



High-Throughput Sequencing-Based Immune Repertoire Study during Infectious Disease

Dongni Hou¹, Cuicui Chen¹, Eric John Seely², Shujing Chen¹ and Yuanlin Song^{1*}

¹Department of Pulmonary Medicine, Zhongshan Hospital, Fudan University, Shanghai, China, ²Department of Medicine, Division of Pulmonary and Critical Care Medicine, University of California San Francisco, San Francisco, CA, USA

The selectivity of the adaptive immune response is based on the enormous diversity of T and B cell antigen-specific receptors. The immune repertoire, the collection of T and B cells with functional diversity in the circulatory system at any given time, is dynamic and reflects the essence of immune selectivity. In this article, we review the recent advances in immune repertoire study of infectious diseases, which were achieved by traditional techniques and high-throughput sequencing (HTS) techniques. HTS techniques enable the determination of complementary regions of lymphocyte receptors with unprecedented efficiency and scale. This progress in methodology enhances the understanding of immunologic changes during pathogen challenge and also provides a basis for further development of novel diagnostic markers, immunotherapies, and vaccines.

Keywords: immune repertoire, high-throughput sequencing, infection, lymphocyte, bioinformatics

INTRODUCTION

The adaptive immune system is composed of B and T cells that form a highly selective guard against evolving pathogens. The foundation of the adaptive immune response is based on the enormous diversity of T and B cell antigen receptors that can recognize epitopes from a near infinite number of different internal and external antigens. This profound diversity of T (TCRs) and B cell receptors (BCRs) is generated by V–D–J gene recombination of the TCR/BCR locus and subsequent somatic hypermutation and class-switching recombination of B cells after antigen stimulation. Thus, study of the immune repertoire, portrayed as the antigen-specific information within lymphocytes, has been a key to understanding the response of adaptive immunity during infection.

Despite extensive efforts using traditional techniques, analysis of the immune repertoire with high resolution has remained difficult. Several sequencing strategies, for example, Sanger sequencing, have been implemented to determine cDNA segments encoding variable regions of immunoglobulin (or TCRs) (1, 2). However, these low-throughput techniques lack the power to provide a broad picture of the full immune repertoire. During the past two decades, however, technical advances in high-throughput sequencing (HTS), also known as next-generation sequencing (NGS), along with evolving bioinformatic and statistical tools, have provided a new approach capable of analyzing the immune repertoire at the single sequence level. These methods create an unprecedentedly high-resolution picture of the immune repertoire and also provide massive data that cover each lymphocyte from the sample, in theory, dispensing with limitation of sequencing quantity (3).

Considering the extremely important role of the adaptive immune system in defending against infectious agents, HTS has great potential to aid in the discovery novel infectious agents and also

OPEN ACCESS

Edited by:

Rene De Waal Malefyt,
Merck, USA

Reviewed by:

Christian Schönbach,
Nazarbayev University, Kazakhstan
Mark Larché,
McMaster University, Canada

*Correspondence:

Yuanlin Song
yhsong70@163.com,
song.yuanlin@zs-hospital.sh.cn

Specialty section:

This article was submitted to
T Cell Biology,
a section of the journal
Frontiers in Immunology

Received: 06 January 2016

Accepted: 19 August 2016

Published: 31 August 2016

Citation:

Hou D, Chen C, Seely EJ, Chen S
and Song Y (2016) High-Throughput
Sequencing-Based Immune
Repertoire Study during
Infectious Disease.
Front. Immunol. 7:336.
doi: 10.3389/fimmu.2016.00336

offers new approaches for antibody or vaccine development. In this review, we introduce the implementation of HTS to the study of the immune repertoire and review the associated bioinformatic tools required for data processing and analysis. We then focus on the success of this technology in facilitating the exploration of infection-related immune repertoires for clinical diagnosis, treatment, and prevention.

GENERATION OF A DIVERSE IMMUNE REPERTOIRE

Amazing diversity makes the immune system the most effective system to fight against a broad scope of disease causing pathogens. This repertoire is generated by a complex series of genetic events (4). For T cells, the variable region of each TCR chain consists of three complementary determining regions (CDRs) and four frame regions (FRs). CDRs are the variable portion of

the receptor and determine the antigen specificity. While CDR1 and CDR2 are formed by variable (V) gene, CDR3 is generated by random selection and recombination of variable (V), diversity (D), and joining (J) gene segments in the heavy chain (V and J region gene segments in light chain) (5, 6) (Figure 1). Thus, CDR3 is the most diverse component of a receptor, which binds MHC molecules and (or) antigens. Construction of the TCR with an alpha chain and a beta chain is also a process that contributes to receptor diversity.

The formation and revision of the T and B cell lymphocyte receptor repertoire is a highly dynamic process. The number of each lymphocyte clone changes dramatically and depends on cell specificity and the history of antigen exposure. When encountering exogenous antigens, T cells that express receptors capable of binding to a specifically compatible peptide–MHC (pMHC) complex will expand, resulting in a massive population of antigen-specific T cells that initiate the adaptive immune response

