REVIEW

History, applications, and challenges of immune repertoire research

Xiao Liu · Jinghua Wu

Received: 14 April 2017 / Accepted: 14 February 2018 / Published online: 27 February 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract The diversity of T and B cells in terms of their receptor sequences is huge in the vertebrate's immune system and provides broad protection against the vast diversity of pathogens. Immune repertoire is defined as the sum of T cell receptors and B cell receptors (also named immunoglobulin) that makes the organism's adaptive immune system. Before the emergence of high-throughput sequencing, the studies on immune repertoire were limited by the underdeveloped methodologies, since it was impossible to capture the whole picture by the low-throughput tools. The massive paralleled sequencing technology suits perfectly the researches on immune repertoire. In this article, we review the history of immune repertoire studies, in terms of technologies and research applications. Particularly, we discuss several aspects of challenges in this field and highlight the efforts to develop potential solutions, in the era of high-throughput sequencing of the immune repertoire.

Keywords Immune repertoire · TCR · BCR · High-throughput sequencing

The normal function of adaptive immunity relies on the generation of diverse T cell receptors (TCRs) that are expressed on the cell membrane of T lymphocytes, and B cell receptors (BCRs), also known as immunoglobulin

(Ig) or antibody in the secreting form, which are expressed by B lymphocytes. The diversified TCRs can recognize and bind to the various epitopes of the antigens presented by the major histocompatibility complex (MHC) molecule of the antigen presenting cells and form a TCR-antigen-MHC structure complex. On the other hand, antibodies can bind to and neutralize the antigens directly. Historically, scientific reports on the TCR and BCR repertoire are unlikely to grasp the whole picture due to the unavailability of high-throughput and sensitive tools. The emergence of high-throughput sequencing (HTS) technique enables scientists to investigate the repertoire in an unprecedented level. However, challenges still remain both in the experimental pipeline and the analytical tools to handle the huge amount of data, and various potential solutions are being developed to fulfill the comprehensive application and translation of this technique.

The generation of TCR and antibody diversity

The adaptive immunity consists of two major components, cellular immunity and humoral immunity, which are highly specific to particular pathogens. T and B lymphocytes are two prolific and functionally important cells that play pivotal roles. A powerful adaptive immune system relies on the generation of diversified TCRs and antibodies to recognize the enormous variety of antigens. In addition, pressures from both the internal and the outside environment, such as pathogen exposure, tumor formation, and autoimmune reaction, could affect the diversity of TCR and antibody by expanding

X. Liu (✉) · J. Wu
BGI-Shenzhen, Shenzhen 518083, China
e-mail: liuxiao@genomics.cn

the “pressure”-associated T/B cells. Therefore, the diversity of TCR and antibody repertoire could reflect the adaptive immune status to some extent. Bearing the significant functions in the immune system, T and B lymphocytes are derived, differentiated, and mature in different organs and tissues. Both T and B cells originate from the pluripotent stem cells in the bone marrow. The precursor of T cells migrate through peripheral blood to the thymus, where they finish the differentiation and the maturation, while the maturation of B cells from the progenitors occurs mostly in the bone marrow.

Generation of antibody diversity

Antibody, also known as immunoglobulin (Ig), is the molecule produced by B cell that can specifically recognize and neutralize a variety of antigens. Antibodies are encoded by a very unique protein-coding system, in which somatic rearrangements happen in the coding genes. Antibody is a Y-shaped protein composed of two identical large heavy chains (encoded by the gene IGH) and two identical small light chains (λ or κ) (encoded by the gene IGL or IGK) (Fig. 1a). The variable region diversity of BCR determines the diversity and specificity of antigen recognition. The human IGH, IGL, and IGK genes locate on the chromosomes 14, 22, and 2, respectively. The immunoglobulin heavy chain variable region is encoded by three gene families, including variable (V), Diversity (D), and Joining (J) genes, which contain numerous gene copies with high sequence homology (Fig. 1b). For light chains, only V and J gene families are involved in recombination. The number of functional genes of each gene family for human and mouse collected by IMGT (<http://www.imgt.org/>) is listed in Table 1. During the process of B cell development and maturation, the variable coding genes of BCR undergo a somatic recombination procedure called V(D)J recombination (rearrangement) (Fig. 1b). During this process, randomly chosen V, (D), and J gene segments are brought into contiguity mediated by the coordinated activities of several enzymes, such as recombination-activating genes (RAG1/RAG2) and terminal deoxynucleotidyl transferase (TdT) (Gauss and Lieber 1996; Schatz and Swanson 2011), and the rearranged variable domain includes four framework regions (FR) and three complementarity-determining regions (CDR) (Fig. 1b). The variety of V, D, and J genes that could be used to recombine provides the basis for the diversity of antibodies, and the random

nontemplated nucleotide additions and deletions in the V-D and D-J (V-J for light chain) joining regions also contribute largely to the tremendous diversity of BCR (Nezlin 2001; Rose 1982). Further diversity is introduced by somatic hypermutation (SHM), which introduces mainly single nucleotide substitutions to the rearranged coding genes of antibody. SHM happens at an extremely high rate of about 10^{-3} per base pair per division in B cells (Kleinstein et al. 2003) by the function of Activation-Induced (Cytidine) Deaminase (AID) and uracil-DNA glycosylase (Larson and Maizels 2004; Teng and Papavasiliou 2007), thus increases the BCR diversity largely (Li et al. 2004; Nezlin 2001). Beyond that, gene transcription and heavy/light chain pairing also improve the diversity of functional antibody and finally yield an astronomical number of possibilities. The human immunoglobulin repertoire can reach to 10^{13} in theory (Schroeder 2006), which exceeds the number of total B cells in human body (10^{11}) (Trepel 1974), and this vast diversity of potential antibodies enables their abilities to recognize a variety of antigens.

Generation of diversity in TCR

T cells can be divided into α/β T cells and γ/δ T cells according to their TCR genes. For human, most of the T cells are α/β TCRs, and γ/δ T cells comprise 0.5–16% of total T cells (mean approximately 4%) (Chien et al. 2014). For α/β T cell, TCR is heterodimer and composed of one alpha and one beta chain, and for γ/δ T cell, its TCR consists of one gamma and one delta chain. The human TCR alpha gene (TRA) and delta gene (TRD) locus locate on chromosome 14, and beta chain gene (TRB) and gamma chain gene (TRG) locus locate on chromosome 7. The numbers of functional gene of each TCR gene family for human and mouse collected by IMGT (<http://www.imgt.org/>) are listed in Table 1. In the development of T cells, the TCR undergoes essentially the same recombination process that assembles the BCR. The TCR locus rearrangement begins from the TCR β chain in the double negative progenitor T cell. The T cells successfully expressing the TCR β chain will turn on the expression of CD4 and CD8 on the T cell membrane surface to form the double positive cells, and at the same time, the alpha chain is rearranged and paired with the beta chain to form $\alpha\beta$ TCR (Germain 2002; Koch and Radtke 2011; Rothenberg et al. 2008). Likewise, the vast number of TCR gene segments that serve as potential candidates for somatic recombination, the junction

Fig. 1 The structure and recombination process of the antibody. **a** The structure of antibody. The antibody is composed of two identical heavy chains (generated by recombination of V, D, and J segments) and two identical light chains (generated by recombination of V and J segments). **b** Somatic VDJ rearrangement of the Immunoglobulin heavy locus (IGH). The top box is a schematic of the germline DNA in this locus with variable segments (VH), diversity segments (DH), joining segments (JH), and constant segment (CH). A V segment consists of 5' UTR, 3' UTR, L-region,

and V-region, which can be divided into six functional regions, three framework regions (FRs), and three complementarity-determining regions (CDRs). The middle line is the completed rearranged IGH gene. During somatic rearrangement, one each of the V, D, and J segments is rearranged to form a continuous VDJ segment. The bottom line is the transcribed and spliced mRNA, which can be translated into the antibody heavy chain protein in Fig. 1a

diversity, and the possibilities of alpha/beta chain pairing contribute to the diversified TCR repertoire to recognize the MHC/antigen complex.

