



## OPEN ACCESS

## EDITED BY

Kenji Mizuguchi,  
Health and Nutrition, Japan

## REVIEWED BY

Hiroki Shirai,  
RIKEN Center for Computational  
Science, Japan  
Sofia Kossida,  
Université de Montpellier, France

## \*CORRESPONDENCE

Daron M. Standley,  
standley@biken.osaka-u.ac.jp

## SPECIALTY SECTION

This article was submitted to Drug  
Discovery in Bioinformatics,  
a section of the journal  
Frontiers in Bioinformatics

RECEIVED 15 September 2022

ACCEPTED 11 October 2022

PUBLISHED 20 October 2022

## CITATION

Xu Z, Ismanto HS, Zhou H, Saputri DS,  
Sugihara F and Standley DM (2022),  
Advances in antibody discovery from  
human BCR repertoires.  
*Front. Bioinform.* 2:1044975.  
doi: 10.3389/fbinf.2022.1044975

## COPYRIGHT

© 2022 Xu, Ismanto, Zhou, Saputri,  
Sugihara and Standley. This is an open-  
access article distributed under the  
terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) 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.

# Advances in antibody discovery from human BCR repertoires

Zichang Xu<sup>1</sup>, Hendra S. Ismanto<sup>1</sup>, Hao Zhou<sup>1</sup>, Dianita S. Saputri<sup>1</sup>,  
Fuminori Sugihara<sup>2</sup> and Daron M. Standley<sup>1,3\*</sup>

<sup>1</sup>Department of Genome Informatics, Research Institute for Microbial Diseases, Osaka University, Suita, Japan, <sup>2</sup>Core Instrumentation Facility, Immunology Frontier Research Center, Osaka University, Suita, Japan, <sup>3</sup>Department Systems Immunology, Immunology Frontier Research Center, Osaka University, Suita, Japan

Antibodies make up an important and growing class of compounds used for the diagnosis or treatment of disease. While traditional antibody discovery utilized immunization of animals to generate lead compounds, technological innovations have made it possible to search for antibodies targeting a given antigen within the repertoires of B cells in humans. Here we group these innovations into four broad categories: cell sorting allows the collection of cells enriched in specificity to one or more antigens; BCR sequencing can be performed on bulk mRNA, genomic DNA or on paired (heavy-light) mRNA; BCR repertoire analysis generally involves clustering BCRs into specificity groups or more in-depth modeling of antibody-antigen interactions, such as antibody-specific epitope predictions; validation of antibody-antigen interactions requires expression of antibodies, followed by antigen binding assays or epitope mapping. Together with innovations in Deep learning these technologies will contribute to the future discovery of diagnostic and therapeutic antibodies directly from humans.

## KEYWORDS

antibody, antigen, B cell sorting, B cell receptor (BCR), next generation sequencing (NGS), repertoire analysis, epitope, machine learning

## 1 Introduction

Antibodies, which are the extracellular portion of B cell receptors (BCRs), play a critical role in adaptive immune responses. An antibody consists of two chains, heavy and light, each of which is composed of a constant and a variable region (Figure 1). The six complementarity determining regions (CDR) of the variable region are responsible for binding a specific antigen with high affinity (Pons et al., 2002; Davila et al., 2022). Antibodies are widely used for both disease diagnosis and treatment.

Traditional therapeutic antibody discovery approaches utilized animals, usually mice, to generate polyclonal antibodies against a target antigen. In this approach, candidate monoclonal antibodies (mAbs) are selected and engineered to minimize immunogenicity in humans, while maintaining target specificity and desired pharmacokinetics. The first blockbuster therapeutic antibody (anti-CD3 OKT3), which was engineered in this manner, was approved by the FDA in 1986. Animal-based antibody discovery had a huge impact on the pharmaceutical industry through the 1990's and motivated the

development of new antibody discovery platforms. By the mid-2000's, approximately one-half of therapeutic antibodies were fully human through the use of transgenic mice or phage display platforms utilizing human BCR genes (Nelson et al., 2010; Ju et al., 2020).

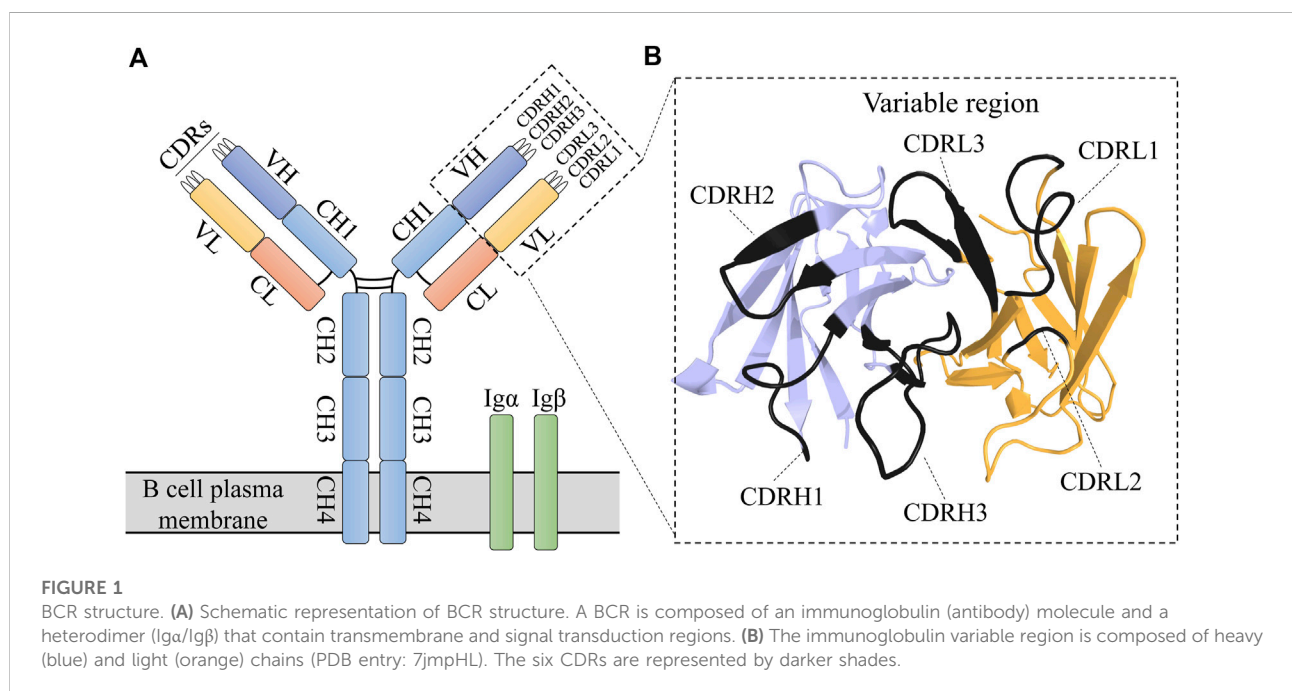
In the past decade, a number of technological breakthroughs have enabled the discovery of antigen-specific mAbs directly from human donors (Pedrioli and Oxenius, 2021). Up to the mid-2000s, mining human B cell receptor (BCR) repertoires for mAbs specific to an antigen of interest was primarily done in academic research labs (Truck et al., 2015; Wang et al., 2015; Goldstein et al., 2019). However, the COVID-19 pandemic brought with it an urgent need for creative ways of targeting the SARS-CoV-2 virus quickly. Remarkably, within months of the pandemic, multiple research groups reported the discovery of neutralizing antibodies from the BCR repertoires of COVID-19 patients (Cao et al., 2020; Hansen et al., 2020; Ju et al., 2020; Pinto et al., 2020; Robbiani et al., 2020; Seydoux et al., 2020; Wang et al., 2020; Zost et al., 2020; Baum et al., 2021). Due to the overwhelming need for a response to the pandemic, along with the rapid availability of resources for COVID-19 related research, many of the mAbs were quickly tested for safety and efficacy in the clinic. The Antibody Society currently lists 35 anti-SARS-CoV-2 mAbs or mAb cocktails undergoing clinical trials (<https://www.antibodysociety.org/covid-19-biologics-tracker>).

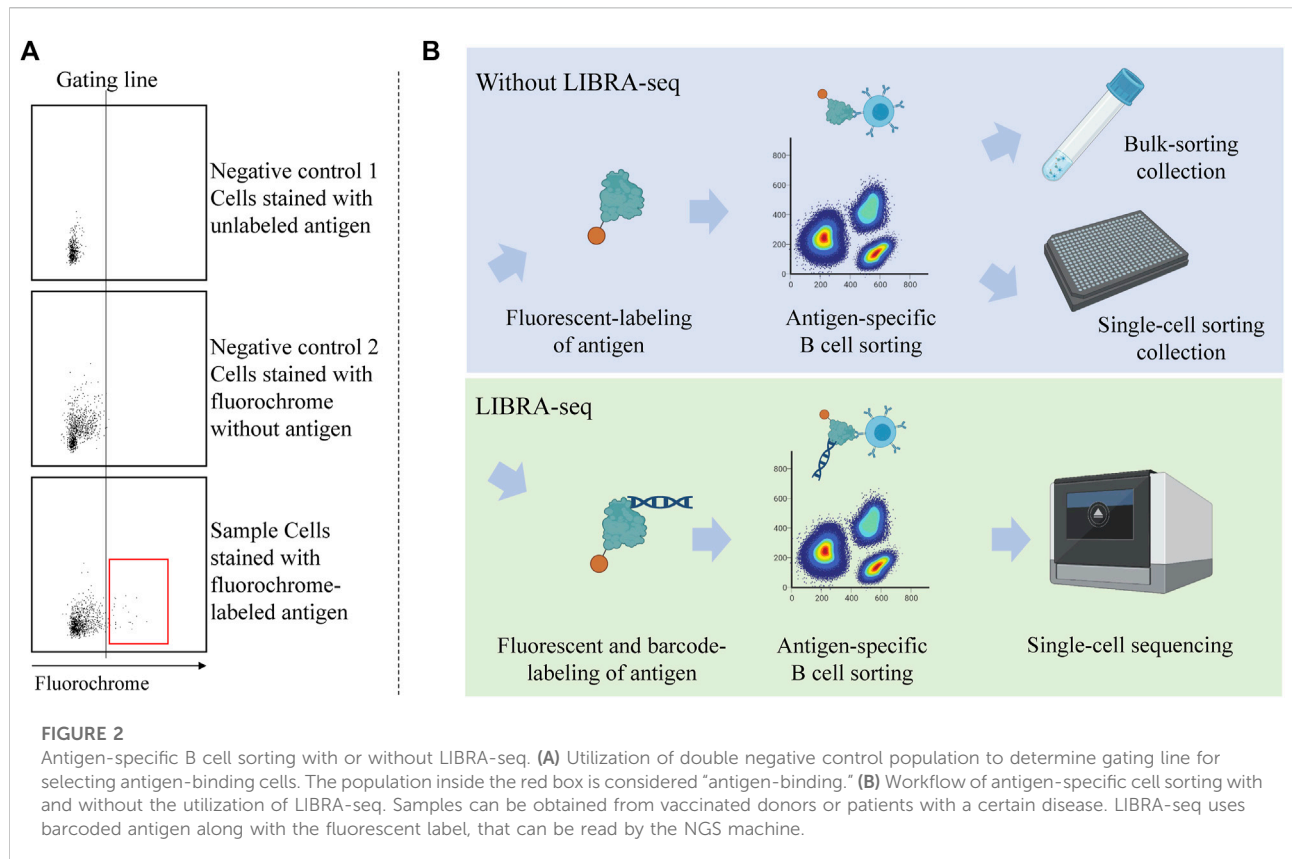
Although it is important not to over-generalize the development of anti-SARS-CoV-2 antibodies to other disease areas, the intensity of research on COVID-19 has refocused attention on the technological innovations that enabled the discovery of antigen-specific antibodies from human BCR

repertoires so quickly. Here we review four main areas of innovation: B Cell sorting, BCR sequencing, BCR repertoire analysis, and experimental validation of antigen binding. Although each of these areas are active research topics on their own, the greatest impact on the pharmaceutical industry will come through synthesis into integrated experimental and computational pipelines. Given the recent breakthroughs in computational biology, including antibody-specific machine-learning methods (Akbar et al., 2022), we can expect rapid growth in this area as data generation merges with data analysis in the context of antibody discovery.

