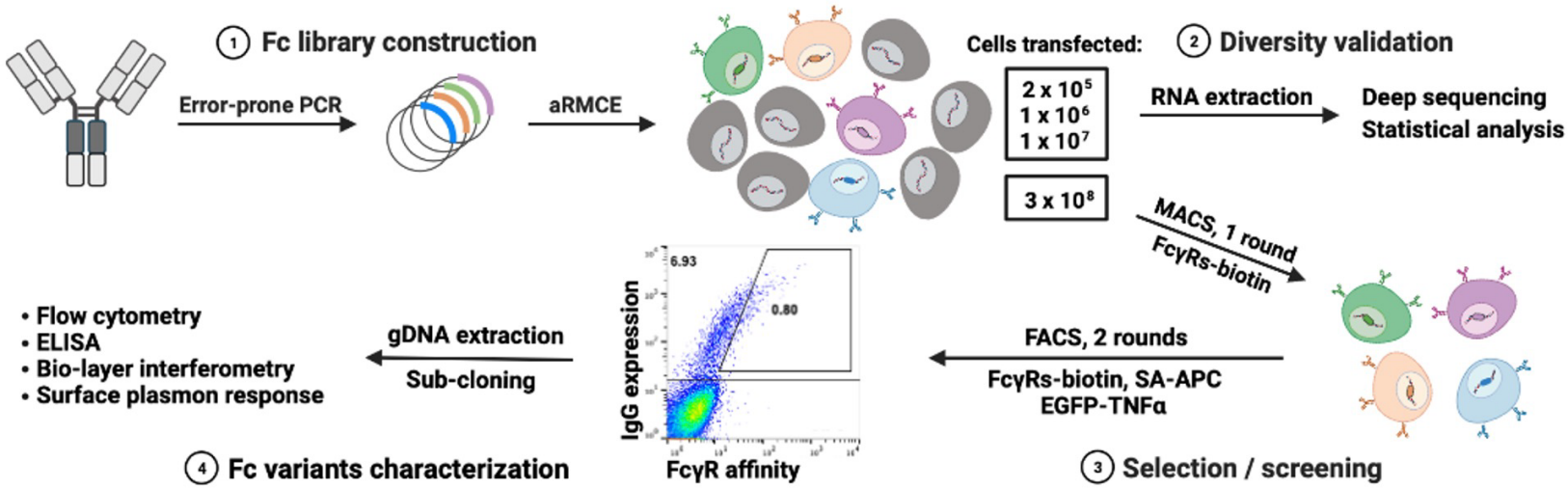


Figure 4.1 Application of aRMCE mammalian combinatory library system for Fc engineering. **(1)** Fc libraries were generated by error-prone PCR on the CH2 domain, and gnomically integrated in a locus-specific and monoclonal manner via aRMCE. **(2)** For libraries derived from 10^5-10^7 transfected CHO cells, their diversities were validated by deep sequencing and statistical analysis. **(3)** A Fc library of ~10^7 diversity was constructed with transfection of 3x10^8 cells and subjected to 1 round of MACS and 2 rounds of FACS for high affinity binders towards the corresponding FcγRs. **(4)** Fc variants were identified and characterize

4.2 Materials and Methods

4.2.1 Error-prone PCR and library construction in *E. coli*

The hinge-C_H2 region (C200-K340) was amplified with IgG1 heavy chain as the template by error-prone PCR using the following recipe: 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 6.5 mM $MgCl_2$, 0.5 mM $MnCl_2$, 1 mM dCTP, 1 mM dTTP, 0.2 mM dATP, 0.2 mM dGTP, 0.3 μM 5′ primer, 0.3 μM 3′ primer, 25.6 pM template DNA, and 0.05 U/μl Taq DNA polymerase. The thermal cycling program was 95 °C for 2 min; 25 cycles of 95 °C for 15 s, 61 °C for 20 s, 68 °C for 50 s; final extension at 68 °C for 5 min and storage at 4 °C. The obtained PCR fragment was digested with BstXI and BsrGI-HF (NEB) and ligated into the RMCE exchange plasmid pEx-3f8HC. The ligation mixture was then transformed into *E. coli* SS320 competent cells, resulting in 1.5×10^8 transformants. Dozens of colonies were randomly picked for miniprep and Sanger sequencing.

4.2.2 Generation of CHO-Puro-3f8LC single clone stable cell lines

mAb 3f8 light chain gene was fused with IRES followed by a zeocin resistant gene with overlapping PCR. The product (3f8LC-IRES-Zeo) was cloned into pCIW with HindIII and XhoI. 5×10^5 cells of CHO-Puro B2 clone was transfected with 2 μg of obtained pCIW-3f8LC-IRES-Zeo and 8 μg PEI 40K. After 24 h, cells were detached and cultured in a 10 cm plate with 1 μg/ml puromycin and 700 μg/ml zeocin for 7 days. Cells were cultured with 100 μg/ml zeocin for 3 more days, and the survived cells were seeded

into 96-well plates at a density of 0.5 cell/well. After 7-day culture, single cell clones stably expressing 3f8 light chain were identified and expanded for further study.

4.2.3 aRMCE transfection and mammalian library construction

For typical aRMCE transfections, 5×10^4 cells of CHO-Puro-3f8LC clone 8 in 1 ml growth media were seeded in one well of a 24-well plate. After cultivation for ~24 h allowing cell doubled, the culture supernatant was replaced with 450 µl growth media containing 650 nM nocodazole, and cells were transfected with 250 ng pF2AC, 250 ng exchange plasmid (e.g., pEx-3f8HC) and 2 µg PEI 40K prepared in 50 µl Expi293 media. Following incubation for 24 h, cells were washed with PBS, detached with 0.25% trypsin-EDTA, and cultured in a 6-well plate with 2 ml fresh growth media until flow cytometric analysis. For library constructions, the aRMCE transfections of pEx-3f8HC library plasmids were conducted similarly as described above, with the amounts of all reagents (DNA, PEI, nocodazole, etc) scaled proportionally according to the areas of culture surface. Specifically, one well of 24-well plates, one well of 6-well plates, one 15 cm plates and 30 of 15 cm plates were used for transfection of 2×10^5, 1×10^6, 1×10^7 and 3×10^8 cells with 0.25, 1, 15, 450 µg pEx-3f8HC library plasmids. Cells were also transfected with pEx-3f8HC carrying Fc WT as controls.

4.2.4 Flow cytometry analysis

After cultivation, transfected cells were harvest, stained with 1 µg/ml mouse anti-human IgG (Fab)-PE (Invitrogen, Cat#MA1-10377) and analyzed on a S3e cell sorter. PE was excited at 488 nm and detected with a 593/40 nm filter (FL2 channel).

4.2.5 Sample preparation and Illumina sequencing

All transfected cells were harvested on day 2 post transfection for total RNA extraction respectively with the RNeasy Plus Mini kit (Qiagen, Cat#74134). Heavy chain cDNA synthesis and amplification were performed with 2 µg purified total RNA as the template in a 100 µl reaction volume by using a SuperScript IV One-Step RT-PCR kit (Thermo Fisher, Cat#12594025). The RT-PCR thermal cycling was: 50 °C for 10 min, 98 °C for 2 min; 25 cycles of 98 °C for 10 s, 67.3 °C for 10 s, 72 °C for 40 s; final extension at 72 °C for 5 min and storage at 4 °C. The products were purified by DNA electrophoresis separately to avoid cross contamination, and the ~1.2 kb bands were extracted (QIAquick Gel Extraction Kit, Cat#28704) for BstXI/BsoBI double digestion and gel purification. For the library derived from 2×10^5 transfected cells, the generated hinge-C_H2 fragments were directly ligated with assembled Illumina index adapters and flow cell adapters at a 1:5:10 molar ratio. For the libraries derived from 1×10^6 and 1×10^7 transfected cells, the generated hinge-C_H2 fragments were ligated with modified Illumina index adapters carrying 0/2/4/6 nucleotide offsets (**Fig 4.5B**). The ligation products were gel purified and PCR amplified to introduce flow cell adapters at 5' and 3' ends using Q5® High-Fidelity DNA polymerase (NEB) with the following thermal cycling: 98 °C for 30 s; 2 cycles of 98 °C for 10 s, 53 °C for 20 s, 72 °C for 30 s; 6 cycles of 98 °C for 10 s, 65 °C for 20 s, 72 °C for 30 s; final extension at 72 °C for 5 min and storage at 4 °C. For multiplexing, the illumina I5 indexes [N/S/E] 501, 502, 505 and I7 indexes N703, 704, 709 were chosen for libraries derived from 2×10^5, 1×10^6 and 1×10^7 transfected cells, respectively. Final PCR products (596-608 bp) were gel purified

(Zymo Research), and their purities and concentrations were determined by spectrophotometry, Agilent 2100 Bioanalyzer and qPCR. The quality confirmed DNA sample derived from 2×10^5 transfected cells was sequenced on an Illumina MiSeq using the reagent kit nano v2 (2×250 cycles); the ones derived from 1×10^6 and 1×10^7 transfected cells were pooled at desired concentrations for optimal read returns. The pooled samples were further verified by Agilent 2100 Bioanalyzer and qPCR before sequencing.

4.2.6 Sequencing data processing and analysis

The pass filter reads were demultiplexed to generate the raw paired FASTQ files which were then quality checked with FASTQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Quality trimming was performed by Sickle (www.github.com/najoshi/sickle) with a minimum average quality score of 20 within the window size 25 nt. Specifically, the window slides along the quality values from the 5' end and remove the read and quality strings from the 3'-end when the average quality score in the window drops below the threshold 20. Processed reads were quality checked again using FASTQC, and paired reads were merged with PANDAseq using the default (simple Bayesian) merging algorithm[15]. Further sequencing analysis was performed with scripts custom-coded using Python 3.8.1, for unique sequences identification, alignments, data trimming and library profiling. Notably, the first 27 nt of Read 1 and the first 9 nt of Read 2 were not subjected for error-prone mutagenesis (Fig. 4.5B), and thus excluded from mutation profiling and diversity calculation.

For the library derived from 10^7 transfected cells, three attempts of Illumina sequencing with varied depth were carried out on the same cDNA replication sample. Using Lincoln-Petersen capture-recapture model, the diversity was estimated as $T = \frac{m \times n}{k}$ (**Table 4.3**), where T is the total clone number, m and n are the clone numbers in the capture and the recapture, and k is the number of the clones observed in both capture and recapture.

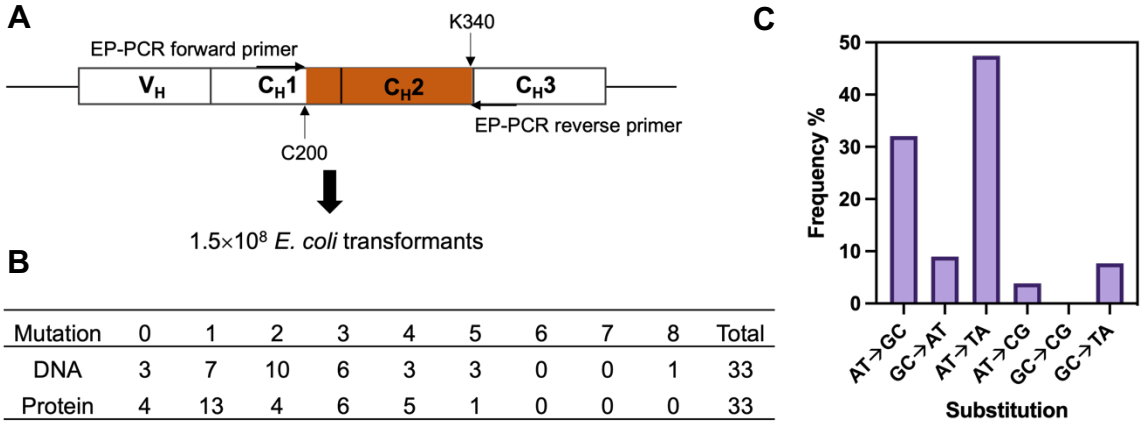
For the library derived from 10^7 transfected cells, three attempts of Illumina sequencing with varied depth were carried out on the same cDNA replication sample. Using Lincoln-Petersen capture-recapture model, the diversity was estimated as $T = \frac{m \times n}{k}$ (**Table 4.3**), where T is the total clone number, m and n are the clone numbers in the capture and the recapture, and k is the number of the clones observed in both capture and recapture.

4.3 Results

4.3.1 Construction of an error prone Fc library in *E. coli*

Structural biology studies suggested that C_H2 domain and the lower hinge region of IgG molecules are responsible for their binding with Fc gamma receptors[17–19]. Aiming to generate Fc variants with improved affinities towards specific FcγRs, the hinge-C_H2 region (C200-K340) was focused for Fc engineering by a combinatorial approach. Random mutagenesis was introduced by error-prone PCR, and electroporation of *E. coli*

competent cells with ligated Fc library plasmids resulted in 1.5×10^8 transformants (**Fig. 4.2A**). Sanger sequencing results of randomly picked colonies suggested that 94% (33/35) tested clones carried full-length heavy chain sequences with 0-8 mutations at DNA level (2.4 bp in average) and 0-5 mutations at amino acid level (1.9 residues in average) (**Fig 4.2B**), a mutation rate of 5.7 bp/kb consistent with the experimental design targeting at 4 bp/kb. Four tested clones had no amino acid mutations including one silent mutation. Besides, we found 4 clones had stop codons within the reading frame. Overall, 25 functional mutation sequences were identified out the 35 tested clones, equivalent to a library efficiency of 71%. Notably, EP-PCR by *Taq* DNA polymerase has well-pronounced preference to AT → GC transitions and AT → TA transversions[20,21]. Although an unbalanced dNTP concentration set was employed to alleviate the mutagenesis bias[2022], our mutation profiling revealed AT → GC transitions (32%) and AT → TA transversions (47%) were still dominant (**Fig 4.3C**). Characterized library plasmids were produced to μg scales and ready for mammalian cell transfection.



B

Mutation	0	1	2	3	4	5	6	7	8	Total
DNA	3	7	10	6	3	3	0	0	1	33
Protein	4	13	4	6	5	1	0	0	0	33

Figure 4.2 Generation of an error-prone Fc library in E. coli. **(A)** The hinge-CH2 region of mutagenesis (C200-K340, highlighted). **(B)** Mutation rate analysis at both DNA and AA levels in 33 E. coli transformants by Sanger sequencing. **(C)** Mutation profiling by Sanger sequencing.

4.3.2 Generation of single cell lines ready for Fc library aRMCE

As the efficiency of transfection and RMCE is reversely proportional to the size of integrated fragment, an exchange plasmid encoding heavy chain of anti-TNFα mAb 3f8[23] without its light chain was constructed (**Fig 4.3A**). Accordingly, the light chain of mAb 3f8 was integrated to genome of CHO-Puro cells by zeocin selection for sable expression of its light chain (**Fig 4.3B**). The obtained CHO-Puro-3f8LC polyclonal cells exhibited 9.3% recombination efficiency when transfected via aRMCE with the 3f8 heavy chain exchange plasmid (**Fig 4.3C**). To achieve a uniform transcription, CHO-Puro-3f8LC single cell clones were selected with zeocin following serial dilutions. Four single cell clones with more than 10% RMCE efficiency were identified (**Fig 4.3D**) among which clone #8 was chosen for future experiments.

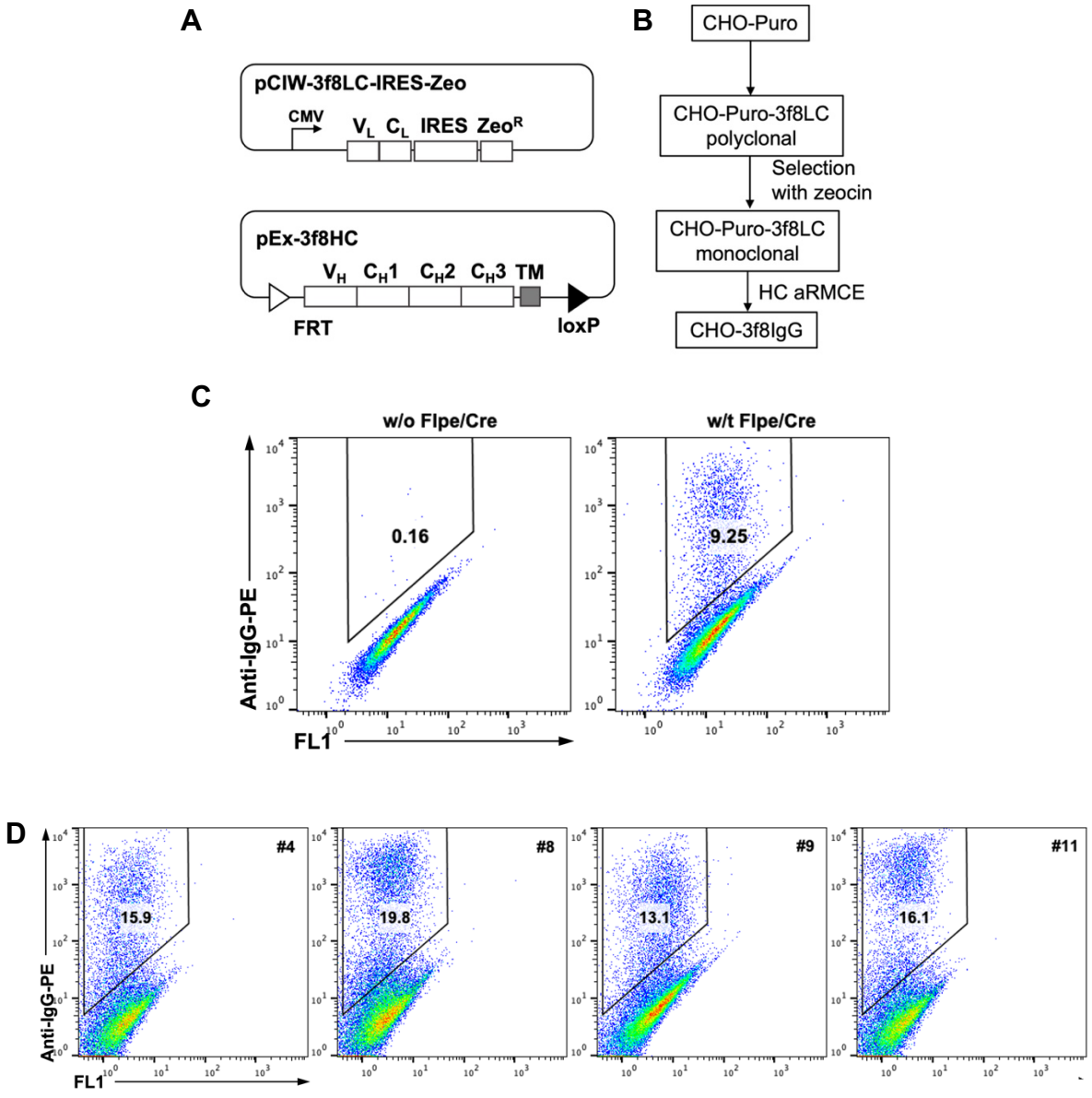


Figure 4.3 Selection of CHO monoclonal cells stably expressing anti-TNFα mAb 3f8 light chain (CHO-3f8LC). **(A)** Diagrams of light chain integration plasmid (pCIW-3f8LC-IRES-Zeo) and heavy chain exchange plasmid (pEx-3f8HC). **(B)** Flowchart on the generation of CHO cells constitutionally expressing 3f8 light chain (CHO-Puro-3f8LC) ready for heavy chain aRMCE. **(C)** Flow cytometric analysis of CHO-Puro-3f8LC multiclonal cells by transfection with pEx-3f8HC via aRMCE. **(D)** Flow cytometric analysis of four CHO-Puro-3f8LC single clones. Cells were stained with 1μg/mL anti-IgG-PE.

4.3.3 Mammalian Fc library construction

To test robusticity of the system, mammalian aRMCE was carried out with numbers of cells across four orders of magnitude, *i.e.,* 2×10^5, 1×10^6, 1×10^7 and 3×10^8 CHO-Puro-3f8LC clone #8 cells were transfected by aRMCE with 0.25, 1, 15, 450 μg, respectively, 3f8 heavy chain exchange plasmids carrying the error-prone Fc library. The surface areas used for cell culturing were also scaled up proportionally (refer to 4.2.3 for details). Considering the capacity of current Illumina platforms, the libraries derived from 2×10^5, 1×10^6 and 1×10^7 transfected cells were subjected to deep DNA sequencing under the condition that their diversities can be covered and thus fully validated. The library derived from 3×10^8 cells will be used for isolation affinity improved Fc variants in **Chapter 5**.