Immune repertoire research prior to the HTS technology

The prior technologies used to investigate the immune repertoire

Considering the importance of diversification of TCR and BCR repertoire to reflect the adaptive immune status, a lot of molecular biology techniques have been developed to assess the repertoire diversity as reviewed by Six et al. (2013). Sanger sequencing was the most popular technique to read the nucleotide sequences before the emergence of HTS; however, due to its low capacity (the need to clone the sequence to a plasmid vector and one sequence per read in a reaction), it is

Table 1 The numbers of functional gene of each BCR and TCR gene family for human and mouse collected by IMGT

Genes	Human	Mouse
IGHV	55	181
IGHD	23	25
IGHJ	6	4
IGLV	32	2–9
IGLJ	5	3
IGKV	34	91
IGKJ	5	4
TRAV	45	82
TRAJ	50	40
TRBV	48	22
TRBD	2	2
TRBJ	13	11
TRDV	3	15
TRDD	3	2
TRDJ	4	2
TRGV	6	7
TRGJ	5	4

extremely difficult for Sanger sequencing to grasp the enormous sequence diversity of TCRs and BCRs. The largest throughput in the literature was to sequence hundreds of TCRs/BCRs by Sanger sequencing (Hindley et al. 2011; Pacholczyk et al. 2006). Besides, several low-resolution techniques had been exploited to visualize the diversity of TCR/BCR repertoire, such as TCR spectratyping or TCR immunoscope (Gorski et al. 1994; Pannetier et al. 1993). TCR spectratyping employed gel electrophoresis to dissect CDR3 amplicons after V-J gene-specific polymerase chain reaction based on the CDR3 length. This method is quick and cheap compared to Sanger sequencing; however, it cannot give any information on the nucleotides or amino acids of CDR3 loops, which is crucial to infer their molecular structure and uncover how TCRs/BCRs recognize the MHC and antigens. On the protein level, flow cytometry had been applied from the early days to quantify the T cells that expressed the α/β TCRs with any particular V gene families. The T cell population can be sorted and quantified by a panel of stained or fluorescence-tagged TCR V β and V α -specific monoclonal antibodies, either individually or in a multiplex format (McLean-Tooke et al. 2008; van den Beemd et al. 2000). The IOTest beta Mark TCR Vbeta Repertoire Kit (IM3497, Beckman Coulter) consists of monoclonal antibodies targeting 24 distinct human TCR V β families. The limitation of this method is that its resolution is quite low. It can only assess whether the TCR repertoire is biased towards a particular V gene family, rather than exploring the prolific diversity of the TCR clonotypes. Additionally, to simplify the TCR model, several groups had developed TCR^{mini} transgenic mice model to reduce the enormous diversity of the TCR repertoire. TCR^{mini} mice are the genetically engineered mice which have knocked out part of V/D/J germline genes. It has a limited size of rearranged TCR repertoire but maintains the normal development and function of the T cells and the adaptive immune response (Lathrop et al. 2011; Pacholczyk et al. 2006). Overall, all these efforts and technologies had paved the way and made it possible to study the great diversity of TCR and BCR repertoire before the emergence the high-throughput sequencing.

Prior research directions and exemplifications

The above technologies have been employed for more than two decades to investigate the immune repertoire

across a full spectrum of cellular and humoral immune response in animal models and human beings. Generally speaking, researches in this field have been focusing on two directions. Firstly, the VDJ pattern of the TCR/BCR as the specific T/B cell identifier can be utilized to compare and track the origin of the cells in a particular environment or/and of certain subtypes/groups, such as the regulatory T cells. Combined TCR^{mini} mice model with several traditional techniques, including spectratyping and Sanger sequencing, researchers found the majority of regulatory T cells originated from the immature cells in the thymus rather than were recruited and converted from the mature T cells in the periphery (Pacholczyk et al. 2006). Additionally, they found that the diversity of the regulatory T cells exceeded the naïve T cells in the TCR^{mini} mice model (Pacholczyk et al. 2006). Another study investigated the origin of regulatory T cells infiltrated in the carcinoma and found that regulatory T cells expressed distinct TCRs from other conventional T cells and from the antigen-experienced effector/memory cells, which implied tumor-infiltrating lymphocytes (TIL) mostly contained natural regulatory T cells rather than induced or converted cells (Hindley et al. 2011).

Secondly, the VDJ pattern of TCR/antibody repertoire and specific TCRs/BCRs in association with various diseases have been extensively studied (Miles et al. 2011). If one were to assume that VDJ recombination is mostly randomized, the probability of two individuals sharing the same TCR/BCR to a specific antigen or peptide-MHC (pMHC) antigen is vanishingly small. However, after the first reports of TCR bias in the 1990s from the studies of human beings exposed to influenza A virus and Epstein-Barr virus (EBV) (Argaet et al. 1994; Moss et al. 1991), growing reports have emerged to describe the predictable distortions of TCR/BCR repertoire due to specific antigens exposure. The terms to describe these distortions include VDJ bias/skewing and “public” sequences, which refer to the TCR/BCR sequences found in multiple individuals. People have discovered various forms and extents of VDJ bias with regard to all kinds of exposures/diseases. Among them, the most obvious examples were found in the field of infections and vaccinations. Some pathogens or vaccines can produce superantigens which are proteins that bind to certain antibody V domains, thus will result in skewed antibody repertoires or even shared antibodies among individuals (Silverman and Goodyear 2006). In contrast, public TCRs have MHC

restrictions (Garcia et al. 2009), which refers to the property that TCRs always recognize pMHC in the context of a specific allele (or set). Abundant pathogens have been found to induce public TCRs in the background of certain pMHC, including influenza A virus (Moss et al. 1991), EBV (Price et al. 2005), cytomegalovirus (Khan et al. 2002), hepatitis B virus (Sing et al. 2001), hepatitis C virus (Umemura et al. 2000), and human immunodeficiency virus (Yu et al. 2007).

Biased TCR/BCR repertoire in malignancy has also been found. The tumor-associated antigens (TAAs) in a tumor tissue of certain malignancy often involve numerous abnormal proteins expressed by the mutated genes. The shared TAAs among patients have the potential to drive the formation of shared repertoire, and the situation could be discrepant between the lymphoid malignancies and other carcinomas. The lymphoblastic leukemia and lymphoma which are induced by malignant T cells or B cells have been found to carry stereotyped repertoire among individuals (Agathangelidis et al. 2012). It infers that these stereotyped cell receptors might not just be the independent proxies of the malignant T/B cells but also interact with and are selected by some antigenic factors that are involved in the tumorigenesis of the malignant lymphocytes. On the other hand, the lymphocytes infiltrated in the solid tumors have been recurrently reported to have diagnostic and prognostic values and play pivotal roles in the entire process of tumorigenesis. The regulatory T cells infiltrated in the solid tumor have been found to skew towards public sequences, which strongly imply that the interaction between the lymphocytes and TAAs has led to the selection of these public cells (Sainz-Perez et al. 2012). Another interesting phenomenon has implied this bias has not solely been driven by TAAs but can also be triggered by certain pathogens in some viral-associated cancer (Miles et al. 2011).

Autoimmune disorders have also been extensively studied for the public TCRs. Most of the self-antigens of autoimmunity have not been clarified thus far. As a result, the targets of the self-antibodies and TCRs are always difficult to define, both in terms of peptide specificity and MHC restriction. Regardless of the unclear mechanism, people have found the TCR alpha and beta chains are conserved inter-individually in systemic lupus erythematosus (Luo et al. 2008). Besides, several studies have reported the TCR bias and MHC restriction in multiple sclerosis (Oksenberg et al. 1993; Wucherpfennig et al. 1990).

Application of HTS in the TCR and BCR repertoire study

The experimental technologies of high-throughput immune repertoire sequencing

If the entire T/B cell repertoire and the antigen-specific T/B cells are to be treated as a haystack and a needle, the emergence of high-throughput sequencing (HTS) provides a powerful weapon to identify/discover a needle in a haystack. The characteristics of HTS, such as enormous reads number and cost-effectiveness per read, fit ideally the investigation on TCR and BCR repertoire. The earliest HTS platform used on this area is the Roche 454 system, which provides on average 500-bp read length and modest read number per run. It is able to capture the entire VDJ regions of TCR and BCR, and can sequence millions of molecules per repertoire, which significantly surpass the Sanger sequencing technique. In 2009, the first study utilizing 454 to sequence the zebrafish antibody repertoire was published in *Science* (Weinstein et al. 2009) and ignited the passion in this research area. Then, two studies adopted 454 sequencer to cover the entire VDJ recombined region of the human antibody and TCR repertoires (Boyd et al. 2009; Wang et al. 2010). The first study utilizing Illumina sequencer to cover the most variable CDR3 regions was published in *Genome Research* to investigate the TCR beta chain characteristics in the human periphery (Freeman et al. 2009). Compared to 454, the Illumina Hiseq platform provides shorter read length but significantly higher throughput of reads and lower cost per read, which is brilliantly suitable to deep sequence a quite complex repertoire. In the current applications, read length and throughput are still two factors that investigators should compromise and do trade-off. For those studies which require reading the entire VDJ region, such as antibody discovery, Illumina Miseq was the best choice due to the long read length. Although Roche 454 can produce long reads, we do not recommend it due to its high indel error and off the shelf in 2013. For TCR, in which CDR3 regions are mainly focused, Illumina Hiseq platform is quick and cost-effective.

