

Chapter 5

Conclusions and Future Perspectives

We developed and validated a new technology for high throughput sequencing of paired antibody heavy and light chains at very high cell throughput. This work was undertaken specifically to address a critical deficiency in currently available high throughput antibody sequencing techniques, namely that the pairing information of heavy and light chains is irreversibly lost during traditional high-throughput sequencing. We first reported a microarray-based technique with capacity for up to 10^5 cells per analysis [1], then translated the same general workflow into emulsion droplets for interrogation of up to 10^7 single B cells per operator in a single day [2]. We determined the VH:VL pairing accuracy of our technique to be >97% [2] and applied our technology for antibody discovery [1, 3–5], proteomic analysis of vaccine responses [3, 6], mechanistic investigation of HIV broadly neutralizing antibody development [4], and to generate novel immunological insight related to the composition and development of antibody repertoires in healthy human donors [2, 7].

A primary future direction for paired heavy:light sequencing is to analyze the features of various B cell subsets. The work presented here analyzed the sequence and structural differences between naïve and antigen-experienced B cells [7], uncovering a number of unreported features regarding the development and maturation of the human B cell repertoire in healthy adults. However, many more distinct B cells have been identified that remain to be analyzed, including earlier developmental stages (e.g. immature B cells) and antigen-experienced subsets such as plasmablasts, plasma cells (including long-lived and short-lived plasma cell subsets), “double-negative” or tissue-like memory, and more recently identified subsets which are still being refined and developed [8–12]. The high-throughput sequence and structural analyses of these B cell populations may reveal their distinct developmental origins and functional contributions to effective (or occasionally, ineffective) adaptive immunity. The antigen-presenting role of B cells has been recognized to have greater importance in recent years [13], and paired heavy:light sequencing may reveal additional insights for the antigen-presenting influence of B cells on T cell responses as well. A major topic of interest is in the identification of

the B cell precursor subset that leads to long-lived plasma cells [12, 14, 15], which has the potential to significantly accelerate the development of highly effective vaccines. Enhanced understanding of B cell development will help us design vaccines with more effective long-term protection by enhanced elicitation of long-lived plasma cells.

Another critical application of paired heavy:light sequencing is in the sequence analysis of vaccine and disease responses. Recent reports have identified convergent genetic signatures that were elicited in response to influenza vaccination [16], dengue infection [17], HIV infection [18, 19], and other diseases [20, 21]. However, most of these signatures rely on the heavy chain identification only, which results in poor predictive capacity and an inability to experimentally test identified antibodies for binding specificity. In other cases, such as VRC01-class HIV broadly neutralizing antibody identification, the gene signatures are based on paired heavy:light sequences which cannot be detected at high-throughput by any other technology [18]. Paired heavy:light sequence information will help to better elucidate the genetic convergence among antibody repertoires of distinct individuals, as it provides a much higher resolution for the identification of genetically similar antibodies than separate heavy-only and light-only antibody sequencing can provide.

Similar to the identification of genetic convergence in vaccines and disease responses, a number of genetic similarities have been observed in autoimmune antibodies [20–23]. Paired heavy:light sequencing in the context of autoimmunity may also provide enhanced resolution of autoimmune B cell features. The additional information contained in paired heavy:light sequences could one day enable diagnosis of the precise mechanism of autoimmunity based on a single blood sample. Furthermore, the confident identification of autoimmune mechanistic targets would highlight potential insights as to how autoimmunity develops and enable targeted strategies for antigen-specific therapies.

Paired heavy:light sequencing has a reasonable efficiency (approximately one antibody VH:VL cluster per 15–20 input cells, as reported in Chap. 4), however it is likely that significant opportunities for improvement remain as the technology matures. Several features of the paired heavy:light sequencing workflow could be further optimized. Perhaps most importantly, the RT-PCR enzyme mix could be further improved. Recently developed protocols in molecular biology have generated enhanced reverse transcription and PCR enzymes [24], and these proteins or comparative analysis of multiple RT-PCR kits may enhance the yield of emulsion PCR reactions. Additionally, further optimization of the mRNA capture step may be possible. Variation of the number of beads, size of emulsion droplets, and lysis buffer compositions could reveal an optimized protocol to enhance the amount of mRNA captured using paired heavy:light sequencing. Additional non-specific B cell stimulation technologies could also increase the amount of immunoglobulin mRNA transcribed to further enhance antibody sequence recovery [4]. Finally, advances in sequence analysis throughput and/or error rates could improve the yield of data and bioinformatic algorithms and enhance the amount of information that can be obtained from a single paired heavy:light sequencing run. The paired

heavy:light sequencing workflow currently operates with acceptable efficiency to obtain far greater yield than any other technology and address a wide range of scientific problems; still, the above opportunities for improvement would provide even greater utility and a broader range of applications for this antibody sequencing platform.

A further important application beyond the analysis of antibody repertoires is in the high-throughput sequencing of paired T cell receptor (TCR) β chain: α chain sequences. Paired β : α sequencing is an analogous problem to heavy:light sequencing, with a few key differences: (i) T cells have lower T-cell receptor expression levels than most B cells, (ii) the TCR genes are much more diverse than B cell genes, (iii) T cells do not show somatic hypermutation, and (iv) T cells show a higher rate of multiple α chains than B cells with allelically included light chains. These features make T cell analysis somewhat more difficult experimentally due to the lower expression levels and higher size of multiplex primer libraries, and will require similar bioinformatic protocols with slight differences as compared to antibody bioinformatic analysis (iii, iv). Paired β : α T cell receptor sequencing is underway with promising results, and the ability to analyze paired β : α repertoires will dramatically improve our ability to understand T cell responses in the setting of infectious diseases, vaccination, and autoimmunity.