Following aRMCE transfection and cell culture, cells were stained with anti-IgG-PE and analyzed by flow cytometry. Results suggested that the percentage of IgG displayed cells was $2.7 \pm 1.1\%$ for these four libraries (**Fig 4.4A**). However, when 10^6 cells were aRMCE transfected in parallel with the 3f8 heavy chain exchange plasmid carrying WT Fc, 11% cells were IgG positive (**Fig 4.4B**). Presumably, this disparity between WT Fc and Fc library can be caused by (1) stop codon introduced by EP-PCR (12%, suggested by small scale sanger sequencing), and (2) low protein expression level due to deleterious mutations.

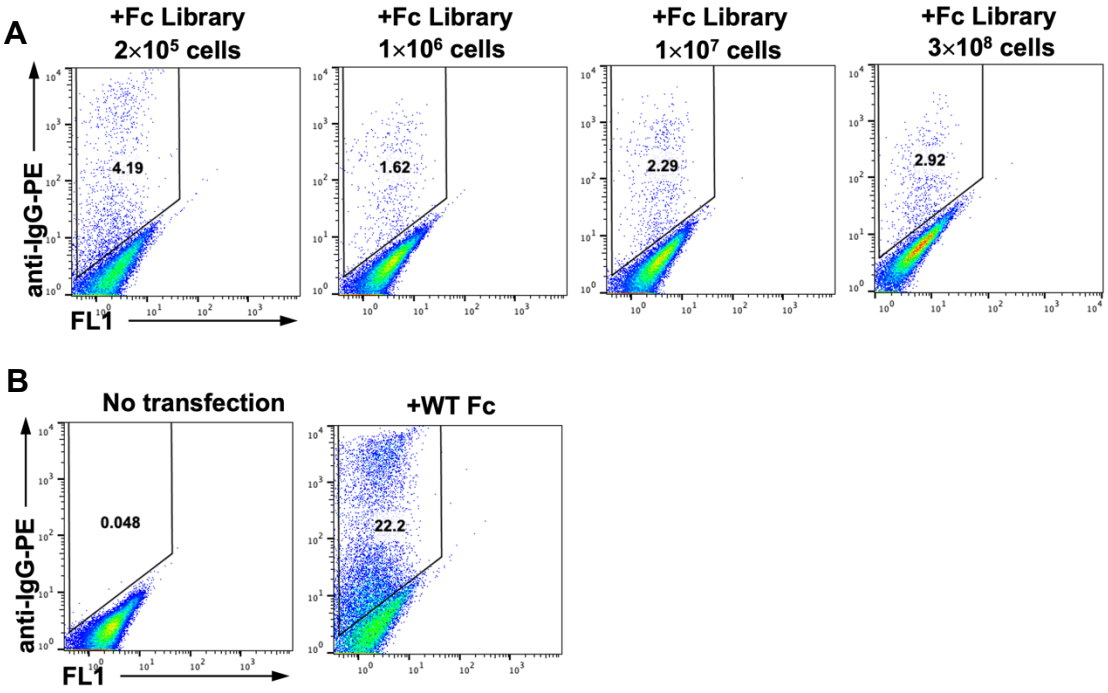


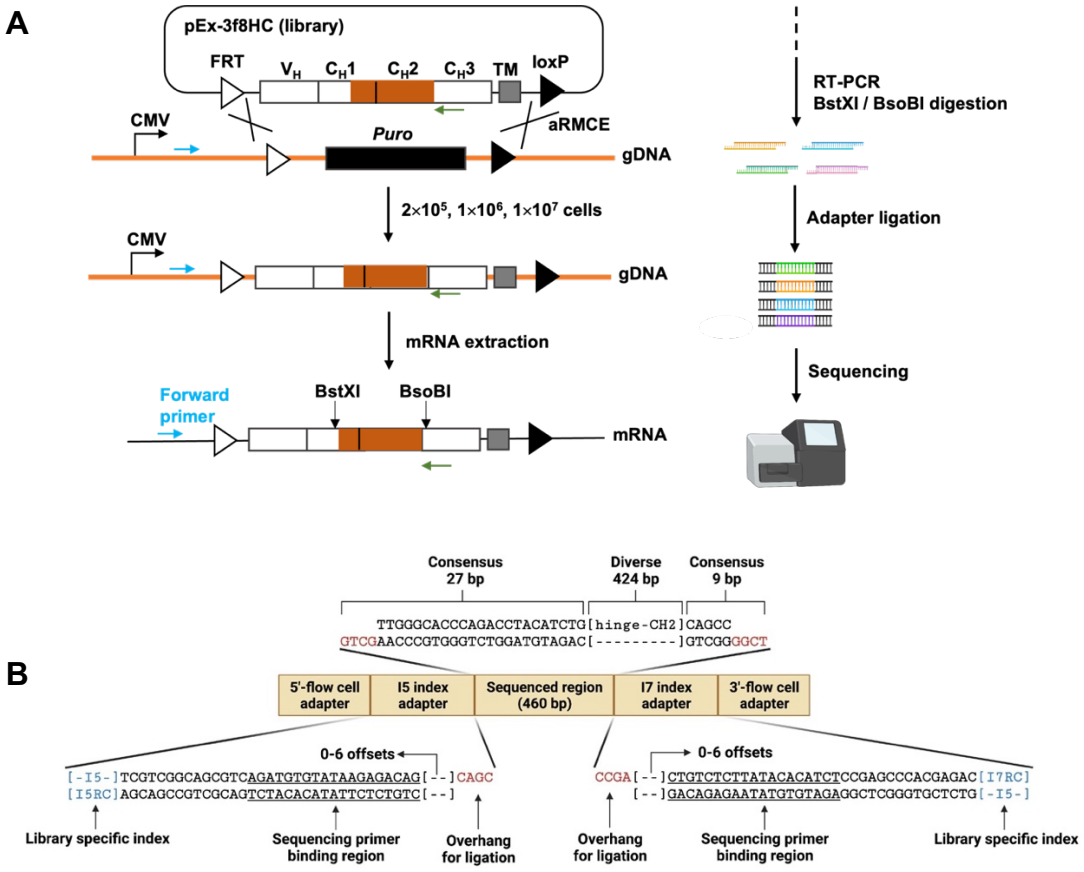
Figure 4.4 Flow cytometric analysis of mammalian Fc library constructed via aRMCE. (**A**) 2×10^5, 1×10^6,1×10^7 and 3×10^8 CHO-Puro-3f8LC cells were transfected with pEx-3f8HC Fc library exchange plasmids. Cells were stained with 1 μg/ml anti-IgG-PE before flow cytometry. (**B**) CHO-Puro-3f8LC cells without transfection and transfected with the Fc WT exchange plasmid were used as controls.

4.3.4 Library DNA sample preparation

Experiments were designed to conserve the diversity and authenticity of constructed libraries. To reduce template loss, total RNA instead of gDNA was extracted and used in the followed RT-PCR amplification, due to the low copy number of gDNA. To ensure only integrated Fc genes but not the ones on exchange plasmid to be amplified, the forward primer was designed to anneal at upstream of the 5' RMCE recognition site on the genome, and the reverse primer anneals between the two recognition sites (**Fig. 4.5A**). In addition, samples from different libraries were prepared separately to avoid

cross contamination. Fourth, non-palindromic restriction enzymes BstXI and BsoBI were used to ensure the hinge-C_H2 fragments ligate to the index adapters only.

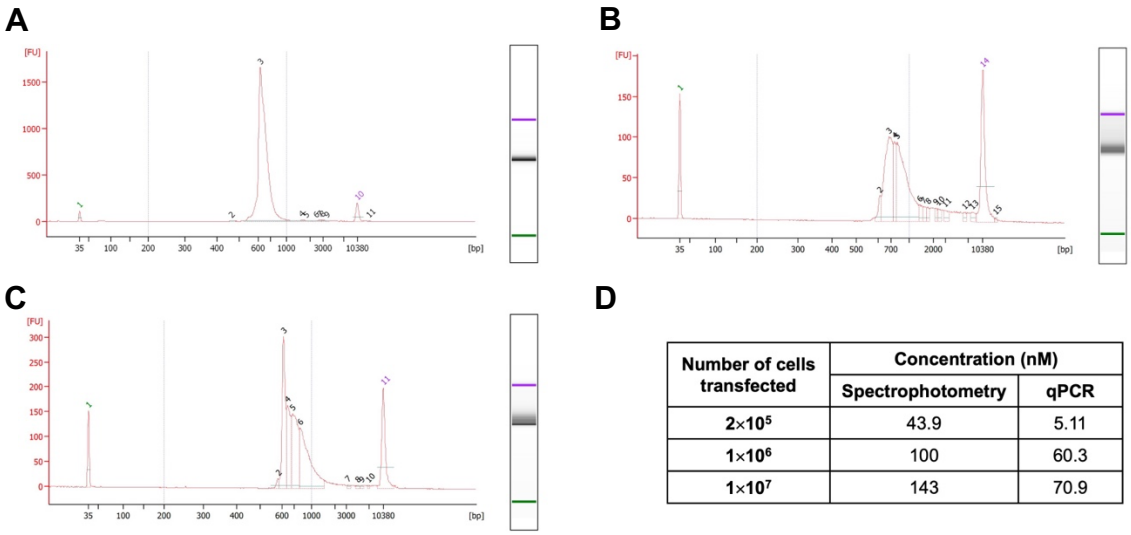
For the library derived from 2×10^5 transfected cells, the RT-PCR product was digested with restriction enzymes and the generated overhung hinge-C_H2 fragments were directly ligated with assembled Illumina index adapters and flow cell adapters, a method able to reduce biases and errors caused by extensive PCR amplification. Bioanalyzer exam on the ligated product after purification showed a sharp peak at 615 bp, similar to the target size 596 bp (**Fig. 4.6A**). Quantitative PCR (qPCR) performed with the primers specific to the flow cell adapters confirmed the adapter ligations and a concentration of 5.1 nM was measured, significantly less than the concentration measured by spectrophotometry (44 nM) (**Fig. 4.6D**), suggesting ligation was not efficient. Therefore, the hinge-C_H2 fragments from 1×10^6 and 1×10^7 transfected cells were ligated with the index adapters followed by an 8-cycle PCR to introduce flow cell adapters. Additionally, 2, 4, or 6 nucleotides were added as "offset" between the overhangs and the sequencing primer binding region to stagger the highly homologous error-prone fragments and even the sequencing burden (**Fig. 4.5B**). Owing to this, the bioanalyzer results showed a serial of peaks between 615 bp to 742 bp (**Fig. 4.6BC**). The qPCR derived concentrations of the libraries from 1×10^6 and 1×10^7 transfected cells were 60 nM and 78 nM, similar to those measured by spectrophotometry (100 nM and 143 nM) (**Fig. 4.6D**), respectively, indicating the majority can bind to the flow cell efficiently.



C

Number of cells transfected	I5 Index		I7 Index	
	Name	Sequence	Name	Sequence
2×10^5	[N/S/E]501	TAGATCGC	N703	AGGCAGAA
1×10^6	[N/S/E]502	CTCTCTAT	N704	TCCTGAGC
1×10^7	[N/S/E]505	GTAAGGAG	N709	GCTACGCT

Figure 4.5 Sample preparation for deep sequencing. **(A)** Gene structures and sample preparation procedure. The forward primer anneals at downstream of the CMV promotor while before the 5' recombination site FRT. The reverse primer anneals at C-terminal of diversified Fc region, between two recombination sites. **(B)** Design of upstream and downstream adapters accommodating flow cell adapters, indexes (in blue), sequencing primer (SP) binding regions, offsets, and overhang (in red) fragments. The 0/2/4/6 random nucleotide offsets were included to even the sequencing burden on the highly homologous samples generated by error-prone PCR. The 5'- and 3'-overhangs allow direct ligations of the adapters to the digested double-stranded hinge-C_H2 library DNA fragments without PCR amplification. **(C)** Indexes used for each library. The I5 and I7 indexes are for the upstream and downstream adapters, respectively, and chosen to maintain a 1:1 ratio of (A+C) : (G+T).



Number of cells transfected	Concentration (nM)	
	Spectrophotometry	qPCR
2×10^5	43.9	5.11
1×10^6	100	60.3
1×10^7	143	70.9

Figure 4.6 Quality control of samples for deep sequencing. Bioanalyzer results of libraries derived from **(A)** 2×10^5, **(B)** 1×10^6 and **(C)** 1×10^7 transfected cells. 0/2/4/6 offsets were added for library derived from 1×10^6 and 1×10^7 transfected cells, resulting in varied size of peaks. **(D)** Library concentrations measured by spectrophotometry and qPCR. 5' and 3' adapters were ligated with the hinge-C_H2 fragments derived from 2×10^5 transfected cells, leading to a huge disparity between the results from spectrophotometry and qPCR. Therefore, PCR was used for the hinge-C_H2 fragments derived from 1×10^6 and 1×10^7 transfected cells.

4.3.5 Sequencing data size and quality

MiSeq data derived from 2×10^5 transfected cells yielded 670,001 pass filter pair-end reads exhibiting high quality with FastQC scores >33 for both Read 1 and Read 2 (**Fig. 4.7**). Read 1 and Read 2 sequence pairs were then assembled based on their overlap region using PANDAseq, and 666,942 merged reads were generated (**Table 4.1**). For the raw data from 1×10^6 transfected cells, FastQC indicated quality scores below 28 at many positions in diverse regions for 3,610,500 pair-end reads. After quality trimming by Sickle, most of the positions had quality scores in the diverse regions above 28 (**Fig 4.8**),

and 3,462,010 fully assembled reads were resulted. For the library DNA derived from 1×10^7 transfected cells, as the diversity was expected to be large, three independent MiSeq runs were performed for accumulating more reads and facilitating the capture-recapture statistical analysis in **4.3.7**. After the similar quality trimming (**Fig 4.9**) and assembly pipeline, 7,461,882 merged raw reads were obtained from the library of 1×10^7 transfected cells.

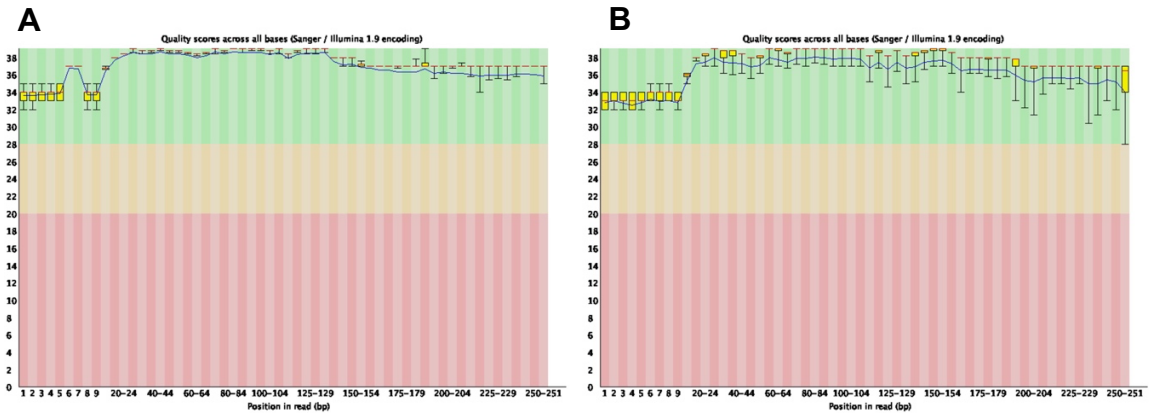


Figure 4.7 FASTQC report of the **(A)** Read 1 and **(B)** Read 2 of the library derived from 2×10^5 transfected cells.

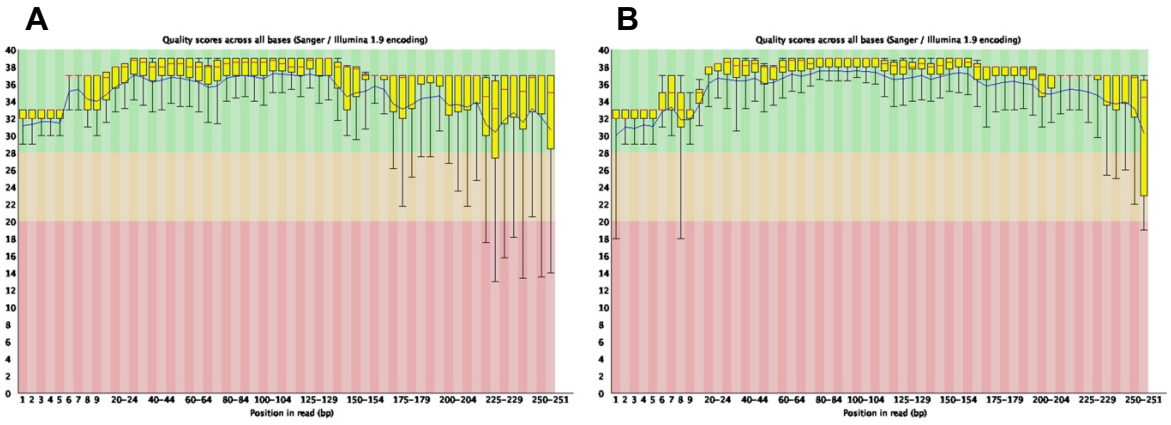


Figure 4.8 FASTQC report of the the quality-trimmed **(A)** Read 1 and **(B)** Read 2 of the library derived from 1×10^6 transfected cells.

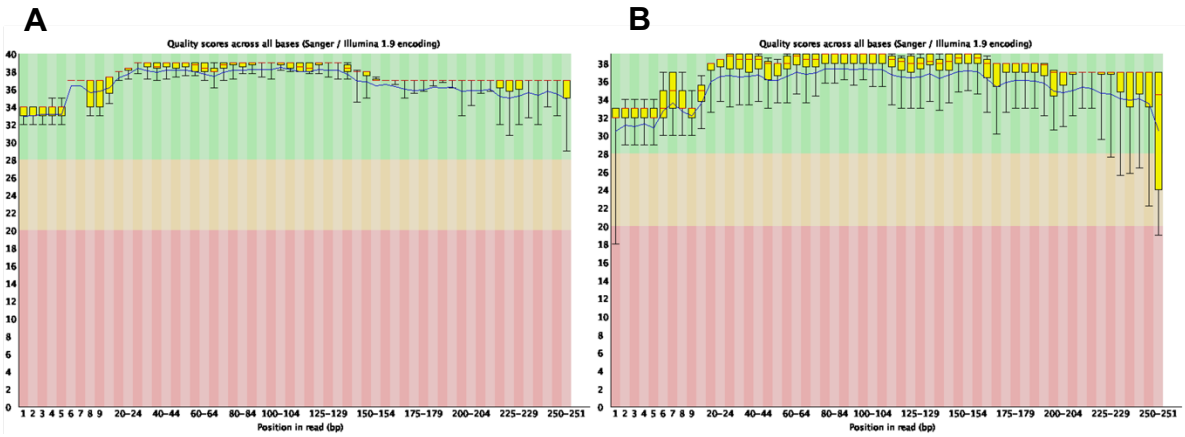


Figure 4.9 FASTQC report of the quality-trimmed **(A)** Read 1 and **(B)** Read 2 of the library derived from 1×10^7 transfected cells. 3 batches deep sequencing were performed and only Batch 3 was shown here.

Table 4.1 Deep sequencing results

Cells transfected	Raw reads passed quality filters		Fully assembled[a]	Unique[b]	Unique, reads ≥ 2	Trimming[c]		aRMCE efficiency (NGS confirmed / cells transfected)
						DNA alignment	AA alignment	
2×10^5	670,001		666,942	138,035	18,781	18,714	18,678	**9.3%**
1×10^6	3,610,500		3,462,010	1,070,543	113,652	112,734	112,406	**11%**
1×10^7	Batch 1	514,073	508,915	296,166	19,095	18,929	18,919	Not determined due to low coverage
	Batch 2	1,832,752	1,814,934	523,961	128,523	127,335	126,999	
	Batch 3	5,323,088	513,8033	1,533,570	238,430	236,922	236,346	
	Total	7,669,913	7,461,882	2,022,299	307,713	305,344	304,476	

[a] Number of pair-end reads after quality trimming and merging.
[b] 460 nt length.
[c] < 10 mismatches.

4.3.6 Unique clones and mutation profiling

Counterfeits could be present in the vast amount of sequencing data due to the sequencing errors. Prior to library profiling, we set up a workflow to identify the authentic unique clones in the library from 2×10^5 transfected cells.