Both genomics DNA (gDNA) and mRNA have been reasonably selected as the research material in various studies. Whether to choose gDNA or mRNA to analyze depends on the aim of the investigation. gDNA of TCR and BCR exists in single copy in the T and B cells, thus

represents the cellular proportion and heterogeneity of T/B cell population. Therefore, gDNA is suitable for calculating the proportion of antigen-specific or target T and B cells, such as vaccine-specific clones or leukemia/lymphoma clones, and for studying the functional/phenotypic evolution of specific TCR/BCR clonotypes. On the other hand, as the expression of TCR/BCR varied significantly among different types/status of T/B cells, such as on naïve cells and plasmablasts, the quantification of TCR/BCR mRNA would be more correlated with the cell function/activation. Technically, gDNA extracted from a T/B cell population usually contains the pre- and post-recombined VDJ fragments and the non-productive VDJ rearrangements, which would serve as the background to interfere with the proper amplification of the TCR/BCR repertoire. Nevertheless, gDNA would always be easier to access than the mRNA, including in storage and shipment.

There are various ways to prepare the sequencing library of TCR/BCR repertoire from both gDNA and mRNA samples. Due to the rareness of the specific VDJ molecules especially the low-frequent copies, the amplification is usually adopted and bias comes along with it. Multiplex PCR is the most convenient and straightforward amplification approach for DNA samples (Klarenbeek et al. 2010; Wang et al. 2010), which usually incorporates V gene and J gene (or C gene for mRNA) specific multiplex primers to amplify the full recombined variable region or CDR3 region of the receptor gene, depending on the region of the V genes that the primers targeted. This one-tube reaction is convenient, but the different efficiencies and cross-reactivities of the various primers will usually bring into certain extent of bias to the amplified products. In order to standardize the procedure to detect the clonally rearranged Immunoglobulin and TCR in a diagnostic setting, such as minimal residual disease detection, the European BIOMED-2 collaborative study has developed and validated standard multiplex PCR primers for amplifying rearranged antibodies and TCR genes early before the emergence of HTS (van Dongen et al. 2003). Several groups are still using the BIOMED-2 primers or the improved versions based on them. Other groups, however, are trying to implement various approaches to minimize the gene-specific bias, such as using nested PCR (Wang et al. 2010), using synthetic templates to design an unbiased PCR reaction (Carlson et al. 2013), or incorporating random barcodes in the primers to adjust the bias and errors (Shugay et al. 2014;

Vollmers et al. 2013). Nevertheless, through all these efforts, the PCR bias can only be minimized but hardly be removed. The other experimental method enriching the rearranged antibodies or TCRs is rapid amplification of cDNA 5' ends (5'RACE), which can theoretically avoid V gene bias as it only incorporates C-gene primers. 5'RACE can only work on mRNA but not gDNA. Primers are designed against a known region in the 3'mRNA, which is the constant gene region in the TCR/antibodies repertoire analysis. The mRNA is reverse transcribed, and a homopolymeric tail is synthetically added to the 3'ends of the cDNA using dNTP and TdT or just using the TdT properties of the reverse transcriptase through the template-switch process (Mamedov et al. 2013; Peters et al. 1999). Another 5' RACE is ligation-anchored PCR, which ligates a linker to the 3' end of cDNA after reverse transcription using RNA ligase (Gao and Wang 2015; Heather et al. 2015). Therefore, the products can be amplified by the constant gene primer and the universal linker primer. Some alterations have been put into this process in order to shorten the product length to get it sequenced by Illumina Hiseq system. For instance, biotins are labeled to the constant gene primers, and the cDNA products are randomly fragmented to an appropriate length to be sequenced by single reads or paired-end reads of Hiseq machine. The biotin-labeled fragments are specifically captured by the magnetic beads to enrich the targeted fragments which contain CDR3 and the constant gene region. The RACE approach has been extensively adopted in the current TCR/BCR studies (Bashford-Rogers et al. 2014; Li et al. 2013; Warren et al. 2011). Multiplex PCR and 5' RACE are two most popular approaches in the current community; however, as analyzed by other review papers (Georgiou et al. 2014; Robins 2013), both have their pros and cons. To provide data for method selection, our group did a systematic evaluation and comparison of the both methodologies (Liu et al. 2016).

In terms of the enrichment of VDJ recombinations, there are also other approaches such as using probes/baits that are designed against the V/J germline gene sequences to capture the rearranged VDJ fragments (He et al. 2011). This approach, however, requires random shearing of gDNA/cDNA before the capturing step and will induce the infrequent complete rearrangements being fragmented and lost in capturing and sequencing. Therefore, it is applicable only in recovering the abundant rearrangements in cases such as lymphoma or leukemia.

The chain pairing information of $V_H:V_L$ for BCR and $V_\alpha:V_\beta$ for TCR is critical to elucidate the function of T/B cells. Previous approaches amplify each chain separately, or enrich two chains simultaneously in the same reaction. However, in a population level, these approaches could not tell every $V_H:V_L$ or $V_\alpha:V_\beta$ pair. Until recently, the investigation on native chain pairs has been conducted through single-cell cloning by limiting dilution and Sanger sequencing on individual cells. However, this strategy is low-throughput and expansive. After incorporating with HTS, new strategies have been developed to enhance the throughput and feasibility to uncover native chain pairing information. A cell-based emulsion RT-PCR approach was developed to allow the selective fusion of the native pairs of TCR $V_\alpha:V_\beta$ genes, and more than 700 pairs could be recovered in a single emulsion experiment (Turchaninova et al. 2013). DeKosky et al. developed a $V_H:V_L$ pairing technology that relies on sequestering single B cells into subnanoliter volume wells, lysing the cells, capturing RNA on poly-dT beads, and generating amplicons encoding linked $V_H:V_L$ segments by emulsion overlap extension PCR (DeKosky et al. 2013). Thousands of unique endogenous $V_H:V_L$ pairs with 97% validated pairing accuracy in a 1-day experiment were achieved. Recently, Howie et al. reported a pairSEQ approach in which a fixed number of T cells are randomly allocated to each well on a 96-well plate, and mRNA was reverse transcribed, attached with well-specific barcodes, amplified, pooled together, and sequenced. Any pairs that share a unique set of wells (barcodes) are considered to come from the same clone, since the probability of two clones sharing the same set of wells is minimal (Howie et al. 2015). Approaches with higher throughput pairing combining emulsion PCR and HTS are under development, and we can anticipate the routine adoption of TCR/BCR pairing sequencing in this field in the near future. Besides the pairing information, single-cell approach can also link the T/B cell phenotype such as gene expression signatures with the receptor pairing sequences, which could help to elucidate the T/B cell evolution and heterogeneity in function (Han et al. 2014).

Bioinformatic tools for processing immune repertoire sequencing data

Feasible bioinformatic tools had been developed to align and analyze TCR and BCR sequences before the

emergence of HTS sequencing, including iHMMune-align (Gaeta et al. 2007), IMGT/V-QUEST (Brochet et al. 2008), and NCBI's IgBLAST. These tools are very useful to determine the recombined V/D/J genes, the CDR region annotation, and the characteristics of the TCR/BCR sequences. HTS provides us with unprecedented depth and diversity of immune repertoire, and more powerful tools/software are warranted to accurately and speedily handle these enormous data. The standard pipeline for immune repertoire data analysis and visualization has been suggested (Greiff et al. 2015b; Yaari and Kleinstein 2015) and some new or updated computational tools have been published to analyze HTS data, such as IMGT/HighV-QUEST (Li et al. 2013), new IgBLAST (Ye et al. 2013), Decombinator (Thomas et al. 2013), pRESTO (Vander Heiden et al. 2014), and MiTCR (Bolotin et al. 2013), which has been further updated to MiXCR to mine both TCR and BCR data (Bolotin et al. 2015). These established tools are available for VDJ gene assignment, CDR and FR annotation, CDR3 length identification, insertion and deletions analysis, and mutation spectrum analysis (BCR). The characteristics of some above analytical tools have been reviewed by Greiff et al. (2015b). Several technical aspects should be considered to improve the analytical tools, including identifying PCR and sequencing errors from the true biological variations and adjusting the PCR bias. While experimental optimization is helpful in these aspects, in silico methods could also be beneficial. Taking them into account, we have established our own bioinformatic tools entitled *IMonitor*, which is short for Immune Monitor, to analyze both TCR and BCR repertoire data (Zhang et al. 2015). We have validated *IMonitor* on both the simulated data and the biological data and compared its accuracy with some existing tools.