Importantly, our method for sequencing multiple mRNAs from single cells has a number of applications beyond the sequencing of antibody heavy and light chain pairs. Importantly, the poly(dT)-based single-cell capture method collects all mRNAs at the single-cell level, which includes far more than the immunoglobulin variable region gene sequences, and future efforts in this area will incorporate additional mRNAs of interest. For example, one could pair heavy chain sequences with transcription factors implicated in B-cell development such as Blimp-1[39] to determine both B cell maturity and VH:VL sequences in a single experiment without the need for fluorescence-activated cell sorting (FACS). As FACS is an expensive and time-consuming task that requires skilled operators, the ability to omit FACS for separate analysis of cell subsets may enable faster, cheaper, and therefore more effective investigations of human adaptive immune repertoires. Another approach is to use barcoded beads or hydrogel droplets for analysis of the entire transcriptome of single cells at very high throughput [25, 26]. We will also analyze paired antibody VH:VL sequences of cells with surface expression of a particular phenotype, for example B-cell receptor affinity to an antigen of interest for extremely rapid and high-throughput antibody discovery.

In summary, our accessible technology for sequencing the paired antibody VH:VL repertoire has enabled rapid interrogation of the immune response and can be applied to investigate B-cell responses in a variety of clinical and research settings. In particular, the suite of new techniques presented here is enhancing high-throughput, high-resolution analysis of human vaccine responses, providing new ways to test vaccine efficacy and inform vaccine design. The high-throughput identification and cloning of paired VH:VL antigen-specific antibodies from responding B cells will enable rapid generation of novel diagnostic, therapeutic or prophylactic antibodies and catalyze further high-impact research in the origins and

development of humoral immunity. As DNA sequencing technologies continue to progress, low-cost high-throughput single-cell antibody sequencing can enable paired antibody repertoire analysis at great depth in large study cohorts and clinical patients and in turn provide unprecedented insights into humoral responses associated with vaccine development, autoimmunity, infectious diseases and other human disease states.

References

1. DeKosky BJ et al (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotech* 31:166–169
2. DeKosky BJ et al (2015) In-depth determination and analysis of the human paired heavy—and light-chain antibody repertoire. *Nat Med* 21:86–91
3. Lavinder JJ et al (2014) Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proc Natl Acad Sci USA* 111:2259–2264
4. Doria-Rose NA et al (2014) Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature* 509:55–62
5. Wang B et al (2015) Facile discovery of a diverse panel of Anti-Ebola Virus antibodies by immune repertoire mining. *Sci Rep* 5:13926
6. Lee J et al (2016) Quantitative, molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nat Med* (Accepted)
7. DeKosky BJ et al (2016) Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. *Proc Natl Acad Sci USA* 113:E2636–E2645
8. Wirths S, Lanzavecchia A (2005) ABCB1 transporter discriminates human resting naive B cells from cycling transitional and memory B cells. *Eur J Immunol* 35:3433–3441
9. Jackson SM, Wilson PC, James, JA, Capra JD (2008) In: Frederick W, Alt KFATHFMJWU, Emil RU (eds) *Advances in immunology*. Academic Press, vol 98, pp 151–224
10. Moir S et al (2008) Evidence for HIV-associated B cell exhaustion in a dysfunctional memory B cell compartment in HIV-infected viremic individuals. *J Exp Med* 205:1797–1805
11. Kaminski DA, Wei C, Qian Y, Rosenberg AF, Sanz I (2012) Advances in human B cell phenotypic profiling. *Front Immunol* 3:302
12. Halliley JL et al (2015) Long-lived plasma cells are contained within the CD19⁺CD38^{hi}CD138⁺ Subset in Human Bone Marrow. *Immunity* 43:132–145
13. Nanton MR, Way SS, Shlomchik MJ, McSorley SJ (2012) Cutting edge: B cells are essential for protective immunity against salmonella independent of antibody secretion. *J Immunol* 189:5503–5507
14. Amanna IJ, Carlson NE, Slifka MK (2007) Duration of humoral immunity to common viral and vaccine antigens. *N Engl J Med* 357:1903–1915
15. Amanna IJ, Slifka MK (2010) Mechanisms that determine plasma cell lifespan and the duration of humoral immunity. *Immunol Rev* 236:125–138
16. Jackson KJL et al (2014) Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell Host Microbe* 16:105–114
17. Parameswaran P et al (2013) Convergent antibody signatures in human dengue. *Cell Host Microbe* 13:691–700
18. Zhou T et al (2013) Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-Class antibodies. *Immunity* 39:245–258
19. Gorman J et al (2016) Structures of HIV-1 Env V1V2 with broadly neutralizing antibodies reveal commonalities that enable vaccine design. *Nat Struct Mol Biol* 23:81–90

20. Lindop R et al (2011) Molecular signature of a public clonotypic autoantibody in primary Sjögren's syndrome: a 'forbidden' clone in systemic autoimmunity. *Arthritis Rheum* 63:3477–3486
21. Rounds WH et al (2014) The antibody genetics of multiple sclerosis: comparing next-generation sequencing to sanger sequencing. *Front Neurol* 5:166
22. Schoettler N, Ni D, Weigert M (2012) B cell receptor light chain repertoires show signs of selection with differences between groups of healthy individuals and SLE patients. *Mol Immunol* 51:273–282
23. Kalinina O et al (2014) Light chain editors of anti-DNA receptors in human B cells. *J Exp Med* 211:357–364
24. Ellefson JW et al (2016) Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science* 352:1590–1593
25. Klein AM et al (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161:1187–1201
26. Macosko EZ et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–1214