## 2 B cell sorting

A repertoire of BCRs refers to a snapshot of all the B cells produced in a given donor at a given time. When studying repertoires, separating cells of interest by cell sorting is commonly used for isolating natural B cells with a specific phenotype or antigen specificity. This is one of the first steps in discovering antibodies from human donors. Common methods used for cell sorting include FACS (Fluorescence-activated cell sorting), MACS (Magnetic-activated cell sorting), or combinations of both. In FACS, fluorescently-labeled antigens are used as probes to isolate antigen-binding B cells, collect them into tubes or plates, and continue further processes such as bulk or single cell BCR gene amplification and sequencing (Gieselmann et al., 2021). Fluorescent-labeling of an antigen is a critical step and can be done *via* covalent chemical conjugation, expression of a recombinant antigen-fluorescent fusion protein,





or by biotinylating the antigen and adding fluorochrome-conjugated streptavidin to make an antigen tetramer, which increases avidity to the antibody. It must be kept in mind that such labeling may occlude some part of the epitope and potentially disturb the B cell antigen recognition process (Boonyaratanakornkit and Taylor, 2019). Despite this potential complication, utilization of fluorescent-labeled antigens is a promising approach to the collection of antigen-specific B cells.

A relatively new technology, MACS, utilizes direct (primary antibody-conjugated microbeads) or indirect magnetic labeling (primary antibody plus a secondary antibody-conjugated microbead) of cells prior to separation through a magnetic field. Although it lacks sensitivity and is not compatible with multiple-marker profiles, the cell throughput, viability, and time requirements for MACS are comparable to FACS (Sutermaster and Darling, 2019). Some researchers combine these two methods to do enrichment of antigen-specific B cells (Galson et al., 2015a; Banach et al., 2022).

Isolating antigen-specific human B cells is nevertheless quite challenging. Memory B cells express large amounts of antigen receptors on their surfaces but are present in very low numbers in peripheral blood, the most accessible repertoire compartment of the human body (Waltari et al., 2019). The other most commonly-studied subset of antigen-specific B cells consists of

antibody-secreting cells (ASCs). ASCs can be found in higher numbers in peripheral blood, especially after vaccination or infection; however, ASCs, especially those of the IgG isotype, are thought to have limited immunoglobulin surface expression. This might be a reason why many previous studies of ASCs did not utilize antigen-based sorting, but rather collected all ASCs and screened for antigen-specificity downstream after culturing the cells *in vitro* and stimulating antibody secretion before sequencing (Lavinder et al., 2014; Galson et al., 2015a; Acquaye-Seedah et al., 2018; Pedrioli and Oxenius, 2021). However, IgA and IgM isotype ASCs retain expression of surface immunoglobulin (Pinto et al., 2013; Blanc et al., 2016), making it relatively straightforward to sort these subsets in an antigen-specific manner.

Another challenge in antigen-based cell sorting is related to the specificity that is, whether or not the selected cells are truly positive binders or just appear through nonspecific binding to the fluorochrome, streptavidin, or any added linkers (Doucett et al., 2005; Boonyaratanakornkit and Taylor, 2019). During the sorting process, it is necessary to reduce these background signals as much as possible. Due to the limitation of sample quantity and a low number of antigen-specific cells in the sample, we often lack ideal positive control cells from which a positive threshold for fluorescence (gate) can be used to define antigen-binding cells. Thus, in general, one must rely on a negative

control population (which can be cells or decoys that are stained by unlabeled antigen or fluorochrome without the antigen) to set the gate for antigen binders (Figure 2A). Additionally, to increase specificity, a double fluorescent staining strategy can be used to label the antigen probe. Here, an antigen is labeled with two different fluorescent labels and the double positive cells are deemed positive (Amanna and Slifka, 2006). However, this approach still leaves some opportunity for non-binders to be recruited, as seen in a previous report that only 80% of sorted cells were positively bound to the antigen after being recombinantly produced and tested by ELISA (Attaf et al., 2020). This observation suggests that experimental validation (discussed in Section 5) is a requirement for any antibody discovery workflow based on antigen-based B cell sorting. This can be a potential bottleneck, since the production of recombinant antibodies following the acquisition of antibody sequences by conventional cloning and expression in mammalian cells can be labor intensive (Pedrioli and Oxenius, 2021).

To overcome these challenges, and to increase the throughput of antigen-based sorting, the Linking B-cell Receptor to Antigen Specificity through Sequencing (LIBRA-seq) method was introduced in recent years (Setliff et al., 2019). LIBRA-seq is a modification of antigen-based cell sorting that makes use of next-generation sequencing (NGS) technology. In LIBRA-seq, in addition to the fluorescent label, the antigen probe is coupled to a unique DNA barcode that is readable in the sequencing stage. The B cells are thus enriched for antigen-binding cells by FACS; then, the specific antigen is mapped to the B cell by the expression level of the barcode (Figure 2B). This allows simultaneous capture of several antigen probes, tagged by the same fluorescent color but different barcodes. Each cell will have scores for each antigen in the screening library. These scores are a function of the unique molecular identifiers (UMIs) for the respective antigen barcodes (Setliff et al., 2019). Several studies utilized this method to efficiently discover SARS-CoV-2 specific antibodies (He et al., 2021; Kramer et al., 2021; Shiakolas et al., 2021; Kramer et al., 2022; Shiakolas et al., 2022; Suryadevara et al., 2022). The latest version of LIBRA-seq allows epitope mapping by barcoding several variants of the antigen, each with known epitopes mutated (Walker et al., 2022).

### 3 B cell receptor sequencing

Due to the unique phenomenon of gene rearrangement in the generation of BCR coding sequences, BCR diversity at the amino acid sequence level is believed to be in the range of  $10^{16}$ – $10^{18}$  (Briney et al., 2019). With the development of NGS, High-throughput sequencing-based (HTS) sequencing has been used to analyze both T cell receptor (TCR) and BCR repertoires (Yaari and Kleinstein, 2015). The first use of HTS technology for immune repertoire analysis was made by Campbell (Campbell

et al., 2008) in 2008 using the Roche454 platform to explore IGH hypermutation variants carried in patients with chronic B-lymphocytic leukemia at the DNA level. Since this time, a number of new technologies have emerged. These can be divided roughly into two groups: bulk and single-cell sequencing. In bulk sequencing the pairing between heavy and light chains is lost; in single-cell sequencing, this pairing is maintained.

#### 3.1 Bulk B cell receptor sequencing

Bulk sequencing provides in-depth information on the frequency of single chains, which gives a high-resolution view of diversity (a measure of the range and distribution of certain features within a given population (Xu et al., 2020)) and clonal expansion (the proliferation of lymphocytes activated by clonal selection in order to produce a clone of identical cells (Polonsky et al., 2016)), as entire cell populations can be sequenced in a single pipeline (Kovaltsuk et al., 2017). Two starting materials can be used as initial templates for repertoire sequence: genomic DNA (gDNA) and messenger RNA (mRNA). gDNA has the advantage of stability and a constant initial gene copy number between cells (Chaudhary and Wesemann, 2018). mRNA as a template requires reverse transcription, during which UMIs can be added, a step that helps in identifying duplicate or/and cloned sequences generated by PCR, which circumvents PCR bias or sequencing errors (Turchaninova et al., 2016; Rosati et al., 2017). In addition, synthetic repertoires utilize long (~500 bp) oligonucleotide synthesis and high-throughput sequencing to generate a template for every possible V/J combination for minimization of PCR amplification bias, and additional computational normalization to remove residual bias. (Carlson et al., 2013). Multiplex-PCR (m-PCR) and 5'RACE approaches are the two main methods used for amplification. m-PCR has the advantage that only one-step PCR is required, regardless of whether adaptors are included in the primers or not; when the material selected is gDNA, the downstream primers are restricted to several J gene segments, due to the existence of introns (Bashford-Rogers et al., 2014). 5'RACE requires only one set of oligonucleotides, and designing primers from the C gene increases specificity and greatly reduces PCR bias (Yeku and Frohman, 2011), but has a relatively complex workflow for the library building. In addition to BCR information, it is often desirable to obtain the phenotypes or specific subsets of the B cells. Information on immune receptor libraries can be extracted from RNA-seq data, as BCRs are part of bulk RNA-seq data. However, the sensitivity of such an approach is low because of the under-expression of genes at the transcriptional level and also because large-scale RNA-seq usually results in a mixture of cellular gene expression profiles in the sample. Therefore, RNA-seq usually requires pre-targeted protein labeling of cells with fluorescently labeled antibodies to purify the cell types in the sample (Picot et al., 2012).



### 3.2 Single-cell B cell receptor sequencing

To obtain paired heavy-light chain sequences, single B cell resolution is required, since the mRNAs of each chain are physically separate. When coupled with single-cell RNA-seq, BCR sequencing can also provide important phenotype information on the cells (Haque et al., 2017). Goldstein and co-workers showed that single B cell sequencing can recover a higher number of antibody lineages compared to hybridoma technology (Goldstein et al., 2019). Single-cell sequencing has also been used to identify SARS-CoV-2 specific Abs (Woodruff et al., 2020; He et al., 2021), tumor-specific Abs (Buus et al., 2021), and autoimmune disease-specific Abs (Sulen et al., 2020; Jin et al., 2021). Single-cell sequencing is now readily available from several companies, including 10x Genomics and Takara Bio.

Single-cell sequencing combines multiple levels of information, not limited to intracellular gene expression and BCR pairing information; by adding specific oligonucleotide barcode-associated antibodies, thereby allowing surface proteins to be characterized, similar to flow cytometry. Single cells can be isolated in microtiter plates or droplets and then physically linked by overlapping extension RT-PCR in the variable regions of heavy and light chains. Although the potential to obtain BCR pairing information at high throughput has been demonstrated, this technique requires custom equipment and does not yield full-length variable region sequence information (Goldstein et al., 2019). Although full-length sequences can be inferred by the assembly, there is uncertainty in this process (DeKosky et al., 2013; DeKosky et al., 2015; McDaniel et al., 2016). RAGE-seq (repertoire and gene expression by sequencing) combines the genomic technologies of Oxford Nanopore Technologies' long reads, 10x Genomics, Illumina's short reads, and CaptureSeq4 major platforms to enrich RNA from single B cell, and then assembles full-length sequences computationally (Singh et al., 2019).

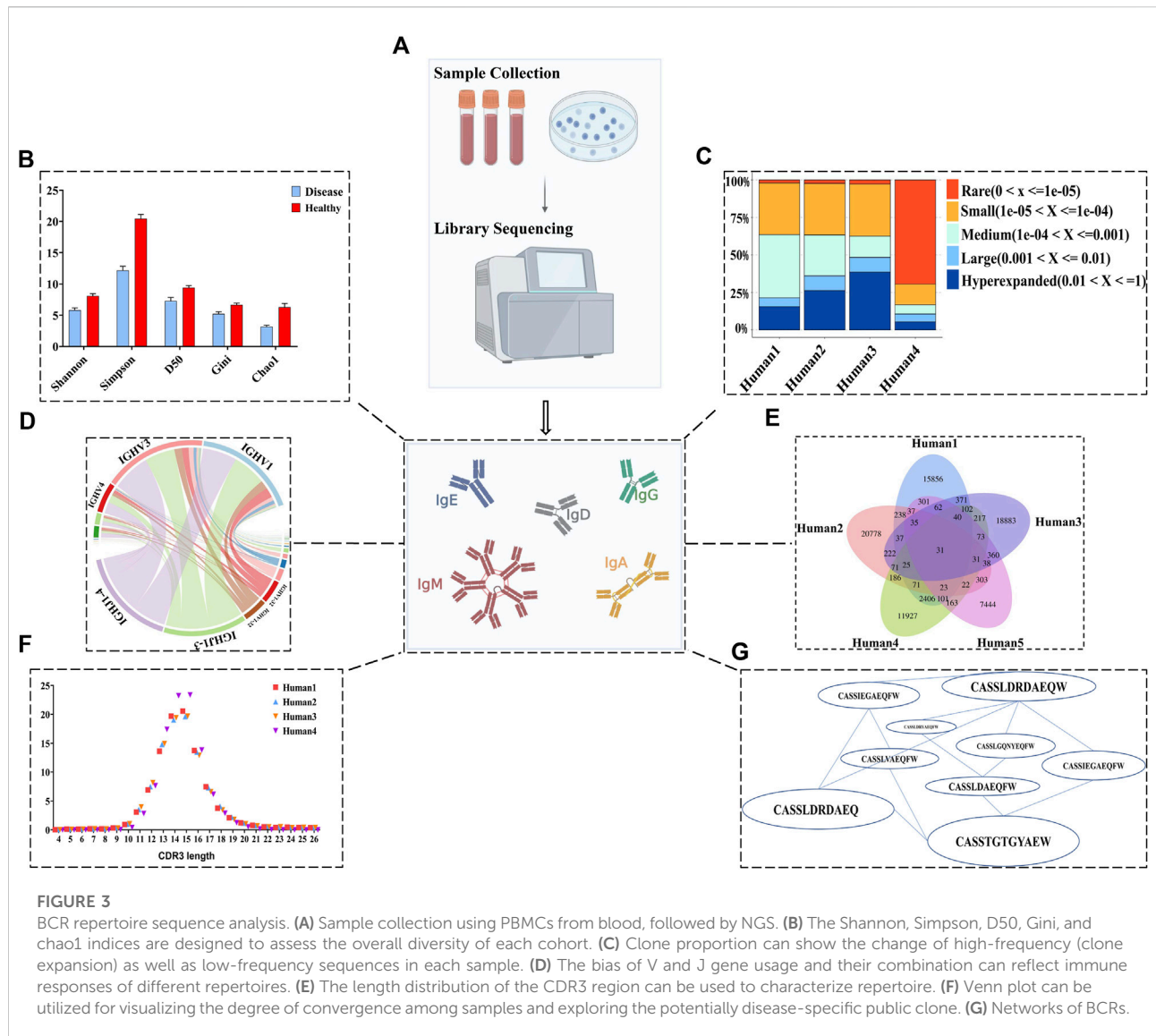
The main limitation to current single-cell sequencing is the tradeoff between sequencing depth and cost. Sequencing depth is the number of transcripts detected from each cell which should be controlled together with the number of cells to get enough coverage (average number of reads that align to specific locus in a reference genome to "cover" reference bases) for confident sequence assignment. The minimum sequencing depth for single cell VDJ analysis is around 5,000 paired reads per cell, while gene expression analysis requires a minimum 20,000 reads per cell, which can be increased depending on needs to analyze a greater number of genes. In comparison with bulk sequencing, which usually obtains 1 million reads per sample, single-cell sequencing depth depends on the desired number of cells in one sample. For example, we recently obtained approximately 20 million reads from one thousand cells with 20,000 reads per