On top of the basic analytical tools, additional computational methods have been developed to dig into the data in greater details. The metrics of immune repertoire diversity can reflect the extent of clonal selection and expansion and may serve as a biomarker telling an individual's immune status. Shannon entropy index and Simpson's diversity index have been introduced from ecology to estimate the repertoire diversity but bear obvious limitations including low sensitivity and dependence upon sequencing depth. New bioinformatic frameworks have been established to link the immunological status with the diversity indices (Greiff et al. 2015a) and to estimate the species richness and

distribution (Laydon et al. 2014). For BCR, somatic hypermutation brings antibody affinity maturation and more diversified BCR; therefore, the concept of molecular evolution was used to describe the dynamics of BCR within individuals. B cell lineage trees were constructed to infer the ancestral relationship between individual cells, and some computational pipelines have been developed to analyze the B cell lineage trees (Bashford-Rogers et al. 2013; Jiang et al. 2013).

We have summarized in a table depicting in details the functions of some existing tools, including chain pairing and germline V/J gene prediction, and whether they can be used with single-cell RNA-seq data (Table 2).

Basic and translational research of immune repertoire by HTS

TCR/antibody spectrum profiling by HTS has been applied to basic and translational research in these years, and has a wide range of clinical and healthcare applications, which will be discussed in details in next several parts.

Basic research using the immune repertoire HTS

Since HTS has forwarded the TCR/BCR repertoire research into a unprecedented level, the species-specific repertoire characteristics of various vertebrate species have been investigated in these years, including zebrafish (Weinstein et al. 2009), mouse (Madi et al. 2014), rhesus monkey (Sundling et al. 2012), and camel (Li et al. 2016). The most crucial problem for studying other species' TCR/BCR repertoire is the lack of V/D/J gene reference sequences for alignment and annotation, which requires more comprehensive and refined genome sequencing and assembly for the non-rearranged germline genes. These animal models are widely used in immunological research, such as rhesus macaque being used in vaccine research and preclinical test, and camel being adopted for the unique VHH antibody discovery. Therefore, we can anticipated that in combination with the traditional animal model studies, immune repertoire data collection and analysis on these non-human species will complement and extend our understanding in their immunological status and eventually benefit the therapeutics on human disease.

Subsets of periphery T/B cells in human have also been investigated for their repertoire heterogeneity by

the HTS deep sequencing. T cells in periphery have been studied for their repertoire size (Warren et al. 2011) and the different subsets have been compared for the identical TCRs to undercover the cell fate determination (Wang et al. 2010). Human memory T cells have been found to consist mainly of unexpanded T cells with a relatively small expanded proportion (Klarenbeek et al. 2010). With regards to T cell fate driven by pathogens, a paper published in *Science* has revealed the functional heterogeneity of memory T cells primed by pathogens or vaccine, and naïve T cells could be induced to multiple fates by the pathogens in vitro (Becattini et al. 2015).

It is an interesting scientific question how and to what extent the immune diversity is determined and affected by the genetic background and antigen exposure. Some studies indicated that the antibody diversity was largely driven by non-heritable influences (Brodin et al. 2015; Jiang et al. 2011). Studies on monozygotic twins have revealed that the naïve antibody repertoire is highly heritable (Glanville et al. 2011), but the situation for the TCR repertoire is more complex (Zvyagin et al. 2014). Using mouse model, Greiff et al. found a dynamic balance of genetic background and antigen-driven antibody repertoire predetermination and also indicated the importance of stochastic variation in repertoire diversity (Greiff et al. 2017a). Except the development and diversity generation of adaptive immune cells, the trafficking and tissue localization of those immune cells are also crucial for immune system to perform immune surveillance and effector functions. In order to define the pattern of B cell distribution in human, Meng et al. investigated the antibody repertoire of eight different anatomic compartments from six human organ donors using IGH HTS and demonstrated that the B cell clones can be parted into two major networks, the blood-rich tissues (blood, bone marrow, spleen, lung) and the gastrointestinal tract (jejunum, ileum, and colon) (Meng et al. 2017).

Immune repertoire HTS has also been used to investigate the pathogenesis of some immune-related disease, such as SLE (Sui et al. 2015; Thapa et al. 2015; Tipton et al. 2015), multiple sclerosis (MS) (Muraro et al. 2014; Stern et al. 2014; von Budingen et al. 2012), and psoriasis (Harden et al. 2015). The B cell isotypes in periphery blood of allergy were also investigated and the IgE was found to be more frequently switched from IgG1 rather than IgM (Looney et al. 2015).

Table 2 The comparison of different bioinformatic tools

Tool name	Input data	Chain	VDJ alignment	Sequencing error	Unique molecular identifier	TRA/TAB or IGH/IGKL pairing	Infer Novel V germline	Infer Novel J germline
Common tools for immune repertoire sequences								
IMGT/HighV--QUEST	FASTA	BCR/TCR	Yes	No	–	–	–	–
IgBLAST	FASTA	BCR/TCR	Yes	No	–	–	–	–
Decombinator (v3.1)	FASTQ	TCR	Yes	Yes	Yes	–	–	–
MiTCR	FASTQ	TCR	Yes	Yes	–	–	–	–
MiXCR	FASTA/FASTQ	BCR/TCR	Yes	Yes	–	–	–	–
IMonitor	FASTA/FASTQ	BCR/TCR	Yes	Yes	–	–	–	–
RTCR	FASTQ	TCR	Yes	Yes	Yes	–	–	–
Specific for unique molecular identifier								
MIGEC	FASTQ	BCR/TCR	Yes	Yes	Yes	–	–	–
Specific for RNA-seq data								
TRUST	BAM	TCR	Yes	–	–	–	–	–
V'Djer	BAM	BCR	Yes	–	–	–	–	–
Specific for single-cell RNA data								
TraCeR	FASTQ	TCR	Yes	–	–	Yes	–	–
BASIC	FASTQ	BCR	No	–	–	Yes	–	–
Inferring novel germline								
TiGER	FASTA	BCR	–	–	–	–	Yes	No
IMPre	FASTA	BCR/TCR	–	–	–	–	Yes	Yes
IgDiscover	FASTA/FASTQ	BCR	–	–	–	–	Yes	No

Antibody discovery

Traditional techniques for antibody discovery rely on hybridoma preparation and large-scale screening on cells with high affinity. This method is reliable but time-consuming and labor-intensive. With direct sequencing on antibody repertoire of immunized animals, we are able to capture the entire antibody response in the periphery or the splenic B cells. By analytical algorithm, the potential antibodies with high affinity to the immunogen can be predicted by their abundance, heavy/light chain pairing, and somatic evolution (Reddy et al. 2010). Combining with protein profiling by mass spectrum in the serum, the antibody proteins can be more accurately identified and quantified and potentially provide better predictions and outcomes (Cheung et al. 2012). The antibodies with known sequences can then be directly cloned, expressed, and tested of affinities.

Vaccine design and evaluation

There are abundant unanswered questions regarding the magnitude and persistence of protection for the new vaccines designed to market. Recently, significant efforts have been put into the understanding of innate immunity and T/B cell response induced by the vaccines. The hallmarks of vaccine-induced immune response are essentially antibodies secreted by B lymphocytes and various involvements of different T lymphocytes. Both the antibody responses including primary low-affinity antibody secreting and affinity maturation together with the secondary responses of the memory recall, and the antigen-specific effector and memory T cells coordinated with each other and other cell types, provide effective protection of the vaccines to the host.