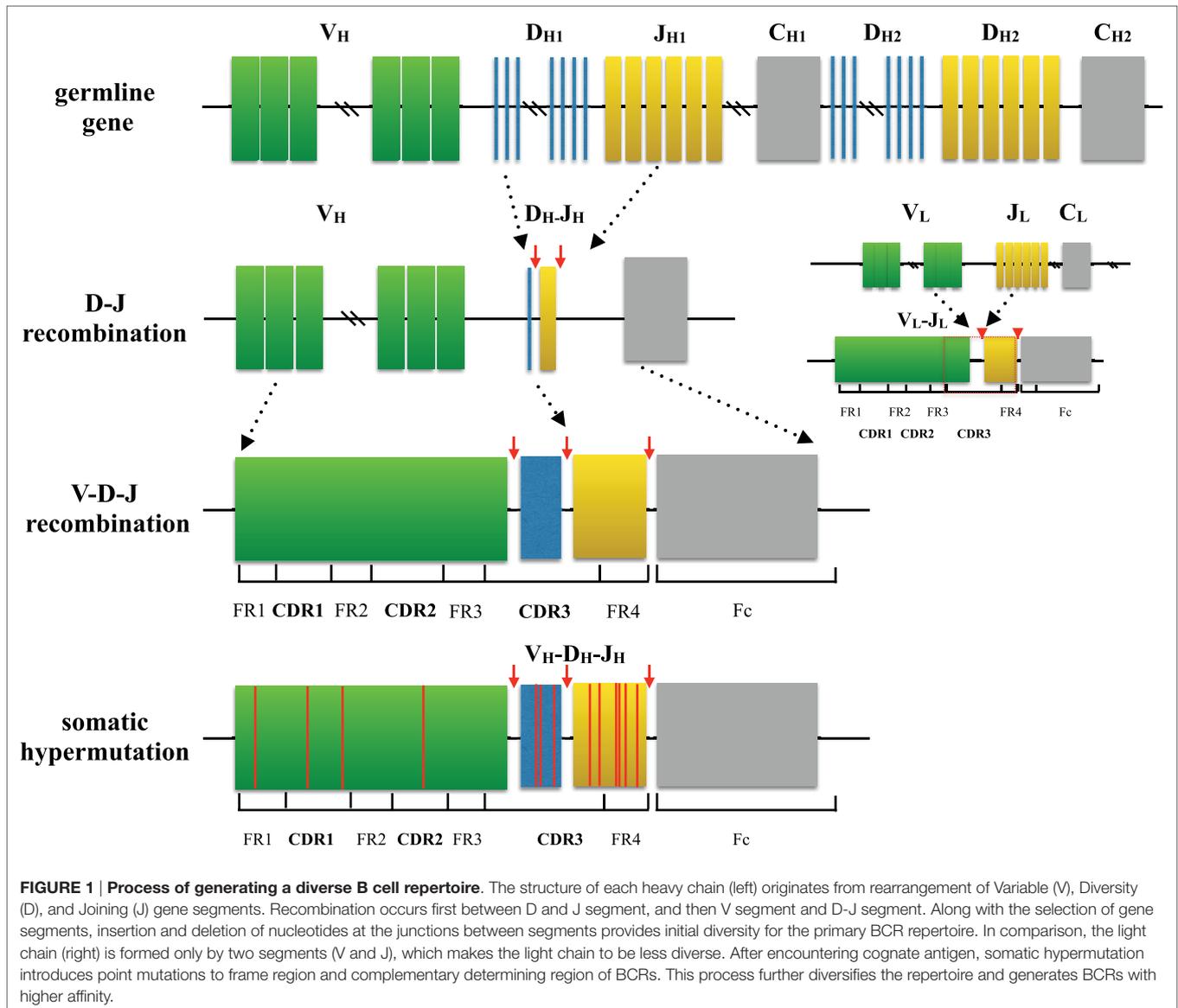


FIGURE 1 | Process of generating a diverse B cell repertoire. The structure of each heavy chain (left) originates from rearrangement of Variable (V), Diversity (D), and Joining (J) gene segments. Recombination occurs first between D and J segment, and then V segment and D-J segment. Along with the selection of gene segments, insertion and deletion of nucleotides at the junctions between segments provides initial diversity for the primary BCR repertoire. In comparison, the light chain (right) is formed only by two segments (V and J), which makes the light chain to be less diverse. After encountering cognate antigen, somatic hypermutation introduces point mutations to frame region and complementary determining region of BCRs. This process further diversifies the repertoire and generates BCRs with higher affinity.

(7–11). This antigen-driven proliferation process of T cells is distinct between CD4+ and CD8+ T cells after initial antigenic stimulus. Although these two types of T cells show comparable protein expression, proliferation rate, and transcriptome features after 2 days of non-infective stimulation, subsequent division of T cells differently depends on continuous existence of self-pMHC complexes. CD4+ T cells proliferate in a limited pattern, and its subsequent response requires persistent stimulation from antigen-presenting cells. CD8+ T cell is “programmed” to extensive expansion after short stimulation even when transferred into antigen-free hosts (12).

The post-antigen stimulation response of B cells is more complicated because it is accompanied by somatic hypermutation and class-switch recombination that offer additional diversification of the B cell repertoire (6). Somatic hypermutation is the process of introducing point mutations at CDR1, CDR2, CDR3, and FR3 to produce B cells of higher affinity to target antigens. These higher affinity clones are then selected and expanded, which is called affinity maturation. Additionally, in class-switch recombination, the gene loci encoding the C region of BCRs are excised and replaced by a series of new constant gene segments, resulting in functional differences of IgG, IgE, or IgA that participate in different immune mechanisms during pathogen elimination.

HIGH-THROUGHPUT SEQUENCING – A NEW STRATEGY FOR IMMUNE REPERTOIRE ANALYSIS

Traditional Strategies for Studying the Immune Repertoire

Prior to HTS, many strategies were developed to explore post-infection immune repertoires (13). Immunoscope spectratyping has been used to investigate TCR/BCR repertoires since the 1990s (5, 14). In this technique, using one (for B cell) or more (for T cell) V or J gene specific primer pairs, the length of CDR3 can be determined (15). CDR3 length in healthy population shows a bell-shaped pattern, indicating a polyclonal repertoire. However, the unusual peaks in infected patients imply a perturbed oligoclonal repertoire with clonal expansion. As such, CDR3 spectratyping provides robust information on the complexity and stability of circulating T/B cell repertoires and insights into the immune repertoire after infection (16–20). Even if it is relatively easy and cheap, the nucleotide sequences of CDR3 remain obscure, and the extent of heterogeneity within a particular CDR3 length cannot be assessed.

Detailed nucleotide sequences of gene segments encoding the variable region can be determined by traditional DNA sequencing techniques such as Sanger sequencing (21, 22). Flow cytometry and CDR3 spectratyping help to isolate T/B cells of interest, which complements weaknesses of Sanger sequencing in quantity limitation. Single-cell sequencing is able to identify sequences of several B cells that produce monoclonal antibodies specific to certain virus, which contributes greatly to analyzing genetic features of the antibodies in the process of antibody discovery (2, 22–24). After collecting peripheral blood samples, the B cells or memory B cells are isolated and immortalized to produce

antibodies. According to the HAI titers and neutralizing titers determined by ELISA, the virus-specific B cells can be identified, which helps to narrow B cell candidates for sequencing. These functional test-based antibody discovery strategies are successful but laborious. Despite this, these strategies are well designed for targeted antibody searching; however, it is insufficient for creating a high-resolution picture of the human immune repertoire.

High-Throughput Sequencing of Lymphocyte Repertoires

High-throughput sequencing has recently become a novel and powerful tool to investigate the immune repertoire. The depth and comprehensiveness of high-throughput immune repertoire sequencing are greater than ever, and the enormous sequencing data of disease-specific TCR/BCR clones provide great potential for the revealing dynamic changes in clonality during infectious states.

Establishing a lymphocyte repertoire database starts from sample collection from carefully selected populations and isolation of interested T cell or B cell subgroups. Due to the well-acknowledged heterogeneity of TCRs and BCRs between individuals, longitudinal studies tracking dynamic alterations in certain population help to reduce difficulties in data interpretation at unraveling the infection course. Classification of subgroups of T cell and B cells, e.g., naive and memory T/B cells, CD4+, and CD8+ T cells, is necessary if distinct behavior of these subgroups is considered in detail.

The methodology of library preparation and amplification need careful design since it affects accuracy of the ultimate sequencing data. Due to the difference of V and J gene segments, a common primer does not apply to sequencing of CDR3. Multiplex PCR is capable to amplify multiple loci simultaneously and, however, is likely to introduce bias. It is because of non-specific amplification, primer-dimer formation, and uneven reaction conditions. More precise and quantitative multiplex PCR may be achieved through primer concentration adjustment and bias filtering using amplification bias among the templates as controls (23). Another alternative PCR method is 5'RACE PCR, which provides a less biased PCR library using primers that bind downstream of the variable domain (24).

Sequencing techniques are evolving continuously to be deeper and more precise, and there are three prevalent HTS platforms available today. The comparison of mechanisms, sequencing depth, and other critical features of each platform is shown in **Table 1**. The Illumina and Roche 454 platforms have been most commonly used during immune repertoire analysis. The outputs of each platform must be analyzed with caution because of notable platform-specific sequencing error (25–28). Insertions and deletions of nucleotides, resulting from imperfect interpretation of homopolymeric stretches, are considerable for the Roche 454 platform (29), while substitution errors are predominant in Illumina platform (30, 31). The overall error rate of Illumina platform is lowest while that of Ion Torrent is highest among the three (32). In an attempt to correct sequencing errors, three algorithms are most commonly used including k-mer spectrum, multiple sequence alignment, and suffix tree (26). Based on

TABLE 1 | High-throughput sequencing platforms used for Lymphocyte Repertoire Study.

HTS platform	Mechanism	Read depth	Read length	Typical throughput	Covered region	Application	Accuracy
Roche 454 GS FLX Titanium XLR70	Pyro-sequencing	~1 million	Up to 600 bp	450 Mb	FR1-C region	BCR (long read), TCR (long read)	Consensus accuracy 99.995%
Illumina HiSeq	Dye terminator sequencing	~2 billion	2 × 250 bp	500–1000 Gb	FR3-C region	TCR (CDR3), BCR (long read), TCR (long read)	>80% bases above Q30
Illumina MiSeq	Dye terminator sequencing	~25 million	2 × 150 bp 2 × 300 bp	13.2–15 Gb	FR1-C region	TCR (CDR3)	>80% bases higher than Q30 at 2 × 150 bp >75% bases higher than Q30 at 2 × 250 bp
Ion Torrent/Life Technologies	Semiconductor sequencing	~1 billion	Up to 200 bp	30 Mb–2 Gb	FR3-C region	TCR	>97% bases

these algorithms, bioinformatic tools are designed for different platforms, for example, BFC (33), HiTEC (34), Lighter (35), Reptile (36), and ECHO (37) are for Illumina platforms, and PyroNoise (38), DeNoiser (39), and HECTOR (40) are for 454 platforms. Their main approaches, correction functions, and qualities are compared in Ref. (27).