First of all, we extracted the diverse region of the length 424 nt same as the WT (**Fig. 4.5B**). As a result, 138,035 unique clones were obtained. Second, we assume the ones with no repeat was counterfeits. After removing those, the unique clone number was decreased dramatically to 18,781 (**Table 4.1**). Third, we performed global pairwise sequence alignment (PSA) to wild type (WT) DNA sequence and found a noticeable number of clones bear up to 55 mismatches likely resulted from sequencing error (**Fig 4.10**). Therefore, the unique clone number was trimmed down to 18714 with an allowance of 10 mismatches given that most of the clones were distributed within 10 mismatches and one clone with 8 mutations was observed in the small-scale sanger sequencing. Furthermore, we have found in the merged reads of length longer or shorter than 424 nt that insertions or deletions tended to occur over the regions containing consecutively repeated nucleotides, which were probably generated from sequencing errors. We hypothesized that the merged reads of length 424 nt could have similar insertions and deletions in equal number, leading to a significant reading frame shift reflected by a low protein alignment score. Therefore, to remove the artifacts, we performed global PSA on the protein sequences translated from the DNA sequences of

the 18714 unique clones. Allowing 10 mismatches, the unique clone number reduced slightly to 18678 (**Table 4.1**).

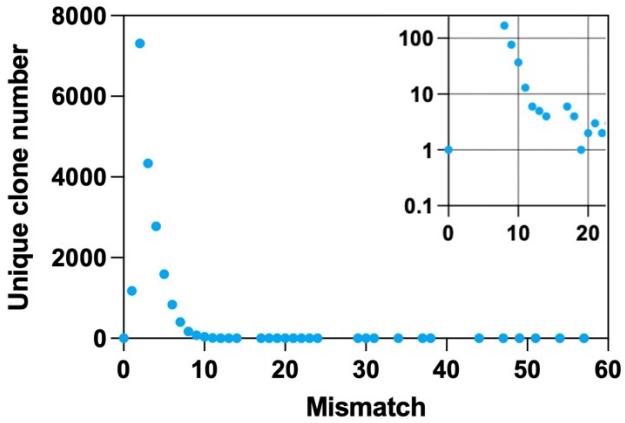


Figure 4.10 DNA mismatch distribution of the unique clones of length 424 with more than 1 read.

Moving forward, we performed library profiling on the 18678 unique clones. Results showed that the unique clones had a significantly uneven abundancy distribution (**Fig. 4.11A**). 8056 and 2254 individual clones had only 2 and 3 repeats, accounting for 43% and 12% of all the unique clones, respectively. As expected, WT was the most abundant clone in the library with over 10^5 repeats, composing approximately 17% of the total reads. The second most abundant clone with 4029 repeats had a single mutation C→G at the first position of the error-prone PCR region, which is surprising considering no C→G mutation was found in sanger sequencing result.

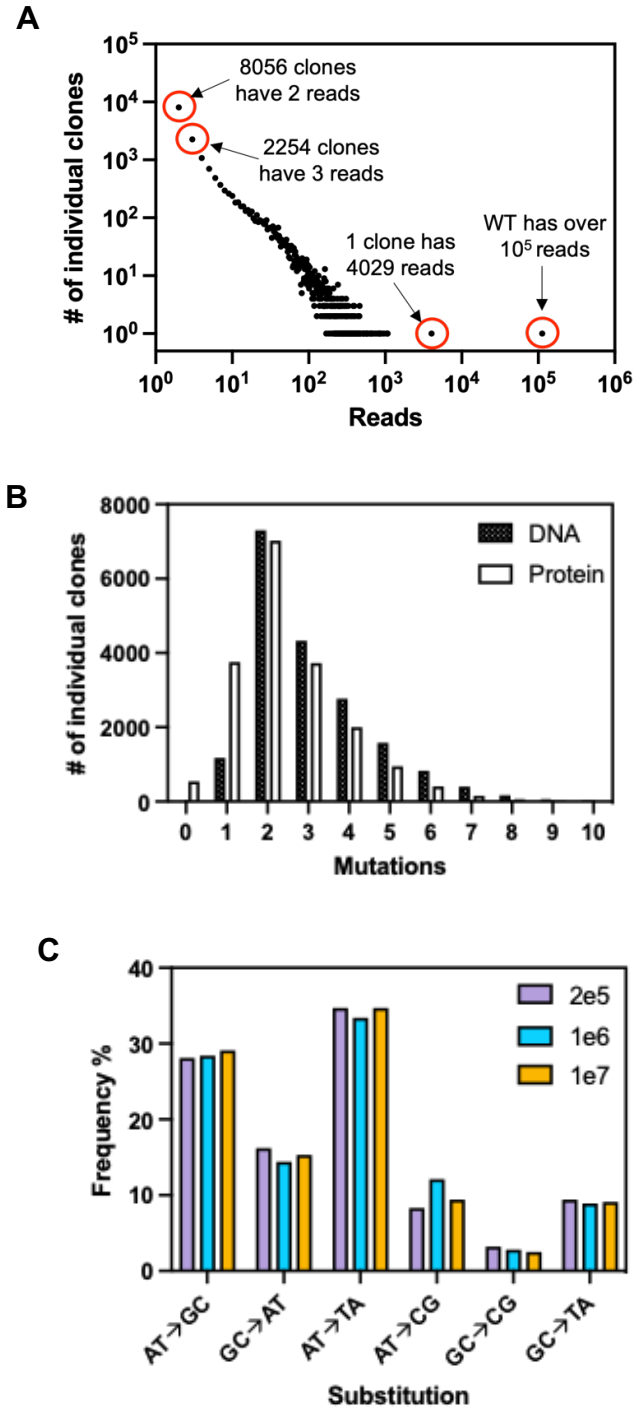


Figure 4.11 Fc Library profiling in mammalian cell by deep sequencing. **(A)** Coverage distribution of the 18678 unique clones in the library from 2×10^5 transfected cells. **(B)** Mutation rate analysis at both DNA and AA levels of the 18678 unique clones in the library from 2×10^5 transfected cells. **(C)** Mutation profiling.

The mutation rate analysis was performed at both DNA and amino acid levels without considering the abundance. It is obvious that the maximum mutation number was 10 since we removed the clones with over 10 mismatches through the alignment. Majority of the clones had 2-4 mutations at DNA level (3.1 bp in average) and 1-3 mutations at AA level (2.5 residues in average) (**Fig. 4.11B**), marginally higher than those from Sanger sequencing (**Fig. 4.2B**). The predominant mutations generated by Taq polymerase were still AT → GC transitions (28%) and AT → TA transversions (35%) (**Fig. 4.11C**), consistent with the results from Sanger sequencing (**Fig. 4.2C**). Furthermore, we found that the 11% of the unique clones contained stop codons, which agrees well with the result from Sanger sequencing (12%). Due to the silent mutations, the unique protein clone was decreased by around 30% to 13647.

We performed similar analysis on the libraries from 1×10^6 and 1×10^7 transfected cells, of which the total merged reads were 3,462,010 and 7,461,882, respectively. As the criteria for the clone authenticity became more stringent, the unique clone number remained at the level of 1×10^5 and 3×10^5 (**Table 4.1**). The library profiling barely showed difference (**Fig. 4.11C**) which are not repeated herein.

4.3.7 Diversity estimation

We define the number of the unique clones on DNA level as diversity. Therefore, the diversities of the Fc libraries from 2×10^5 and 1×10^6 transfected cells were reckoned as 18,678 and 112,406, corresponding to 9.3% and 11% aRMCE efficiency. In contrast, the aRMCE efficiency in the 1×10^7 transfected cells was derived as 3% based on the 304,476

unique clones. With such discrepancy, we further applied capture-recapture analysis on the 3 batches of sequencing results, which has been used in ecology to estimate population sizes (**Table 4.2**). Each batch of deep sequencing was assumed as a capture and the other as a recapture. Without the clone abundancy distribution considered, it yielded an estimate of 374,368 – 579,576, corresponding to a 3.7% - 5.8% RMCE efficiency (**Table 4.3**). We believe the true diversity of the library from 1×10^7 transfected cells was more underestimated than others. Assuming a coherent aRMCE efficiency, the clone number could be potentially 1×10^6 whereas the raw reads only yielded 7.4×10^6. With merely 7 times coverage and a naturally uneven clone abundancy, there could be a noticeable number of clones omitted. Based on the estimated RMCE in the lower bound, the Fc library constructed from 3×10^8 transfected cells could have 10^7 clones.

Table 4.2 Lincoln-Petersen capture-recapture model

	Capture	**Recapture**
Clone #	m	n
Overlap clone #	k	
Diversity	$\frac{m\times n}{k}$	

Table 4.3 Diversity estimations of the library derived from 1×10^7 transfected cells

	Batch 1	Batch 2	Batch 2	Batch 3	Batch 1	Batch 3
Confirmed unique clones	18,919	126,999	126,999	236,346	18,919	236,346
Overlapped clones	6,418		78,623		7,715	
Diversity estimation	**374,368**		**381,767**		**579,576**	
RMCE efficiency	**3.7%**		**3.8%**		**5.8%**	

4.4 Discussion and Conclusions

Here we report a deep sequencing workflow to validate the diversities of libraries constructed via aRMCE in CHO cells of which the numbers spanned over 3 orders of magnitude. We made sure the authenticity of the library with cautious sample preparation, including utilizing the extracted total RNA as the template and one primer only complementary to the mRNA, etc. We avoided overestimation of the library diversity with strict criteria such as restricting the numeration to the unique DNA clones with the same length (424 nt) as WT, at least 2 repeats, and up to 10 amino acid mismatches to WT after translation. In addition, we examined the DNA clones of the length 423 nt bearing one deletion which tends to occur in the consecutively repetitive nucleotide regions. With correction of such deletions, we found the unique clone number

with more than 1 read only increased by 0.6%. Since the contribution is trivial, we did not include such clones into the pool for further analysis.

The estimated aRMCE efficiency, approximately 10%, is reproducible, as indicated by comparable results between the libraries from 2×10^5 and 1×10^6 transfected cells with the similar and sufficient sequencing read coverage. The aRMCE efficiency in 1×10^7 transfected cells were estimated to be 3.7% - 5.8% with a simple model in capture-recapture analysis without considering the clone abundancy distribution. In light of these results, we reckon that the Fc library constructed from 3× 108 transfected cells has a diversity of 10^7 in a conservatively lower bound. This is so far the first reported large mammalian cell library with single copy per cell using RMCE-based system.

The deep sequencing provided a consistent library profiling compared to Sanger sequencing. Results showed that the error-prone PCR generated library has a few limitations. For instance, significantly uneven abundance distribution of different clones was observed with WT occurring over 105 times and 8056 clones only having 2 reads. In addition, 11% - 12% of the unique clones contained stop codons and unique protein diversity accounting for only 70%. On one hand, this suggests that the aRMCE-based mammalian cell library could potentially have a larger size when the donor library is fully diverse with evenness. On the other hand, the error-prone PCR may be resorted to only when the structure or function information is missing.

4.5 References

1. Wheeler, D. A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–876 (2008).

2. Levy, S. & Strausberg, R. L. Individual genomes diversify. *Nature* 456, 49–51 (2008).

3. Cloonan, N. *et al.* Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5, 613–619 (2008).

4. Eid, J. *et al.* Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 323, 133–138 (2009).

5. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6, 377–382 (2009).

6. Rothberg, J. M. & Leamon, J. H. The development and impact of 454 sequencing. *Nat Biotechnol* 26, 1117–1124 (2008).

7. Metzker, M. L. Sequencing technologies — the next generation. *Nat Rev Genet* 11, 31–46 (2010).

8. Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* 32, 158–168 (2014).

9. Wrenbeck, E. E., Faber, M. S. & Whitehead, T. A. Deep sequencing methods for protein engineering and design. *Curr Opin Struc Biol* 45, 36–44 (2017).

10. Weinstein, J. A., Jiang, N., White, R. A., Fisher, D. S. & Quake, S. R. High-Throughput Sequencing of the Zebrafish Antibody Repertoire. *Science* 324, 807–810 (2009).

11. Reddy, S. T. *et al.* Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat Biotechnol* 28, 965–969 (2010).

12. DeKosky, B. J. *et al.* High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* 31, 166–169 (2013).

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

14. Zhang, S.-Q. *et al.* High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nat Biotechnol* 36, 1156–1159 (2018).

15. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *Bmc Bioinformatics* 13, 31 (2012).

16. Cormack, R. M. & Seber, G. A. F. The Estimation of Animal Abundance and Related Parameters. *Biometrics* 40, 570 (1984).

17. Ramsland, P. A. *et al.* Structural basis for Fc gammaRIIa recognition of human IgG and formation of inflammatory signaling complexes. *J Immunol Baltim Md 1950* 187, 3208–17 (2011).

18. Mimoto, F. *et al.* Engineered antibody Fc variant with selectively enhanced FcγRIIb binding over both FcγRIIaR131 and FcγRIIaH131. *Protein Engineering, Design and Selection* 26, 589–98 (2013).

19. Ferrara, C. *et al.* Unique carbohydrate–carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose. *Proc National Acad Sci* 108, 12669–12674 (2011).

20. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *Genome Res* 2, 28–33 (1992).

21. Vanhercke, T., Ampe, C., Tirry, L. & Denolf, P. Reducing mutational bias in random protein libraries. *Anal Biochem* 339, 9–14 (2005).

22. Wilson, D. S. & Keefe, A. D. Random mutagenesis by PCR. *Curr Protoc Mol Biology Ed Frederick M Ausubel Et Al* Chapter 8, Unit8.3 (2001).

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

Chapter 5: Isolation of Fc Variants with Improved FcγR Binding from Mammalian Combinatorial Libraries

Abstract

Optimized Fc has proven benefits for therapeutic antibodies on treatments of cancer and infectious diseases. The current approaches of site-directed mutagenesis and computational design are limited by low throughput while the high-throughput phage, bacteria and yeast display technologies are suboptimal for Fc engineering because the N-glycosylation at site 297 is essential for its biological functions via binding to FcγRs. Here, we apply aRMCE mammalian cell library platform developed in **Chapter 3** for Fc engineering. In **Chapter 4**, we have constructed and validated, by deep sequencing, a monoclonal library of mammalian cells displaying full-length IgGs encompassing $>10^7$ Fc mutation clones. In this Chapter, we screened the large Fc library via MACS and FACS for optimized binding towards FcγRIIIa^{F176}, FcγRIIa and FcγRIIb, respectively. Following characterizations by flow cytometry, ELISA, biolayer interferometry and surface plasmon resonance, numerous unique and novel Fc variants with improved affinity and selectivity have been identified. Particularly, the isolated Fc variant 3a-7 (218I/231V/K334E) showed over 5-fold binding affinity improvement compared to wild type (WT) Fc towards FcγRIIIa^{F176}; 2a-10 (G236E/K288R/K290W/K320M) exhibits enhanced binding towards FcγRIIa^{R131} whereas decreased binding to FcγRIIb; 2b-1 (K222I/V302E/L328F/K334E) were identified with more binding affinity improvement towards FcγRIIb than FcγRIIa^{R131} and decreased binding to FcγRIIIa^{F176}. Using Fc

engineering as a demonstration, we expect that the aRMCE mammalian cell library approach developed in my thesis will have a broad applicability in biologics discovery and engineering.

5.1 Introduction

Monoclonal antibodies (mAbs) have proven to be remarkably versatile therapeutics for the treatment of autoimmune, infectious and malignant diseases[1–4]. This is not only due to their Fab portion in binding to specific antigens but also their Fc portion in regulating immune responses through interacting with Fc gamma receptors (FcγRs) expressed on the surface of immune effector cells. The low affinity activating receptors FcγRIIIa and FcγRIIa are the main contributors to antibody-dependent cell-mediated cytotoxicity (ADCC) and antibody-dependent cell-mediated phagocytosis (ADCP)[5,6], respectively. It has been well established that antibodies with increased binding of Fc to FcγRIIIa or FcγRIIa have enhanced ADCC or ADCP, leading to better efficacy in cancer and antiviral treatments[7–10]. Notably, FcγRIIIa has two allotypes with single amino acid mutations at position 176, i.e., V176 and F176. Clinical evidence has shown that rituximab exhibits better performance in lymphoma patients of FcγRIIIa^{V176} than those of FcγRIIIa^{F176} which correlates with a stronger binding affinity of V176 allotype to IgG1 Fc[11–14]. Similarly, it has been found that FcγRIIa^{R131} is associated with greater susceptibility to infectious disease due to its lower affinity to IgG1 Fc compared to FcγRIIa^{H131} [15–17]. For the inhibitory receptor FcγRIIb, its affinity with Fc is crucial for the

antitumor activity of agonistic mAbs[18,19]. Motivated by these findings, numerous studies have explored different approaches to enhance the interaction of Fc to FcγRs.

With computational designs, the investigators at Xencor identified Fc mutant DEL (S239D/I332E/A330L) which showed 58-fold affinity enhancement toward FcγRIIIa^{F176} compared to trastuzumab WT[20]. By screening over 900 variants, Fc clones GA (G236A) and GAALIE (G236A/I332E/A330L) with up to 6-fold improvement for binding to FcγRIIa^{R131} were isolated [17,21]. Similarly, by comprehensive mutagenesis design, production and screening, clone PD (P238D) and V12 (E233D/G237D/P238D/H268D/P271G/A330R) with selectively enhanced FcRγIIb binding were identified out of 500 Fc variants[22]. Despite of great success, these studies all are limited to a small number of mutagenesis positions which could omit important mutations. To address this issue, yeast surface display was utilized for screening from a random library of ~ 1×10^7 Fc mutants[23] and identified LPL (F243L/R293P/Y300L) with 4.6-fold lower k_{off} and F243L/R292P/Y300L/V305I/P396L with over 10-fold lower K_D to FcγRIIIa^{F176}. In addition, combinatorial libraries of aglycosylated Fc have been constructed and periplasmically displayed in *E. coli*, resulting in the isolation of the Fc variant E382V/M428I binding to FcγRI with nanomolar affinity[24].

One problem associated with microbial display for Fc engineering is the different glycosylation pattern from mammalian cells, given that the N-glycosylation at the site 298 on the Fc region is essential for binding to FcγRs. In contrast, mammalian cell display could be an ideal platform to screen Fc variants with correct folding and post

translational modifications. Recently, lentivirus-based mammalian cell display has been used to isolate Fc variants with improved binding affinity towards FcγRIIIa^{F176} and FcγRIIb. However, this method requires low multiplicity of infection (MOI) to achieve single copy per cell, which compromised the library size to 7× 10^5. In **Chapter 4** of this thesis, we have demonstrated the aRMCE-enabled construction of large Fc library, encompassing over 10^7 diversities with full length IgG displayed on the surface at single copy per cell level. In this Chapter, we successfully isolated Fc variants from this large library with improved affinity and selectivity towards FcγRIIIa^{F176}, FcγRIIIa^{R131} and FcγRIIb, respectively.

5.2 Materials and Methods

5.2.1 Plasmid construction

Genes encoding the extracellular domains of human Fcγ receptors (FcγRs), including FcγRIIIa^{F176}, FcγRIIa^{R131} and FcγRIIb were chemically synthesized (IDT) while FcγRIIIa^{V176} and FcγRIIa^{H131} were generated by site-directed mutagenesis using overlapping primers as well as FcγRIIIa^{F176} and FcγRIIa^{R131} as templates, respectively. The gene encoding FcγRIIIa^{F176} was fused with human azurocidin preproprotein (MTRLTVLALLAGLLASSRA)[25] as the signal peptide at N-terminus and an NsiI cutting site and a 6 × His tag was introduced to the C-terminus by extension PCR, of which the whole gene fragment was cloned to pCIW-3f8LC-IRES-Zeo with HindIII and EcoRI, giving pCIW-FcRIIIaF-His-IRES-Zeo. The gene fragment including the elements

$(GGGGS)_2$ linker, AviTag™ [26] and a 6 × His tag was generated by oligos annealing with NsiI and EcoRI overhungs at 5' and 3' prime ends, respectively, and cloned to pCIW-FcRIIIaF-IRES-Zeo to give FcγRIIIa^{F176} production plasmid pCIW-FcRIIIaF-Avitag-His-IRES-Zeo. Other FcγRs expression plasmids were generated by replacing the FcγR genes with EcoRV and NsiI.

The gene of *E. coli* biotin ligase (BirA) was amplified from pET21a-BirA (Addgene #20857), fused with an endoplasmic reticulum (ER) retention signal DYKDEL[27–29] (DYKD is the truncated form of the FLAG epitope tag) at its C-terminus[30,31] by extension PCR, and cloned to pCIW-3f8LC-IRES-Zeo via XhoI and HpaI. The zeocin resistance gene was replaced with hygromycin resistance gene by overlapping PCR, giving pCIW-BirA-IRES-HygR.

C_H2 fragments encoding Fc mutants, including LPL(F243L/R293P/Y300L), DEL (S239D/I332E/A330L), GA (G236A), GAALIE (G236A/I332E/A330L), PD (P238D), V12 (E233D/G237D/P238D/H268D/ P271G/A330R), were generated by site-directed mutagenesis using overlapping primers, then cloned to pEx-3f8HC with BstXI and BsrGI to give associated heavy chain exchange plasmids.