cell (unpublished results). The cost of single-cell sequencing mainly comes from the library preparation step which can be 10–20 times higher than for traditional bulk sequencing. Current developments are focused on how to reduce the cost and increase the sequencing depth of single-cell sequencing (Haque et al., 2017; Upadhyay et al., 2018; Wu et al., 2018).

### 3.3 Annotation of raw sequence data

Annotation includes defining the V, D, and J genes for a given BCR, inferring the accurate amino acid sequence, and assigning the CDR boundaries. These are nontrivial tasks. Several numbering schemes to define CDRs have been proposed including Kabat, Chothia, Martin, Gelfand, IMGT, and AHO (Dondelinger et al., 2018). Meanwhile, several tools have been developed to streamline the process of annotation to use these numbering schemes. For the assignment of CDRs ANARCI is a reliable and user-friendly tool (Dunbar and Deane, 2016). For gene and amino acid assignment, the strengths of the various tools have been systematically discussed in several previous publications (Heather et al., 2018; Lopez-Santibanez-Jacome et al., 2019; Smakaj et al., 2020). Here, we will describe several of tools for the analysis of bulk- and single-cell sequence data. IMGT is one of the most widely used annotation platforms today. High-quality germline sequence information for most species is assembled in IMGT, and therefore reference libraries for the vast majority of sequence annotation tools are derived from IMGT (Lefranc et al., 2005; Manso et al., 2022). In 2011, IMGT developed a platform for HTS T/B repertoire data, supporting raw sequence uploads in FASTA and FASTQ formats (Alamyar et al., 2012; Li et al., 2013). IgBLAST was originally developed as a tool for analyzing immunoglobulin sequences using BLAST, a local alignment method (Ye et al., 2013). Although data can be uploaded directly through the webpage, it does not show many advantages in the analysis of HTS data or presentation of results. MiXCR is another widely used stand-alone package for BCR repertoire annotation, as there is no restriction on the number of sequences. It uses an improved k-mer chaining algorithm for sequence alignment, and an error correction procedure can be performed based on the quality of the sequences (Bolotin et al., 2015).

For scRNA-seq data generated through the 10x Genomics platform, pre-processing similar to bulk sequencing is required before downstream analysis. The company offers Cell Ranger (Zheng et al., 2017), which is recommended for its ability to process both gene expression and paired TCR/BCR data. The development of tools to reconstruct immune repertoire information from single-cell or bulk RNA-seq is an active area, including tools such as BASIC, BRACER, and BALDR (Haque et al., 2017; Upadhyay et al., 2018; Wu et al., 2018).



## 4 B cell receptor repertoire analysis

BCR repertoire sequence data is growing rapidly. In this section, we will first describe methods to analyze diversity, clonal composition, or the specificity of BCRs from different cohorts. Next, we introduce databases and platforms to store repertoire data. Lastly, we will briefly discuss antibody structural modeling and epitope/paratope prediction.

### 4.1 B cell receptor repertoire sequence analysis

Sequence analysis of BCR repertoire data is a rapidly evolving field that can be roughly divided into three main components: diversity, clonal composition (the relative abundance of specific

clones (ImmunoMind, 2019)), and antigen/disease binding specificity. There are a large number of tools and packages that can be used to analyze BCR diversity, clonal frequency, and networks of BCRs. Some well-used tools like Immcantation (Vander Heiden et al., 2014; Gupta et al., 2015), and Immunarch (ImmunoMind, 2019) allows visualization of results after direct import of data from Cell Ranger or MiXCR. Some representative visualizations of results are shown in Figure 3. Many metrics in repertoire analysis are general methods used beyond BCRs or TCRs, such as Shannon and Simpson diversity (Leinster and Cobbold, 2012; Greiff et al., 2015a). Shannon diversity correlates with increasing sequence uniformity, whereas Simpson diversity assigns greater weight to dominant sequences. Clonal abundance is another measurement to quantify sequence distribution that can be used to determine the difference in the ratio of high-frequency to low-frequency sequences between healthy and

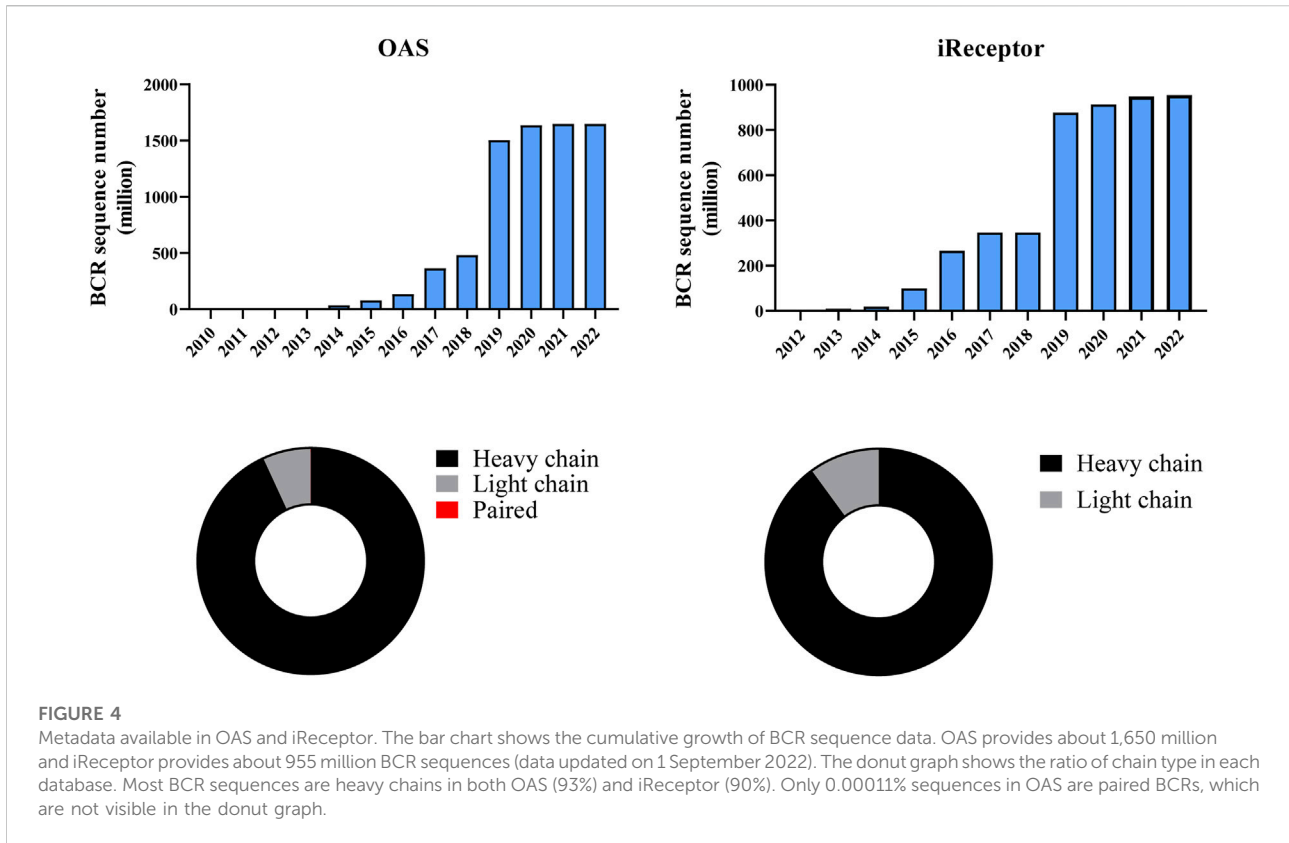
diseased patients (Yaari and Kleinstein, 2015). High expression or low expression of V, D, and J BCR genes can indicate immune responses (Lee et al., 2021; Kotagiri et al., 2022). Meanwhile, the length and amino acid usage of CDR3 regions is often used to characterize repertoires in terms of a few dependent parameters. For example, one study reported that the average CDRH3 lengths of IGHG1, IGHA1, and IGHA2 were significantly greater in COVID-19 patients than healthy cohorts (Galson et al., 2020). Moreover, the same authors identified CDRH3 amino acid sequence signatures within COVID-19 patients with different symptoms (Galson et al., 2020). Also, diversity caused by high-frequency mutations in somatic cells is another important feature of BCR sequences. In general, mutation analysis shows the extent of differentiation compared to germline sequences and indicates antigen-driven affinity maturation. One group utilized the frequency of somatic hypermutation (SHM) of the heavy chain as a feature to identify HIV patients possessing broadly neutralizing antibodies (Roskin et al., 2020).

Identification patterns relating to specific antigens or disease cohorts is a major challenge in BCR repertoire analysis. Previous studies have observed BCR sharing among HIV patients, HBV vaccination donors, Influenza vaccination donors, and COVID-19 patients (Jackson et al., 2014; Galson et al., 2015a; Setliff et al., 2018; Kim et al., 2021; Voss et al., 2021). In order to quantify the convergence (the existence of similar or identical BCR sequences among donors in a common cohort) of BCRs, clonotypes (same V, J gene and identical amino acids in CDR3 region) analysis is widely used in both bulk and single-cell analyses (Soto et al., 2019; Raybould et al., 2021b). Furthermore, clustering such clonotypes (e.g. with 80% or greater sequence identity in their CDR3) is often used (Galson et al., 2020; Nielsen et al., 2020). Similar CDR3 sequences that dominate the immune response in different individuals following antigen stimulation are often referred to as a “convergent” or “public” (Truck et al., 2015). Experimental validation of clustering will be described in detail in section 5.2.

## 4.2 Repertoire databases and data mining

Due to continuous advances in sequencing technology, BCR repertoire sequence data, especially bulk data, has grown rapidly in recent years. Most data associated with published repertoire research is stored in public databases such as the Sequence Read Archive (SRA) or European Nucleotide Archive (ENA), in the form of raw NGS reads. Since SRA and ENA do not allow sequence-level searches, such analysis must be performed on specialized repertoire web servers or by using command-line tools. In this section, we describe a number of databases for antibody sequences, structures or both, that can help in the mining of antigen-specific BCR sequences.