PCR and sequencing errors inevitably result in overestimate of repertoire diversity. The common statistical strategy for both PCR and sequencing error removal is eliminating low abundance and low-quality sequences (with low Phred score), but it leads to a great loss of sequencing information. To rescue these sequences, low-quality CDR3 sequences with no more than three low-quality nucleotides can be mapped to “core clonotypes” derived from high-quality sequences with allowed mismatches at low-quality position. Then, the low-abundance core clonotypes are merged with the high-abundance core clonotypes with less than three allowed mismatches at V (≤ 2 mismatches), D (≤ 1 mismatches), and J (≤ 2 mismatches) gene segments to correct PCR errors. This integrated algorithm based on sequence quality abundance could efficiently correct artificial errors and avoid information loss, thus providing more reliable estimation of repertoire diversity (28). Unique molecular identifiers that label each starting molecule help to reduce both PCR and sequencing errors (24, 41). Combined with this experimental strategy, molecular identifier groups-based error correction (MIGEC) corrects PCR and sequencing errors more efficiently than other quality- and frequency-based strategies (42).

Determination of the V–D–J gene segments from which the CDR3s are rearranged, as well as identification of point mutations, is often achieved using the ImMunoGeneTics database (<http://www.imgt.org>) (43), despite the controversies about its validity (44–46). New V–D–J gene annotation tools based on various algorithms are reported, such as IgBLAST (47), iHMMune-align (48), and Decombinator (49, 50) (Table 2). In addition, many integrated bioinformatic tools (MiTCR, LymAnalyzer, Change-O, etc.) for data processing are developed recently (51–62) (Table 2), which provides various statistical approaches for diversity estimation, repertoire comparison, clustering analysis, and somatic hypermutation analysis (Table 3). Despite these tools, standardized bioinformatic analysis and visualization strategy is lacking, which remains the main obstacle for comparison of researches from different investigators.

PROGRESS IN INFECTION-RELATED IMMUNE REPERTOIRE

High-throughput sequencing techniques have revolutionized the study of the immune repertoire. Utilizing HTS, many important insights into mechanisms of immune response have been gained. It is also the cornerstone for potential clinical applications of repertoire analysis, including identification of diagnostic biomarkers, design of therapeutic antibodies, and development of new vaccines.

Assessing Dynamic Changes in the Immune Repertoire after Antigen Stimulation

Estimating the diversity of a TCR/BCR repertoire is necessary for estimating the theoretical size of the repertoire and for tracking changes in clonal populations during the clinical course of infection. Several different methods may be used to describe the diversity of lymphocyte repertoires at different levels – VDJ recombination diversity (90), Simpson diversity index (91, 92), and some non-parametric methods. Decrease in the overall diversity of the immune repertoire have been observed after various antigen exposures, including HIV, influenza, and human herpes virus, which implies expansion of particular T/B cell clones (67, 88, 92, 93). Our group compared changes in the diversity of the TCR beta chain and BCR heavy chain after H7N9 virus infection. Interestingly, these results show that the diversity of the BCR heavy chain starts to increase 2 weeks after H7N9 infection, while the TCR beta chain repertoire continues to contract. In addition, a more diverse BCR repertoire and a less diverse TCR beta chain repertoire in convalescent phase correlate with improved prognoses, implying differences in the response process of humoral and cellular immunity.

The immune response to vaccination has been used as an ideal model for antibody repertoire research due to the convenience drawing blood samples at well-defined time points (94). Studies using vaccines, such as influenza and TT, have revealed dynamic changes in the size and diversity of antibody repertoires before and after antigen stimulation (41, 90, 95, 96). Comparison of post-vaccination responses suggests divergent repertoire properties among individuals, different age groups, and successive

TABLE 2 | Annotation and integrated bioinformatic tools for Lymphocyte Repertoire Data Analysis.

Bioinformatic tools	Models	Key feature	Limitation	Cases of infection study	Reference
For VDJ gene assignment					
IMGT/HighV-QUEST	–	Designed for NGS data; most commonly used	Online version; results sent to user after 1–2 weeks	MERS-CoV (63), influenza (64)	(65)
iHMUNE-Align	Hidden Markov model (HMM)	Most probable gene segment; accurate IGHD gene identification	Nucleotide insertions or deletions at the gene junctions but not within germline genes	Influenza vaccine (66); EBV, CMV (67)	(48)
SoDA2	Hidden Markov model (HMM)	Probability model; ≥ 1 probable rearrangement candidates	–	HIV (68)	(50)
IgBlast	BLAST algorithm	Open source; gene databases search simultaneously	Relatively low throughput (<1000 per batch)	Dengue virus (69); HIV (70)	(47)
Decombinator	Aho–Corasick algorithm	Increased speed	Better matches between query sequence and target sequence needed; only for TCR	<i>M. tuberculosis</i> (71) HIV (72)	(49)
Integrated toolkits					
O-Change	Integrated	Functions unique to BCR analysis (somatic hypermutation analysis, quantifying selection pressure, and calculating sequence chemical properties)	–	West Nile Virus (73)	(54)
LymAnalyzer	Integrated	Accurate and complete assignment; polymorphism and SHM analysis	–	–	(56)
tcR	Integrated	Based on R; diversity calculation; repertoire comparison; visualization	For output of MiTCR	–	(57)
iMonitor	Integrated	Re-alignment to improve accuracy; visualization; PCR and sequencing errors correction	No polymorphism and SHM analysis	–	(59)
MiTCR	Integrated	Accurate and higher speed; PCR and sequencing errors correction; sequence quality filter; low quality sequencing data rescue	Strongly related to sequencing quality; only for TCR	CMV (74)	(58)
ImmunExplorer (IMEX)	Integrated	Simple statistical analysis and visualization; repertoire comparison	For IMGT/HighV-QUEST outputs	–	(55)
VDJtools	Integrated	Diversity assessment, clustering analysis, CDR3 region analysis, and data visualization; simple and user-friendly	Only for TCR; compatible annotation tools needed	–	(61)
sciReptor	Integrated	Single-cell sequencing data and HTS data; supports flow cytometry index data; clustering and SHM analysis	–	–	(60)
Others					
Tool for Ig genotype elucidation <i>via</i> Rep-Seq (TigGER)		Novel V alleles identification; personalized germline database construction	Complementary tools for IMGT; only for novel alleles not distantly related; only for TCR		(75)

immunization of the same individual with different influenza vaccines (TIV and LAIV) (66, 90, 95). The maximum clonal response has been found to occur 7 days after vaccination, but the magnitude of response varies between individuals despite an identical immune challenge, which may be influenced by previous exposure, age, and other concurrent immune responses. In addition, study of B cell memory can be achieved by repeated sequencing of samples taken from the same individual in separate immune responses to an antigen. Vollmers and colleagues identified a group of B-cell clones as a recall response to two substantially different vaccine compositions, implying the possibility of identifying cross-specific antibody using repertoire analysis (41).

Immune responses recorded by sequencing data have also been useful in testing the role of adjuvants in eliciting broad-spectrum

antibodies. Wiley and co-workers tested the immune response of mice immunized with malaria vaccine by analyzing IgG repertoires. They found that TLR agonist used as adjuvant increases the diversity of IgG variable region, which is related to improved ability of the antibodies to recognize a broad spectrum of epitopes (97). These studies exemplify a new level of details assessing vaccine response and pioneered HTS implications in vaccine design.

Signatures of TCR/BCR Sequences for the Diagnosis of Infectious Diseases

In infected patients, antigen-specific T/B cell repertoires form in response to antigen exposure in both circulation and peripheral tissues. Immune repertoire sequencing provides broad information including crucial antigen-specific clones, which have the potential to halt the spread of pathogens (98). Diagnostic marker

TABLE 3 | Statistical algorithms and models for Lymphocyte Repertoire Data Analysis.