For cloning isolated Fc variants, the genomic DNA was extracted from the collected cell populations by using Wizard Genomic DNA Purification Kit (Promega, Cat#A1120). The heavy chain genes were then amplified by PCR with 1 μg of the extracted genomic DNA as the template in a 100-μl total reaction by using Q5-High Fidelity DNA polymerase (NEB). The thermal cycling program was 98 °C for 30 s; 30 cycles of 98 °C for 10 s, 63 °C for 20 s, 72 °C for 40 s; final extension at 72 °C for 2 min and storage at

4 °C. The PCR products were gel purified, digested with BstXI and BsrGI, and cloned into the exchange plasmid pEx-3f8HC.

Genes encoding trastuzumab (Herceptin; Roche/Genentech) heavy chain and light chain were chemically synthesized (IDT). The light chain was cloned to pCIW-BirA-IRES-Hy with XhoI and HpaI, giving pCIW-TrastLC-IRES-Hy. The heavy chain was fused with chimeric intron at the N-terminus with overlapping PCR and cloned into pCIW-3f8LC-IRES-Zeo with NheI and PmlI, yielding pCIW-TrastHC-IRES-Zeo. The heavy chain was fused with $(GGGGS)_2$ linker followed by Avitag with overlapping PCR and cloned to pCIW-TrastHC-IRES-Zeo via BsrGI and KpnI, giving pCIW-TrastHC-Avitag-IRES-Zeo. The Fc variants were sub-cloned from pEx-3f8HC to the WT trastuzumab HC or HC-Avitag expression plasmids via BstXI and BsrGI.

5.2.2 Transfection

a. Transfection for protein production

Expi293F cells were seeded at 2×10^6/ml in Expi293 media for transfection which was performed with PEI and total plasmids in 3.5:1 (w/w) ratio. For production of biotinylated FcγRs-Avitag, cells were co-transfected with the corresponding pCIW-FcR-Avitag-His-IRES-Zeo and pCIW-BirA-IRES-Hy (1:1, w/w); for production of FcγRs, cells were transfected with corresponding pCIW-FcR-His-IRES-Zeo.

For production of soluble IgG, Expi293F cell stably expressing the light chain was generated first. Specifically, Expi293F was transfected with pCIW-TrastLC-IRES-Hy. One day after the transfection, the cells were cultured in fresh Expi293 media

supplemented with 350 μg/ml hygromycin for 7 days, then maintained with 50 μg/mL hygromycin, giving Expi293-TrastLC cells. For production of trastuzumab or its Fc variants, the cells were transfected with the corresponding pCIW-TrastHC-IRES-Zeo; for production of biotinylated trastuzumab or its Fc variants, the cells were co-transfected with the corresponding pCIW-TrastHC-Avitag-IRES-Zeo and pCIW-BirA-IRES-Hy (1:1, w/w).

b. RMCE transfection for IgG display on CHO cell surface

CHO-Puro-3f8LC cells were transfected with the exchange plasmids pEx-3f8HC carrying Fc WT or Fc variants as described in **Chapter 4**, section **4.2.3**.

5.2.3 Protein expression and purification

Five days post transfection, the cell culture was centrifuged at 1000 × g for 10 min, then the supernatant was subjected to a second time centrifugation at 10000 × g for 20 min. The supernatant was filtered by 0.45 μm MCE membrane (MilliporeSigma, Cat#HAWP04700) prior to loading to the column. FcγRs were purified by nickel affinity chromatography with His-Pure Ni-NTA resin (Thermo Scientific, Cat#88222), IgGs were purified by protein A resin (GenScript, Cat#L00210).

5.2.4 MACS enrichment

On day 2 post transfection, cells were harvested and subjected to MACS isolation. Specifically, total cells were washed with 0.5% PBSA twice, incubated with a mixture of biotinylated FcγRs-Avitag for 30 min at 4 °C, including 100 nM FcγRIIIa^{F176}, 50 nM FcγRIIIa^{V176}, 50 nM FcγRIIa^{H131}, 50 nM FcγRIIa^{R131} and 200 nM FcγRIIb. Subsequently,

the cells were washed twice with 0.5% PBSA, resuspended in it with Streptavidin Microbeads (Miltenyi), and incubated at 4 °C for 15 min with occasional inversion. Following incubation, cells were washed with 0.5% PBSA once and then resuspend in 3 ml 0.5% PBSA. The resuspended cells were loaded to a pre-rinsed LS magnetic column attached to a QuadroMACS separator, washed three times with 3 ml 0.5% PBSA. After flowing through, the column was removed from the magnetic separator and the cells were flushed directly into a collection tube with 5 ml PBSA. Total cell number was counted by hemocytometer as ~ 2×10^6. Finally, the cell elute was centrifuged to remove the excessive magnetic beads to improve the cell viability, and the cell pellet was resuspended in 10 ml of F12K complete media cultured in a 10 cm dish. When MACS-isolated cells were expanded to the number of 2×10^8, 9/10 of the cells were saved as stocks of 2×10^7 cells per cryogenic vial for future use and the remaining cells were subjected to further expansion for FACS selection towards improved affinity to each type of FcγRs.

5.2.5 Flow cytometry analysis and FACS

a. Stable cell line with Fc WT or variants reported on literatures displayed

On day 2 post transfection, cells were stained with 1 μg/ml mouse anti-human IgG (Fab)-PE (Invitrogen, Cat#MA1-10377) and the top 2% PE^+ cells were sorted on single mode by a S3e cell sorter (Bio-Rad).

b. Dual-color flow cytometry analysis and FACS

Cells were collected and resuspended in 0.5% PBSA at a concentration of 1×10^6/100 µl and incubated with 60 nM of the antigen EGFP-TNFα and the corresponding biotinylated FcγRs at 4 °C for 30 min. Subsequently, the cells were washed twice with 0.5% PBSA, resuspended in it with 1 µg/ml streptavidin-APC (Invitrogen Cat# 17-4317-82), and incubated at 4 °C for 15 min. Finally, the cells were washed twice with 0.5% PBSA and ready for flow cytometry analysis. EGFP and APC were excited by 488 nm and 640 nm laser, respectively. EGFP was detected with 526/48 nm filter on FL1 channel, APC 670/30 nm filter on FL3 channel. For affinity comparison between different Fc variants, the concentrations of the biotinylated FcγRs-Avitag were listed: 40 nM FcγRIIIaF176; 45 nM FcγRIIaH131 and 55 nM FcγRIIaR131; 80 nM FcγRIIb. For FACS selection, 1st round: 80 nM FcγRIIIaF176; 90 nM FcγRIIaH131, 110 nM FcγRIIaR131; 160 nM FcγRIIb; 2nd round: 50 nM FcγRIIIaF176; 60 nM FcγRIIaH131, 70 nM FcγRIIaR131; 130 nM FcγRIIb. The cells with a higher signal ratio of APC to EGFP represented a higher affinity towards the FcγR under same display level, which were sorted with enrich mode for the 1st round and purity mode for the 2nd round of FACS. Among the EGFP-TNFα positive cells, the median fluorescence intensity (MFI) of EGFP-TNFα (MFITNFα) and FcγRs (MFIFcγR) were obtained in the software Flowjo 10.4. The MFIs of FcγRs were normalized to the IgG expression with the equation $\frac{\mathrm{MFIFc\gamma R}}{\mathrm{MFTNF\alpha}} \times 100$.

5.2.6 ELISA

Maxisorp 96-well plates (Thermo Scientific, Cat#439454) were coated with 100 µl of 5 µg/ml streptavidin (NEB) at 4 °C overnight and then blocked with 2% BSA in PBS at room temperature (RT) for 2 h. 100 µl of 2 µg/ml biotinylated IgGs were added for incubation at RT for 45 min, and plates were washed 4 times with PBS supplemented with 0.05% Tween-20 (PBST) and twice with PBS. FcγRIIIa^{F176}, FcγRIIa^{H131} and FcγRIIa^{R131} starting at 3 µM and FcγRIIb starting at 6 µM were 1:3 serially diluted with 2% BSA in PBS and added by 100 µl to each well. After 1 h incubation at RT and washing as described above, 100 µl anti-His-HRP at 1:10,000 dilution (Abcam, Cat#ab1187) was added. Plates were incubated at RT for 1 h and washed. 50 µl TMB (Thermo Scientific, Cat# 34028) was added each well to develop the signals. Reaction was terminated by adding 50 µl 2M H_2SO_4, and absorbance at 450 nm was measured with an Epoch microplate reader (BioTek). ELISA experiments were performed in duplicates.

5.2.7 Bio-layer interferometry (BLI)

Binding kinetics were measured by bio-layer interferometry using BLItz on streptavidin biosensors (FortéBio, Cat# 18-5019) with PBS as the buffer. 5 µg/ml of biotinylated FcRs were loaded on sensors for 2 min. Association was monitored for 1 min with purified IgGs at three decreasing concentrations, and dissociation was monitored with PBS for 2 min. IgG concentrations used to test the binding to both FcγRIIIa allotypes are 1000 nM, 500 nM, 250 nM; to both FcγRIIa allotypes, 500 nM, 250 nM,

125 nM; to FcRIIb, 4000 nM, 2000 nM, 1000 nM except 3000 nM, 1500 nM, 750 nM for Fc variant 3a-2, PD, 2b-5, 2b-7, 2b-10, 2400 nM, 1200 nM, 600 nM for LPL and 2000 nM, 1000 nM, 500 nM for 2b-1. Kinetic analysis and curve fitting were performed with the software BLItz Pro. The sensorgrams were obtained with GraphPad Prism 9.

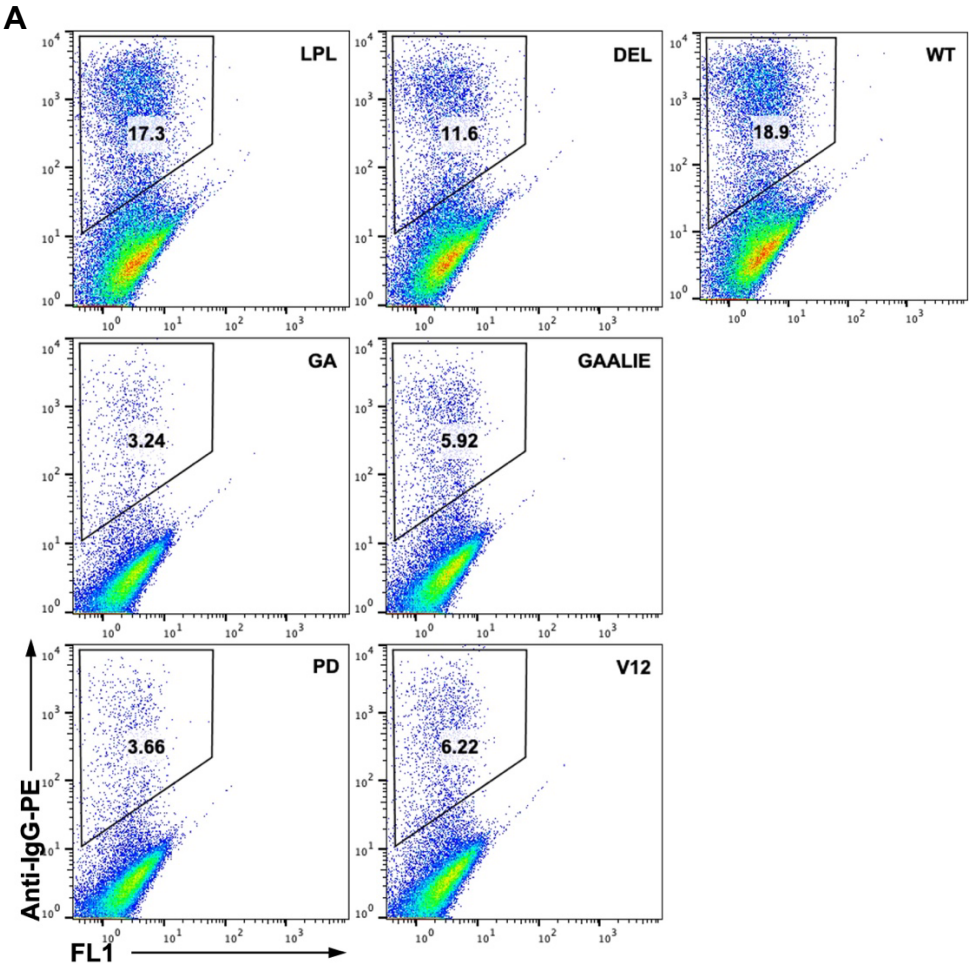
5.2.8 Surface plasmon resonance (SPR)

SPR experiments were performed using Biacore T200 (GE Healthcare) in PBS at 25 °C. Biotin CAPture Reagent (Cytiva, Cat#28920234), modified with streptavidin, was immobilized on a Series S Sensor Chip CAP (Cytiva, Cat#28920234) through deoxyribooligonucleotide interaction. 500 nM biotinylated FcγRIIIa^{F176} and FcγRIIa^{R131} were then captured in different flow cells. The purified trastuzumab IgGs carrying Fc WT and Fc variants were injected at a flow rate of 30 μL/min for 2 min over the FcγR-captured surfaces for association coefficient measurement. IgG concentrations used to test the binding to FcγRIIIa^{F176} are 1000 nM, 500 nM, 250 nM; to FcγRIIa^{R131} allotypes, 400 nM, 200 nM, 100 nM. In dissociation phase, PBS was injected at 30 μL/min for 10 min. Between each assay cycle, the sensor chip surface was regenerated with 8 M guanidine-HCl and 1 M NaOH at 3:1 volume ratio for 30 s at a flow rate of 30 μL/min. K_D values were calculated using the 1:1 binding kinetics model of the Biacore T200 Evaluation software. The sensorgrams were obtained with GraphPad Prism 9.

5.3 Results

5.3.1 Construction of stable cell lines of anti-TNFα IgGs carrying Fc WT and variants

To determine whether this system could discriminate between Fc variants with different affinities towards FcRs and facilitate the downstream screening of Fc variants by FACS, we chose a few Fc variants reported previously and displayed them on stable cell lines via aRMCE (**Fig. 5.1A**). Compared to WT, LPL was reported to have a 4.6-fold lower k_{off} towards FcRIIIa^{F176}. Here it showed a rightward tilt than WT when stained with FcRIIIa^{F176} (**Fig. 5.1B**). DEL identified by rational design had a 58-fold FcRIIIa^{F176} affinity enhancement than WT in the context of trastuzumab by AlphaScreen analysis[20]. Displayed on the CHO-Puro 3f8LC cell surface, it showed a significant rightward tilt than LPL and WT by flow cytometry analysis (**Fig. 5.1B**). GA and GAALIE were reported to have a K_D value of 488 nM and 819 nM[21], respectively. Despite of the small difference, they were still able to be differentiated by flow cytometry with GA having a more rightward tilt than GAALIE. V12 was engineered based on PD, resulting in 62-fold enhanced binding to FcRγIIb[22], of which the degree agrees well with what we observed through flow cytometry analysis (**Fig. 5.1B**). These results provided sufficient confidence to proceed to Fc engineering using the constructed mammalian cell Fc library. Moreover, the stable cell lines with WT and these Fc variants displayed will be used as positive controls for FACS gating strategies to enrich positive binders from the pool.


A
LPL
17.3
DEL
11.6
WT
18.9
GA
3.24
GAALIE
5.92
PD
3.66
V12
6.22
Anti-IgG-PE
FL1

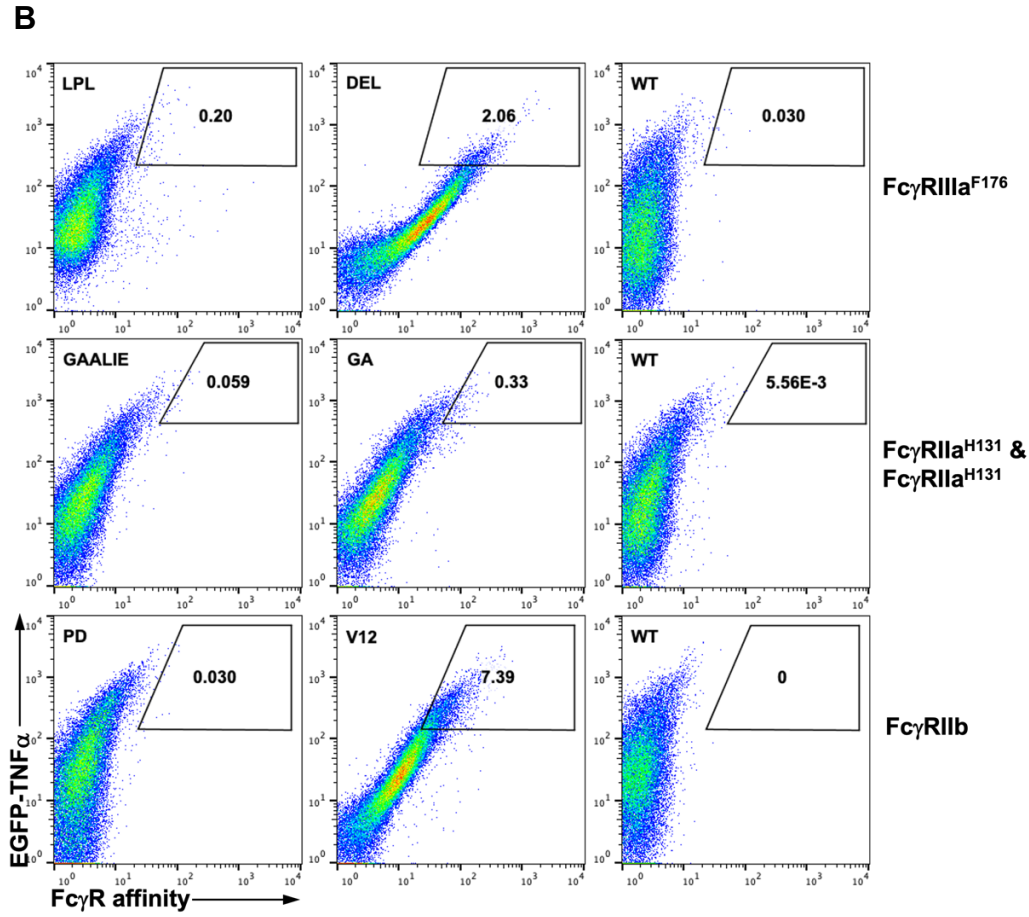


Figure 5.1 Fine affinity discrimination between Fc WT and reported Fc variants. Fc variants included LPL and DEL for FcγRIIIa, GA and GAALIE for FcγRIIa, and PD and V12 for FcγRIIb. **(A)** Generation of cell lines by transfection of CHO-Puro-3f8LC cells with 3f8 HC exchange plasmids via aRMCE. Cells were stained with anti-IgG-PE, and the top 2% PE^+ cells were sorted. **(B)** Characterization of obtained 3f8 IgG stable cell lines. Cells were stained with EGFP-TNFα for IgG expression, and biotinylated corresponding FcγRs and streptavidin-APC for FcγR binding.

5.3.2 MACS and FACS enriched high affinity Fc variants

The Fc library of 10^7 clones constructed in **Chapter 4** was enriched by one round of MACS using a mixture of biotinylated FcγRs-Avitag and streptavidin conjugated

magnetic beads. Additional flow cytometry analysis showed that 2.92% of the cells displayed IgG before MACS (**Fig 4.4A** in **Chapter 4**) and 0.44% of the cells were able to bind with the FcγRs at defined concentrations (data not shown). After MACS, cells were counted and found approximately 2×10^6 cells isolated, which agrees with the low percentage of FcγR+ cell population 0.44% before MACS. After MACS, isolated cells were expanded and stained with the antigen EGFP-TNFα to test the enrichment. Around 19 % of the cells were TNFα+ **(Fig. 5.2, post MACS, gate not shown)**, suggesting a successful MACS enrichment.