- Observed Antibody Space (OAS) is a comprehensive and frequently updated website and database (Kovaltsuk et al., 2018; Olsen et al., 2022) Although OAS also contains paired data, to our knowledge, it is the first organized collection of bulk BCR sequences. Metadata such as study, species, disease, vaccine, B cell source, and subset can be searched (Figure 4).
- The iReceptor (Corrie et al., 2018) platform allows sharing and comparing adaptive immune receptor repertoire (AIRR)-seq data. It has two key components: a data repository that focuses on AIRR data, and a web-based Scientific Gateway that allows researchers to discover, federate, explore, and analyze AIRR-seq data (Figure 4).
- SAbDab (Dunbar et al., 2014) is a frequently updated resource containing all publicly available antibody structures and, similar to OAS, is convenient to search using metadata, including species, experimental method, resolution, or amino acid at a given position using canonical numbering.
- IMGT/3Dstructure-DB (Ehrenmann et al., 2010) is a three-dimensional structure database of IMGT entries that stores the structures of immunoglobulins, TCRs, and major histocompatibility complex proteins of humans and other vertebrate species. A related database, IMGT/2Dstructure-DB, stores the amino acid sequences from INN/WHO and Kabat databases (Ehrenmann and Lefranc, 2011). IMGT/3Dstructure-DB contains 8,437 entries as of 3 June 2022.
- huARdb (Wu et al., 2022) is a versatile and user-friendly web interface consisting of data from 444,794 high confidence T or B cells with full-length TCR/BCR sequences and transcriptomes from 215 datasets, which have been subjected to a uniform workflow.
- PIRD (Zhang et al., 2020) is a multi-species BCR dataset that contains 5 main information modules, including project information, sample information, raw sequencing data, annotated TCR or BCR repertoires, and a database of TCRs and BCRs targeting known antigens (TBAbd). PIRD can also carry out analyses, including biased gene usage, the length distribution of CDR3, and the diversity index for each dataset directly.
- ImmPort (Bhattacharya et al., 2018) is one of the largest repositories of open immunology data. It hosts data from more than 300 clinical and mechanistic studies in humans and immunological studies on model organisms, categorized as Private Data, Shared Data, Data Analysis, and Resources, with a focus on allergy, autoimmune disease, infection response, transplantation, and vaccine response.
- cAb-Rep (Guo et al., 2019) contains 306 immunoglobulin repertoires from a database consisting of 121 healthy, vaccinated, or autoimmune disease donors. The database contains 267.9 million IGH and 72.9 IGL full- or nearly



full-length transcripts that have been annotated according to isotype, somatic hypermutation (SHM), and other biological characteristics.

- ImmuneDB (Rosenfeld et al., 2017; Monasterio et al., 2018) is a high-throughput immune receptor sequencing data system that integrates data storage and analysis. The developers demonstrated that ImmuneDB and MiXCR have comparable performance in annotating raw data. Output includes selection pressure, lineage mapping, novel allele detection, etc. ImmuneDB states that their method can quickly identify more potential sequences compared with IMGT/High-Vquest and that IT performs similar to MiXCR on the same input data.
- VDJSer (Christley et al., 2018) integrates large data storage and analysis. The advantage of VDJSer is that sequence annotations can be performed after quality control screening of raw data.
- Our lab has recently launched InterClone (Wilamowski et al., 2022), a resource that contains both BCR and TCR repertoire data, along with tools to store, search or cluster the data. Distinguishing features of InterClone include: the ability of users to control the visibility of their data; efficient encoding of CDR regions to allow flexible searches or clustering using user-specified similarity thresholds for CDRs; a large amount of

BCR data, particularly for COVID-19, influenza, HIV, and healthy donors.

- There are also databases created for specific diseases. CoV-AbDab (Raybould et al., 2021a) currently contains 10,005 antibodies and nanobodies from published papers/patents that bind to at least one betacoronaviruses (last updated: 26th July 2022). This database is the first known integration of antibodies that bind SARS-Cov2 and other betacoronaviruses, including SARS-CoV1 and MERS-CoV. It contains evidence of cross-neutralization, the origin of antibody nanobodies, full-length variable structural domain sequences, germline assignments, epitope regions, PDB codes (if relevant), homology models, and literature references.
- CATNAP (Yoon et al., 2015) is a web server for HIV data, including antibody sequences from the authors' own and published studies. As input, users can select specific antibodies or viruses, a panel from published studies, or search using local data. The output overlays neutralization panel data, viral epidemiology data, and viral protein sequence comparison on a single page with further information and analysis. Users can highlight alignment positions, or select antibody contact residues and view position-specific information from the HIV database.



### 4.3 Antibody structure prediction

Protein structure prediction is one of the areas of computational biology that has progressed most rapidly in recent years, owing to breakthroughs in Deep learning (Baek et al., 2021; Jumper et al., 2021). Before these breakthroughs, a plethora of antibody modeling tools existed that performed similarly well for all regions except CDRH3 loops (Almagro et al., 2014). However, it is likely that going forward, all state-of-the-art methods for antibody modeling will utilize some aspects of current Deep learning-based protein structural modeling methods. An important first step in this direction is DeepAb, which convincingly out-performed traditional template-based methods, including our own Repertoire Builder, in terms of antibody structural accuracy (Ruffolo et al., 2022). In a recent assessment, we found that the average CDRH3 root-mean-square deviation (RMSD) dropped from 4.38 to 3.44 Å for AlphaFold compared with our own, previously state-of-the-art, Repertoire Builder in a large and diverse set of 620 antibodies (Xu et al., 2022). Therefore, for antibodies without bound antigens, the current Deep learning approach appears to be a significant improvement.

Unfortunately, it is becoming accepted that the multi-chain extension of AlphaFold, AlphaFold multimer, does not work well for antibody-antigen complexes (Evans et al., 2022). This problem probably arises in part from the fact that AlphaFold uses overall sequence similarity to construct multiple sequence alignments, which are, in turn, used as feature vectors. Many antibodies that target different antigens will be aligned in this process, resulting in a noisy signal. Indeed, when we assessed the complex modeling performance of AlphaFold multimer using a small benchmark of 25 antibody-antigen complexes, we found that the vast majority were docked to the wrong epitope (Standley et al., 2022). Therefore, it will be interesting to see if a more careful selection of sequences and structural templates within the Deep learning workflow will lead to more coherent antibody-antigen complex modeling. CDR-based clustering is one of the functions repertoire databases (see Section 4.2) can provide. Another interesting direction is to couple antibody-antigen complex modeling with epitope prediction.

### 4.4 Epitope prediction

As of 19 July 2022, there were only 9,811 recorded antibody-antigen structures available in the Protein Data Bank (Berman et al., 2008; Raybould et al., 2020). Due to the time-consuming and labor-intensive process of experimental methods to investigate antibody-antigen interactions experimentally (see Section 5), there is a need for computational approaches that can quickly predict the epitope and paratope from sequence or structure information. Compared to the difficulty of epitope prediction (where almost any surface patch of antigen could

be an epitope for some antibody), the paratope prediction problem is relatively easy. Most paratopes are located within the six CDRs in the variable fragment of heavy and light chains. Many published tools like Parapred (Liberis et al., 2018) and proABC-2 (Ambrosetti et al., 2020) can achieve satisfactory performance in paratope prediction. Thus, in this section, we will focus on epitope prediction.

In recent decades, many tools have been developed in order to predict continuous/linear B-cell epitopes using antigen sequence information or discontinuous/conformational B-cell epitope using antigen structure information. These methods generally adopt machine learning approaches (support vector machines, random forests, linear regression, and neural networks) to learn epitope features from known complex structures (Table 1). One problem with many epitope prediction tools is that they only use features of the antigen, whereas we are generally interested in antibody-specific epitopes (Sela-Culang et al., 2015). The direction to solve this problem is to introduce antibody features into the process of epitope prediction. Some tools, including PECAN (Pittala and Bailey-Kellogg, 2020) and Pinet (Dai and Bailey-Kellogg, 2021) used Deep learning to extract antibody and antigen features for use in epitope prediction. Other tools, like EpiPred (Krawczyk et al., 2014), MAbTope (Bourquard et al., 2018), and AbAdapt (Davila et al., 2022), incorporate antibody-antigen docking-based features; these studies have demonstrated that the inclusion of the antibody features improves epitope prediction. As expected, antibody-antigen docking is sensitive to antibody model quality (Davila et al., 2022). We recently incorporated the more accurate antibody models produced by AlphaFold (Jumper et al., 2021) into the AbAdapt pipeline (Xu et al., 2022). We observed significant improvement in docking, paratope prediction, and antibody-specific epitope prediction compared with the default AbAdapt pipeline. In a realistic case, using an anti-SARS-CoV-2 RBD antibody complex benchmark, the use of AlphaFold resulted in higher epitope prediction accuracy than all other tested tools.

It is also worth noting that the combination of different deep or machine learning models is becoming a general trend. A Deep learning framework was developed to extract local features around target residues and global features of the full antigen sequence using Graph Convolutional Networks (GCNs) and Attention-Based Bidirectional Long Short-Term Memory (Att-BLSTM) networks separately (Lu et al., 2022). The local and global features from two networks were combined to predict the epitope and demonstrate that global features play a critical role in structure-based epitope prediction (Lu et al., 2022). Moreover, recent work introduced general protein language models that not only focus on the reported antigen-antibody complex to capture binding patterns, but also used the deep transformer based protein language model, ESM-1b (Rives et al., 2021), to achieve more accurate epitope prediction only using the

TABLE 1 Summary of the epitope prediction tool.

Catalog	Names	Availability	Method	References
linear B-cell epitope	ABCPred	<a href="https://webs.iitd.edu.in/raghava/abcpred/index.html">https://webs.iitd.edu.in/raghava/abcpred/index.html</a>	Recurrent neural network	Saha and Raghava (2006)
linear B-cell epitope	AAPred	<a href="https://www.bioinf.ru/aappred/">https://www.bioinf.ru/aappred/</a>	Support vector machine	Davydov and Tonevitsky (2009)
linear B-cell epitope	FBCPred/BCPREDS	<a href="https://ailab.cs.iastate.edu/bcpreds/">https://ailab.cs.iastate.edu/bcpreds/</a>	Two machine learning approaches	El-Manzalawy et al. (2008)
linear B-cell epitope	COBEpro	<a href="https://scratch.proteomics.ics.uci.edu">https://scratch.proteomics.ics.uci.edu</a>	Support vector machine	Sweredoski and Baldi (2009)
linear B-cell epitope	BepiPred-2.0	<a href="https://services.healthtech.dtu.dk/service.php?BepiPred-2.0">https://services.healthtech.dtu.dk/service.php?BepiPred-2.0</a>	Random forest	Jespersen et al. (2017)
linear B-cell epitope	Lbtope	<a href="http://crdd.osdd.net/raghava/lbtope/">http://crdd.osdd.net/raghava/lbtope/</a>	Support vector machine and k-nearest neighbor	Singh et al. (2013)
linear B-cell epitope	DRREP	<a href="https://github.com/CorticalComputer/DRREP">https://github.com/CorticalComputer/DRREP</a>	Deep neural network	Sher et al. (2017)
linear B-cell epitope	SVMTriP	<a href="http://sysbio.unl.edu/SVMTriP">http://sysbio.unl.edu/SVMTriP</a>	Support vector machine	Yao et al. (2012)
linear B-cell epitope	LBEEP	<a href="https://github.com/brsaran/LBEEP">https://github.com/brsaran/LBEEP</a>	Support vector machine and AdaBoost-random forest	Saravanan and Gautham (2015)
linear B-cell epitope	EPMLR	<a href="http://www.bioinfo.tsinghua.edu.cn/epitope/EPMLR/">http://www.bioinfo.tsinghua.edu.cn/epitope/EPMLR/</a>	Multiple linear regression	Lian et al. (2014)
linear B-cell epitope	iBCE-EL	<a href="http://thegleelab.org/iBCE-EL">http://thegleelab.org/iBCE-EL</a>	Randomized tree and gradient boosting classifiers	Manavalan et al. (2018)
linear B-cell epitope	iLBE	<a href="http://kurata14.bio.kyutech.ac.jp/iLBE/">http://kurata14.bio.kyutech.ac.jp/iLBE/</a>	Random forest	Hasan et al. (2020)
linear B-cell epitope	EpiDope	<a href="http://github.com/mcollatz/EpiDope">http://github.com/mcollatz/EpiDope</a>	Deep neural network	Collatz et al. (2021)
Conformational B-Cell epitope	EliPro	<a href="http://tools.iedb.org/ellipro/">http://tools.iedb.org/ellipro/</a>	Clustering of neighboring residues based on protrusion index	Ponomarenko et al. (2008)
Conformational B-Cell epitope	PEPITO	<a href="http://pepito.proteomics.ics.uci.edu/">http://pepito.proteomics.ics.uci.edu/</a>	Linear combination	Sweredoski and Baldi (2008)
Conformational B-Cell epitope	CBTOPE	<a href="http://www.imtech.res.in/raghava/cbtope/">http://www.imtech.res.in/raghava/cbtope/</a>	Support vector machine	Ansari and Raghava (2010)
Conformational B-Cell epitope	DiscoTope 2.0	<a href="https://services.healthtech.dtu.dk/service.php?DiscoTope-2.0">https://services.healthtech.dtu.dk/service.php?DiscoTope-2.0</a>	Epitope propensity scores	Kringelum et al. (2012)
Conformational B-Cell epitope	SEPPA 3.0	<a href="http://www.badd-cao.net/seppa3/index.html">http://www.badd-cao.net/seppa3/index.html</a>	Logistic regression and clustering coefficient	Zhou et al. (2019)
Conformational B-Cell epitope	CluSMOTE	<a href="https://github.com/BSolihah/conformational-epitope-predictor">https://github.com/BSolihah/conformational-epitope-predictor</a>	Support vector machine and decision tree	Solihah et al. (2020)
Combining antibody feature	EpiPred	<a href="http://opig.stats.ox.ac.uk/webapps/newsabdab/sabpred/epipred/">http://opig.stats.ox.ac.uk/webapps/newsabdab/sabpred/epipred/</a>	Combing the conformational matching of structures and a specific score	Krawczyk et al. (2014)
Combining antibody feature	MABtope	Lead corresponding contact	Integration of docking-based prediction method and experimental steps	Bourquard et al. (2018)
Combining antibody feature	PECAN	<a href="https://github.com/vamships/PECAN">https://github.com/vamships/PECAN</a>	Paratope and epitope prediction with graph convolution attention network	Pittala and Bailey-Kellogg (2020)
Combining antibody feature	Pinet	<a href="https://github.com/FTD007/Pinet">https://github.com/FTD007/Pinet</a>	Geometric deep neural network	Dai and Bailey-Kellogg (2021)
Combining antibody feature	AbAdapt	<a href="https://sysimm.org/abadapt/">https://sysimm.org/abadapt/</a>	Combining docking-based features to predict antibody-specific epitope	Davila et al. (2022)

antigen sequence information in BepiPred-3.0 (Clifford et al., 2022). Recently, Robert and co-workers used simulated antibody-antigen data in order to circumvent the lack of experimentally-determined antibody-antigen structure complexes and focused on the challenging problem of learning antigen and epitope specificity features from antibody sequences (Robert et al., 2022). The potential advantage of these later methods is that they circumvent the time-consuming docking step.