Statistical methods	Functions	Reference
Maximum entropy models	Statistical properties of the repertoire	(76)
Capture–recapture analysis	Diversity assessment	(77, 78)
Shannon index	Diversity assessment	(79)
Gini–Simpson index	Diversity assessment	(79, 80)
Chao1 algorithm	Richness (maximum number of clones)	(81)
Abundance-based coverage estimator	Richness (maximum number of clones)	(82)
Chao2	Lower bound of species richness	(83) (84)
Wu–Kabat index		
Poisson abundance model	Diversity and clone distribution estimation	(85, 86)
Morisita–Horn similarity index	Repertoire similarity comparison	(87)
Clonality score	Convergence-level estimation	(67, 84, 88)
Hidden Markov model	V and J assignment (probability based)	(48, 50)
Aho–Corasick algorithm	V and J assignment	(49)
Fast-tag-searching algorithm	V and J assignment (Hamming distance based)	(56)
Single nucleotide polymorphism (SNP) calling algorithm	Identify novel VDJ gene alleles	(89)

discovery using sequencing data relies on these antigen-specific clones with stereotyped features in the post-infection repertoires. These features are assessed at different levels such as gene rearrangement, identical or similar CDR3 sequence overlap, and certain CDR3s length.

After influenza H1N1 vaccination, the dominant clonotype of Ig heavy chains has the same V–J gene rearrangement, CDR3 length, and somatic mutation position in CDR1 and CDR3 with previously reported influenza antibodies (66). However, in this study, the convergent dominant sequence is only found in one individual. Further researches in a broader population including non-dominant sequences are needed. A more successful example is reported in Ig repertoires related to dengue virus infection. Using cross validation and other approaches, stereotyped CDR3 sequences or CDR3 lengths that have high prevalence in the acute dengue samples are found to be specific to acute dengue infection, which are absent or of low prevalence in healthy and post-convalescent population (88).

Identification of pathogen-specific sequences also helps in differential diagnosis between infectious and non-infectious diseases. According to Dziubianau et al., comparing PBMCs-derived T cell clonotypes specific to a given virus with T cells from different origins (allograft-derived and urine-derived lymphocytes) provides a new methodology for differential diagnosis of two post-transplant complications – BKV-associated nephropathy and acute cellular rejection, which shows a glimpse of applications of T cell sequencing in diagnosis (99). In addition, a recent study searched sequencing data for CDR3 amino acid motifs that have been reported to be specific for a particular pathogen and succeeded in identifying CDR3 sequences identical or similar to these motifs in post-vaccination volunteers (100). According to

these results, interestingly, it is low frequency sequences that possess the probability of becoming promising biomarkers instead of the dominant ones.

However, immune responses show dramatic differences in CDR3 sequences responding to same pathogens across individuals and age groups. This intrinsic divergence between individuals is the major obstacle in finding “public” sequences as optimal biomarkers. Nevertheless, we hold a promise for the application of HTS data in differential diagnosis because it provides a large number of candidate sequences for biomarker investigation. Instead of using single biomarkers such as PSA or AFP in diagnosis, a combination panel of selected sequences may establish a pathogen-specific sequence library for diagnosis, which holds the potential of unprecedented sensitivity and specificity.

Identification of Antigen-Specific Antibodies and T Cells Based on High-Throughput Sequencing Data

Recombinant monoclonal antibodies have great potential in the treatment of specific infections. In recent years, several strategies, including phage display libraries, single B cell expression, and B cell immortalization, have been used to discover antibodies against specific antigens (101). HTS of the antibody repertoire, combined with subsequent bioinformatic tools or traditional screening tools, has facilitated the identification of antigen-specific sequences (102–105).

The methodology predicting antigen specificity completely from analysis of BCR sequences has not been possible yet, albeit it has considerable potential in immune repertoire studies. Nevertheless, there have been many efforts made in mining the HTS data for functional antibodies (106). A relatively direct method for identifying such sequences is based on the similarity of amino acid sequence to previously reported antibodies. Researchers have successfully found sequences of high identity with the broadly neutralizing antibodies and strain-specific antibodies from established antibody repertoires of patients with influenza infection or vaccination (66). Some of these sequences have proven to have neutralizing activity, validating the potential of deep sequencing-based antibody identification. Success of this work also suggests the possibility of monoclonal antibody synthesis without cell cloning for treatment (66). Furthermore, another method using the frequency rank of heavy chain and light chain sequences to predict the function of antibody sequences has been reported successful in mouse models (107).

Exciting work by Kwong and coworkers demonstrates the feasibility of identifying neutralizing active clones through bioinformatic analysis from HIV patients (108, 109). They established several steps for interrogating variants of known neutralizing antibody classes from HIV-infected patients, with or without previous knowledge if the patient had antibodies belonging to this family. First, the heavy or light chain sequences, derived from the germline IGHV or IGLV gene same as template antibody, are isolated from a new donor. Then, these sequences were compared with the germline gene for “divergence” and the template antibody sequence for “identity,” generating a contour

plot called “divergence/identity plot.” The sequences segregate into clusters in the plot, from which the high divergence and high identity sequences were selected as candidates for neutralizing antibodies. This process is called “Grid-based strategy.” Then, the germline sequence, candidate sequences, and template antibody sequences of the same class are merged to build a phylogenetic tree rooted by the template sequence, which is called “cross-donor analysis.” The heavy chain sequences from new donor clustered in the subtree of the template sequences are then expressed with template light chain to generate antibodies. Most of these sequences have neutralizing activity to HIV-1. The outstanding efficiency of this method was also demonstrated when compared to sequence-based strategy and prevalence-based strategy (108, 109). Lu et al. also validated this phylogenetic method in identifying functional anti-*Staphylococcus aureus* antibodies (110). These strategies start from previously reported antibody sequences. However, such antibody sequences are not always available, especially during poorly characterized viral infections such as H7N9.

Pairing the heavy and light chains as an integrated antibody has been another challenge for HTS-based immune repertoire analysis. In most cases, researchers only focus on the heavy chain, which causes a critical loss of antibody integrity and leads to problems in following synthesis of artificial monoclonal antibody. Two strategies have been reported to correctly pair the heavy chain and light chain sequences based on the frequency or evolution models. Reddy and colleagues (107) have pioneered pairing based on the frequency ranks, using plasma cells isolated from bone marrow of immunized mice and matching the two chains of similar rank order. Monoclonal antibodies expressed in this way did show antigen specificity. Due to the linkage of heavy chain and light chain as an integrated protein, their evolution undergoes the same enzymatic mutation process, and they evolve together to bind the same antigen with high affinity. Based on this theoretical foundation, phylogenetic analysis has been used as another method to compare the evolutionary topography of the heavy chain and light chain after bioinformatic identification of transcripts related to a known HIV neutralizing antibody (109, 111). Reconstituted novel antibodies consist of phylogenetically matched chains showing similar neutralizing function but less auto-reactivity compared to the mismatched ones.

Several groups have recently achieved advances in the technology of paired sequencing of antibodies. Single-cell PCR has been utilized to create a two-dimensional bar-coded primer matrix to link two chains of the BCR (112). Using this technique, Busse and coworkers analyzed paired sequences of over 46,000 B cells in one experiment and accomplished subsequent antibody gene cloning and expression. At the same time, Turchaninova and coworkers performed pioneering research in emulsion-based technology for sequencing antibody repertoires of paired chains (113). They used water-in-oil emulsions for cell-based overlap expansion RT-PCR, although its yield was relatively low yield. Another high-throughput paired sequencing method by DeKosky et al. used micro well plates to isolate B cells and magnetic beads to capture mRNAs (114). Very recently, DeKosky’s group combined

and improved these previous techniques, and developed a cost-effective and efficient methodology to establish a more precisely paired repertoire (115).

Predicting T cell specificity based on TCR heterodimer sequence is more difficult than antibodies because of the highly variable nature of each of the components of the TCR–peptide–MHC complex (116). Due to the challenges posed by the highly variable CDR3 loop of the TCR and the complexity of predicting protein–protein interactions (117, 118), experimental functional tests for mining antigen-specific T cells might be a more fruitful approach (119).

Implementation of Immune Repertoire Analysis in Vaccine Development

Recent advances in HTS-based antibody sequencing may provide the biggest benefit for the field of vaccine development. Over the years, efforts to elicit protective immune responses to HIV by immunization have not been successful. During acute viral infections, high-affinity neutralizing antibodies develop in just weeks. However, generating effective broadly neutralizing antibodies during chronic infections, such as HIV, takes significantly longer time. Furthermore, the neutralizing power of these antibodies is often variable due to impairment of the host immune function, unusual features of Env, and co-evolution of the virus in response to the host antibody response (120, 121).