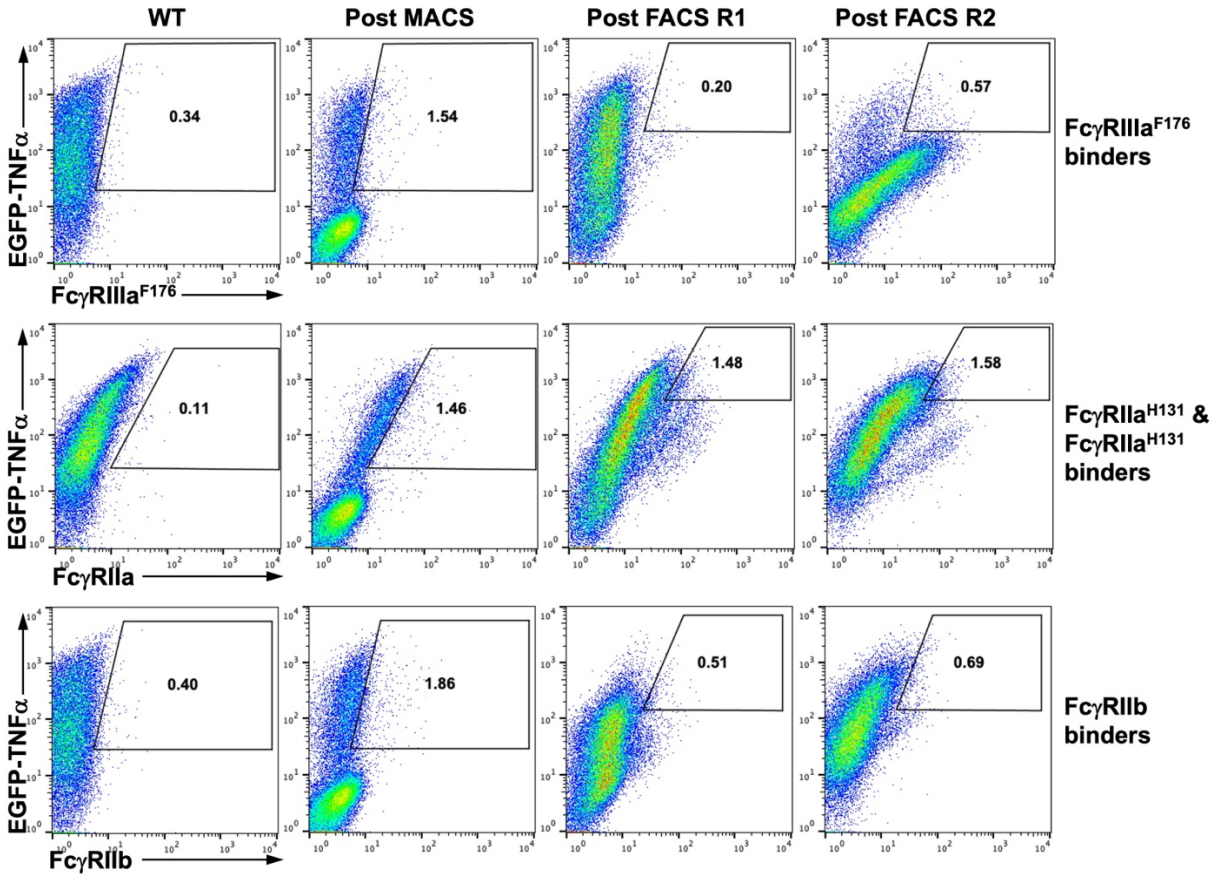


Figure 5.2 Enrichment of high affinity binders towards FcγRs by two successive rounds of FACS. Library cells were sorted towards FcγRIIIa^{F176}, FcγRIIa (both H131 and R131 allotypes) and FcγRIIb, respectively. Cells were stained with EGFP-TNFα for IgG expression and biotinylated corresponding FcγRs and streptavidin-APC for FcγR binding.

Following MACS, two rounds of FACS enrichment were performed for FcγRIIIa^{F176}, FcγRIIa^{R131}/IIa ^{H131} and FcγRIIb, respectively, using an equilibrium screening strategy with reduced concentrations in the 2nd round. As a result, all the three enriched Fc libraries after round 2 showed a more rightward tilt than before the selection by flow cytometry (**Fig. 5.2**), suggesting the majority population had the improved affinity towards the corresponding FcγRs.

5.3.3 Identification of isolated Fc variants

Genomic DNA was thus extracted from each of the enriched Fc libraries; the heavy chain genes were recovered by PCR amplification and cloned into the pEx-3f8-HC exchange vector. 48, 48 and 32 single colonies were randomly picked from the enriched Fc libraries toward FcγRIIIa^{F176}, FcγRIIa and FcγRIIb for Sanger sequencing, respectively. 24 to 34 unique Fc variants were identified from each of enriched Fc libraries, labeled as 3a, 2a, 2b mutants, respectively. These Fc variants harbor over 40 unique amino acid changes spanning the lower C_H1, hinge and C_H2 regions (**Fig. 5.3**, only C_H2 shown). Several identified sites (236, 266, 267, 305, 326, 328, 332, 334) or mutations (V266M, K326E, K334E) had also been reported previously by protein design algorithm or yeast surface display library. For example, the site 236 was identified for improved binding to FcγRIIa with the mutation as A, while we identified here as E. The site 305, 332 and the mutations K326E, K334E have been identified by rational design, yeast display or lentivirus-based mammalian cell display for FcγRIIIa. The sites 266, 267 and the mutations L328F/I were identified for FcγRIIb. Novel mutations and combinations were also found, which will be discussed in following sections.

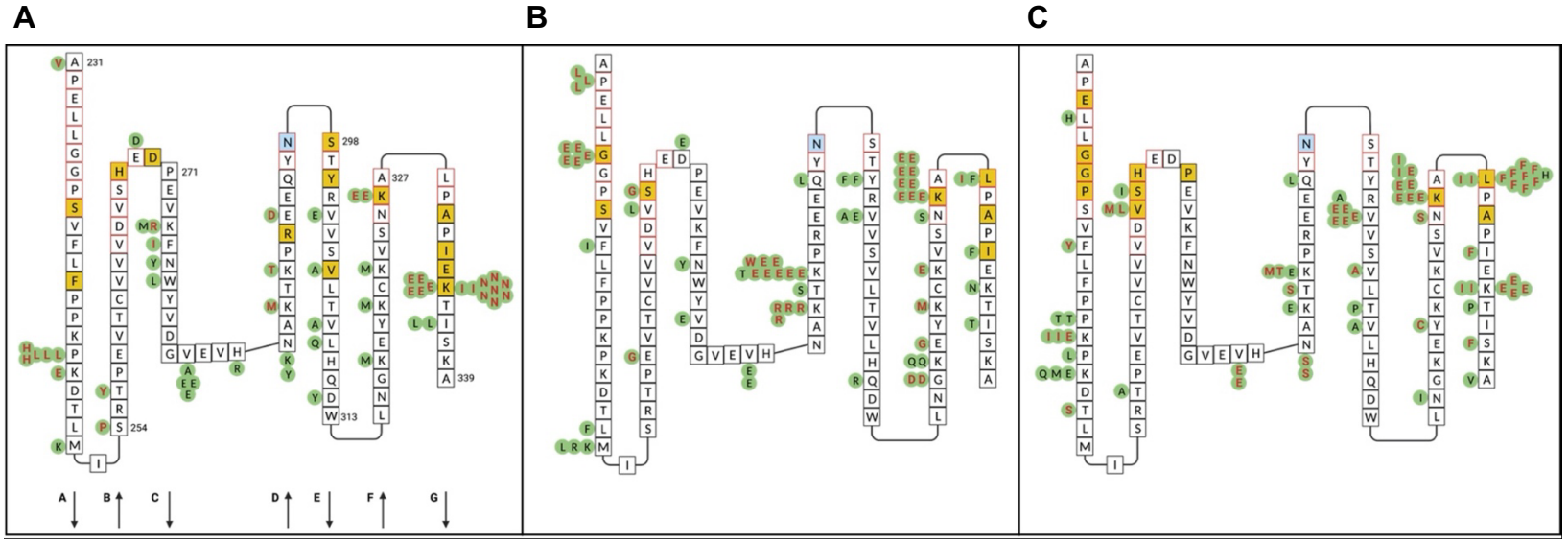
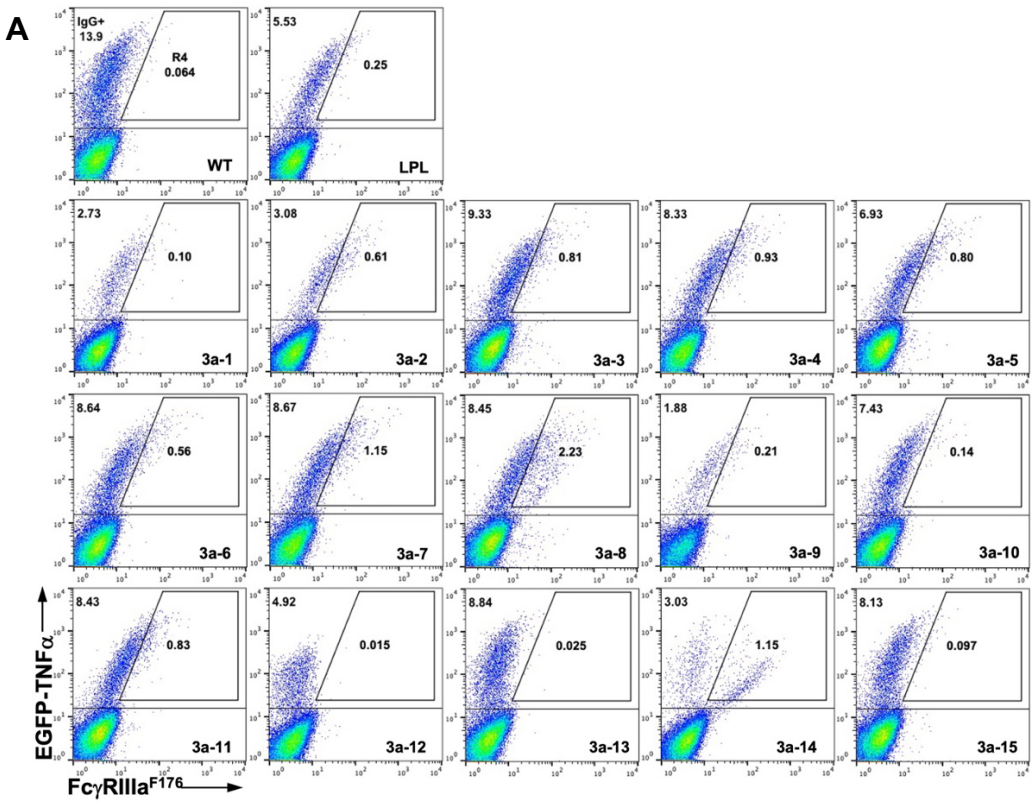


Figure 5.3 Two-dimensional Collier de Pearles representations of the IgG1 C_H2 domain showing identified mutations towards **(A)** FcγRIIIa^{F176}, **(B)** FcγRIIa (both H131 and R131 allotypes) and **(C)** FcγRIIb. Green circles represent all mutations identified in this study, among which the ones with improved affinity confirmed by ELISA are shown with red letters; yellow shades represent the beneficial mutation sites reported in literatures; red boxes represent the residues interact with FcγRs; blue shades represent the glycosylation site. Antiparallel β strands within the C_H2 domain fold are also shown.

5.3.4 Characterizations of Fc variants with improved FcγRIIIa affinity

To test whether the identified Fc variants have improved binding to FcγRIIIa^{F176}, we transfected the pEx plasmids carrying individual Fc variants back to CHO-Puro 3f8 LC and performed an initial screening by flow cytometry. We found that most of the Fc variants showed a rightward tilt same as the positive control, indicating an improved affinity towards FcγRIIIa^{F176} (**Fig. 5.4A**). Moreover, we quantitatively compared their affinities to WT based on the median fluorescence intensity (MFI) of FcγR normalized to the antibody surface expression (**Table 5.1**). The positive control LPL has an MFI ratio of 7.5, at least 2-fold of the WT. 13 of the selected Fc variants showed a higher value of the normalized MFI than WT and most of them were even higher than LPL. We also found that the Fc variants showed a slightly higher affinity toward FcγRIIb (**Fig. 5.4A, Table 5.1**), because the extracellular domain of FcγRIIIa^{F176} shares 50% homology with FcγRIIb.


A
IgG+
13.9
R4
0.064
WT
5.53
0.25
LPL
2.73
0.10
3a-1
3.08
0.61
3a-2
9.33
0.81
3a-3
8.33
0.93
3a-4
6.93
0.80
3a-5
8.64
0.56
3a-6
8.67
1.15
3a-7
8.45
2.23
3a-8
1.88
0.21
3a-9
7.43
0.14
3a-10
8.43
0.83
3a-11
4.92
0.015
3a-12
8.84
0.025
3a-13
3.03
1.15
3a-14
8.13
0.097
3a-15
EGFP-TNFα
FcγRIIIa^F176

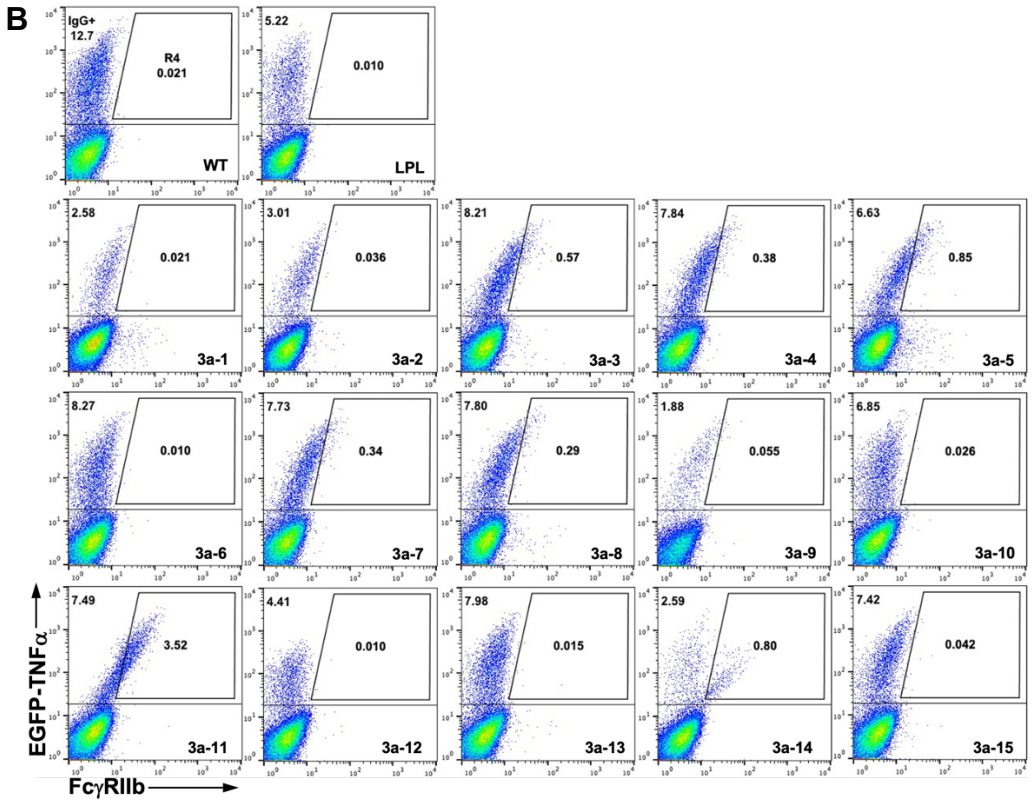


Figure 5.4 Flow cytometric analysis of isolated IIIa variants with **(A)** FcγRIIIa^{F176} and **(B)** FcγRIIb. Cells were stained with EGFP-TNFα for IgG expression and biotinylated corresponding FcγRs and streptavidin-APC for FcγR binding. WT and LPL clones were used as controls. Quantitative analysis of cell populations in R4 gates were shown in **Table 5.1**.

Table 5.1 Quantitative flow cytometry analysis of isolated FcγRIIIa^{F176} binders.

Clone	Mutations	FcγRIIIa^{F176} MFI[a] / TNFα MFI	FcγRIIb MFI / TNFα MFI
WT	-	3.4	2.6
LPL	F243L R292P Y300L	7.5	2.6
3a-1	P247L	8.0	5.2
3a-2	P247L K248E K334I	11.1	5.8
3a-3	P247H S254P	9.3	6.9
3a-4	P247L T256Y K290T	8.6	6.1
3a-5	P247H	9.6	7.1
3a-6	K334E	8.2	3.9
3a-7	K218I A231V K334E	8.9	6.1
3a-8	K334N	13.2	5.5
3a-9	F275I K288M K334I	9.2	5.0
3a-10	E293D K334E	6.9	3.5
3a-11	K274R K326E	9.8	12.9
3a-12	W277L	6.2	5.9
3a-13	V211D M252K H285R V305A L309Q K320M	3.9	4.0
3a-14	E269D K334E	16.6	10.8
3a-15	V282E	5.0	4.0

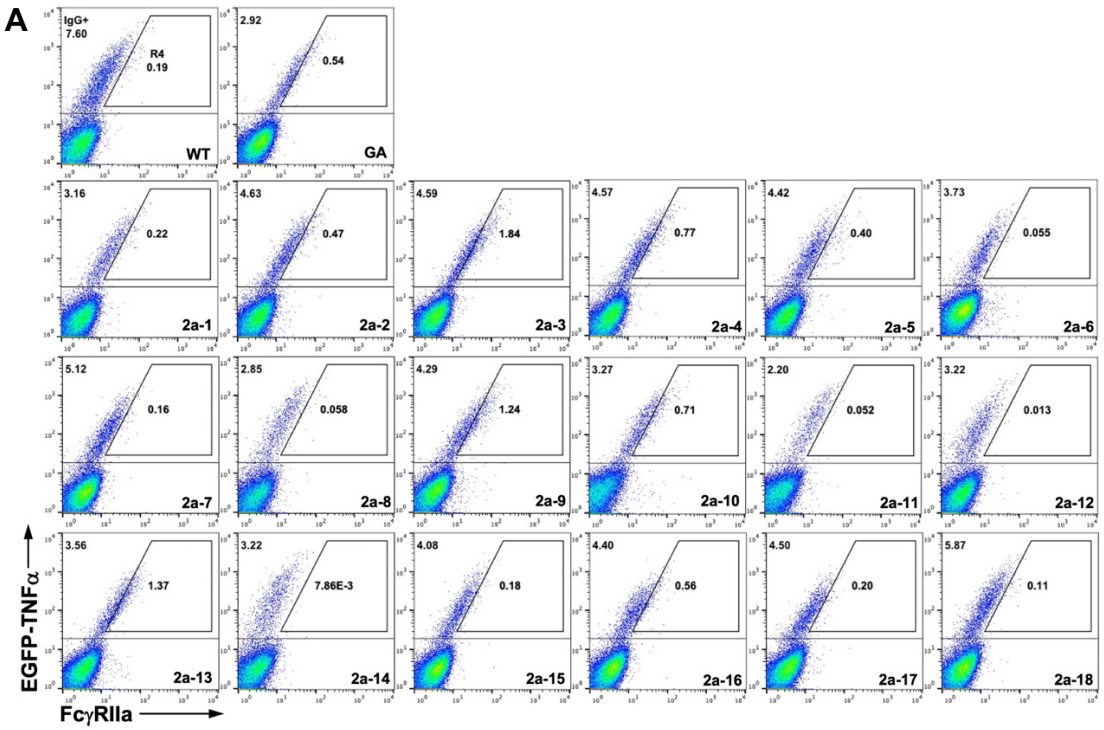
[a] MFI, median fluorescence intensity. Relative binding strength on FcγR was normalized with IgG expression, calculated as $\frac{MFIFc\gamma R}{MFITNF\alpha} \times 100$.

To examine the affinity of Fc variants to different FcγRs in soluble format, we conducted ELISA on the selected 11 Fc variants which were incorporated into trastuzumab (Herceptin; Roche/Genentech). All trastuzumab variants demonstrated enhanced affinity towards FcγRIIIa^{F176} (**Fig 5.7A**). We further characterized 7 trastuzumab variants with a higher FcγRIIIa^{F176}/FcγRIIb binding ratio by bio-layer interferometry (Blitz). All variants showed different levels of decrease on K_D towards both FcγRIIIa^{F176} and FcγRIIIa^{V176} (**Tabel 5.4**). 3a-1 (P247L) and 3a-5 (P247H) have the

different mutations at the same site, but they share a similar K_D value, 250 nM and 228 nM, respectively. 3a-2 (P247L/K248E/K334I) had 2 more mutations than 3a-1, which further decreased the K_D to 190 nM. 3a-7 (K218I/A231V/K334E) demonstrated the highest affinity improvement on FcγRIIIa^{F176} with over 4-fold decrease in k_{off} than WT (**Fig. 5.8A**). Additional SPR analysis was performed for WT, LPL, 3a-2 and 3a-9, of which the results were consistent with that of BLI (**Table 5.4**, **Fig. 5.9A**).