## 5 Experimental validation

### 5.1 Epitope discovery

A prerequisite to therapeutic antibody discovery is to identify the epitope for a given antibody-antigen pair. There are several well-established experimental approaches to elucidate the epitope information including X-ray crystallography, nuclear magnetic resonance (NMR), peptide-based microarrays,

mutagenesis, and cryo-electron microscopy (Cryo-EM). X-ray crystallography is the gold standard to determine the precise binding between antibody and antigen (Holcomb et al., 2017). However, X-ray crystallography has disadvantages in terms of throughput and cost; moreover, flexible or membrane-bound antigens are notoriously difficult to crystallize (Abbott et al., 2014). Nuclear magnetic resonance can also be utilized to obtain detailed epitope mapping information (Blech et al., 2013). But this method has relatively low sensitivity, requires high purity and solubility, and small size for the proteins (Wüthrich, 1990; Pan et al., 2016). Although peptide-based microarrays are high-throughput, and can sometimes identify epitopes with high sensitivity, peptide-based microarray performance is limited by various factors: affinity of the peptides, immobilization methods, and conformational constraints induced by the immobilization. Furthermore, the limitation of linear epitopes is a major concern (Qi et al., 2019). Mutagenesis allows the investigation of epitopes without the need for structure determination. For example, using the alanine shotgun approach, epitopes for difficult proteins such as membrane proteins can be quickly identified. One disadvantage of this approach is that it is hard to clarify whether the mutation has disrupted the folding (Peng et al., 2011).

Due to the combined requirement to quickly and precisely identify epitopes, many studies combine traditional epitope discovery strategies with cutting-edge technologies. Antibody binding epitope Mapping (AbMap), a combination of phage-displayed peptide libraries with next-generation sequencing, was developed to determine 200 antibody-specific epitopes in a single run (Qi et al., 2021). Additionally, microarrays consisting of 648 overlapping peptides that cover the four major structural proteins of the SARS-CoV-2 virus have been constructed: Spike, Nucleocapsid, Membrane, and Envelope (Hotop et al., 2022). This microarray fingerprint of positive serum samples was learned by a machine learning model and the epitopes were used to diagnose COVID-19 positive and negative donors (Hotop et al., 2022).

Among the wide range of experimental epitope mapping methods, we will focus on two promising approaches: Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS) and Deep Mutational Scanning (DMS), which have the potential to perform medium-throughput epitope discovery.

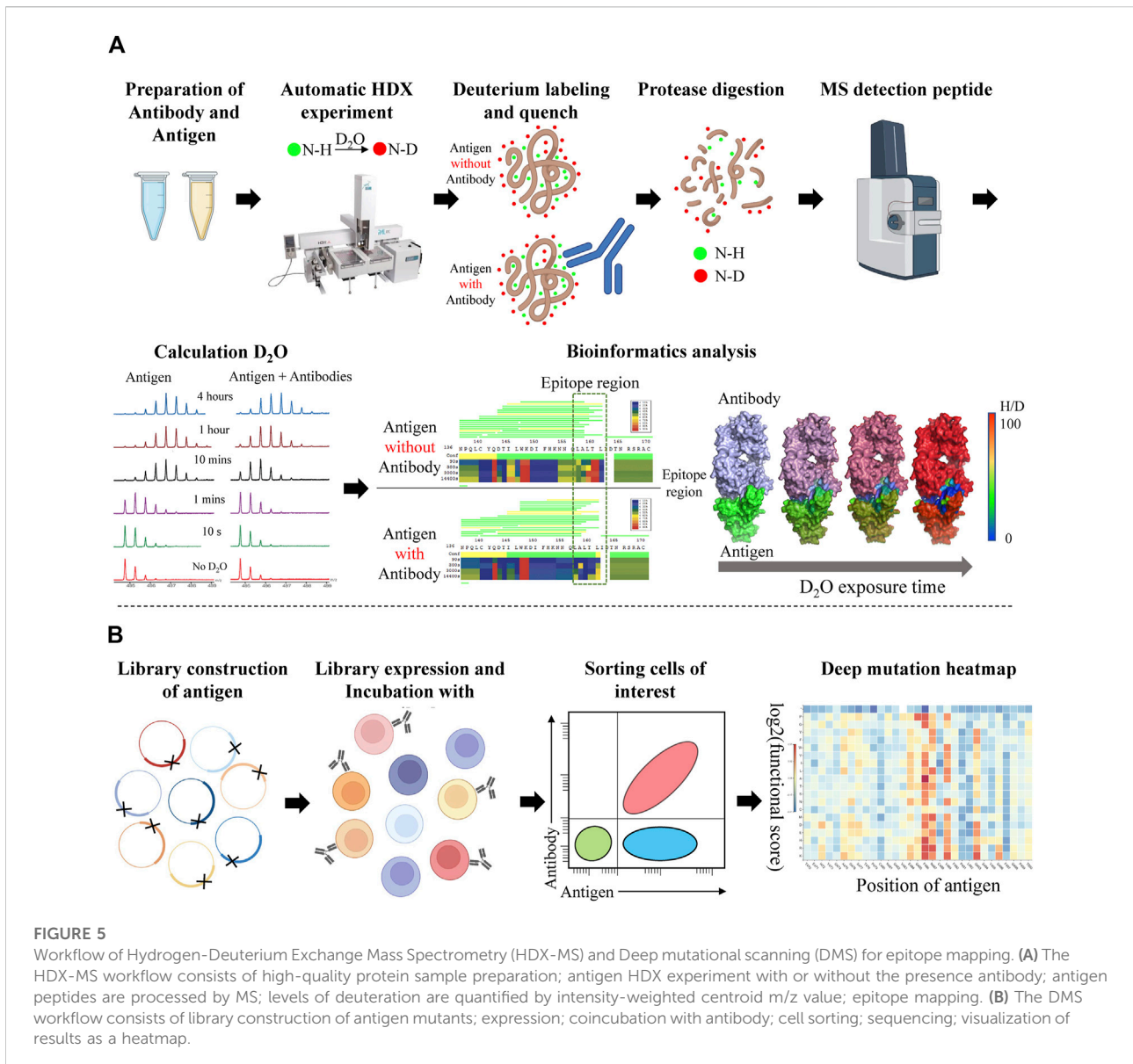
HDX-MS measures changes in the mass of a protein by isotope exchange between amide hydrogens of the protein backbone and its surrounding solvent. The folded state of the protein and its dynamics will affect the rate of this exchange (Masson et al., 2019). In recent years, HDX-MS has been increasingly used for epitope and paratope mapping of antibody-antigen complexes due to its speed and small sample size requirements, and insensitivity to protein size. A semi-automated HDX-MS workflow was used to perform epitope mapping of Fab-CR6261 with diverse influenza Hemagglutinin subtypes (Puchades et al., 2019). Similarly, an in-house HDX-MS

system was constructed to explore the binding of birch Bet v1 protein, a native pollen allergen, in the presence of four antibodies that target non-redundant epitopes (Zhang et al., 2018). Two uncontentious epitope loops of TL1A with anti-TL1A monoclonal antibody 1 were identified by HDX-MS (Huang et al., 2018).

The routine workflow of HDX-MS epitope mapping is performed by using antigen alone as a reference and in the presence of antibodies. The antigen and antibody are labeled in D<sub>2</sub>O buffer under equilibrium conditions at several time points. Compared with antigen alone, the contact of antibody lowers the solvent exposure of antigen residues in the epitope region and leads to the reduction of deuterium incorporation. After protease digestion, proteolytic peptides are desalted and separated on a mass spectrum (MS) system. The fingerprint of antigen alone and antigen-antibody complex will be captured and analyzed by downstream bioinformatics analysis (Figure 5A) (Tran et al., 2022). However, HDX-MS also has some limitations. The major limitation is that HDX-MS can only capture peptide-level information and the individual residue contribution among peptides remains uncertain. Another limitation is the insufficient sequence coverage of peptides spanning the whole protein sequence (Masson et al., 2019). HDX-MS also can't capture the information of prolines which do not have an amide hydrogen group for deuterium exchange (Huang et al., 2018). Recent studies have incorporated peptide-level information from HDX-MS with antibody-antigen docking to overcome the drawbacks of either method alone (Bennett et al., 2019; Fields et al., 2021).

DMS makes use of massive (typically 1 million) mutant versions of a protein in a single experiment to reveal their intrinsic properties by analyzing large-scale phenotype readouts (Fowler and Fields, 2014). By incorporating NGS, the DMS method can observe the effect of individual mutants in a large population. A typical DMS workflow for epitope mapping includes library construction and mutation design of the antigen; library expression and incubation with antibody; sorting cells of interest by FACS or measuring the binding affinity; sequencing the selected mutations and constructing the data of deep mutation heatmap through bioinformatics analysis. (Figure 5B).

DMS has been used to investigate various disease-related antigens. In one study, DMS was utilized to precisely map the epitopes of a panel of cross-neutralizing nanobodies against H1N1 and H5N1 (Gaiotto and Hufton, 2016). In another study, functional constraints and comprehensive mutations of the Zika virus envelope (E) protein were constructed and the effects of viral growth as well as viral neutralization by two monoclonal antibodies were measured (Sourisseau et al., 2019). Additionally, a platform that combines immunoprecipitation of phage peptide libraries and DMS (Phage-DMS) was constructed. Through Phage-DMS, the authors designed all possible amino acid variants of the HIV Envelope and performed fine mapping of epitopes using four well-



characterized HIV antibodies (Garrett et al., 2020). In a recent report, all mutations to the SARS-CoV-2 RBD were first measured by DMS and the effect of expression and affinity for ACE2 were also evaluated (Starr et al., 2020). Meanwhile, DMS was also used to systematically mutate Wuhan-Hu-1, Alpha, Beta, Delta, and Eta variant RBDs and identified some substitutions that cause epistatic shifts during viral evolution (Starr et al., 2022). Also, many studies have reported the utilization of DMS to investigate hotspots of SARS-CoV-2 RBD that enable escape from neutralizing antibodies (Greaney et al., 2021a; Greaney et al., 2021b; Starr et al., 2021; Tsai et al., 2021). These applications convincingly demonstrate that DMS can facilitate the understanding of antigen function and systematically evaluate antibody escape.

In addition to epitope analysis, DMS can be applied to antibodies themselves to identify paratopes or for optimizing other phenotypes. In one case, DMS was used to identify many affinity-enhancing mutations at the variable light-heavy chain interface of an anti-lysozyme antibody; a variant with tenfold higher affinity as well as substantially improved stability were identified (Warszawski et al., 2019). Furthermore, a fully automated design protocol, AblIFT, was established for improving molecular interactions across the variable light-heavy interface and applied to anti-VEGF/QSOX1 antibodies to improve affinity, stability, and expression (Warszawski et al., 2019). In another application, DMS was combined with Deep learning to optimize the affinity, viscosity, clearance, solubility, and immunogenicity of trastuzumab (Mason et al., 2021).