Deep sequencing analysis has identified rare variants of known HIV-neutralizing antibodies and has elucidated the ontogeny of these neutralizing antibodies (108, 109, 111, 122). These findings have cast a light on antibody guided vaccine development. In following studies, the HTS-based phylogenetic strategy greatly facilitated study in co-evolution of neutralizing antibodies and virus mutants (123). Combined with long-term follow-up studies, these results illustrate how mutations in some sites allow the virus to escape some neutralizing antibodies, and how the virus, with the help of secondary neutralizing antibodies, becomes sensitive to the neutralizing antibody (98, 124, 125). These studies suggest a promising pathway to elicit broadly neutralizing antibodies by sequential immunization with selected immunogens (123, 126). Furthermore, structure studies of the neutralizing antibody family provide candidates for future vaccine designs (127).

CONCLUSION

High-throughput sequencing has been a breakthrough technology for the study of the immune repertoire and has already had a profound effect on our knowledge of the immune systems physiology during health and disease. In particular, HTS has transformed our understanding of immune repertoire formation during infection, malignancy, and autoimmunity. Advances in this field of research will rely on progress of similar laboratory techniques. Collection of more precise sequencing data can be anticipated with consistent improvement of sequencing techniques and error correction strategies. An increasing number of researchers have shown interest in this area in recent years, and this has provided vast quantities of data that could provide answers to important existing questions. These data repositories

should be effectively utilized. Establishing a public database and collecting deep sequencing data for collective collaboration will facilitate information exchange and the investigation of the varieties of repertoires across gender, age, race, and healthy state. In addition, development of standardized bioinformatic tools will be indispensable for harnessing HTS output.

Although the expected potential of immune repertoire studies in clinical use is enormous, more work remains to be done to incorporate the observed dynamic changes and sequence signatures with clinical features and outcomes. Questions remain about how the severity of certain infections is related to alterations of the immune repertoire response and various manifestations of CDR3 sequences, and how to predict abundance of protective immunoglobulins or T cell from a given sequence library. In terms of therapeutic discoveries, identification and production of functional antibodies and T cells will promote the development of passive immune therapies and vaccines. Traditional and recently reported large-scale screening strategies may contribute greatly to this process. Advances in HTS of the immune repertoire during health and disease will provide comprehensive views of the adaptive immune response in very near future and will open the door to more rationale immunotherapy for infection.

REFERENCES

- Ehlich A, Martin V, Muller W, Rajewsky K. Analysis of the B-cell progenitor compartment at the level of single cells. *Curr Biol* (1994) 4(7):573–83. doi:10.1016/S0960-9822(00)00129-9
- Wilson PC, Andrews SF. Tools to therapeutically harness the human antibody response. *Nat Rev Immunol* (2012) 12(10):709–19. doi:10.1038/nri3285
- Woodsworth DJ, Castellarin M, Holt RA. Sequence analysis of T-cell repertoires in health and disease. *Genome Med* (2013) 5(10):98. doi:10.1186/gm502
- Gearhart PJ. Immunology: the roots of antibody diversity. *Nature* (2002) 419(6902):29–31. doi:10.1038/419029a
- Pannetier C, Cochet M, Darche S, Casrouge A, Zoller M, Kourilsky P. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. *Proc Natl Acad Sci U S A* (1993) 90(9):4319–23. doi:10.1073/pnas.90.9.4319
- Tonegawa S. Somatic generation of antibody diversity. *Nature* (1983) 302(5909):575–81. doi:10.1038/302575a0
- van Stipdonk MJ, Hardenberg G, Bijker MS, Lemmens EE, Droin NM, Green DR, et al. Dynamic programming of CD8+ T lymphocyte responses. *Nat Immunol* (2003) 4(4):361–5. doi:10.1038/ni912
- Gett AV, Hodgkin PD. A cellular calculus for signal integration by T cells. *Nat Immunol* (2000) 1(3):239–44. doi:10.1038/79782
- Alexander-Miller MA, Leggett GR, Berzofsky JA. Selective expansion of high- or low-avidity cytotoxic T lymphocytes and efficacy for adoptive immunotherapy. *Proc Natl Acad Sci U S A* (1996) 93(9):4102–7. doi:10.1073/pnas.93.9.4102
- Busch DH, Pamer EG. T cell affinity maturation by selective expansion during infection. *J Exp Med* (1999) 189(4):701–10. doi:10.1084/jem.189.4.701
- Foulds KE, Zenewicz LA, Shedlock DJ, Jiang J, Troy AE, Shen H. Cutting edge: CD4 and CD8 T cells are intrinsically different in their proliferative responses. *J Immunol* (2002) 168(4):1528–32. doi:10.4049/jimmunol.168.4.1528
- Rabenstein H, Behrendt AC, Ellwart JW, Naumann R, Horsch M, Beckers J, et al. Differential kinetics of antigen dependency of CD4+ and CD8+ T cells. *J Immunol* (2014) 192(8):3507–17. doi:10.4049/jimmunol.1302725

AUTHOR CONTRIBUTIONS

DH was involved in study design, wrote the first draft of the manuscript, conducted the literature search, reviewed the abstracts, performed the analysis, and contributed to the final draft; CC and SC were involved in study design and reviewed the abstracts. ES revised the manuscript. YS designed and supervised the study, revised the final draft, and contributed to the analysis. All authors have read and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (81170056, 81490533, 81100046, and 81570028) and by grant B115 from Shanghai Leading Academic Discipline Project. YS was supported by the State Key Basic Research Program (973) project (2015CB553404), Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning and Key Medical grant from Shanghai Science and Technology Committee (11411951102 and 12JC1402300), and supported by Doctoral Fund of Ministry of Education of China (20130071110044).

- Six A, Mariotti-Ferrandiz ME, Chaara W, Magadan S, Pham HP, Lefranc MP, et al. The past, present, and future of immune repertoire biology – the rise of next-generation repertoire analysis. *Front Immunol* (2013) 4:413. doi:10.3389/fimmu.2013.00413
- Gorski J, Yassai M, Zhu X, Kissella B, Kissella B, Keever C, et al. Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR3 size spectratyping. Correlation with immune status. *J Immunol* (1994) 152(10):5109–19.
- Pannetier C, Even J, Kourilsky P. T-cell repertoire diversity and clonal expansions in normal and clinical samples. *Immunol Today* (1995) 16(4):176–81. doi:10.1016/0167-5699(95)80117-0
- Balamurugan A, Ng HL, Yang OO. Rapid T cell receptor delineation reveals clonal expansion limitation of the magnitude of the HIV-1-specific CD8+ T cell response. *J Immunol* (2010) 185(10):5935–42. doi:10.4049/jimmunol.1002236
- Musette P, Bureau JF, Gachelin G, Kourilsky P, Brahic M. T lymphocyte repertoire in Theiler's virus encephalomyelitis: the nonspecific infiltration of the central nervous system of infected SJL/J mice is associated with a selective local T cell expansion. *Eur J Immunol* (1995) 25(6):1589–93. doi:10.1002/eji.1830250618
- Sourdive DJ, Murali-Krishna K, Altman JD, Zajac AJ, Whitmire JK, Pannetier C, et al. Conserved T cell receptor repertoire in primary and memory CD8 T cell responses to an acute viral infection. *J Exp Med* (1998) 188(1):71–82. doi:10.1084/jem.188.1.71
- Ademokun A, Wu YC, Martin V, Mitra R, Sack U, Baxendale H, et al. Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages. *Ageing Cell* (2011) 10(6):922–30. doi:10.1111/j.1474-9726.2011.00732.x
- Collette A, Bagot S, Ferrandiz ME, Cazenave PA, Six A, Pied S. A profound alteration of blood TCRB repertoire allows prediction of cerebral malaria. *J Immunol* (2004) 173(7):4568–75. doi:10.4049/jimmunol.173.7.4568
- Klein U, Rajewsky K, Kuppers R. Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. *J Exp Med* (1998) 188(9):1679–89. doi:10.1084/jem.188.9.1679