5.3.5 Characterizations of Fc variants with improved FcγRIIa affinity

We conducted similar analysis for the Fc variants from the enriched Fc library towards FcγRIIa. 16 of 18 selected Fc variants showed improved affinity towards FcγRIIa by flow cytometry (**Fig. 5.5A**). 11 Fc variants were incorporated into trastuzumab and evaluated by ELISA (**Fig. 5.7B**). 2a-8 showed no improvement towards either of the FcγRIIa consistent with the results from flow cytometry whereas 2a-7 showed opposite result. One mutant 2a-3 surprisingly showed more improved affinity towards FcγRIIIa^{F176} than FcγRIIa. 2a-9, 10, and 11 all carrying one mutation G236E showed good specificity toward FcγRIIa with reduced affinity towards FcγRIIIa and none to slightly increased affinity towards FcγRIIb. We further examined 2a-9, 10 and 11 by BLI (**Table 5.4**, **Fig. 5.8B**). 2a-9 and 2a-11 showed an affinity improvement on FcγRIIa^{R131} and reduced affinity towards H131 allotype. 2a-10 exhibited the most improved affinity toward FcγRIIa with a K_D value of 141 nM and reduced affinity towards FcγRIIb. In addition, the K_D of 2a-10 by SPR analysis was 150 nM, no significant difference from the one measured by BLI (**Table 5.4**, **Fig. 5.9B**).


A
IgG+
7.60
R4
0.19
WT
2.92
0.54
GA
3.16
0.22
2a-1
4.63
0.47
2a-2
4.59
1.84
2a-3
4.57
0.77
2a-4
4.42
0.40
2a-5
3.73
0.055
2a-6
5.12
0.16
2a-7
2.85
0.058
2a-8
4.29
1.24
2a-9
3.27
0.71
2a-10
2.20
0.052
2a-11
3.22
0.013
2a-12
3.56
1.37
2a-13
3.22
7.86E-3
2a-14
4.08
0.18
2a-15
4.40
0.56
2a-16
4.50
0.20
2a-17
5.87
0.11
2a-18
EGFP-TNFα
FcγRIIa

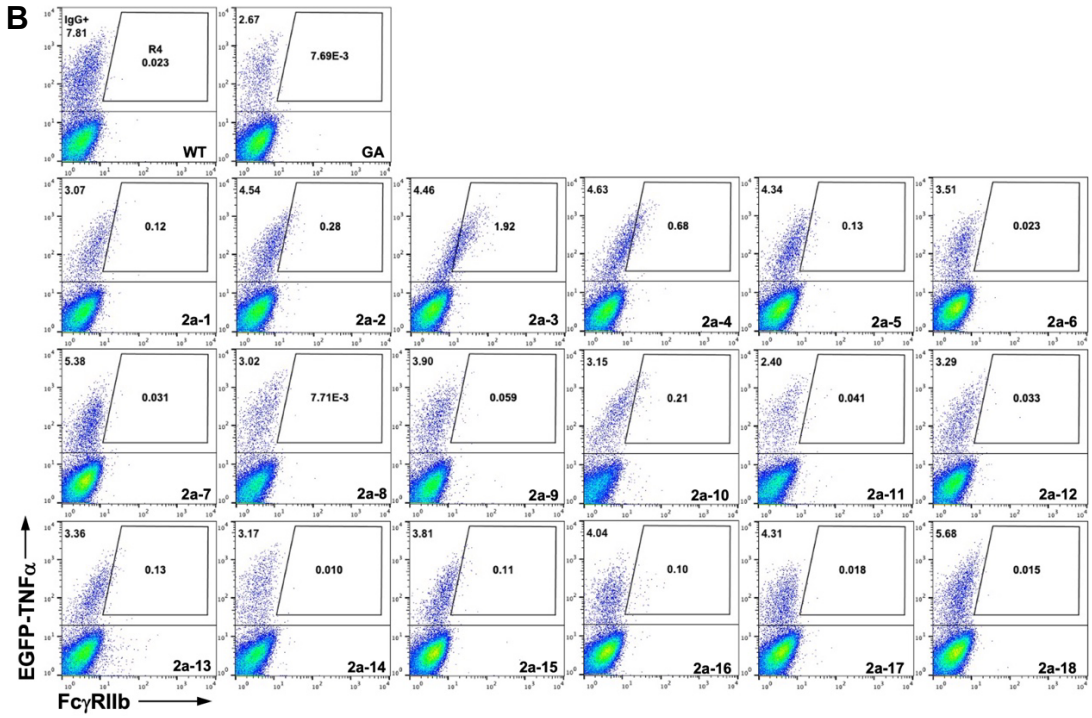


Figure 5.5 Flow cytometric analysis of isolated IIa variants with **(A)** FcγRIIa and **(B)** FcγRIIb. Cells were stained with EGFP-TNFα for IgG expression and biotinylated corresponding FcγRs and streptavidin-APC for FcγR binding. WT and GA clones were used as controls. Quantitative analysis of cell populations in R4 gates were shown in **Table 5.2**.

Table 5.2 Quantitative flow cytometry analysis of isolated FcγRIIa binders.

Clone	Mutations	FcγRIIa[a] MFI[b] TNFα MFI	FcγRIIb MFI TNFα MFI
WT		8.4	2.8
GA	G236A	17.6	2.6
2a-1	P232L	12.5	5.9
2a-2	P232L G316D	14.4	6.1
2a-3	K290E K326E	19.8	11.7
2a-4	K326E	15.5	7.3
2a-5	E258G L328I	13.1	5.8
2a-6	S267G K290E	10.4	4.5
2a-7	M252K K290E Q311R	13.2	5.4
2a-8	V284E K290E	7.7	3.2
2a-9	G236E	17.5	3.6
2a-10	G236E K288R K290W K320M	14.8	4.0
2a-11	G236E E318G K322E	15.2	4.0
2a-12	N325S	6.9	4.6
2a-13	E216G K222R	21.2	6.5
2a-14	K288R	5.4	2.8
2a-15	L251F N276Y K290E	14.3	7.5
2a-16	C200W G236E	15.5	5.5
2a-17	G236E M252R	15.3	5.5
2a-18	V302E	10.4	4.5

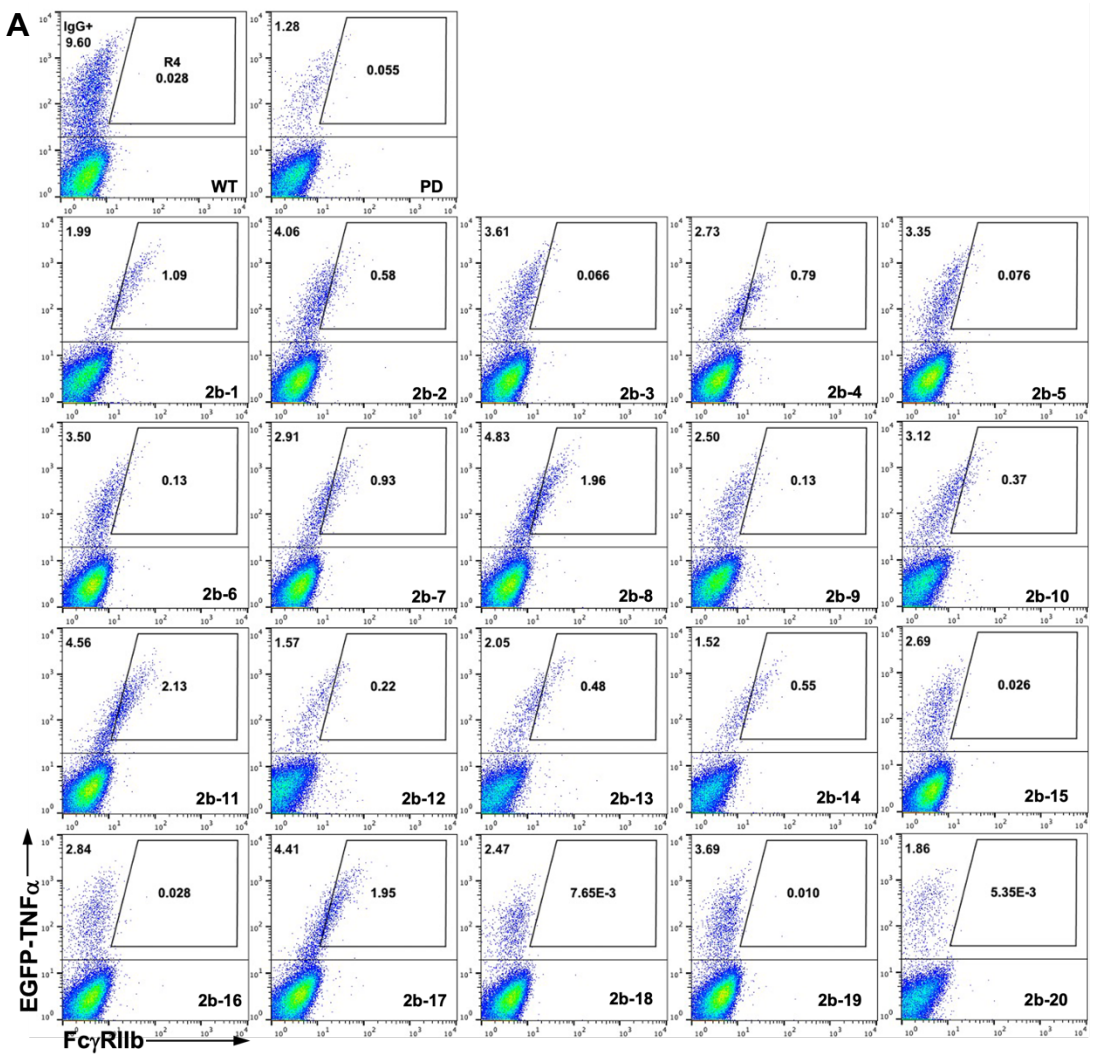
[a] FcγRIIa^{R131}: FcγRIIa^{H131}= 1:1

[b] MFI, median fluorescence intensity. Relative binding strength on FcγR was normalized with IgG expression, calculated as $\frac{MFIFc\gamma R}{MFITNF\alpha} \times 100$.

5.3.6 Characterizations of Fc variants with improved FcγRIIb affinity

Similar flow cytometry, ELISA, BLI analysis were conducted for the Fc variants targeting FcγRIIb. 18 out of 20 selected Fc variants showed increased MFI of FcγR on

flow cytometer (**Fig. 5.6A**). 12 out of 14 trastuzumab Fc variants showed improved affinity towards FcγRIIb in ELISA, though the specificity was varied (**Fig. 5.7C**. 2b-1, 5, 7 and 10 were further examined by BLI (**Table 5.4**, **Fig. 5.8C**). PD, previously reported FcγRIIb enhanced Fc variant, was shown 450 nM of K_D towards FcγRIIb whereas reduced affinity towards both FcγRIIa and FcγRIIIa^{F176}. The 4 selected variants all showed lower K_D towards FcγRIIb compared to PD and WT. At the meantime, the affinities of the 4 variants towards FcγRIIa was slightly increased while those towards FcγRIIIa were reduced or barely changed.


A
IgG+
9.60
R4
0.028
WT
1.28
0.055
PD
1.99
1.09
2b-1
4.06
0.58
2b-2
3.61
0.066
2b-3
2.73
0.79
2b-4
3.35
0.076
2b-5
3.50
0.13
2b-6
2.91
0.93
2b-7
4.83
1.96
2b-8
2.50
0.13
2b-9
3.12
0.37
2b-10
4.56
2.13
2b-11
1.57
0.22
2b-12
2.05
0.48
2b-13
1.52
0.55
2b-14
2.69
0.026
2b-15
2.84
0.028
2b-16
4.41
1.95
2b-17
2.47
7.65E-3
2b-18
3.69
0.010
2b-19
1.86
5.35E-3
2b-20
EGFP-TNFα
FcγRIIb

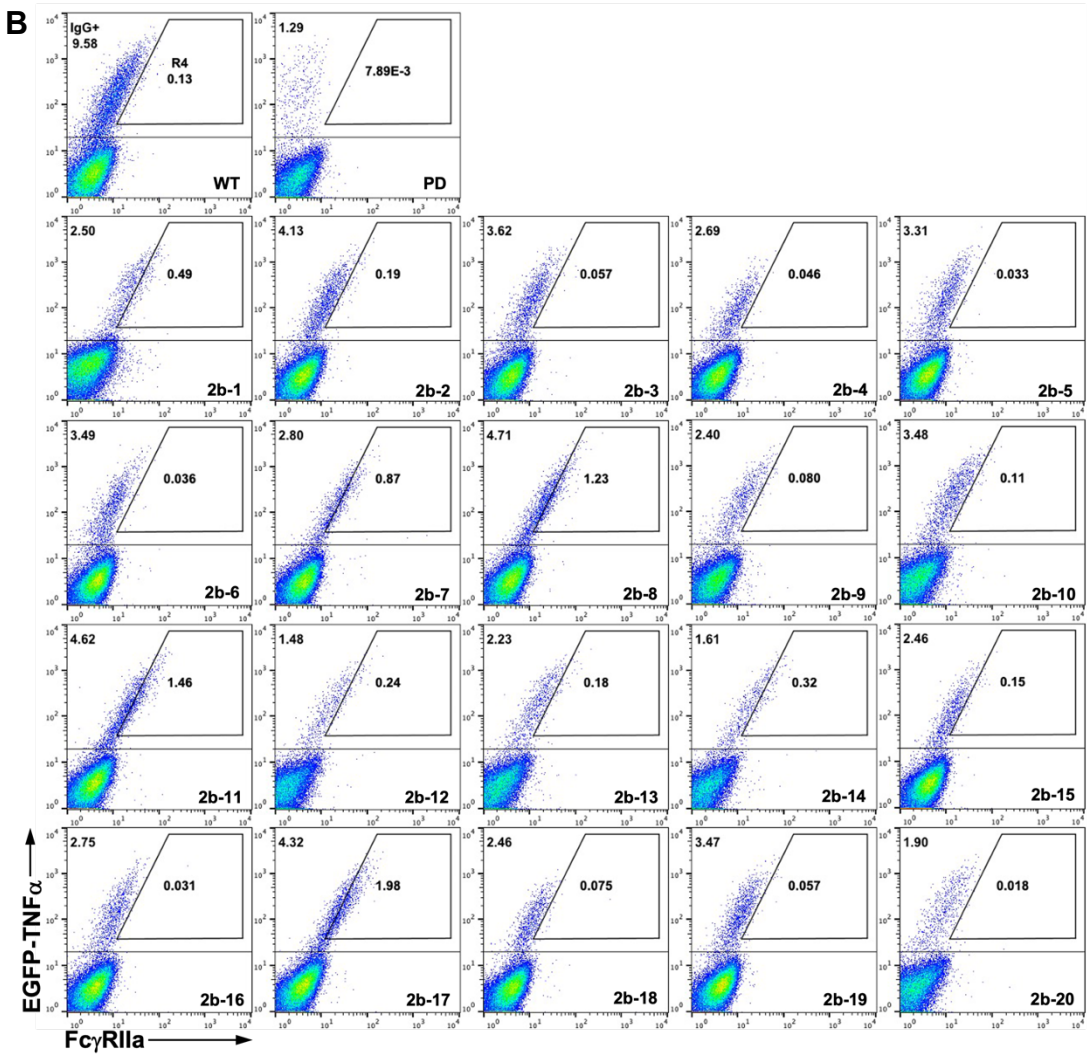


Figure 5.6 Flow cytometric analysis of isolated IIb variants with **(A)** FcγRIIb and **(B)** FcγRIIa. Cells were stained with EGFP-TNFα for IgG expression and biotinylated corresponding FcγRs and streptavidin-APC for FcγR binding. WT and PD clones were used as controls. Quantitative analysis of cell populations in R4 gates were shown in **Table 5.3**.

Table 5.3 Quantitative flow cytometry analysis of isolated FcγRIIb binders.

Clone	Mutations	FcγRIIa[a] MFI[b] / TNFα MFI	FcγRIIb MFI / TNFα MFI
WT		7.4	3.4
PD	P238D	2.1	4.7
2b-1	K222I V302E L328F K334E	12.0	11.1
2b-2	V284E N315I K326I L328H	10.2	7.8
2b-3	K326I S337F	8.3	5.5
2b-4	P228R L234H K248Q S267I V302E	11.2	12.9
2b-5	E216G K290M Y319C N325S	6.9	5.8
2b-6	V266M	8.2	6.9
2b-7	K246E T250S V305A K326E	16.0	9.5
2b-8	F241Y V284E K326E K334I	15.0	10.8
2b-9	L328F	8.0	5.5
2b-10	N286S L328I	7.1	5.5
2b-11	T250S K326E	15.3	12.1
2b-12	T289S K290T I332F	10.1	5.5
2b-13	V266L	8.6	6.6
2b-14	K246I K326E	11.3	8.3
2b-15	A339V	12.6	5.1
2b-16	P245T	8.4	3.6
2b-17	T209S K290E K326E	18.0	11.8
2b-18	C200W T256A	13.0	5.2
2b-19	V202A T307P V308A	9.3	4.8
2b-20	K248M T335P	6.5	1.8

[a] FcγRIIa^{R131}: FcγRIIa^{H131}= 1:1

[b] MFI, median fluorescence intensity. Relative binding strength on FcγR was normalized with IgG expression, calculated as $\frac{MFIFc\gamma R}{MFITNF\alpha} \times 100$.

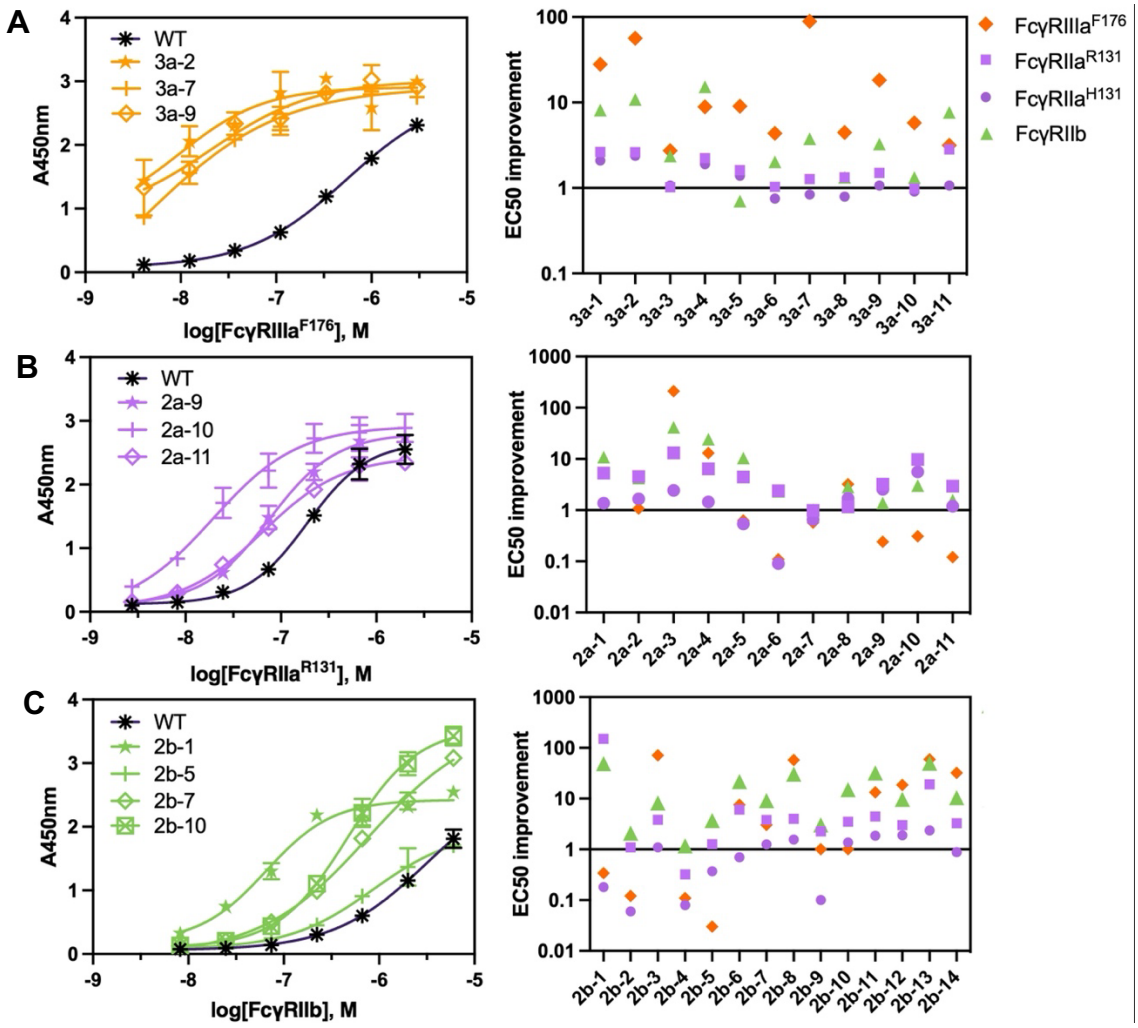
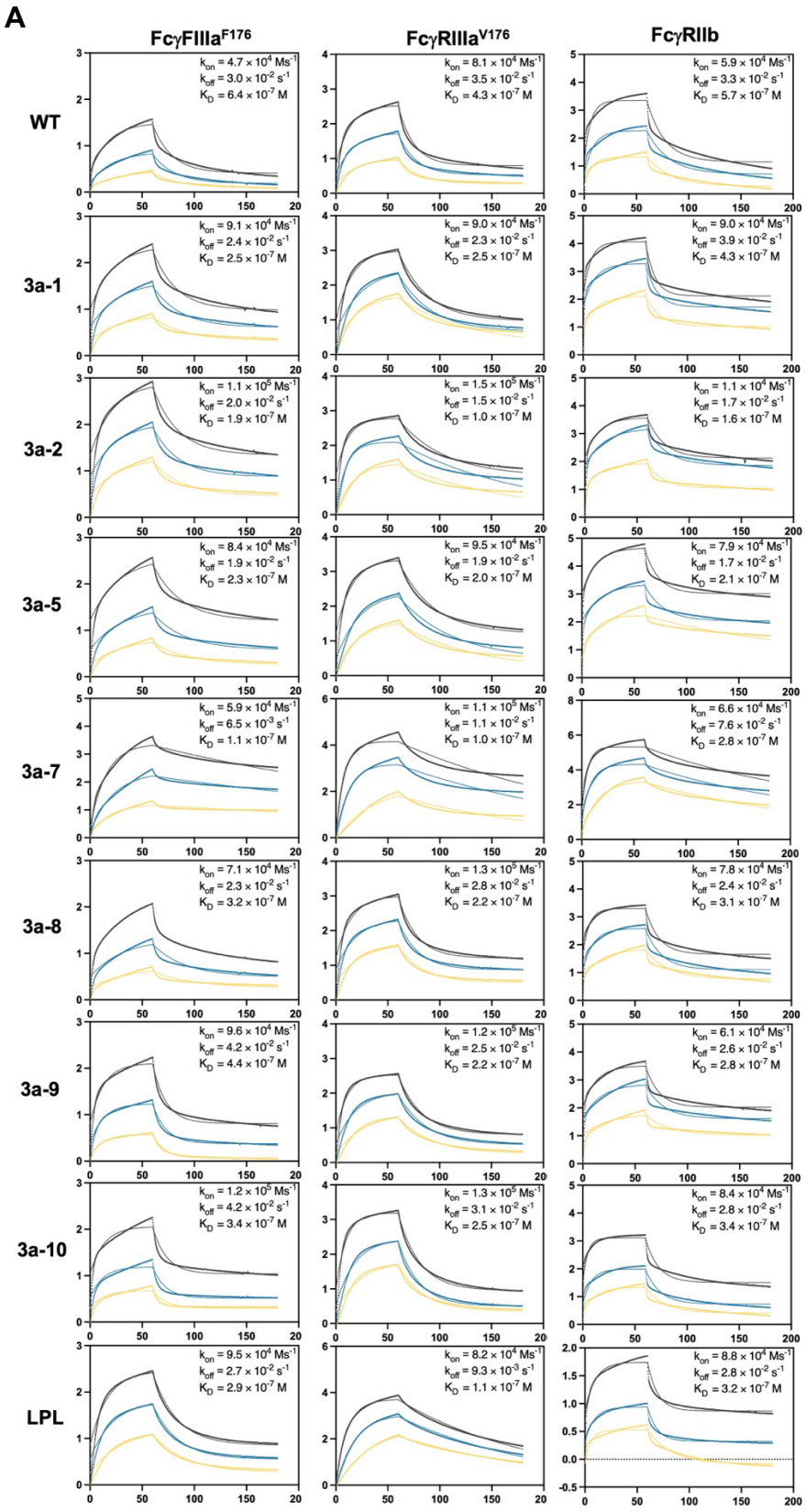