Reasonably, the dynamic antibody and T cell responses induced by vaccines can benefit substantially from the TCR/BCR repertoire HTS studies. The induction and proliferation of antigen-specific T/B cells, the class-switching of antibody isotypes, and the somatic hypermutations acquired in the selection process of affinity maturation can be investigated in details by analyzing the TCR/antibody repertoire sequences. Together, the magnitude and breadth of responses can be assessed by comparing the change of the repertoire among different time-points during vaccination process and different vaccine groups. Non-human primates of rhesus as the preclinical animal model for vaccine

design have been utilized to investigate the presence and development of antigen-specific antibodies post-immunization of a HIV vaccine, with the prior knowledge of identified HIV neutralizing antibodies (Dai et al. 2015). Human antibody responses post-influenza vaccination has been traced for antigen-specific antibody clones and a certain extent of their convergent rearrangement has been identified (Jackson et al. 2014). A longitudinal study on dynamics of antibody response following years post-vaccination in three volunteers has demonstrated significant antibody proliferation and accumulation of somatic hypermutations on isotype-switched antibodies (Laserson et al. 2014). Magnitude of isotype switching from IgM to IgG or IgA has been compared between two different vaccines (Jiang et al. 2013) and lineage analysis and comparison between repeated yearly vaccinated samples reveal signatures of memory B cell activation (Vollmers et al. 2013). Most studies on antibody responses include analysis on breadth and diversity of antibodies induced by vaccination and TLR adjuvants have been found to expand the B cell repertoire following malaria vaccination (Wiley et al. 2011). The heritability of antibody responses post-dose of vaccines has been investigated from identical twin pairs and IgM instead of IgG B cells are more similar between twins and show higher level of heritability (Wang et al. 2015). The aging effect to vaccine responses has also been studied (Jiang et al. 2013). Currently, most studies on vaccine effects have focused on antibody responses, while Becattini et al. have revealed that CD4+ memory T cells induced by distinct vaccines and pathogens have functional heterogeneity and multiple fates post-priming (Becattini et al. 2015).

Biomarkers for cancer diagnosis, subtyping, and prognosis

One important field for translational research and medicine of TCR/BCR HTS sequencing is the minimal residual disease (MRD) detection for lymphatic hematological tumors, including T/B cell-derived leukemia and lymphoma. The sensitivity and prognosis use for NGS in comparison with the traditional techniques such as flow cytometry and allelic-specific PCR have been recurrently investigated ever since the beginning of this field (Boyd et al. 2009; Logan et al. 2011; Pui et al. 2015; Pulsipher et al. 2015; Wu et al. 2012, 2016). NGS MRD detection has been demonstrated to correlate well

with previous techniques but with much higher sensitivity. Recently, MRD detection for various subtypes of lymphoma has been investigated, and the plasma has exhibited superior detection power than the cellular repertoire (Kurtz et al. 2015; Roschewski et al. 2015; Weng et al. 2013).

However, though the technological sensitivity and reproducibility of NGS based MRD detection have been validated in a handful of studies, there are several unanswered questions regarding the superiorities of the new technique, before the translation can take a step further. (1) Does the higher detection sensitivity of the MRD level benefit the early detection of relapse and the longer survival? (2) How does the immune reconstruction benefit the patients? (3) Will the malignant lymphocyte clones undertake evolution in their cell receptor sequences (TCR/BCR) and does the evolution occur under the pressure of therapy and correlate with relapse?

Regarding to the first question, studies have been undertaken and the data seem to be supportive of the positive correlation between the clinical benefits and the higher sensitivity (Logan et al. 2014; Pui et al. 2015). However, prospective clinical trials with larger sample size are still required for clinical validation. As for question 2, the repertoire sequencing facilitates the clinicians the evaluation the immune reconstruction after transplantation or other therapies, but limited studies have pursued to quantify and qualify the immune reconstruction with the NGS data. Our investigation of child B-ALL not only tried to demonstrate the higher sensitivity of NGS MRD compared with flow cytometry but also evaluated the emergence and dynamics of evolved IGH clones during treatment, aiming to provide contributions to question 3 (Wu et al. 2016). Other studies have also evaluated the profile and consequences of clonal evolution in B-ALL (Bashford-Rogers et al. 2016; Gawad et al. 2012).

Tumor-infiltrating lymphocytes are good prognostic biomarkers and have been extensively studied on various solid tumors for decades. Immune repertoire HTS has been used to investigate the intra-tumoral and inter-tumoral heterogeneity (Emerson et al. 2013; Gerlinger et al. 2013; Wang et al. 2017) to analyze the tumor deeply or help to subtype the tumor. For immune checkpoint therapies, only part of patients respond to the therapy and it is difficult to predict the result at the beginning of the therapy. Studies on CTLA4 blockade treatment on the periphery T cells revealed that the treatment could broaden the TCR repertoire and

increase its diversity, and the responders maintained a more stable repertoire with less expanded or contracted TCR clones compared with the non-responders (Cha et al. 2014; Robert et al. 2014). Taken together, the application of HTS to the TCR/BCR discovery and identification has facilitated its broader use as biomarkers in the disease prognosis and monitoring. Furthermore, people including us are trying to construct large database for disease-associated repertoire and keen to apply it to the disease prediction, subtyping, and prognosis. As for the antibody discovery and the vaccine development, both of them are tens-of-billion-dollar industry and HTS will definitely demonstrate its bright future in these fields.

Challenges and problems to be solved in the era of HTS

Though HTS has demonstrated its power to probe the tremendous diversity of TCR and BCR repertoire, challenges still remain ahead for its usage in wider research and clinical applications. The first problem comes from the differences of data production and processing in different laboratories. For data production, the input samples could be genomic DNA or mRNA, the gene fragment could be enriched by multiplex PCR or 5' RACE, and different research groups use different PCR primers and different sequencing platform. Different experimental approaches harbor their own characteristics and bias, and the current divergent selection of methods in different labs makes the published results difficult to compare and do meta-analysis. Consequently, it is critically required to get experimental approaches as uniform or standard as possible to be used in different laboratories. For the data processing, although abundant analytical tools have been developed to uncover the biological meaning of the diversified TCR and BCR data, none of the existing analytical tools have satisfied all the needs and obviously surpassed the others, and the latter published tools have become superior in speed and accuracy. Perhaps, this is one of the reasons that researchers working in this area generally prefer to use custom-designed bioinformatics pipelines. Therefore, only the standardized and open-resource computational tools can facilitate the reproducing and meta-analysis of results generated from different laboratories, and this is in urgent need with the area matures. In summary, a centralized public database storing

interchangeable and standardized data, validated open-resource algorithms for data analysis and standards for experimental description could be eventually beneficial to the application in this field (Georgiou et al. 2014).

The second problem in this field is to distinguish the biological variations from the errors and the bias introduced in the different steps. The improved sensitivity of HTS sequencing to read the individual sequences would also amplify the errors. These errors could be derived from the reverse transcription (for mRNA samples), the PCR amplification, and the HTS sequencing. The latter two are the major sources of various types of chimeras and nucleotide errors. The PCR process is carried along the whole pipeline, including amplification of VDJ rearrangement and the sequencing library construction. The mis-annealing of PCR primers to the templates could introduce unspecific amplification and the mis-incorporation of PCR templates would introduce chimeric artifacts. In addition, The DNA polymerase will introduce certain extent of amplification errors in every PCR cycles. All these chimeras and nucleotide errors will accumulate with the augment of PCR cycles. Hence, some PCR errors incorporated in the early cycles could be retained in relative higher frequencies in the final HTS data.

On the other hand, HTS sequencing errors have been an issue since the very beginning. Currently, several platforms have been utilized to sequence the immune repertoire. Data produced by Roche 454, Life PGM, and Proton are usually enriched with indel errors in the nucleotide positions with homopolymers. Algorithms have been developed to correct these polymer indels with certain success (Bolotin et al. 2012; Prabakaran et al. 2011). However, the high frequency of indel errors sometimes is difficult to be distinguished from biological variations and thus significantly reduces the rate and accuracy of alignment to the germline V/D/J genes. Alternatively, Illumina Hiseq and Miseq are dominated with base substitutions (Nguyen et al. 2011), and the overall error rate is lower than the pyrosequencing technology. Collectively, the Illumina platform is more suitable to sequence the TCR/BCR repertoire with the combination of lower base-calling error rate and cost.