22. Kupperts R, Zhao M, Hansmann ML, Rajewsky K. Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. *EMBO J* (1993) 12(13):4955–67.
23. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun* (2013) 4:2680. doi:10.1038/ncomms3680
24. He L, Sok D, Azadnia P, Hsueh J, Landais E, Simek M, et al. Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. *Sci Rep* (2014) 4:6778. doi:10.1038/srep06778
25. Nguyen P, Ma J, Pei D, Obert C, Cheng C, Geiger TL. Identification of errors introduced during high throughput sequencing of the T cell receptor repertoire. *BMC Genomics* (2011) 12:106. doi:10.1186/1471-2164-12-106
26. Yang X, Chockalingam SP, Aluru S. A survey of error-correction methods for next-generation sequencing. *Brief Bioinform* (2013) 14(1):56–66. doi:10.1093/bib/bbs015
27. Laehnemann D, Borkhardt A, McHardy AC. Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Brief Bioinform* (2016) 17(1):154–79. doi:10.1093/bib/bbv029
28. Bolotin DA, Mamedov IZ, Britanova OV, Zvyagin IV, Shagin D, Ustyugova SV, et al. Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. *Eur J Immunol* (2012) 42(11):3073–83. doi:10.1002/eji.201242517
29. Gilles A, Meglecz E, Pech N, Ferreira S, Malausa T, Martin JF. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics* (2011) 12:245. doi:10.1186/1471-2164-12-245
30. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* (2012) 13:341. doi:10.1186/1471-2164-13-341
31. Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol* (2011) 12(11):R112. doi:10.1186/gb-2011-12-11-r112
32. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. *Genome Biol* (2013) 14(5):R51. doi:10.1186/gb-2013-14-5-r51
33. Li H. BFC: correcting Illumina sequencing errors. *Bioinformatics* (2015) 31(17):2885–7. doi:10.1093/bioinformatics/btv290
34. Ilie L, Fazayeli F, Ilie S. HiTEC: accurate error correction in high-throughput sequencing data. *Bioinformatics* (2011) 27(3):295–302. doi:10.1093/bioinformatics/btq653
35. Song L, Florea L, Langmead B. Lighter: fast and memory-efficient sequencing error correction without counting. *Genome Biol* (2014) 15(11):509. doi:10.1186/PREACCEPT-9663167051308943
36. Yang X, Dorman KS, Aluru S. Reptile: representative tiling for short read error correction. *Bioinformatics* (2010) 26(20):2526–33. doi:10.1093/bioinformatics/btq468
37. Kao WC, Chan AH, Song YS. ECHO: a reference-free short-read error correction algorithm. *Genome Res* (2011) 21(7):1181–92. doi:10.1101/gr.111351.110
38. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, et al. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* (2009) 6(9):639–41. doi:10.1038/nmeth.1361
39. Reeder J, Knight R. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat Methods* (2010) 7(9):668–9. doi:10.1038/nmeth0910-668b
40. Wirawan A, Harris RS, Liu Y, Schmidt B, Schroder J. HECTOR: a parallel multistage homopolymer spectrum based error corrector for 454 sequencing data. *BMC Bioinformatics* (2014) 15:131. doi:10.1186/1471-2105-15-131
41. Vollmers C, Sit RV, Weinstein JA, Dekker CL, Quake SR. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc Natl Acad Sci U S A* (2013) 110(33):13463–8. doi:10.1073/pnas.1312146110
42. Shugay M, Britanova OV, Merzlyak EM, Turchaninova MA, Mamedov IZ, Tuganbaev TR, et al. Towards error-free profiling of immune repertoires. *Nat Methods* (2014) 11(6):653–5. doi:10.1038/nmeth.2960
43. Lefranc MP. IMGT, the international ImMunoGeneTics database(R). *Nucleic Acids Res* (2003) 31(1):307–10. doi:10.1093/nar/gkg085
44. Lee CE, Gaeta B, Malming HR, Bain ME, Sewell WA, Collins AM. Reconsidering the human immunoglobulin heavy-chain locus: I. An evaluation of the expressed human IGHD gene repertoire. *Immunogenetics* (2006) 57(12):917–25. doi:10.1007/s00251-005-0062-5
45. Wang Y, Jackson KJ, Sewell WA, Collins AM. Many human immunoglobulin heavy-chain IGHV gene polymorphisms have been reported in error. *Immunol Cell Biol* (2008) 86(2):111–5. doi:10.1038/sj.icb.7100144
46. Collins AM, Wang Y, Singh V, Yu P, Jackson KJ, Sewell WA. The reported germline repertoire of human immunoglobulin kappa chain genes is relatively complete and accurate. *Immunogenetics* (2008) 60(11):669–76. doi:10.1007/s00251-008-0325-z
47. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* (2013) 41(Web Server issue):W34–40. doi:10.1093/nar/gkt382
48. Gaeta BA, Malming HR, Jackson KJ, Bain ME, Wilson P, Collins AM. iHMMune-align: hidden Markov model-based alignment and identification of germline genes in rearranged immunoglobulin gene sequences. *Bioinformatics* (2007) 23(13):1580–7. doi:10.1093/bioinformatics/btm147
49. Thomas N, Heather J, Ndifon W, Shawe-Taylor J, Chain B. Decombinator: a tool for fast, efficient gene assignment in T-cell receptor sequences using a finite state machine. *Bioinformatics* (2013) 29(5):542–50. doi:10.1093/bioinformatics/btt004
50. Munshaw S, Kepler TB. SoDA2: a hidden Markov model approach for identification of immunoglobulin rearrangements. *Bioinformatics* (2010) 26(7):867–72. doi:10.1093/bioinformatics/btq056
51. Rempala GA, Seweryn M, Ignatowicz L. Model for comparative analysis of antigen receptor repertoires. *J Theor Biol* (2011) 269(1):1–15. doi:10.1016/j.jtbi.2010.10.001
52. Eisenstein M. Personalized, sequencing-based immune profiling spurs startups. *Nat Biotechnol* (2013) 31(3):184–6. doi:10.1038/nbt0313-184b
53. Mamedov IZ, Britanova OV, Zvyagin IV, Turchaninova MA, Bolotin DA, Putintseva EV, et al. Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Front Immunol* (2013) 4:456. doi:10.3389/fimmu.2013.00456
54. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* (2015) 31(20):3356–8. doi:10.1093/bioinformatics/btv359
55. Schaller S, Weinberger J, Jimenez-Heredia R, Danzer M, Oberbauer R, Gabriel C, et al. ImmunExplorer (IMEX): a software framework for diversity and clonality analyses of immunoglobulins and T cell receptors on the basis of IMGT/HighV-QUEST preprocessed NGS data. *BMC Bioinformatics* (2015) 16:252. doi:10.1186/s12859-015-0687-9
56. Yu Y, Ceredig R, Seoighe C. LymAnalyzer: a tool for comprehensive analysis of next generation sequencing data of T cell receptors and immunoglobulins. *Nucleic Acids Res* (2016) 44(4):e31. doi:10.1093/nar/gkv1016
57. Nazarov VI, Pogorelyy MV, Komech EA, Zvyagin IV, Bolotin DA, Shugay M, et al. tCR: an R package for T cell receptor repertoire advanced data analysis. *BMC Bioinformatics* (2015) 16:175. doi:10.1186/s12859-015-0613-1
58. Bolotin DA, Shugay M, Mamedov IZ, Putintseva EV, Turchaninova MA, Zvyagin IV, et al. MiTCR: software for T-cell receptor sequencing data analysis. *Nat Methods* (2013) 10(9):813–4. doi:10.1038/nmeth.2555
59. Zhang W, Du Y, Su Z, Wang C, Zeng X, Zhang R, et al. IMonitor: a robust pipeline for TCR and BCR repertoire analysis. *Genetics* (2015) 201(2):459–72. doi:10.1534/genetics.115.176735
60. Imkeller K, Arndt PF, Wardemann H, Busse CE. sciReceptor: analysis of single-cell level immunoglobulin repertoires. *BMC Bioinformatics* (2016) 17(1):67. doi:10.1186/s12859-016-0920-1
61. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, et al. VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput Biol* (2015) 11(11):e1004503. doi:10.1371/journal.pcbi.1004503
62. Calis JJ, Rosenberg BR. Characterizing immune repertoires by high throughput sequencing: strategies and applications. *Trends Immunol* (2014) 35(12):581–90. doi:10.1016/j.it.2014.09.004
63. Ying T, Prabakaran P, Du L, Shi W, Feng Y, Wang Y, et al. Junctional and allele-specific residues are critical for MERS-CoV neutralization by an exceptionally potent germline-like antibody. *Nat Commun* (2015) 6:8223. doi:10.1038/ncomms9223