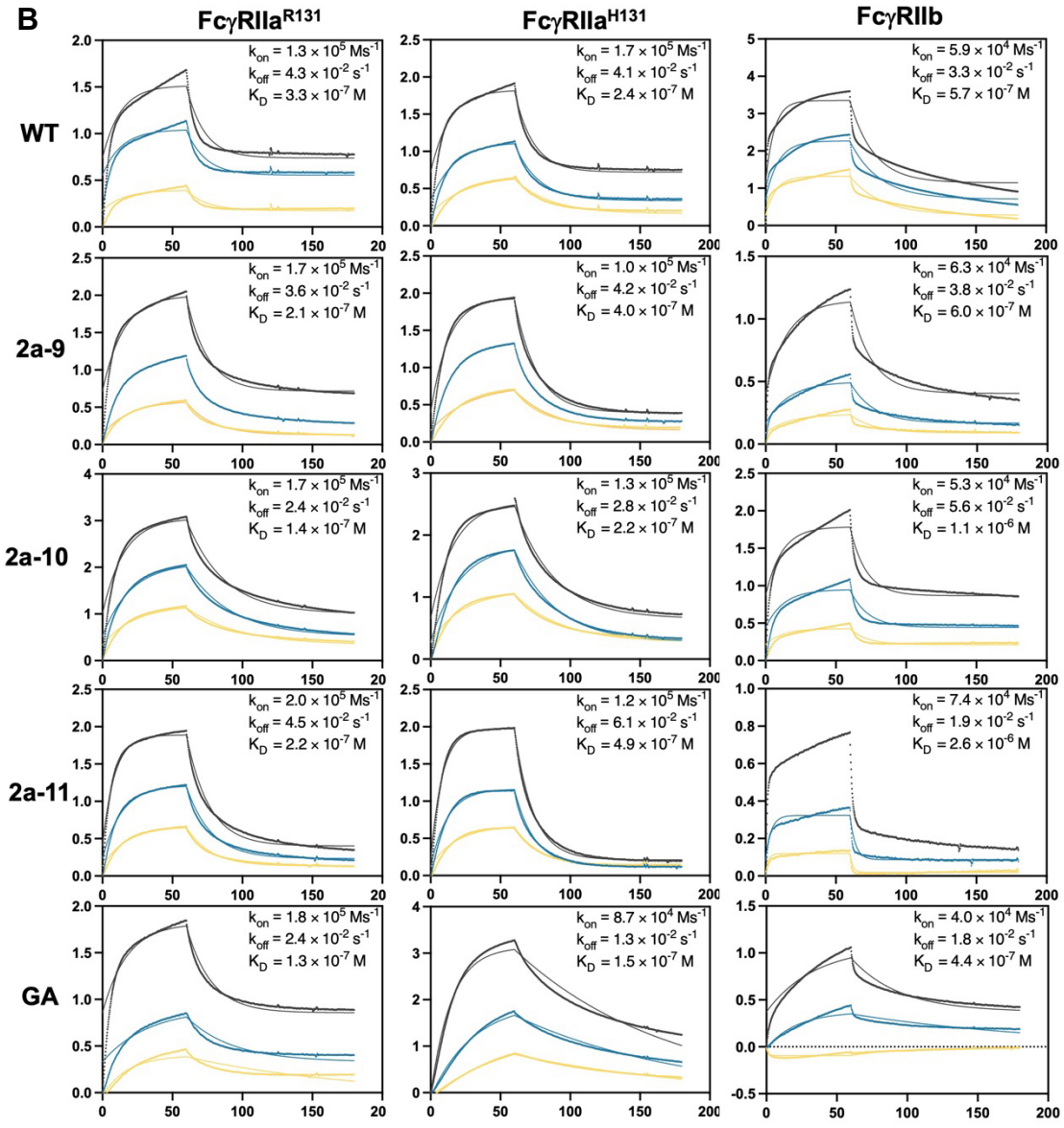
Figure 5.7 ELISA results of isolated Fc variants. **(A)** IIIa clones, **(B)** IIa clones and **(C)** IIb clones. All tested Fc variants were produced as trastuzumab IgGs. **(left)** Sigmoidal curves of representative Fc variants (n = 2). **(right)** EC_{50} improvement calculated by the equation $\frac{EC50(WT)}{EC50(Fc\ variant)}$, showing affinity improvement (value >1) and affinity reduction (value <1).

Table 5.4 Binding affinity of isolated Fc variants.

Clone	Mutations	FcγRIIIa^{F176}	FcγRIIIa^{V176}	FcγRIIa^{R131}	FcγRIIa^{H131}	FcγRIIb
WT	-	640 ± 76 (430)	430 ± 50	330 ± 83 (280)	240	570 ± 64
3a-1	P247L	250	250			430
3a-2	P247L K248E K334I	190 (180)	100			160
3a-5	P247H	230	200			210
3a-7	K218I A231V K334E	110	100			280
3a-8	K334N	320	220			310
3a-9	F275I K288M K334I	440 (400)	220			280
3a-10	E293D K334E	340	250			340
LPL	F243L R292P Y300L	290 (100)	110			320
2a-9	G236E			210	400	600
2a-10	G236E K288R K290W K320M			140 (150)	220	1100
2a-11	G236E E318G K322E			220	490	2600
GA	G236A			130	150	440
2b-1	K222I V302E L328F K334E	800		160		180
2b-5	E216G K290M Y319C N325S	2500		240		430
2b-7	K246E T250S V305A K326E	530		190		210
2b-10	N286S L328I	530		170		280
PD	P238D	> 5000		1250		450

K_D (nM) values were measured by bio-layer interferometry and surface plasmon response (shown in parenthesis).


A
FcγFIIIa^F176
FcγRIIIa^V176
FcγRIIb
WT
3a-1
3a-2
3a-5
3a-7
3a-8
3a-9
3a-10
LPL
k_on = 4.7 × 10^4 Ms^-1
k_off = 3.0 × 10^-2 s^-1
K_D = 6.4 × 10^-7 M
k_on = 8.1 × 10^4 Ms^-1
k_off = 3.5 × 10^-2 s^-1
K_D = 4.3 × 10^-7 M
k_on = 5.9 × 10^4 Ms^-1
k_off = 3.3 × 10^-2 s^-1
K_D = 5.7 × 10^-7 M
k_on = 9.1 × 10^4 Ms^-1
k_off = 2.4 × 10^-2 s^-1
K_D = 2.5 × 10^-7 M
k_on = 9.0 × 10^4 Ms^-1
k_off = 2.3 × 10^-2 s^-1
K_D = 2.5 × 10^-7 M
k_on = 9.0 × 10^4 Ms^-1
k_off = 3.9 × 10^-2 s^-1
K_D = 4.3 × 10^-7 M
k_on = 1.1 × 10^5 Ms^-1
k_off = 2.0 × 10^-2 s^-1
K_D = 1.9 × 10^-7 M
k_on = 1.5 × 10^5 Ms^-1
k_off = 1.5 × 10^-2 s^-1
K_D = 1.0 × 10^-7 M
k_on = 1.1 × 10^4 Ms^-1
k_off = 1.7 × 10^-2 s^-1
K_D = 1.6 × 10^-7 M
k_on = 8.4 × 10^4 Ms^-1
k_off = 1.9 × 10^-2 s^-1
K_D = 2.3 × 10^-7 M
k_on = 9.5 × 10^4 Ms^-1
k_off = 1.9 × 10^-2 s^-1
K_D = 2.0 × 10^-7 M
k_on = 7.9 × 10^4 Ms^-1
k_off = 1.7 × 10^-2 s^-1
K_D = 2.1 × 10^-7 M
k_on = 5.9 × 10^4 Ms^-1
k_off = 6.5 × 10^-3 s^-1
K_D = 1.1 × 10^-7 M
k_on = 1.1 × 10^5 Ms^-1
k_off = 1.1 × 10^-2 s^-1
K_D = 1.0 × 10^-7 M
k_on = 6.6 × 10^4 Ms^-1
k_off = 7.6 × 10^-2 s^-1
K_D = 2.8 × 10^-7 M
k_on = 7.1 × 10^4 Ms^-1
k_off = 2.3 × 10^-2 s^-1
K_D = 3.2 × 10^-7 M
k_on = 1.3 × 10^5 Ms^-1
k_off = 2.8 × 10^-2 s^-1
K_D = 2.2 × 10^-7 M
k_on = 7.8 × 10^4 Ms^-1
k_off = 2.4 × 10^-2 s^-1
K_D = 3.1 × 10^-7 M
k_on = 9.6 × 10^4 Ms^-1
k_off = 4.2 × 10^-2 s^-1
K_D = 4.4 × 10^-7 M
k_on = 1.2 × 10^5 Ms^-1
k_off = 2.5 × 10^-2 s^-1
K_D = 2.2 × 10^-7 M
k_on = 6.1 × 10^4 Ms^-1
k_off = 2.6 × 10^-2 s^-1
K_D = 2.8 × 10^-7 M
k_on = 1.2 × 10^5 Ms^-1
k_off = 4.2 × 10^-2 s^-1
K_D = 3.4 × 10^-7 M
k_on = 1.3 × 10^5 Ms^-1
k_off = 3.1 × 10^-2 s^-1
K_D = 2.5 × 10^-7 M
k_on = 8.4 × 10^4 Ms^-1
k_off = 2.8 × 10^-2 s^-1
K_D = 3.4 × 10^-7 M
k_on = 9.5 × 10^4 Ms^-1
k_off = 2.7 × 10^-2 s^-1
K_D = 2.9 × 10^-7 M
k_on = 8.2 × 10^4 Ms^-1
k_off = 9.3 × 10^-3 s^-1
K_D = 1.1 × 10^-7 M
k_on = 8.8 × 10^4 Ms^-1
k_off = 2.8 × 10^-2 s^-1
K_D = 3.2 × 10^-7 M


B
FcγRIIa^R131
FcγRIIa^H131
FcγRIIb
WT
2a-9
2a-10
2a-11
GA
k_on = 1.3 × 10^5 Ms^-1
k_off = 4.3 × 10^-2 s^-1
K_D = 3.3 × 10^-7 M
k_on = 1.7 × 10^5 Ms^-1
k_off = 4.1 × 10^-2 s^-1
K_D = 2.4 × 10^-7 M
k_on = 5.9 × 10^4 Ms^-1
k_off = 3.3 × 10^-2 s^-1
K_D = 5.7 × 10^-7 M
k_on = 1.7 × 10^5 Ms^-1
k_off = 3.6 × 10^-2 s^-1
K_D = 2.1 × 10^-7 M
k_on = 1.0 × 10^5 Ms^-1
k_off = 4.2 × 10^-2 s^-1
K_D = 4.0 × 10^-7 M
k_on = 6.3 × 10^4 Ms^-1
k_off = 3.8 × 10^-2 s^-1
K_D = 6.0 × 10^-7 M
k_on = 1.7 × 10^5 Ms^-1
k_off = 2.4 × 10^-2 s^-1
K_D = 1.4 × 10^-7 M
k_on = 1.3 × 10^5 Ms^-1
k_off = 2.8 × 10^-2 s^-1
K_D = 2.2 × 10^-7 M
k_on = 5.3 × 10^4 Ms^-1
k_off = 5.6 × 10^-2 s^-1
K_D = 1.1 × 10^-6 M
k_on = 2.0 × 10^5 Ms^-1
k_off = 4.5 × 10^-2 s^-1
K_D = 2.2 × 10^-7 M
k_on = 1.2 × 10^5 Ms^-1
k_off = 6.1 × 10^-2 s^-1
K_D = 4.9 × 10^-7 M
k_on = 7.4 × 10^4 Ms^-1
k_off = 1.9 × 10^-2 s^-1
K_D = 2.6 × 10^-6 M
k_on = 1.8 × 10^5 Ms^-1
k_off = 2.4 × 10^-2 s^-1
K_D = 1.3 × 10^-7 M
k_on = 8.7 × 10^4 Ms^-1
k_off = 1.3 × 10^-2 s^-1
K_D = 1.5 × 10^-7 M
k_on = 4.0 × 10^4 Ms^-1
k_off = 1.8 × 10^-2 s^-1
K_D = 4.4 × 10^-7 M

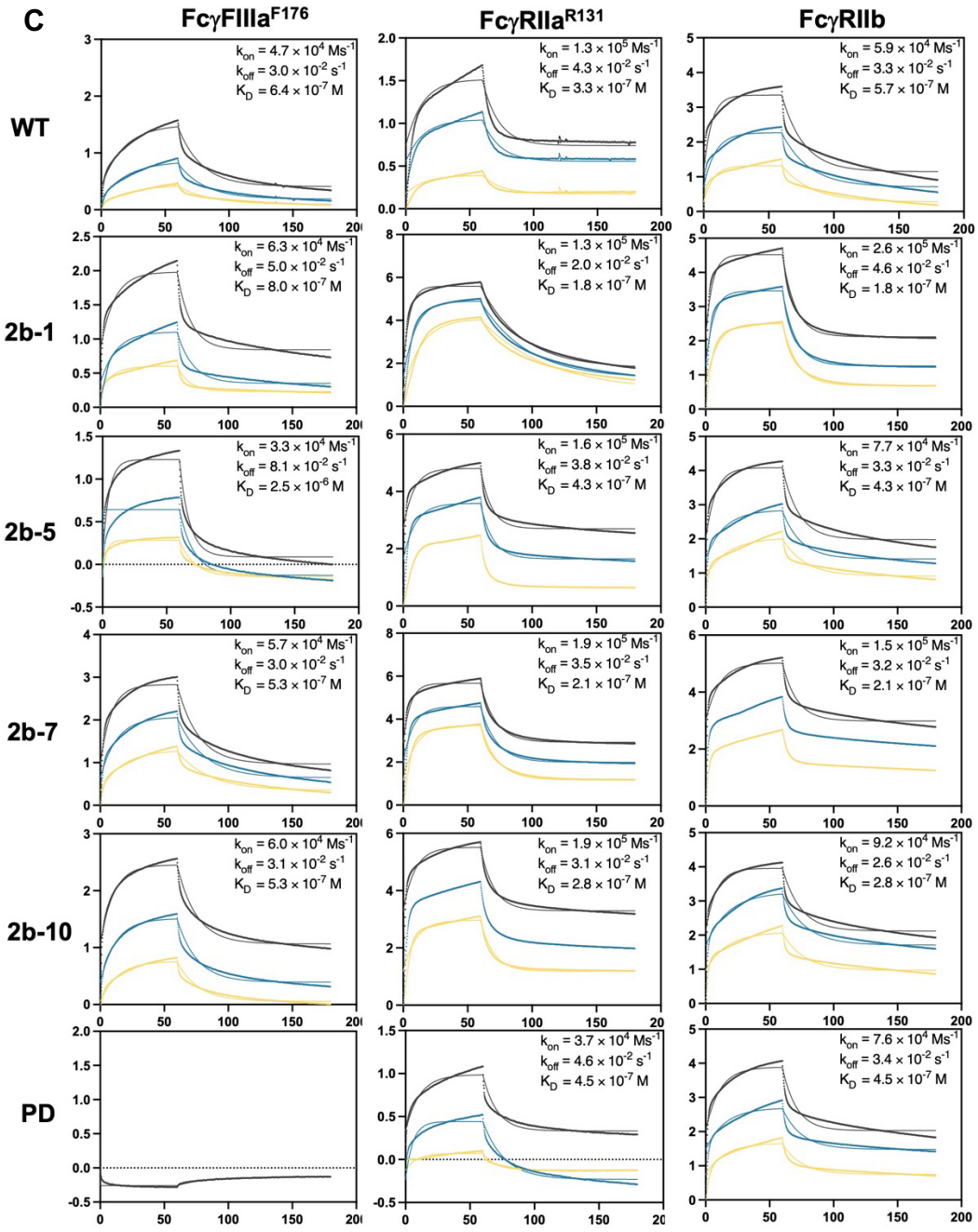


Figure 5.8 Binding kinetics measured by bio-layer interferometry. **(A)** IIIa clones, **(B)** IIa clones and **(C)** IIb clones. Tested Fc variants were produced as trastuzumab IgGs. x-axis is time (s); y axis is binding (nm).