Recently, by leveraging DMS technology, researchers engineered a nanobody initially specific for SARS-CoV-1 RBD in order to bind SARS-CoV-2 RBD (Laroche et al., 2022). Our group contributed to the DMS-based engineering of an ACE2 decoy that could neutralize the SARS-CoV-2 Omicron variant and proved the decoy prevented escape for each single-residue mutation in the RBD of SARS-CoV-2 (Ikemura et al., 2022). We also constructed a database, SpikeDB, that provides changes in infectivity, antigenic escape, ACE2 affinity, and protein expression caused by point mutations in the spike protein of SARS-CoV-2 using DMS (Ikemura et al., 2022).

## 5.2 Validation of repertoire data mining

Repertoire sequencing is generating large amounts of human BCR data. However, most published BCR sequences lack information about targeted antigen or epitope. The development of antigen-specific B cell sorting technologies such as LIBRA-seq could solve the problem of assigning antibodies to their antigens (see section 2) (Setliff et al., 2019). Although antigen-specific B cell sorting is quite a powerful approach, it requires recombinant antigens and specialized cell sorting techniques that are difficult to scale to large numbers of antigens. Computational approaches to identifying antigen-specific antibodies are one way of simplifying very large repertoire sequence data sets.

Various methods are being developed to cluster antibodies that target the same antigen and epitope. Some approaches use information from antibody sequence only, while others use structure information, if available. (Galson et al., 2015b; Xu et al., 2019; Ripoll et al., 2021; Wong et al., 2021). Here, we will focus on sequence-based approaches to search for antibodies with similar target antigens and epitopes.

Clustering of clonotypes was the first method used to group antibodies that possibly target the same antigen (Reddy et al., 2010; Zhu et al., 2013; Galson et al., 2015b; Greiff et al., 2015b). This approach assumes that antibodies with the same V and J genes as well as a given CDR3 amino acid sequence identity (e.g. 80%–100%) in the heavy chain are more likely than other BCRs to target the same antigen and epitope (Galson et al., 2015b; Truck et al., 2015; Soto et al., 2019). Clustering of clonotypes can be applied to single-chain (usually heavy chain) or heavy and light chain paired data (Raybould et al., 2021b). A recent study assembled approximately 8,000 published COVID-19 antibodies from more than 200 donors and demonstrated that antibodies binding to SARS-CoV-2 spike RBD, NTD or S2 possessed distinct convergent clonotype features (Wang et al., 2022). Such clonotype clusters are generally restricted to BCRs with the same CDR3 length (Satpathy et al., 2015); Since many antibodies whose CDR3 length differs by 1-2 amino acids have been found to target the same epitope (D'Angelo et al., 2018; Wong et al., 2019), there is benefit in adding flexibility to

BCR clustering. Moreover, antibodies with different V and J genes or with CDR3 sequence identities below 80% have been found to target the same anti-SARS-CoV-2 RBD epitopes (Barnes et al., 2020; Dejnirattisai et al., 2021) or NTD epitopes (Liu et al., 2021). In the case of the human antibody repertoires, overlap between donors as defined by clonotyping antibody sequences is about 0.3% among three healthy adult donors and 0.1% among three cord blood samples (Soto et al., 2019). In order to increase sensitivity, our group constructed a clustering tool on the InterClone web server that provides a more flexible thresholds CDR similarity (Wilamowski et al., 2022). This method assumes that antibodies within a CDR sequence identity are more likely to target the same epitope. A detailed explanation of this method and the process of a realistic application are described below.

Recently, two groups discovered a set of 11 SARS-CoV-2 infection enhancing antibodies (Li et al., 2021; Liu et al., 2021). We sought to identify such infection-enhancing antibodies in a large-scale antibody repertoire sequence data from COVID-19 patients and healthy donors (Ismanto et al., 2022). Because enhancing antibodies bind their antigen primarily *via* their heavy chain, as captured in the Cryo-EM structure data (PDB ID 7LAB, 7DZX, 7DZY), we focused on heavy chain CDRs. Moreover, since we did not know in advance the safest sequence identity threshold to use for each CDR, we used 80% for CDRH1 and CDRH2 and 60% for CDRH3. Although we could find antigen binders within these thresholds, the false positive rate was quite high (more than 80%) (Ismanto et al., 2022). A safer threshold seems to be 90% for CDRH1 and CDRH2 and 70% for CDRH3. Donor antigen exposure was a critical factor in the false positive rate. We found that the true enhancing antibody rate was approximately 100 times higher in COVID-19 patients than in healthy donors. Unlike other web servers (e.g., Vidjil, AbYsis, OAS, etc.) (Duez et al., 2016; Swindells et al., 2017; Olsen et al., 2022), InterClone hosts a large database of BCR and TCR sequences and allows such flexible search or clustering operations on the stored data. InterClone also allows users to control data visibility.

## 6 Future perspectives

Human BCR repertoires are shaped by antigen exposure. A wide range of diseases, from infection, cancer, and autoimmunity, can shape our repertoires. Aging also has a profound effect on BCR (and TCR) diversity. For these reasons, BCR and TCR repertoires have attracted much attention as potential biomarkers for health, disease, vaccination, or other therapeutic activity. In recent months, two groups have reported the ability to clearly separate COVID-19 patients based on BCR repertoires (Ortega et al., 2021; Chen et al., 2022), and our own, unpublished findings support this ability. Therefore, it is reasonable to anticipate a new



generation of biomarkers based on BCR repertoires that will drive forward technology in the four main areas discussed here (sorting, sequencing, analysis, and validation).

One of the distinguishing features of BCR-based biomarkers from conventional biomarkers is that the antibodies encoded by the BCR sequences are directly involved in the prevention or (as in the case of autoimmunity) mediation of disease. This implies that the downstream application of BCR-based biomarkers is a new generation of therapeutics that closely resemble our body's own defence mechanisms. Computational methods, in particular the ability to accurately predict antibody-antigen interactions from repertoire data, will make a critical difference in these efforts. Given the recent breakthroughs in Deep learning, there is thus much cause for optimism in the coming years for repertoire-based biomarkers and therapeutics.

One topic we have not covered is antibody delivery. Antibodies are traditionally administered directly as proteins. Moreover, mRNA vaccination has proved to be a robust and safe way to induce neutralizing antibodies against COVID-19 (Chaudhary et al., 2021). mRNA vaccinations are now being designed against various antigen targets including Zika, influenza virus, CMV, Respiratory syncytial virus, Ebola, and HIV. At present, most mRNA vaccines encode one or more antigens (Barbier et al., 2022). Two studies have explored the feasibility of expressing antibodies against Respiratory syncytial virus and HIV *via* mRNA vaccination (Tiwari et al., 2018; Lindsay et al., 2020). In one study, researchers expressed whole palivizumab (neutralizing antibody of RSV) in the lung *via* synthetic mRNA delivery by intratracheal aerosol (Tiwari et al., 2018). Cells co-transfected with mRNAs encoding the light and heavy chains at a 1:4 molar ratio could efficiently form whole IgG antibodies and prevent detectable infection (Tiwari et al., 2018). In another study, mRNA that encoded a HIV neutralizing antibody (PGT121) as well as a membrane anchor protein, was used to efficiently localize the antibody to the cell surface and capture simian-HIV (Lindsay et al., 2020). Thus, it will be interesting to see whether antibody delivery can be routinely implemented as RNA.

Of the four technologies reviewed here (B cell sorting, BCR sequencing, BCR repertoire analysis, and Experimental validation), repertoire analysis is the one area that expected to become radically transformed by advances in computational science. While we can cluster antibodies into specificity groups using CDR or gene usage similarity, the sensitivity of such methods is severely limited. Such limitation, in turn, is due to poor coverage of well-annotated BCR sequences, as discussed in a research (Jespersen et al., 2019). Only a few antigens have been well studied, so machine learning models are currently unable to learn the patterns associated with specific epitopes. Therefore, the transformation from clustering based on similarity to the ability to predict epitopes will require steady progress in sorting, sequencing, and validation. Such progress will come

through investment in repertoire-based analysis of various diseases, data sharing and basic infrastructure. The establishment of data standards is one important step. It remains uncertain whether data providers will merge together under government-sponsored institutions (e.g., NIH, EBI, AMED) or remain independently operated. In either scenario, it will be interesting to see how the pharmaceutical industry responds to the growing information contained in human BCR repertoires.

## Author contributions

DS conceived the idea and outlook of this review. ZX, HI, HZ, DS, and FS reviewed the published literature and wrote the manuscript under the supervision of DS. ZX, HZ, DS, and FS made the figures and table. All authors listed participated in the revision and discussions. ZX formulated the draft, FS and DS revised the final submitted version.

## Funding

This work was supported by Japan Agency for Medical Research and Development (AMED) Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research) under Grant Numbers 22ama121025j0001.