64. Cortina-Ceballos B, Godoy-Lozano EE, Tellez-Sosa J, Ovilla-Munoz M, Samano-Sanchez H, Aguilar-Salgado A, et al. Longitudinal analysis of the peripheral B cell repertoire reveals unique effects of immunization with a new influenza virus strain. *Genome Med* (2015) 7:124. doi:10.1186/s13073-015-0239-y
65. Giudicelli V, Chaume D, Lefranc MP. IMG2/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. *Nucleic Acids Res* (2004) 32(Web Server issue):W435–40. doi:10.1093/nar/gkh412
66. Jackson KJ, Liu Y, Roskin KM, Glanville J, Hoh RA, Seo K, et al. Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell Host Microbe* (2014) 16(1):105–14. doi:10.1016/j.chom.2014.05.013
67. Wang C, Liu Y, Xu LT, Jackson KJ, Roskin KM, Pham TD, et al. Effects of aging, cytomegalovirus infection, and EBV infection on human B cell repertoires. *J Immunol* (2014) 192(2):603–11. doi:10.4049/jimmunol.1301384
68. Wiehe K, Easterhoff D, Luo K, Nicely NI, Bradley T, Jaeger FH, et al. Antibody light-chain-restricted recognition of the site of immune pressure in the RV144 HIV-1 vaccine trial is phylogenetically conserved. *Immunity* (2014) 41(6):909–18. doi:10.1016/j.immuni.2014.11.014
69. Godoy-Lozano EE, Tellez-Sosa J, Sanchez-Gonzalez G, Samano-Sanchez H, Aguilar-Salgado A, Salinas-Rodriguez A, et al. Lower IgG somatic hypermutation rates during acute dengue virus infection is compatible with a germinal center-independent B cell response. *Genome Med* (2016) 8(1):23. doi:10.1186/s13073-016-0276-1
70. Phad GE, Vazquez Bernat N, Feng Y, Ingale J, Martinez Murillo PA, O'Dell S, et al. Diverse antibody genetic and recognition properties revealed following HIV-1 envelope glycoprotein immunization. *J Immunol* (2015) 194(12):5903–14. doi:10.4049/jimmunol.1500122
71. Thomas N, Best K, Cinelli M, Reich-Zeliger S, Gal H, Shifrut E, et al. Tracking global changes induced in the CD4 T-cell receptor repertoire by immunization with a complex antigen using short stretches of CDR3 protein sequence. *Bioinformatics* (2014) 30(22):3181–8. doi:10.1093/Bioinformatics/Btu523
72. Heather JM, Best K, Oakes T, Gray ER, Roe JK, Thomas N, et al. Dynamic perturbations of the T-cell receptor repertoire in chronic HIV infection and following antiretroviral therapy. *Front Immunol* (2015) 6:644. doi:10.3389/fimmu.2015.00644
73. Tsioris K, Gupta NT, Ogunniyi AO, Zimmisky RM, Qian F, Yao Y, et al. Neutralizing antibodies against West Nile virus identified directly from human B cells by single-cell analysis and next generation sequencing. *Integr Biol (Camb)* (2015) 7(12):1587–97. doi:10.1039/c5ib00169b
74. Ruggiero E, Nicolay JP, Fronza R, Arens A, Paruzynski A, Nowrouzi A, et al. High-resolution analysis of the human T-cell receptor repertoire. *Nat Commun* (2015) 6:8081. doi:10.1038/ncomms9081
75. Gadala-Maria D, Yaari G, Uduman M, Kleinstein SH. Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *Proc Natl Acad Sci U S A* (2015) 112(8):E862–70. doi:10.1073/pnas.1417683112
76. Mora T, Walczak AM, Bialek W, Callan CG Jr. Maximum entropy models for antibody diversity. *Proc Natl Acad Sci U S A* (2010) 107(12):5405–10. doi:10.1073/pnas.1001705107
77. Weinstein JA, Jiang N, White RA III, Fisher DS, Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. *Science* (2009) 324(5928):807–10. doi:10.1126/science.1170020
78. Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, et al. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc Natl Acad Sci U S A* (2009) 106(48):20216–21. doi:10.1073/pnas.0909775106
79. Galson JD, Truck J, Fowler A, Munz M, Cerundolo V, Pollard AJ, et al. In-depth assessment of within-individual and inter-individual variation in the B cell receptor repertoire. *Front Immunol* (2015) 6:531. doi:10.3389/fimmu.2015.00531
80. Jost L. Partitioning diversity into independent alpha and beta components. *Ecology* (2007) 88(10):2427–39. doi:10.1890/06-1736.1
81. Eren MI, Chao A, Hwang WH, Colwell RK. Estimating the richness of a population when the maximum number of classes is fixed: a nonparametric solution to an archaeological problem. *PLoS One* (2012) 7(5):e34179. doi:10.1371/journal.pone.0034179
82. Pacholczyk R, Ignatowicz H, Kraj P, Ignatowicz L. Origin and T cell receptor diversity of Foxp3+CD4+CD25+ T cells. *Immunity* (2006) 25(2):249–59. doi:10.1016/j.immuni.2006.05.016
83. Chao A, Bunge J. Estimating the number of species in a stochastic abundance model. *Biometrics* (2002) 58(3):531–9. doi:10.1111/j.0006-341X.2002.00531.x
84. Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee JY, et al. Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A* (2014) 111(36):13139–44. doi:10.1073/pnas.1409155111
85. Greene J, Birtwistle MR, Ignatowicz L, Rempala GA. Bayesian multivariate Poisson abundance models for T-cell receptor data. *J Theor Biol* (2013) 326:1–10. doi:10.1016/j.jtbi.2013.02.009
86. Sepulveda N, Paulino CD, Carneiro J. Estimation of T-cell repertoire diversity and clonal size distribution by Poisson abundance models. *J Immunol Methods* (2010) 353(1–2):124–37. doi:10.1016/j.jim.2009.11.009
87. Venturi V, Kedzierska K, Tanaka MM, Turner SJ, Doherty PC, Davenport MP. Method for assessing the similarity between subsets of the T cell receptor repertoire. *J Immunol Methods* (2008) 329(1–2):67–80. doi:10.1016/j.jim.2007.09.016
88. Parameswaran P, Liu Y, Roskin KM, Jackson KK, Dixit VP, Lee JY, et al. Convergent antibody signatures in human dengue. *Cell Host Microbe* (2013) 13(6):691–700. doi:10.1016/j.chom.2013.05.008
89. Yu X, Sun S. Comparing a few SNP calling algorithms using low-coverage sequencing data. *BMC Bioinformatics* (2013) 14:274. doi:10.1186/1471-2105-14-274
90. Laserson U, Vigneault F, Gadala-Maria D, Yaari G, Uduman M, Vander Heiden JA, et al. High-resolution antibody dynamics of vaccine-induced immune responses. *Proc Natl Acad Sci USA* (2014) 111(13):4928–33. doi:10.1073/pnas.1323862111
91. Conrad JA, Ramalingam RK, Duncan CB, Smith RM, Wei J, Barnett L, et al. Antiretroviral therapy reduces the magnitude and T cell receptor repertoire diversity of HIV-specific T cell responses without changing T cell clonotype dominance. *J Virol* (2012) 86(8):4213–21. doi:10.1128/JVI.06000-11
92. Costa AI, Koning D, Ladell K, McLaren JE, Grady BP, Schellens IM, et al. Complex T-cell receptor repertoire dynamics underlie the CD8+ T-cell response to HIV-1. *J Virol* (2015) 89(1):110–9. doi:10.1128/JVI.01765-14
93. Zhu J, Peng T, Johnston C, Phasouk K, Kask AS, Klock A, et al. Immune surveillance by CD8alphaalpha+ skin-resident T cells in human herpes virus infection. *Nature* (2013) 497(7450):494–7. doi:10.1038/nature12110
94. Galson JD, Pollard AJ, Truck J, Kelly DF. Studying the antibody repertoire after vaccination: practical applications. *Trends Immunol* (2014) 35(7):319–31. doi:10.1016/j.it.2014.04.005
95. Jiang N, He J, Weinstein JA, Penland L, Sasaki S, He XS, et al. Lineage structure of the human antibody repertoire in response to influenza vaccination. *Sci Transl Med* (2013) 5(171):171ra19. doi:10.1126/scitranslmed.3004794
96. Lavinder JJ, Wine Y, Giesecke C, Ippolito GC, Horton AP, Lungu OI, et al. Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proc Natl Acad Sci U S A* (2014) 111(6):2259–64. doi:10.1073/pnas.1317793111
97. Wiley SR, Raman VS, Desbien A, Bailor HR, Bhardwaj R, Shakri AR, et al. Targeting TLRs expands the antibody repertoire in response to a malaria vaccine. *Sci Transl Med* (2011) 3(93):93ra69. doi:10.1126/scitranslmed.3002135
98. Robinson WH. Sequencing the functional antibody repertoire – diagnostic and therapeutic discovery. *Nat Rev Rheumatol* (2015) 11(3):171–82. doi:10.1038/nrrheum.2014.220
99. Dziubianau M, Hecht J, Kuchenbecker L, Sattler A, Stervbo U, Rodelsperger C, et al. TCR repertoire analysis by next generation sequencing allows complex differential diagnosis of T cell-related pathology. *Am J Transplant* (2013) 13(11):2842–54. doi:10.1111/ajt.12431
100. Truck J, Ramasamy MN, Galson JD, Rance R, Parkhill J, Lunter G, et al. Identification of antigen-specific B cell receptor sequences using public repertoire analysis. *J Immunol* (2015) 194(1):252–61. doi:10.4049/jimmunol.1401405
101. Hoogenboom HR. Selecting and screening recombinant antibody libraries. *Nat Biotechnol* (2005) 23(9):1105–16. doi:10.1038/nbt1126
102. Zhai W, Glanville J, Fuhrmann M, Mei L, Ni I, Sundar PD, et al. Synthetic antibodies designed on natural sequence landscapes. *J Mol Biol* (2011) 412(1):55–71. doi:10.1016/j.jmb.2011.07.018