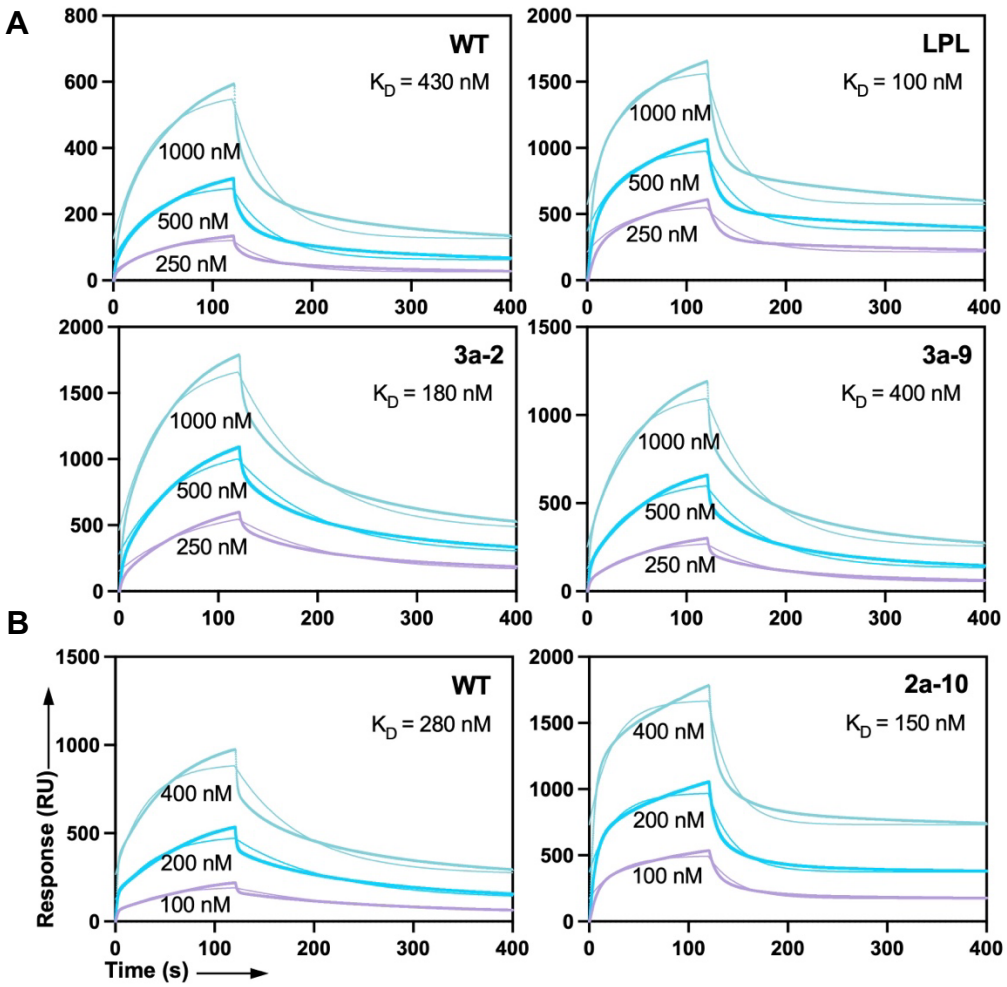


Figure 5.9 Surface plasmon response analysis of trastuzumab IgGs carrying Fc WT and Fc variants on **(A)** FcγRIIIa^{F176} and **(B)** FcγRIIa^{R131}.

5.4 Discussion and Conclusions

Display technology can be used to improve the affinity of lead clones or populations of clones, and typically involves the creation of a library of variants that is subsequently subjected to stringent selections using limiting antigen concentrations. Fc portion of the antibody has been previously displayed on yeast surface for engineering. However, it was so difficult to be detected with soluble FcγRs by flow cytometry that FcγR tetrameric

complexes with streptavidin had to be used to enhance the affinity[32]. By contrast, we did not see this issue in our mammalian cell-based system which is even capable of fine affinity discrimination. One possibility could be that the Fc has more suitable glycosylation in mammalian cells compared to yeast considering the N-glycosylation at the site 298 is crucial for binding to the FcγRs. Moreover, mammalian cell can display full-length IgG on the surface, presenting properly folded Fc in more native conformation. Another explanation associates with the cell surface. Given that yeast cell walls are made of a microfibrillar array of beta-1-3 glucan and beta-1-6 glucan chains, overlaid by mannoproteins, the antibodies are attached to a stiff, hydrophilic and 3-dimensional matrix compared to the glycocalyx membrane of mammalian cells providing a more dynamic environment.

In **Chapter 4**, we have constructed a mammalian cell library of 10^7 Fc variants in the context of full-length IgG. Although the error-prone PCR generated biases, stop codons and silent mutations in the library, here we were still able to isolate Fc variants with improved affinities towards corresponding FcγRs through only one round of MACS and two rounds of FACS selections. Some of the identified mutation sites (232, 236, 266, 267, 326, 328, 334) coincide with the interface between Fc and FcγR indicated by structural analysis as well as the ones reported on literatures[22,23,33–35], suggesting our platform could be beneficial to the target proteins without structural information. Novel mutations and combinations were also identified. Unlike the reported mutations that required combinations, our selected novel variants have shown significant improvement on affinity directly after selections. In addition, we found some mutations in the lower

C_H1 and hinge region could cause low MFI of EGFP-TNFα, indicating a decreased expression level or reduced binding affinity towards the antigen TNFα. This provides valuable information for the design of a more targeted Fc library in the future. Overall, these results indicate the robustness of our aRMCE-based mammalian cell display platform.

It is preferred to have specificity towards each of the FcγRs, albeit difficult due to high sequence homology. We have isolated Fc variants showing decent specificities, such as 3a-1, 2a-10, 2a-11 and 2b-1. In the future, we could include depletion rounds into selection to further improve the specificity. Moreover, antibody drugs with a slow off rate could bind with the target for a long time even after the plasma concentration has decreased to zero. To isolate the Fc variants with lower k_{off}, we could combine kinetic screening and equilibrium screening in the future selection.

It is common that Fc with improved affinity showed enhanced effector functions and thus show better efficacy in mouse models. Experimental results includes that anti-CD40 agonist antibodies with Fc enhanced for binding to FcγRIIb induced robust antitumor activity in humanized mouse models of bladder cancer[36]. Recent studies showed that anti-influenza IgG mAbs for selective binding of Fc to FcγRIIa resulted in enhanced ability to prevent or treat lethal viral respiratory infection in mice with increased maturation of dendritic cells and the induction of protective CD8+ T cell responses[7]; moreover, the anti-SARS-CoV-2 mAbs with Fc optimized for binding to both FcγRIIIa and FcγRIIa resulted in approximately 5-fold reduction in the mAb dose in mouse required to achieve

full therapeutic benefit[10]. Margetuximab, an Fc optimized anti-Her2 mAb have completed a phase I clinical trial for patients with HER2-overexpressing carcinomas[37]. In light of these exciting results, we would like to test our identified Fc variants on the effector cell functions, which could be potentially useful in human disease treatment.

5.5 References

1. Casadevall, A., Dadachova, E. & Pirofski, L. Passive antibody therapy for infectious diseases. *Nat Rev Microbiol* 2, 695–703 (2004).

2. Scott, A. M., Wolchok, J. D. & Old, L. J. Antibody therapy of cancer. *Nat Rev Cancer* 12, 278–287 (2012).

3. Chan, A. C. & Carter, P. J. Therapeutic antibodies for autoimmunity and inflammation. *Nat Rev Immunol* 10, 301–316 (2010).

4. Yasunaga, M. Antibody Therapeutics and Immunoregulation in Cancer and Autoimmune Disease. *Semin Cancer Biol* 64, 1–12 (2019).

5. Nimmerjahn, F. & Ravetch, J. V. Fcγ receptors as regulators of immune responses. *Nat Rev Immunol* 8, 34–47 (2008).

6. Hogarth, P. M. & Pietersz, G. A. Fc receptor-targeted therapies for the treatment of inflammation, cancer and beyond. *Nat Rev Drug Discov* 11, 311–331 (2012).

7. Bournazos, S., Corti, D., Virgin, H. W. & Ravetch, J. V. Fc-optimized antibodies elicit CD8 immunity to viral respiratory infection. *Nature* 588, 1–6 (2020).

8. Liu, R., Oldham, R. J., Teal, E., Beers, S. A. & Cragg, M. S. Fc-Engineering for Modulated Effector Functions—Improving Antibodies for Cancer Treatment. *Antibodies* 9, 64 (2020).

9. Horst, H. J. van der, Nijhof, I. S., Mutis, T. & Chamuleau, M. E. D. Fc-Engineered Antibodies with Enhanced Fc-Effector Function for the Treatment of B-Cell Malignancies. *Cancers* 12, 3041 (2020).

10. Yamin, R. *et al.* Fc-engineered antibody therapeutics with improved anti-SARS-CoV-2 efficacy. *Nature* 1–9 (2021) doi:10.1038/s41586-021-04017-w.

11. Treon, S. P. *et al.* Polymorphisms in FcγRIIIA (CD16) receptor expression are associated with clinical response to rituximab in Waldenstrom's macroglobulinemia. *J Clin Oncol Official J Am Soc Clin Oncol* 22, 6556 (2004).

12. Cartron, G. *et al.* Therapeutic activity of humanized anti-CD20 monoclonal antibody and polymorphism in IgG Fc receptor FcγRIIIa gene. *Blood* 99, 754–758 (2002).

13. Gavin, P. G. *et al.* Association of Polymorphisms in FCGR2A and FCGR3A With Degree of Trastuzumab Benefit in the Adjuvant Treatment of ERBB2/HER2–Positive Breast Cancer: Analysis of the NSABP B-31 Trial. *Jama Oncol* 3, 335 (2016).

14. Bibeau, F. *et al.* Impact of FcγRIIa-FcγRIIIa Polymorphisms and KRAS Mutations on the Clinical Outcome of Patients With Metastatic Colorectal Cancer Treated With Cetuximab Plus Irinotecan. *J Clin Oncol* 27, 1122–1129 (2009).

15. Yee, A. M. F., Phan, H. M., Zuniga, R., Salmon, J. E. & Musher, D. M. Association Between FcγRIIa-R131 Allotype and Bacteremic Pneumococcal Pneumonia. *Clin Infect Dis* 30, 25–28 (2000).

16. Sorge, N. M. V., Pol, W. V. D. & Winkel, J. G. J. V. D. FcγR polymorphisms: Implications for function, disease susceptibility and immunotherapy. *Tissue Antigens* 61, 189–202 (2003).

17. Richards, J. O. *et al.* Optimization of antibody binding to FcγRIIa enhances macrophage phagocytosis of tumor cells. *Molecular Cancer Therapeutics* 7, 2517–2527 (2008).

18. Mayes, P. A., Hance, K. W. & Hoos, A. The promise and challenges of immune agonist antibody development in cancer. *Nat Rev Drug Discov* 17, 509–527 (2018).

19. Li, F. & Ravetch, J. V. Inhibitory Fcγ Receptor Engagement Drives Adjuvant and Anti-Tumor Activities of Agonistic CD40 Antibodies. *Science* 333, 1030–1034 (2011).

20. Lazar, G. A. *et al.* Engineered antibody Fc variants with enhanced effector function. *Proc National Acad Sci* 103, 4005–4010 (2006).

21. Weitzenfeld, P., Bournazos, S. & Ravetch, J. V. Antibodies targeting sialyl Lewis A mediate tumor clearance through distinct effector pathways. *J Clin Invest* 129, 3952–3962 (2019).

22. Mimoto, F. *et al.* Engineered antibody Fc variant with selectively enhanced FcγRIIb binding over both FcγRIIaR131 and FcγRIIaH131. *Protein Engineering, Design and Selection* 26, 589–98 (2013).

23. Stavenhagen, J. B. *et al.* Fc Optimization of Therapeutic Antibodies Enhances Their Ability to Kill Tumor Cells In vitro and Controls Tumor Expansion In vivo via Low-Affinity Activating Fcγ Receptors. *Cancer Research* 67, (2007).

24. Jung, S. T. *et al.* Aglycosylated IgG variants expressed in bacteria that selectively bind FcγRI potentiate tumor cell killing by monocyte-dendritic cells. *Proc National Acad Sci* 107, 604–609 (2010).

25. Kober, L., Zehe, C. & Bode, J. Optimized signal peptides for the development of high expressing CHO cell lines. *Biotechnol Bioeng* 110, 1164–1173 (2013).

26. Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* 8, 921–929 (1999).

27. Munro, S. & Pelham, H. R. B. A C-terminal signal prevents secretion of luminal ER proteins. *Cell* 48, 899–907 (1987).

28. Beerli, R. R., Wels, W. & Hynes, N. E. Autocrine Inhibition of the Epidermal Growth Factor Receptor by Intracellular Expression of a Single-Chain Antibody. *Biochem Bioph Res Co* 204, 666–672 (1994).

29. Beerli, R. R., Wels, W. & Hynes, N. E. Intracellular expression of single chain antibodies reverts ErbB-2 transformation. *J Biol Chem* 269, 23931–23936 (1994).

30. Barat, B. & Wu, A. M. Metabolic biotinylation of recombinant antibody by biotin ligase retained in the endoplasmic reticulum. *Biomol Eng* 24, 283–291 (2007).

31. Howarth, M. & Ting, A. Y. Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat Protoc* 3, 534–45 (2008).

32. JEFFREY, S., SERGEY, G., CHRISTOPHER, R. & NADINE, T. Identification and engineering of antibodies with variant Fc regions and methods of using same.

33. Shields, R. L. *et al.* High Resolution Mapping of the Binding Site on Human IgG1 for FcγRI, FcγRII, FcγRIII, and FcRn and Design of IgG1 Variants with Improved Binding to the FcγR*. *J Biol Chem* 276, 6591–6604 (2001).

34. Chen, D. *et al.* A general Fc engineering platform for the next generation of antibody therapeutics. *Theranostics* 11, 1901–1917 (2021).

35. Liu, Z. *et al.* Asymmetrical Fc Engineering Greatly Enhances Antibody-dependent Cellular Cytotoxicity (ADCC) Effector Function and Stability of the Modified Antibodies. *J Biol Chem* 289, 3571–3590 (2014).

36. Garris, C. S., Wong, J. L., Ravetch, J. V. & Knorr, D. A. Dendritic cell targeting with Fc-enhanced CD40 antibody agonists induces durable antitumor immunity in humanized mouse models of bladder cancer. *Sci Transl Med* 13, eabd1346 (2021).

37. Bang, S. P-041 Conditionally reprogrammed pancreatic cancer cells: A novel patient-derived cell line model for human pancreatic cancer studies. *Ann Oncol* 28, iii24 (2017).

Chapter 6: Conclusions and Future Directions

6.1 Conclusions

For a long time, it has been widely agreed that RMCE-based system is deficient in construction of large mammalian cell libraries due to its low efficiency albeit the precise single copy gene integration. Now, we override this consensus with the two complementary approaches developed in this study.

In Chapter 2, we exhibited the improved RMCE efficiency by simian virus 40 (SV40) large T mediated DNA replication (T-RMCE) which can maintain a high copy number of GOI in transfected cells. The utility of T-RMCE has been illustrated by GFP evolution.

In Chapter 3, we demonstrated that cell cycle arrest at mitosis phase can enhance RMCE efficiency presumably by removing the nucleus barrier for the genome access, temporarily doubling the amount of landing pad, and pausing chromatin condensation. This approach is not gene of interest specific or cell clone dependent.

In Chapter 4, we constructed mammalian cell libraries of up to 10 million clones displaying a repertoire of Fc in full-length IgG format on the cell surface. We validated the library diversities by deep sequencing, bioinformatics and statistical analysis.

In Chapter 5, we presented the robustness of our aRMCE-based system in Fc engineering. Within 3 rounds of MACS/FACS, panels of Fc variants were isolated showing improved affinities towards three FcγRs respectively. Future effector function assay will be performed to show their potential in cancer and infectious disease treatments.

Overall, we developed a platform that enables the construction of an ideal mammalian cell library which encompasses large diversity of transgenes with single copy integrated to a defined genomic locus. This has been supported by the deep sequencing analysis and successful directed evolution of two proteins of interest. Our work will significantly contribute to the protein engineering field and eventually lead to more effective therapeutic strategies.

6.2 Future Directions

SV40 large T and cell cycle arrest could improve the RMCE efficiency from different perspectives, one increasing the transgene copy number by replication, the latter providing easier access. We have constructed large mammalian cell libraries with either of the approaches in different mammalian cell lines. Moving forward, we would like to combine these two approaches and apply in a suspension cell line which offers advantages associated with scalability.

Mammalian cell library construction has an important application in antibody engineering. It is not only because high affinity antibody could be isolated ---- which has been readily achieved by other more convenient display technologies (phage display or yeast display) ---- but also associated with the superior biophysical properties of the antibodies isolated from mammalian cell library[1]. This could help to resolve multiple developability issues during the early stages of lead discovery, such as aggregation propensity, polyreactivity and immunogenicity[2]. Our platform for mammalian cell

display could significantly reduce the risk in the future development of antibody drugs and improve the successful rate in clinical trials.

T cell receptors (TCRs) represent a promising alternative to antibodies for viral infection and cancer treatment[3–7]. The TCR is a disulfide-linked heterodimeric protein normally consisting of the highly variable α and β chains anchored on the T cell surface that are able to recognize peptide presented in complex with MHC molecules on antigen presenting cells. Libraries of TCRs have been displayed on yeast cells as single chains requiring the peptide:MHC complex presented in a multimeric format[8,9]. Chervin[10] et al introduced TCRs by retroviral infection in T cells and an effective library size of less than 10^4 clones was generated. Using our RMCE-based mammalian cell library system, construction of a much larger TCR library would be realistic with αβ heterodimer displayed on mammalian cell surface, expediting the TCR engineering process and resulting in more effective TCR-based treatment strategies.

Besides protein engineering of complex proteins, our platform could enable the drug discovery towards “difficult” targets. For instance, G protein-coupled receptors (GPCRs) play fundamental roles in many physiological processes and thereby represent the largest family of attractive drug targets [11,12]. Unlike the soluble protein, functionally active GPCRs are difficult to purify reliably in sufficient quantity and maintain folded form in a long lifetime[11]. Therefore, it is preferred to screen the library against the GPCR presented in its native context on the whole cell surface[13]. However, binding-based selection can lead to isolation of antibodies with non-specific binding or nonfunctional ones, resulting in expensive, late-stage discovery failures. Therefore, a function-based selection system

would be advantageous in which both the library of full-length antibodies and the GPCR target display on the same cell surface and only the antibodies can be isolated that are able to activate/inactivate the downstream signaling pathway of the GPCR as indicated by a reporter gene inside the cell. Ren[14] et al has recently combined single domain antibody display on GPCR-expressed cells and β-arrestin recruitment-based cell sorting and screening to directly identify agonists and antagonist targeting human apelin receptor. However, one round of phage panning was still required for binder enrichment to accommodate their capacity of mammalian cell library, leading to the loss of potentially functional binders. Our large mammalian cell library construction platform will help avoid this dilemma and pave the wave to the antibody discovery for GPCRs.

6.3 References

1. Dyson, M. R. *et al.* Beyond affinity: selection of antibody variants with optimal biophysical properties and reduced immunogenicity from mammalian display libraries. *Mabs* 12, 1829335 (2020).

2. Jain, T. *et al.* Biophysical properties of the clinical-stage antibody landscape. *P Natl Acad Sci Usa* 114, 944–949.

3. Greenbaum, U., Dumbrava, E. I., Biter, A. B., Haymaker, C. L. & Hong, D. S. Engineered T-cell Receptor T Cells for Cancer Immunotherapy. *Cancer Immunol Res* 9, 1252–1261 (2021).

4. Zhang, Y. & Li, Y. T cell receptor-engineered T cells for leukemia immunotherapy. *Cancer Cell Int* 19, 2 (2019).

5. Godbole, I. *et al.* Engineered chimeric T cell receptor fusion construct (TRuC)-expressing T cells prevent translational shutdown in SARS-CoV-2-infected cells. *Biorxiv* 2021.06.25.449871 (2021) doi:10.1101/2021.06.25.449871.

6. Li, D. *et al.* Genetically engineered T cells for cancer immunotherapy. *Signal Transduct Target Ther* 4, 35 (2019).

7. Wisskirchen, K. *et al.* T cell receptor grafting allows virological control of Hepatitis B virus infection. *J Clin Invest* (2019) doi:10.1172/jci120228.

8. Weber, K. S., Donermeyer, D. L., Allen, P. M. & Kranz, D. M. Class II-restricted T cell receptor engineered in vitro for higher affinity retains peptide specificity and function. *P Natl Acad Sci Usa* 102, 19033–19038 (2005).

9. Holler, P. D. *et al.* In vitro evolution of a T cell receptor with high affinity for peptide/MHC. *Proc National Acad Sci* 97, 5387–5392 (2000).

10. Chervin, A. S., Aggen, D. H., Raseman, J. M. & Kranz, D. M. Engineering higher affinity T cell receptors using a T cell display system. *J Immunol Methods* 339, 175–184 (2008).

11. Hutchings, C. J., Koglin, M., Olson, W. C. & Marshall, F. H. Opportunities for therapeutic antibodies directed at G-protein-coupled receptors. *Nat Rev Drug Discov* 16, 787–810 (2017).

12. Davenport, A. P., Scully, C. C. G., Graaf, C. de, Brown, A. J. H. & Maguire, J. J. Advances in therapeutic peptides targeting G protein-coupled receptors. *Nat Rev Drug Discov* 1–25 (2020) doi:10.1038/s41573-020-0062-z.

13. Jo, M. & Jung, S. T. Engineering therapeutic antibodies targeting G-protein-coupled receptors. *Exp Mol Medicine* 48, e207 (2016).

14. Ren, H. *et al.* Function-based high-throughput screening for antibody antagonists and agonists against G protein-coupled receptors. *Commun Biology* 3, 146 (2020).