Several strategies have been applied to minimize and correct these errors. Using high-fidelity enzymes to enrich the VDJ rearranged fragment can reduce the errors produced by PCR (McInerney et al. 2014). Random barcodes or unique molecular identifiers (UMI) are random sequences of bases used to tag every

molecule. Incorporating random barcodes to the DNA fragments was originally developed to detect the extremely low-frequency mutations, especially in the tumor tissue or circulating tumor DNA (Kinde et al. 2011). By adding the highly diversified barcodes (the unique number of barcodes is designed to surpass the number of unique TCRs and BCRs in the studied repertoire), the set of VDJ fragments with the same barcodes could be grouped together and treated as the same template or molecule in the original repertoire before amplification and sequencing (Shugay et al. 2014; Turchaninova et al. 2016). Thereafter, the UMI approach has been widely used in which the amplification and sequencing errors could be corrected simultaneously (Britanova et al. 2014; Egorov et al. 2015; Khan et al. 2016; Vollmers et al. 2013). Another approach is to apply various algorithms to do the *in silico* correction (Benichou et al. 2012; Vander Heiden et al. 2014; Warren et al. 2011; Zhang et al. 2015).

Thirdly but not last, the available information for the antigenic specificity/reactivity of the TCR/BCR clonotypes is still too limited, which has hampered the deeper biological analysis of the large-scale data generated by HTS. Previous studies have validated and provided thousands of antigen/epitope reactive TCR/BCR sequences through immunological assays such as flow cytometry-based experiments and pMHC multimeric technology, but there are spaces to improve in at least two aspects to meet the growing demands in this field. The first is the technology optimization and development for high-throughput identification of the antigen reactive or antigen-related TCRs/BCRs. Researchers have made some progress, such as combining the immune assays with immune receptor sequencing (Klinger et al. 2015) and developing algorithms to predict antigen-specific TCRs/BCRs (Dash et al. 2017; Glanville et al. 2017), but greater efforts are warranted to keep the pace with the growing TCR/BCR data and decipher their antigen specificities. Recently, Emersom et al. reported they successfully identified the signatures of cytomegalovirus exposure history by statistical classification framework based on the public TCRs of large cohorts (Emerson et al. 2017), which indicated the possibility of classification and screening diseases according to the accurately defined public TCRs (Greiff et al. 2017b).

Besides, though attempts have been made (Shugay et al. 2017; Tickotsky et al. 2017), there are still lack of

centralized databases depositing and updating all the aforementioned antigen-specific TCRs/BCRs, their connected antigens/epitopes, and/or pMHC restrictions (for TCRs only). These kinds of public available knowledge database could serve as valuable resources for the community and accelerating the clinical applications of the huge amount of immune repertoire data.

Concluding remarks

Recently, immune repertoire sequencing has demonstrated its potential to detect the minimal residual disease in leukemia, to unravel the heterogeneity of tumor microenvironment, and to detect the dynamics and broadness of the lymphocyte response to vaccination. With more attempts and solutions being developed to face the challenges and solve the technique problems in this field, and most researchers in the community are to share uniform experimental and analytical pipelines, we envision that more interesting biological discoveries will be reported and biological interpretations will be validated.

Acknowledgements We want to thank the supervision and guidance of Prof. Karsten Kristiansen from the University of Copenhagen and the support from BGI-Shenzhen for this work.

References

- Agathangelidis A, Darzentas N, Hadzidimitriou A, Brochet X, Murray F, Yan XJ, et al. Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood*. 2012;119:4467–75.
- Argaet VP, Schmidt CW, Burrows SR, Silins SL, Kurilla MG, Doolan DL, et al. Dominant selection of an invariant T cell antigen receptor in response to persistent infection by Epstein-Barr virus. *J Exp Med*. 1994;180:2335–40.
- Bashford-Rogers RJ, Palser AL, Huntly BJ, Rance R, Vassiliou GS, Follows GA, et al. Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. *Genome Res*. 2013;23:1874–84.
- Bashford-Rogers RJ, Palser AL, Idris SF, Carter L, Epstein M, Callard RE, et al. Capturing needles in haystacks: a comparison of B-cell receptor sequencing methods. *BMC Immunol*. 2014;15:29.
- Bashford-Rogers RJ, Nicolaou KA, Bartram J, Goulden NJ, Loizou L, Koumas L, et al. Eye on the B-ALL: B-cell receptor repertoires reveal persistence of numerous B-lymphoblastic leukemia subclones from diagnosis to relapse. *Leukemia*. 2016;30:2312–21.
- Becattini S, Latorre D, Mele F, Foglierini M, De Gregorio C, Cassotta A, et al. T cell immunity. Functional heterogeneity of human memory CD4(+) T cell clones primed by pathogens or vaccines. *Science*. 2015;347:400–6.
- Benichou J, Ben-Hamo R, Louzoun Y, Efroni S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology*. 2012;135:183–91.
- 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;42:3073–83.
- 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;10:813–4.
- Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. 2015;12:380–1.
- Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, et al. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci Transl Med*. 2009;1:12ra23.
- Britanova OV, Putintseva EV, Shugay M, Merzlyak EM, Turchaninova MA, Staroverov DB, et al. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. *J Immunol*. 2014;192:2689–98.
- Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res*. 2008;36:W503–8.
- Brodin P, Jovic V, Gao T, Bhattacharya S, Angel CJL, Furman D, et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell*. 2015;160:37–47.
- 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.
- Cha E, Klinger M, Hou Y, Cummings C, Ribas A, Faham M, et al. Improved survival with T cell clonotype stability after anti-CTLA-4 treatment in cancer patients. *Sci Transl Med*. 2014;6:238ra270.
- Cheung WC, Beausoleil SA, Zhang X, Sato S, Schieferl SM, Wieler JS, et al. A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat Biotechnol*. 2012;30:447–52.
- Chien YH, Meyer C, Bonneville M. Gammadelta T cells: first line of defense and beyond. *Annu Rev Immunol*. 2014;32:121–55.
- Dai K, He L, Khan SN, O'Dell S, McKee K, Tran K, et al. Rhesus macaque B-cell responses to an HIV-1 trimer vaccine revealed by unbiased longitudinal repertoire analysis. *MBio*. 2015;6.
- Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature*. 2017;547:89.
- 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;31:166–9.