103. Larman HB, Xu GJ, Pavlova NN, Elledge SJ. Construction of a rationally designed antibody platform for sequencing-assisted selection. *Proc Natl Acad Sci U S A* (2012) 109(45):18523–8. doi:10.1073/pnas.1215549109
104. Larimore K, McCormick MW, Robins HS, Greenberg PD. Shaping of human germline IgH repertoires revealed by deep sequencing. *J Immunol* (2012) 189(6):3221–30. doi:10.4049/jimmunol.1201303
105. Ravn U, Didelot G, Venet S, Ng KT, Gueneau F, Rousseau F, et al. Deep sequencing of phage display libraries to support antibody discovery. *Methods* (2013) 60(1):99–110. doi:10.1016/j.ymeth.2013.03.001
106. Mathonet P, Ullman CG. The application of next generation sequencing to the understanding of antibody repertoires. *Front Immunol* (2013) 4:265. doi:10.3389/fimmu.2013.00265
107. Reddy ST, Ge X, Miklos AE, Hughes RA, Kang SH, Hoi KH, et al. Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat Biotechnol* (2010) 28(9):965–9. doi:10.1038/nbt.1673
108. Zhu J, Wu X, Zhang B, McKee K, O'Dell S, Soto C, et al. De novo identification of VRC01 class HIV-1-neutralizing antibodies by next-generation sequencing of B-cell transcripts. *Proc Natl Acad Sci U S A* (2013) 110(43):E4088–97. doi:10.1073/pnas.1306262110
109. Zhu J, Ofek G, Yang Y, Zhang B, Louder MK, Lu G, et al. Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proc Natl Acad Sci U S A* (2013) 110(16):6470–5. doi:10.1073/pnas.1219320110
110. Lu DR, Tan YC, Kongpachith S, Cai X, Stein EA, Lindstrom TM, et al. Identifying functional anti-*Staphylococcus aureus* antibodies by sequencing antibody repertoires of patient plasmablasts. *Clin Immunol* (2014) 152(1–2):77–89. doi:10.1016/j.clim.2014.02.010
111. Zhu J, O'Dell S, Ofek G, Pancera M, Wu X, Zhang B, et al. Somatic populations of PGT135-137 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinformatics. *Front Microbiol* (2012) 3:315. doi:10.3389/fmicb.2012.00315
112. Busse CE, Czogiel I, Braun P, Arndt PF, Wardemann H. Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes. *Eur J Immunol* (2014) 44(2):597–603. doi:10.1002/eji.201343917
113. Turchaninova MA, Britanova OV, Bolotin DA, Shugay M, Putintseva EV, Staroverov DB, et al. Pairing of T-cell receptor chains via emulsion PCR. *Eur J Immunol* (2013) 43(9):2507–15. doi:10.1002/eji.201343453
114. DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM, et al. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* (2013) 31(2):166–9. doi:10.1038/nbt.2492
115. DeKosky BJ, Kojima T, Rodin A, Charab W, Ippolito GC, Ellington AD, et al. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat Med* (2015) 21(1):86–91. doi:10.1038/nm.3743
116. Newell EW, Davis MM. Beyond model antigens: high-dimensional methods for the analysis of antigen-specific T cells. *Nat Biotechnol* (2014) 32(2):149–57. doi:10.1038/nbt.2783
117. Reiser JB, Darnault C, Gregoire C, Mosser T, Mazza G, Kearney A, et al. CDR3 loop flexibility contributes to the degeneracy of TCR recognition. *Nat Immunol* (2003) 4(3):241–7. doi:10.1038/ni891
118. Birnbaum ME, Mendoza JL, Sethi DK, Dong S, Glanville J, Dobbins J, et al. Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* (2014) 157(5):1073–87. doi:10.1016/j.cell.2014.03.047
119. Pan X, Huang LC, Dong T, Peng Y, Cerundolo V, McGowan S, et al. Combinatorial HLA-peptide bead libraries for high throughput identification of CD8(+) T cell specificity. *J Immunol Methods* (2014) 403(1–2):72–8. doi:10.1016/j.jim.2013.11.023
120. Kwong PD, Mascola JR, Nabel GJ. Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. *Nat Rev Immunol* (2013) 13(9):693–701. doi:10.1038/nri3516
121. Kwong PD, Mascola JR. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* (2012) 37(3):412–25. doi:10.1016/j.immuni.2012.08.012
122. Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, et al. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* (2011) 333(6049):1593–602. doi:10.1126/science.1207532
123. Liao HX, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, et al. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* (2013) 496(7446):469–76. doi:10.1038/nature12053
124. Gao F, Bonsignori M, Liao HX, Kumar A, Xia SM, Lu X, et al. Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. *Cell* (2014) 158(3):481–91. doi:10.1016/j.cell.2014.06.022
125. Wu X, Zhang Z, Schramm CA, Joyce MG, Do Kwon Y, Zhou T, et al. Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. *Cell* (2015) 161(3):470–85. doi:10.1016/j.cell.2015.03.004
126. Haynes BF, Kelsoe G, Harrison SC, Kepler TB. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nat Biotechnol* (2012) 30(5):423–33. doi:10.1038/nbt.2197
127. Zhou T, Lynch RM, Chen L, Acharya P, Wu X, Doria-Rose NA, et al. Structural repertoire of HIV-1-neutralizing antibodies targeting the CD4 supersite in 14 donors. *Cell* (2015) 161(6):1280–92. doi:10.1016/j.cell.2015.05.007

Conflict of Interest Statement: We have no financial or personal relationships with other people or organizations that can inappropriately influence our work; there is no professional or other personal interest of any nature or kind in any product, service, and/or company that could be construed as influencing the position presented in this article.

Copyright © 2016 Hou, Chen, Seely, Chen and Song. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.