## Acknowledgments

We would like to thank all members of the Systems Immunology and Genome Informatics Lab for the very helpful discussion.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Abbott, W. M., Damschroder, M. M., and Lowe, D. C. (2014). Current approaches to fine mapping of antigen-antibody interactions. *Immunology* 142 (4), 526–535. doi:10.1111/imm.12284
- Acquaye-Seedah, E., Reczek, E. E., Russell, H. H., DiVenere, A. M., Sandman, S. O., Collins, J. H., et al. (2018). Characterization of individual human antibodies that bind pertussis toxin stimulated by acellular immunization. *Infect. Immun.* 86 (6). doi:10.1128/IAI.00004-18
- Akbar, R., Bashour, H., Rawat, P., Robert, P. A., Smorodina, E., Cotet, T. S., et al. (2022). Progress and challenges for the machine learning-based design of fit-for-purpose monoclonal antibodies. *MABS* 14 (1), 2008790. doi:10.1080/19420862.2021.2008790
- Alamyar, E., Giudicelli, V., Li, S., Duroux, P., and Lefranc, M.-P. (2012). Imgt/HighV-Quest: The Imgt® web portal for immunoglobulin (ig) or antibody and T cell receptor (tr) analysis from ngs high throughput and deep sequencing. *Immunome Res.* 08 (01). doi:10.4172/1745-7580.1000056
- Almagro, J. C., Teplyakov, A., Luo, J., Sweet, R. W., Kodangattil, S., Hernandez-Guzman, F., et al. (2014). Second antibody modeling assessment (AMA-II). *Proteins*. 82 (8), 1553–1562. doi:10.1002/prot.24567
- Amanna, I. J., and Slikf, M. K. (2006). Quantitation of rare memory B cell populations by two independent and complementary approaches. *J. Immunol. Methods* 317 (1–2), 175–185. doi:10.1016/j.jim.2006.09.005
- Ambrosetti, F., Olsen, T. H., Olimpieri, P. P., Jimenez-Garcia, B., Milanetti, E., Marcatilli, P., et al. (2020). proABC-2: Prediction of AntiBody contacts v2 and its application to information-driven docking. *Bioinformatics* 36 (20), 5107–5108. doi:10.1093/bioinformatics/btaa644
- Ansari, H. R., and Raghava, G. P. (2010). Identification of conformational B-cell Epitopes in an antigen from its primary sequence. *Immunome Res.* 6, 6. doi:10.1186/1745-7580-6-6
- Attaf, N., Cervera-Marzal, I., Dong, C., Gil, L., Renand, A., Spinelli, L., et al. (2020). FB5P-seq: FACS-based 5-prime end single-cell RNA-seq for integrative analysis of transcriptome and antigen receptor repertoire in B and T cells. *Front. Immunol.* 11, 216. doi:10.3389/fimmu.2020.00216
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373 (6557), 871–876. doi:10.1126/science.abj8754
- Banach, M., Harley, I. T. W., McCarthy, M. K., Rester, C., Stassinopoulos, A., Kedl, R. M., et al. (2022). Magnetic enrichment of SARS-CoV-2 antigen-binding B cells for analysis of transcriptome and antibody repertoire. *Magnetochemistry* 8 (2), 23. doi:10.3390/magnetochemistry8020023
- Barbier, A. J., Jiang, A. Y., Zhang, P., Wooster, R., and Anderson, D. G. (2022). The clinical progress of mRNA vaccines and immunotherapies. *Nat. Biotechnol.* 40 (6), 840–854. doi:10.1038/s41587-022-01294-2
- Barnes, C. O., Jette, C. A., Abernathy, M. E., Dam, K. A., Esswein, S. R., Gristick, H. B., et al. (2020). SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* 588 (7839), 682–687. doi:10.1038/s41586-020-2852-1
- Bashford-Rogers, R. J., Palser, A. L., Idris, S. F., Carter, L., Epstein, M., Callard, R. E., et al. (2014). Capturing needles in haystacks: A comparison of B-cell receptor sequencing methods. *BMC Immunol.* 15 (29), 29. doi:10.1186/s12865-014-0029-0
- Baum, A., Fulton, B. O., Wloga, E., Copin, R., Pascal, K. E., Russo, V., et al. (2021). Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* 369 (6506), 1014–1018. doi:10.1126/science.abd0831
- Bennett, M. R., Dong, J., Bombardi, R. G., Soto, C., Parrington, H. M., Nargi, R. S., et al. (2019). Human  $V_{H1-69}$  gene-encoded human monoclonal antibodies against *Staphylococcus aureus* IsdB use at least three distinct modes of binding to inhibit bacterial growth and pathogenesis. *mBio* 10 (5), e02473. doi:10.1128/mBio.02473-19
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2008). The protein Data Bank. *Nucleic Acids Res.* 28 (1), 235–242. doi:10.1093/nar/28.1.235
- Bhattacharya, S., Dunn, P., Thomas, C. G., Smith, B., Schaefer, H., Chen, J., et al. (2018). ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci. Data* 5, 180015. doi:10.1038/sdata.2018.15
- Blanc, P., Moro-Sibilot, L., Barthly, L., Jagot, F., This, S., de Bernard, S., et al. (2016). Mature IgM-expressing plasma cells sense antigen and develop competence for cytokine production upon antigenic challenge. *Nat. Commun.* 7, 13600. doi:10.1038/ncomms13600
- Blech, M., Peter, D., Fischer, P., Bauer, M. M., Hafner, M., Zeeb, M., et al. (2013). One target—two different binding modes: Structural insights into gevokizumab and canakinumab interactions to interleukin-1 $\beta$ . *J. Mol. Biol.* 425 (1), 94–111. doi:10.1016/j.jmb.2012.09.021
- Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V., et al. (2015). MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods* 12 (5), 380–381. doi:10.1038/nmeth.3364
- Boonyaratankornkit, J., and Taylor, J. J. (2019). Techniques to study antigen-specific B cell responses. *Front. Immunol.* 10, 1694. doi:10.3389/fimmu.2019.01694
- Bourquard, T., Musnier, A., Puard, V., Tahir, S., Ayoub, M. A., Jullian, Y., et al. (2018). MABTope: A method for improved epitope mapping. *J. I.* 201 (10), 3096–3105. doi:10.4049/jimmunol.1701722
- Briney, B., Inderbitzin, A., Joyce, C., and Burton, D. R. (2019). Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* 566 (7744), 393–397. doi:10.1038/s41586-019-0879-y
- Buus, T. B., Herrera, A., Ivanova, E., Mimitou, E., Cheng, A., Herati, R. S., et al. (2021). Improving oligo-conjugated antibody signal in multimodal single-cell analysis. *Elife* 10, e61973. doi:10.7554/eLife.61973
- Campbell, P. J., Pleasance, E. D., Stephens, P. J., Dicks, E., Rance, R., Goodhead, I., et al. (2008). Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 105 (35), 13081–13086. doi:10.1073/pnas.0801523105
- Cao, Y., Su, B., Guo, X., Sun, W., Deng, Y., Bao, L., et al. (2020). Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* 182 (1), 73–84. doi:10.1016/j.cell.2020.05.025
- Carlson, C. S., Emerson, R. O., Sherwood, A. M., Desmarais, C., Chung, M. W., Parsons, J. M., et al. (2013). Using synthetic templates to design an unbiased multiplex PCR assay. *Nat. Commun.* 4, 2680. doi:10.1038/ncomms3680
- Chaudhary, N., Weissman, D., and Whitehead, K. A. (2021). mRNA vaccines for infectious diseases: Principles, delivery and clinical translation. *Nat. Rev. Drug Discov.* 20 (11), 817–838. doi:10.1038/s41573-021-00283-5
- Chaudhary, N., and Wesemann, D. R. (2018). Analyzing immunoglobulin repertoires. *Front. Immunol.* 9, 462. doi:10.3389/fimmu.2018.00462
- Chen, Y., Ye, Z., Zhang, Y., Xie, W., Chen, Q., Lan, C., et al. (2022). A deep learning model for accurate diagnosis of infection using antibody repertoires. *J. I.* 208 (12), 2675–2685. doi:10.4049/jimmunol.2200063
- Christley, S., Scarborough, W., Salinas, E., Rounds, W. H., Toby, I. T., Fonner, J. M., et al. (2018). VDJServer: A cloud-based analysis portal and data commons for immune repertoire sequences and rearrangements. *Front. Immunol.* 9, 976. doi:10.3389/fimmu.2018.00976
- Clifford, J., Høie, M. H., Nielsen, M., Deleuran, S., Peters, B., and Marcatilli, P. (2022). *BepiPred-3.0: Improved B-cell epitope prediction using protein language models*. One Bungtown Road, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. doi:10.1101/2022.07.11.499418
- Collatz, M., Mock, F., Barth, E., Holzer, M., Sachse, K., and Marz, M. (2021). EpiDope: A deep neural network for linear B-cell epitope prediction. *Bioinformatics* 37 (4), 448–455. doi:10.1093/bioinformatics/btaa773
- Corrie, B. D., Marthandan, N., Zimonja, B., Jaglale, J., Zhou, Y., Barr, E., et al. (2018). iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol. Rev.* 284 (1), 24–41. doi:10.1111/imr.12666
- D'Angelo, S., Ferrara, F., Naranjo, L., Erasmus, M. F., Hrabec, P., and Bradbury, A. R. M. (2018). Many routes to an antibody heavy-chain CDR3: Necessary, yet insufficient, for specific binding. *Front. Immunol.* 9, 395. doi:10.3389/fimmu.2018.00395
- Dai, B., and Bailey-Kellogg, C. (2021). Protein interaction interface region prediction by geometric deep learning. *Bioinformatics* 37, 2580–2588. doi:10.1093/bioinformatics/btab154
- Davila, A., Xu, Z., Li, S., Rozewicki, J., Wilamowski, J., Kotelnikov, S., et al. (2022). AbAdapt: An adaptive approach to predicting antibody-antigen complex structures from sequence. *Bioinform. Adv.* 2 (1). doi:10.1093/bioadv/vbac015
- Davydov, Y. I., and Tonevitsky, A. G. (2009). Prediction of linear B-cell epitopes. *Mol. Biol. Los. Angel.* 43 (1), 150–158. doi:10.1134/s0026893309010208
- Dejnirattisai, W., Zhou, D., Ginn, H. M., Duyvesteyn, H. M. E., Supasa, P., Case, J. B., et al. (2021). The antigenic anatomy of SARS-CoV-2 receptor binding domain. *Cell* 184 (8), 2183–2200. e2122. doi:10.1016/j.cell.2021.02.032
- DeKosky, B. J., Ippolito, G. C., Deschner, R. P., Lavinder, J. J., Wine, Y., Rawlings, B. M., et al. (2013). High-throughput sequencing of the paired human

- immunoglobulin heavy and light chain repertoire. *Nat. Biotechnol.* 31 (2), 166–169. doi:10.1038/nbt.2492
- DeKosky, B. J., Kojima, T., Rodin, A., Charab, W., Ippolito, G. C., Ellington, A. D., et al. (2015). In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat. Med.* 21 (1), 86–91. doi:10.1038/nm.3743
- Dondelinger, M., Filee, P., Sauvage, E., Quinting, B., Muyldermans, S., Galleni, M., et al. (2018). Understanding the significance and implications of antibody numbering and antigen-binding surface/residue definition. *Front. Immunol.* 9, 2278. doi:10.3389/fimmu.2018.02278
- Doucett, V. P., Gerhard, W., Owler, K., Curry, D., Brown, L., and Baumgarth, N. (2005). Enumeration and characterization of virus-specific B cells by multicolor flow cytometry. *J. Immunol. Methods* 303 (1-2), 40–52. doi:10.1016/j.jim.2005.05.014
- Duez, M., Giraud, M., Herbert, R., Rocher, T., Salson, M., and Thonier, F. (2016). Vidjil: A web platform for analysis of high-throughput repertoire sequencing. *PLoS One* 11 (11), e0166126. doi:10.1371/journal.pone.0166126
- Dunbar, J., and Deane, C. M. (2016). Anarci: Antigen receptor numbering and receptor classification. *Bioinformatics* 32 (2), 298–300. doi:10.1093/bioinformatics/btv552
- Dunbar, J., Krawczyk, K., Leem, J., Baker, T., Fuchs, A., Georges, G., et al. (2014). SAbDab: The structural antibody database. *Nucleic Acids Res.* 42, D1140–D1146. doi:10.1093/nar/gkt1043
- Ehrenmann, F., Kaas, Q., and Lefranc, M. P. (2010). IMGT/3Dstructure-DB and IMGT/DomainGapAlign: A database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res.* 38, D301–D307. doi:10.1093/nar/gkp946
- Ehrenmann, F., and Lefranc, M. P. (2011). IMGT/3Dstructure-DB: Querying the IMGT database for 3D structures in immunology and immunoinformatics (IG or antibodies, TR, MH, RPI, and FPIA). *Cold Spring Harb. Protoc.* 2011 (6), pdb.prot5637–761. doi:10.1101/pdb.prot5637
- El-Manzalawy, Y., Dobbs, D., and Honavar, V. (2008). Predicting flexible length linear B-cell epitopes. *Comput. Syst. Bioinforma. Conf.* 7, 121–132.
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., et al. (2022). *Protein complex prediction with AlphaFold-Multimer*. One Bungtown Road, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. doi:10.1101/2021.10.04.463034
- Fields, J. K., Kihn, K., Birkedal, G. S., Klontz, E. H., Sjostrom, K., Gunther, S., et al. (2021). Molecular basis of selective cytokine signaling inhibition by antibodies targeting a shared receptor. *Front. Immunol.* 12, 779100. doi:10.3389/fimmu.2021.779100
- Fowler, D. M., and Fields, S. (2014). Deep mutational scanning: A new style of protein science. *Nat. Methods* 11 (8), 801–807. doi:10.1038/nmeth.3027
- Gaiotto, T., and Hufton, S. E. (2016). Cross-neutralising nanobodies bind to a conserved pocket in the Hemagglutinin stem region identified using yeast display and deep mutational scanning. *PLoS One* 11 (10), e0164296. doi:10.1371/journal.pone.0164296
- Galson, J. D., Schatzle, S., Bashford-Rogers, R. J. M., Raybould, M. I. J., Kovaltsuk, A., Kilpatrick, G. J., et al. (2020). Deep sequencing of B cell receptor repertoires from COVID-19 patients reveals strong convergent immune signatures. *Front. Immunol.* 11, 605170. doi:10.3389/fimmu.2020.605170
- Galson, J. D., Truck, J., Fowler, A., Clutterbuck, E. A., Munz, M., Cerundolo, V., et al. (2015a). Analysis of B Cell repertoire dynamics following hepatitis B vaccination in humans, and enrichment of vaccine-specific antibody sequences. *EBioMedicine* 2 (12), 2070–2079. doi:10.1016/j.ebiom.2015.11.034
- Galson, J. D., Truck, J., Fowler, A., Munz, M., Cerundolo, V., Pollard, A. J., et al. (2015b). In-depth assessment of within-individual and inter-individual variation in the B cell receptor repertoire. *Front. Immunol.* 6, 531. doi:10.3389/fimmu.2015.00531
- Garrett, M. E., Itell, H. L., Crawford, K. H. D., Basom, R., Bloom, J. D., and Overbaugh, J. (2020). Phage-DMS: A comprehensive method for fine mapping of antibody epitopes. *iScience* 23 (10), 101622. doi:10.1016/j.isci.2020.101622
- Gieselmann, L., Kreer, C., Ercanoglu, M. S., Lehnen, N., Zehner, M., Schommers, P., et al. (2021). Effective high-throughput isolation of fully human antibodies targeting infectious pathogens. *Nat. Protoc.* 16 (7), 3639–3671. doi:10.1038/s41596-021-00554-w
- Goldstein, L. D., Chen, Y. J., Wu, J., Chaudhuri, S., Hsiao, Y. C., Schneider, K., et al. (2019). Massively parallel single-cell B-cell receptor sequencing enables rapid discovery of diverse antigen-reactive antibodies. *Commun. Biol.* 2, 304. doi:10.1038/s42003-019-0551-y
- Greaney, A. J., Starr, T. N., Barnes, C. O., Weisblum, Y., Schmidt, F., Caskey, M., et al. (2021a). Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nat. Commun.* 12 (1), 4196. doi:10.1038/s41467-021-24435-8
- Greaney, A. J., Starr, T. N., Gilchuk, P., Zost, S. J., Binshtein, E., Loes, A. N., et al. (2021b). Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* 29 (1), 44–57. doi:10.1016/j.chom.2020.11.007
- Greiff, V., Bhat, P., Cook, S. C., Menzel, U., Kang, W., and Reddy, S. T. (2015a). A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status. *Genome Med.* 7 (1), 49. doi:10.1186/s13073-015-0169-8
- Greiff, V., Miho, E., Menzel, U., and Reddy, S. T. (2015b). Bioinformatic and statistical analysis of adaptive immune repertoires. *Trends Immunol.* 36 (11), 738–749. doi:10.1016/j.it.2015.09.006
- Guo, Y., Chen, K., Kwong, P. D., Shapiro, L., and Sheng, Z. (2019). cAb-rep: A database of curated antibody repertoires for exploring antibody diversity and predicting antibody prevalence. *Front. Immunol.* 10, 2365. doi:10.3389/fimmu.2019.02365
- Gupta, N. T., Vander Heiden, J. A., Uduman, M., Gadala-Maria, D., Yaari, G., and Kleinstein, S. H. (2015). Change-O: A toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data: Table 1. *Bioinformatics* 31 (20), 3356–3358. doi:10.1093/bioinformatics/btv359
- Hansen, J., Baum, A., Pascal, K. E., Russo, V., Giordano, S., Wloga, E., et al. (2020). Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science* 369 (6506), 1010–1014. doi:10.1126/science.abb0827
- Haque, A., Engel, J., Teichmann, S. A., and Lonnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* 9 (1), 75. doi:10.1186/s13073-017-0467-4
- Hasan, M. M., Khatun, M. S., and Kurata, H. (2020). iLBE for computational identification of linear B-cell epitopes by integrating sequence and evolutionary features. *Genomics Proteomics Bioinforma.* 18 (5), 593–600. doi:10.1016/j.gpb.2019.04.004
- He, B., Liu, S., Wang, Y., Xu, M., Cai, W., Liu, J., et al. (2021). Rapid isolation and immune profiling of SARS-CoV-2 specific memory B cell in convalescent COVID-19 patients via LIBRA-seq. *Signal Transduct. Target. Ther.* 6 (1), 195. doi:10.1038/s41392-021-00610-7
- Heather, J. M., Ismail, M., Oakes, T., and Chain, B. (2018). High-throughput sequencing of the T-cell receptor repertoire: Pitfalls and opportunities. *Brief. Bioinform.* 19 (4), 554–565. doi:10.1093/bib/bbw138
- Holcomb, J., Spellmon, N., Zhang, Y., Doughan, M., Li, C., and Yang, Z. (2017). Protein crystallization: Eluding the bottleneck of X-ray crystallography. *AIMS Biophys.* 4 (4), 557–575. doi:10.3934/biophys.2017.4.557
- Hotop, S. K., Reimering, S., Shekhar, A., Asgari, E., Beutling, U., Dahlke, C., et al. (2022). Peptide microarrays coupled to machine learning reveal individual epitopes from human antibody responses with neutralizing capabilities against SARS-CoV-2. *Emerg. Microbes Infect.* 11 (1), 1037–1048. doi:10.1080/22221751.2022.2057874
- Huang, R. Y., Krystek, S. R., Jr., Felix, N., Graziano, R. F., Srinivasan, M., Pashine, A., et al. (2018). Hydrogen/deuterium exchange mass spectrometry and computational modeling reveal a discontinuous epitope of an antibody/TL1A Interaction. *Mabs* 10 (1), 95–103. doi:10.1080/19420862.2017.1393595
- Ikemura, N., Taminishi, S., Inaba, T., Arimori, T., Motooka, D., Katoh, K., et al. (2022). An engineered ACE2 decoy neutralizes the SARS-CoV-2 Omicron variant and confers protection against infection *in vivo*. *Sci. Transl. Med.* 14 (650), eabn7737. doi:10.1126/scitranslmed.abn7737
- ImmunoMind (2019). Immunarch: An R package for painless bioinformatics analysis of T-cell and B-cell immune repertoires. *Zenodo* 10, 5281. doi:10.5281/zenodo.3367200
- Ismanto, H. S., Xu, Z., Saputri, D. S., Wilamowski, J., Li, S., Nugraha, D. K., et al. (2022). *Landscape of infection enhancing antibodies in COVID-19 and healthy donors*. One Bungtown Road, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. doi:10.1101/2022.07.09.499414
- Jackson, K. J., Liu, Y., Roskin, K. M., Glanville, J., Hoh, R. A., Seo, K., et al. (2014). Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell Host Microbe* 16 (1), 105–114. doi:10.1016/j.chom.2014.05.013
- Jespersen, M. C., Mahajan, S., Peters, B., Nielsen, M., and Marcatili, P. (2019). Antibody specific B-cell epitope predictions: Leveraging information from antibody-antigen protein complexes. *Front. Immunol.* 10, 298. doi:10.3389/fimmu.2019.00298
- Jespersen, M. C., Peters, B., Nielsen, M., and Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* 45 (W1), W24–W29. doi:10.1093/nar/gkx346
- Jin, W., Yang, Q., Peng, Y., Yan, C., Li, Y., Luo, Z., et al. (2021). Single-cell RNA-Seq reveals transcriptional heterogeneity and immune subtypes associated with disease activity in human myasthenia gravis. *Cell Discov.* 7 (1), 85. doi:10.1038/s41421-021-00314-w