- Egorov ES, Merzlyak EM, Shelenkov AA, Britanova OV, Sharonov GV, Staroverov DB, et al. Quantitative profiling of immune repertoires for minor lymphocyte counts using unique molecular identifiers. *J Immunol*. 2015;194:6155–63.
- Emerson RO, Sherwood AM, Rieder MJ, Guenthoer J, Williamson DW, Carlson CS, et al. High-throughput sequencing of T-cell receptors reveals a homogeneous repertoire of tumour-infiltrating lymphocytes in ovarian cancer. *J Pathol*. 2013;231:433–40.
- Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet*. 2017;49:659–65.
- Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res*. 2009;19:1817–24.
- 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;23:1580–7.
- Gao F, Wang K. Ligation-anchored PCR unveils immune repertoire of TCR-beta from whole blood. *BMC Biotechnol*. 2015;15:39.
- Garcia KC, Adams JJ, Feng D, Ely LK. The molecular basis of TCR germline bias for MHC is surprisingly simple. *Nat Immunol*. 2009;10:143–7.
- Gauss GH, Lieber MR. Mechanistic constraints on diversity in human V(D)J recombination. *Mol Cell Biol*. 1996;16:258–69.
- Gawad C, Pepin F, Carlton VE, Klinger M, Logan AC, Miklos DB, et al. Massive evolution of the immunoglobulin heavy chain locus in children with B precursor acute lymphoblastic leukemia. *Blood*. 2012;120:4407–17.
- Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol*. 2014;32:158–68.
- Gerlinger M, Quezada SA, Peggs KS, Furness AJ, Fisher R, Marafioti T, et al. Ultra-deep T cell receptor sequencing reveals the complexity and intratumour heterogeneity of T cell clones in renal cell carcinomas. *J Pathol*. 2013;231:424–32.
- Germain RN. T-cell development and the CD4-CD8 lineage decision. *Nat Rev Immunol*. 2002;2:309–22.
- Glanville J, Kuo TC, von Budingen HC, Guey L, Berka J, Sundar PD, et al. Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. *Proc Natl Acad Sci U S A*. 2011;108:20066–71.
- Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, et al. Identifying specificity groups in the T cell receptor repertoire. *Nature*. 2017.
- 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;152:5109–19.
- Greiff V, Bhat P, Cook SC, Menzel U, Kang W, Reddy ST. A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status. *Genome Med*. 2015a;7:49.
- Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and statistical analysis of adaptive immune repertoires. *Trends Immunol*. 2015b;36:738–49.
- Greiff V, Menzel U, Miho E, Weber C, Riedel R, Cook S, et al. Systems analysis reveals high genetic and antigen-driven predetermination of antibody repertoires throughout B cell development. *Cell Rep*. 2017a;19:1467–78.
- Greiff V, Weber CR, Palme J, Bodenhofer U, Miho E, Menzel U, et al. Learning the high-dimensional immunogenomic features that predict public and private antibody repertoires. *J Immunol*. 2017b;199:2985–97.
- Han A, Glanville J, Hansmann L, Davis MM. Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat Biotechnol*. 2014;32:684–92.
- Harden JL, Hamm D, Gulati N, Lowes MA, Krueger JG. Deep sequencing of the T-cell receptor repertoire demonstrates polyclonal T-cell infiltrates in psoriasis. *F1000Res*. 2015;4:460.
- He J, Wu J, Jiao Y, Wagner-Johnston N, Ambinder RF, Diaz LA Jr, et al. IgH gene rearrangements as plasma biomarkers in non-Hodgkin's lymphoma patients. *Oncotarget*. 2011;2:178–85.
- 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.
- Hindley JP, Ferreira C, Jones E, Lauder SN, Ladell K, Wynn KK, et al. Analysis of the T-cell receptor repertoires of tumour-infiltrating conventional and regulatory T cells reveals no evidence for conversion in carcinogen-induced tumors. *Cancer Res*. 2011;71:736–46.
- Howie B, Sherwood AM, Berkebile AD, Berka J, Emerson RO, Williamson DW, et al. High-throughput pairing of T cell receptor alpha and beta sequences. *Sci Transl Med*. 2015;7:301ra131.
- 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;16:105–14.
- Jiang N, Weinstein JA, Penland L, White RA, Fisher DS, Quake SR. Determinism and stochasticity during maturation of the zebrafish antibody repertoire. *Proc Natl Acad Sci*. 2011;108:5348–53.
- 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;5:171ra119.
- Khan N, Cobbold M, Keenan R, Moss PA. Comparative analysis of CD8+ T cell responses against human cytomegalovirus proteins pp65 and immediate early 1 shows similarities in precursor frequency, oligoclonality, and phenotype. *J Infect Dis*. 2002;185:1025–34.
- Khan TA, Friedensohn S, Gorter de Vries AR, Straszewski J, Ruscheweyh HJ, Reddy ST. Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Sci Adv*. 2016;2:e1501371.
- Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A*. 2011;108:9530–5.

- Klarenbeek PL, Tak PP, van Schaik BD, Zwinderman AH, Jakobs ME, Zhang Z, et al. Human T-cell memory consists mainly of unexpanded clones. *Immunol Lett*. 2010;133:42–8.
- Kleinstein SH, Louzoun Y, Shlomchik MJ. Estimating hypermutation rates from clonal tree data. *J Immunol*. 2003;171:4639–49.
- Klinger M, Pepin F, Wilkins J, Asbury T, Wittkop T, Zheng J, et al. Multiplex identification of antigen-specific T cell receptors using a combination of immune assays and immune receptor sequencing. *PLoS One*. 2015;10:e0141561.
- Koch U, Radtke F. Mechanisms of T cell development and transformation. *Annu Rev Cell Dev Biol*. 2011;27:539–62.
- Kurtz DM, Green MR, Bratman SV, Scherer F, Liu CL, Kunder CA, et al. Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. *Blood*. 2015;125:3679–87.
- Larson ED, Maizels N. Transcription-coupled mutagenesis by the DNA deaminase AID. *Genome Biol*. 2004;5:211.
- 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 U S A*. 2014;111:4928–33.
- Lathrop SK, Bloom SM, Rao SM, Nutsch K, Lio CW, Santacruz N, et al. Peripheral education of the immune system by colonic commensal microbiota. *Nature*. 2011;478:250–4.
- Laydon DJ, Melamed A, Sim A, Gillet NA, Sim K, Darko S, et al. Quantification of HTLV-1 clonality and TCR diversity. *PLoS Comput Biol*. 2014;10:e1003646.
- Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev*. 2004;18:1–11.
- Li S, Lefranc MP, Miles JJ, Alamyar E, Giudicelli V, Duroux P, et al. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nat Commun*. 2013;4:2333.
- Li X, Duan X, Yang K, Zhang W, Zhang C, Fu L, et al. Comparative analysis of immune repertoires between Bactrian Camel's conventional and heavy-chain antibodies. *PLoS One*. 2016;11:e0161801.
- Liu X, Zhang W, Zeng X, Zhang R, Du Y, Hong X, et al. Systematic comparative evaluation of methods for investigating the TCRbeta repertoire. *PLoS One*. 2016;11:e0152464.
- Logan AC, Gao H, Wang C, Sahaf B, Jones CD, Marshall EL, et al. High-throughput VDJ sequencing for quantification of minimal residual disease in chronic lymphocytic leukemia and immune reconstitution assessment. *Proc Natl Acad Sci U S A*. 2011;108:21194–9.
- Logan AC, Vashi N, Faham M, Carlton V, Kong K, Buno I, et al. Immunoglobulin and T cell receptor gene high-throughput sequencing quantifies minimal residual disease in acute lymphoblastic leukemia and predicts post-transplantation relapse and survival. *Biol Blood Marrow Transplant*. 2014;20:1307–13.
- Looney TJ, Lee JY, Roskin KM, Hoh RA, King J, Glanville J, et al. Human B-cell isotype switching origins of IgE. *J Allergy Clin Immunol*. 2015.
- Luo W, Ma L, Wen Q, Wang N, Zhou MQ, Wang XN. Analysis of the interindividual conservation of T cell receptor alpha- and beta-chain variable regions gene in the peripheral blood of patients with systemic lupus erythematosus. *Clin Exp Immunol*. 2008;154:316–24.
- Madi A, Shifrut E, Reich-Zeliger S, Gal H, Best K, Ndifon W, et al. T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Res*. 2014;24:1603–12.
- 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.
- McInerney P, Adams P, Hadi MZ. Error rate comparison during polymerase chain reaction by DNA polymerase. *Mol Biol Int*. 2014;2014:287430.
- McLean-Tooke A, Barge D, Spickett GP, Gennery AR. T cell receptor Vbeta repertoire of T lymphocytes and T regulatory cells by flow cytometric analysis in healthy children. *Clin Exp Immunol*. 2008;151:190–8.
- Meng W, Zhang B, Schwartz GW, Rosenfeld AM, Ren D, Thome JJC, et al. An atlas of B-cell clonal distribution in the human body. *Nat Biotechnol*. 2017;35:879.
- Miles JJ, Douek DC, Price DA. Bias in the alphabeta T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunol Cell Biol*. 2011;89:375–87.
- Moss PA, Moots RJ, Rosenberg WM, Rowland-Jones SJ, Bodmer HC, McMichael AJ, et al. Extensive conservation of alpha and beta chains of the human T-cell antigen receptor recognizing HLA-A2 and influenza A matrix peptide. *Proc Natl Acad Sci U S A*. 1991;88:8987–90.
- Muraro PA, Robins H, Malhotra S, Howell M, Phippard D, Desmarais C, et al. T cell repertoire following autologous stem cell transplantation for multiple sclerosis. *J Clin Invest*. 2014;124:1168–72.
- Nezlin R. Combinatorial events in generation of antibody diversity. *Comb Chem High Throughput Screen*. 2001;4:377–83.
- 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.
- Oksenberg JR, Panzara MA, Begovich AB, Mitchell D, Erlich HA, Murray RS, et al. Selection for T-cell receptor V beta-D beta-J beta gene rearrangements with specificity for a myelin basic protein peptide in brain lesions of multiple sclerosis. *Nature*. 1993;362:68–70.
- Pacholczyk R, Ignatowicz H, Kraj P, Ignatowicz L. Origin and T cell receptor diversity of Foxp3+CD4+CD25+ T cells. *Immunity*. 2006;25:249–59.
- 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;90:4319–23.
- Peters DG, Kassam AB, Yonas H, O'Hare EH, Ferrell RE, Brufsky AM. Comprehensive transcript analysis in small quantities of mRNA by SAGE-lite. *Nucleic Acids Res*. 1999;27:e39.
- Prabakaran P, Streaker E, Chen W, Dimitrov DS. 454 antibody sequencing—error characterization and correction. *BMC Res Notes*. 2011;4:404.
- Price DA, Brenchley JM, Ruff LE, Betts MR, Hill BJ, Roederer M, et al. Avidity for antigen shapes clonal dominance in

- CD8+ T cell populations specific for persistent DNA viruses. *J Exp Med*. 2005;202:1349–61.
- Pui CH, Pei D, Coustan-Smith E, Jeha S, Cheng C, Bowman WP, et al. Clinical utility of sequential minimal residual disease measurements in the context of risk-based therapy in childhood acute lymphoblastic leukaemia: a prospective study. *Lancet Oncol*. 2015;16:465–74.
- Pulsipher MA, Carlson C, Langholz B, Wall DA, Schultz KR, Bunin N, et al. IgH-V(D)J NGS-MRD measurement pre- and early post-allotransplant defines very low- and very high-risk ALL patients. *Blood*. 2015;125:3501–8.
- 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;28:965–9.
- Robert L, Tsoi J, Wang X, Emerson R, Homet B, Chodon T, et al. CTLA4 blockade broadens the peripheral T-cell receptor repertoire. *Clin Cancer Res*. 2014;20:2424–32.
- Robins H. Immunosequencing: applications of immune repertoire deep sequencing. *Curr Opin Immunol*. 2013;25:646–52.
- Roschewski M, Dunleavy K, Pittaluga S, Moorhead M, Pepin F, Kong K, et al. Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. *Lancet Oncol*. 2015;16:541–9.
- Rose DR. The generation of antibody diversity. *Am J Hematol*. 1982;13:91–9.
- Rothenberg EV, Moore JE, Yui MA. Launching the T-cell-lineage developmental programme. *Nat Rev Immunol*. 2008;8:9–21.
- Sainz-Perez A, Lim A, Lemercier B, Leclerc C. The T-cell receptor repertoire of tumor-infiltrating regulatory T lymphocytes is skewed toward public sequences. *Cancer Res*. 2012;72:3557–69.
- Schatz DG, Swanson PC. V(D)J recombination: mechanisms of initiation. *Annu Rev Genet*. 2011;45:167–202.
- Schroeder HW Jr. Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Dev Comp Immunol*. 2006;30:119–35.
- Shugay M, Britanova OV, Merzlyak EM, Turchaninova MA, Mamedov IZ, Tuganbaev TR, et al. Towards error-free profiling of immune repertoires. *Nat Methods*. 2014;11:653–5.
- Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res*. 2017;gkx760–0.
- Silverman GJ, Goodyear CS. Confounding B-cell defences: lessons from a staphylococcal superantigen. *Nat Rev Immunol*. 2006;6:465–75.
- Sing GK, Li D, Chen X, Macnaughton T, Lichanska AM, Butterworth L, et al. A molecular comparison of T lymphocyte populations infiltrating the liver and circulating in the blood of patients with chronic hepatitis B: evidence for antigen-driven selection of a public complementarity-determining region 3 (CDR3) motif. *Hepatology*. 2001;33:1288–98.
- 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.
- Stern JN, Yaari G, Vander Heiden JA, Church G, Donahue WF, Hintzen RQ, et al. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci Transl Med*. 2014;6:248ra107.
- Sui W, Hou X, Zou G, Che W, Yang M, Zheng C, et al. Composition and variation analysis of the TCR beta-chain CDR3 repertoire in systemic lupus erythematosus using high-throughput sequencing. *Mol Immunol*. 2015;67:455–64.
- Sundling C, Li Y, Huynh N, Poulsen C, Wilson R, O'Dell S, et al. High-resolution definition of vaccine-elicited B cell responses against the HIV primary receptor binding site. *Sci Transl Med*. 2012;4:142ra196.
- Teng G, Papavasiliou FN. Immunoglobulin somatic hypermutation. *Annu Rev Genet*. 2007;41:107–20.
- Thapa DR, Tonikian R, Sun C, Liu M, Dearth A, Petri M, et al. Longitudinal analysis of peripheral blood T cell receptor diversity in patients with systemic lupus erythematosus by next-generation sequencing. *Arthritis Res Ther*. 2015;17:132.
- 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;29:542–50.
- Tickotsky N, Sagiv T, Priulsky J, Shifrut E, Friedman N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics*. 2017;33:2924–9.
- Tipton CM, Fucile CF, Darce J, Chida A, Ichikawa T, Gregoretti I, et al. Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus. *Nat Immunol*. 2015;16:755–65.
- Trepel F. Number and distribution of lymphocytes in man. A critical analysis. *Klin Wochenschr*. 1974;52:511–5.
- 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;43:2507–15.
- Turchaninova MA, Davydov A, Britanova OV, Shugay M, Bikos V, Egorov ES, et al. High-quality full-length immunoglobulin profiling with unique molecular barcoding. *Nat Protoc*. 2016;11:1599–616.
- Umemura T, Yoshizawa K, Ota M, Katsuyama Y, Inada H, Tanaka E, et al. Analysis of T cell repertoire in the liver of patients with chronic hepatitis C. *Clin Exp Immunol*. 2000;121:120–6.
- van den Beemd R, Boor PP, van Lochem EG, Hop WC, Langerak AW, Wolvers-Tettero IL, et al. Flow cytometric analysis of the Vbeta repertoire in healthy controls. *Cytometry*. 2000;40:336–45.
- van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 concerted action BMH4-CT98-3936. *Leukemia*. 2003;17:2257–317.
- Vander Heiden JA, Yaari G, Uduman M, Stern JN, O'Connor KC, Hafler DA, et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics*. 2014;30:1930–2.
- 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;110:13463–8.

- von Budingen HC, Kuo TC, Sirota M, van Belle CJ, Apeltsin L, Glanville J, et al. B cell exchange across the blood-brain barrier in multiple sclerosis. *J Clin Invest*. 2012;122:4533–43.
- Wang C, Sanders CM, Yang Q, Schroeder HW Jr, Wang E, Babrzadeh F, et al. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc Natl Acad Sci U S A*. 2010;107:1518–23.
- Wang C, Liu Y, Cavanagh MM, Le Saux S, Qi Q, Roskin KM, et al. B-cell repertoire responses to varicella-zoster vaccination in human identical twins. *Proc Natl Acad Sci U S A*. 2015;112:500–5.
- Wang T, Wang C, Wu J, He C, Zhang W, Liu J, et al. The different T-cell receptor repertoires in breast cancer tumors, draining lymph nodes, and adjacent tissues. *Cancer Immunol Res*. 2017;5:148–56.
- Warren RL, Freeman JD, Zeng T, Choe G, Munro S, Moore R, et al. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res*. 2011;21:790–7.
- Weinstein JA, Jiang N, White RA 3rd, Fisher DS, Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. *Science*. 2009;324:807–10.
- Weng WK, Armstrong R, Arai S, Desmarais C, Hoppe R, Kim YH. Minimal residual disease monitoring with high-throughput sequencing of T cell receptors in cutaneous T cell lymphoma. *Sci Transl Med*. 2013;5:214ra171.
- 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;3:93ra69.
- Wu D, Sherwood A, Fromm JR, Winter SS, Dunsmore KP, Loh ML, et al. High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. *Sci Transl Med*. 2012;4:134ra163.
- Wu J, Jia S, Wang C, Zhang W, Liu S, Zeng X, et al. Minimal residual disease detection and evolved IGH clones analysis in acute B lymphoblastic leukemia using IGH deep sequencing. *Front Immunol*. 2016;7:403.
- Wucherpfennig KW, Ota K, Endo N, Seidman JG, Rosenzweig A, Weiner HL, et al. Shared human T cell receptor V beta usage to immunodominant regions of myelin basic protein. *Science*. 1990;248:1016–9.
- Yaari G, Kleinstein SH. Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med*. 2015;7:121.
- Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res*. 2013;41:W34–40.
- Yu XG, Lichterfeld M, Chetty S, Williams KL, Mui SK, Miura T, et al. Mutually exclusive T-cell receptor induction and differential susceptibility to human immunodeficiency virus type 1 mutational escape associated with a two-amino-acid difference between HLA class I subtypes. *J Virol*. 2007;81:1619–31.
- 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;201:459–72.
- Zvyagin IV, Pogorelyy MV, Ivanova ME, Komech EA, Shugay M, Bolotin DA, et al. Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing. *Proc Natl Acad Sci U S A*. 2014;111:5980–5.