- Ju, B., Zhang, Q., Ge, J., Wang, R., Sun, J., Ge, X., et al. (2020). Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* 584 (7819), 115–119. doi:10.1038/s41586-020-2380-z
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2
- Kim, S. I., Noh, J., Kim, S., Choi, Y., Yoo, D. K., Lee, Y., et al. (2021). Stereotypic neutralizing V H antibodies against SARS-CoV-2 spike protein receptor binding domain in patients with COVID-19 and healthy individuals. *Sci. Transl. Med.* 13 (578), eabd6990. doi:10.1126/scitranslmed.abd6990
- Kotagiri, P., Mescia, F., Rae, W. M., Bergamaschi, L., Tuong, Z. K., Turner, L., et al. (2022). B cell receptor repertoire kinetics after SARS-CoV-2 infection and vaccination. *Cell Rep.* 38 (7), 110393. doi:10.1016/j.celrep.2022.110393
- Kovaltsuk, A., Krawczyk, K., Galson, J. D., Kelly, D. F., Deane, C. M., and Truck, J. (2017). How B-cell receptor repertoire sequencing can be enriched with structural antibody data. *Front. Immunol.* 8, 1753. doi:10.3389/fimmu.2017.01753
- Kovaltsuk, A., Leem, J., Kelm, S., Snowden, J., Deane, C. M., and Krawczyk, K. (2018). Observed antibody Space: A resource for data mining next-generation sequencing of antibody repertoires. *J. I.* 201 (8), 2502–2509. doi:10.4049/jimmunol.1800708
- Kramer, K. J., Johnson, N. V., Shiakolas, A. R., Suryadevara, N., Periasamy, S., Raju, N., et al. (2021). Potent neutralization of SARS-CoV-2 variants of concern by an antibody with an uncommon genetic signature and structural mode of spike recognition. *Cell Rep.* 37 (1), 109784. doi:10.1016/j.celrep.2021.109784
- Kramer, K. J., Wilfong, E. M., Voss, K., Barone, S. M., Shiakolas, A. R., Raju, N., et al. (2022). Single-cell profiling of the antigen-specific response to BNT162b2 SARS-CoV-2 RNA vaccine. *Nat. Commun.* 13 (1), 3466. doi:10.1038/s41467-022-31142-5
- Krawczyk, K., Liu, X., Baker, T., Shi, J., and Deane, C. M. (2014). Improving B-cell epitope prediction and its application to global antibody-antigen docking. *Bioinformatics* 30 (16), 2288–2294. doi:10.1093/bioinformatics/btu190
- Kringelum, J. V., Lundegaard, C., Lund, O., and Nielsen, M. (2012). Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLoS Comput. Biol.* 8 (12), e1002829. doi:10.1371/journal.pcbi.1002829
- Laroche, A., Orsini Delgado, M. L., Chalopin, B., Cuniasso, P., Dubois, S., Sierocki, R., et al. (2022). Deep mutational engineering of broadly-neutralizing nanobodies accommodating SARS-CoV-1 and 2 antigenic drift. *MAbs* 14 (1), 2076775. doi:10.1080/19420862.2022.2076775
- Lavinder, J. J., Wine, Y., Giesecke, C., Ippolito, G. C., Horton, A. P., Lungu, O. I., et al. (2012). Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proc. Natl. Acad. Sci. U. S. A.* 111 (6), 2259–2264. doi:10.1073/pnas.1317793111
- Lee, J. H., Toy, L., Kos, J. T., Safonova, Y., Schief, W. R., Havenar-Daughton, C., et al. (2014). Vaccine genetics of IGHV1-2 VRC01-class broadly neutralizing antibody precursor naive human B cells. *NPJ Vaccines* 6 (1), 113. doi:10.1038/s41541-021-00376-7
- Lefranc, M. P., Giudicelli, V., Kaas, Q., Duprat, E., Jabado-Michaloud, J., Scaviner, D., et al. (2005). IMGT, the international ImMunoGeneTics information system(R). *Nucleic Acids Res.* 33, D593–D597. doi:10.1093/nar/gki065
- Leinster, T., and Cobbold, C. A. (2012). Measuring diversity: The importance of species similarity. *Ecology* 93 (3), 477–489. doi:10.1890/10-2402.1
- Li, D., Edwards, R. J., Manne, K., Martinez, D. R., Schafer, A., Alam, S. M., et al. (2021). *In vitro* and *in vivo* functions of SARS-CoV-2 infection-enhancing and neutralizing antibodies. *Cell* 184 (16), 4203–4219. e4232. doi:10.1016/j.cell.2021.06.021
- Li, S., Lefranc, M. P., Miles, J. J., Alamyar, E., Giudicelli, V., Duroux, P., et al. (2013). IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nat. Commun.* 4, 2333. doi:10.1038/ncomms3333
- Lian, Y., Ge, M., and Pan, X. M. (2014). Epmlr: Sequence-based linear B-cell epitope prediction method using multiple linear regression. *BMC Bioinforma.* 15, 414. doi:10.1186/s12859-014-0414-y
- Liberis, E., Velickovic, P., Sormanni, P., Vendruscolo, M., and Lio, P. (2018). Parapred: Antibody paratope prediction using convolutional and recurrent neural networks. *Bioinformatics* 34 (17), 2944–2950. doi:10.1093/bioinformatics/bty305
- Lindsay, K. E., Vanover, D., Thoresen, M., King, H., Xiao, P., Badial, P., et al. (2020). Aerosol delivery of synthetic mRNA to vaginal mucosa leads to durable expression of broadly neutralizing antibodies against HIV. *Mol. Ther.* 28 (3), 805–819. doi:10.1016/j.yymthe.2020.01.002
- Liu, Y., Soh, W. T., Kishikawa, J. I., Hirose, M., Nakayama, E. E., Li, S., et al. (2021). An infectivity-enhancing site on the SARS-CoV-2 spike protein targeted by antibodies. *Cell* 184 (13), 3452–3466. doi:10.1016/j.cell.2021.05.032
- Lopez-Santibanez-Jacome, L., Avendano-Vazquez, S. E., and Flores-Jasso, C. F. (2019). The pipeline repertoire for ig-seq analysis. *Front. Immunol.* 10, 899. doi:10.3389/fimmu.2019.00899
- Lu, S., Li, Y., Ma, Q., Nan, X., and Zhang, S. (2022). A structure-based B-cell epitope prediction model through combing local and global features. *Front. Immunol.* 13, 890943. doi:10.3389/fimmu.2022.890943
- Manavalan, B., Govindaraj, R. G., Shin, T. H., Kim, M. O., and Lee, G. (2018). iBCE-EL: A new ensemble learning framework for improved linear B-cell epitope prediction. *Front. Immunol.* 9, 1695. doi:10.3389/fimmu.2018.01695
- Manso, T., Folch, G., Giudicelli, V., Jabado-Michaloud, J., Kushwaha, A., Nguéfac Ngoune, V., et al. (2022). IMGT® databases, related tools and web resources through three main axes of research and development. *Nucleic Acids Res.* 50 (D1), D1262–D1272. doi:10.1093/nar/gkab1136
- Mason, D. M., Friedensohn, S., Weber, C. R., Jordi, C., Wagner, B., Meng, S. M., et al. (2016). Optimization of therapeutic antibodies by predicting hydrogen specificity from antibody sequence via deep learning. *Nat. Biomed. Eng.* 5 (6), 600–612. doi:10.1038/s41551-021-00699-9
- Masson, G. R., Burke, J. E., Ahn, N. G., Anand, G. S., Borchers, C., Brier, S., et al. (2019). Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* 16 (7), 595–602. doi:10.1038/s41592-019-0459-y
- McDaniel, J. R., DeKosky, B. J., Tanno, H., Ellington, A. D., and Georgiou, G. (2019). Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nat. Protoc.* 11 (3), 429–442. doi:10.1038/nprot.2016.024
- Monasterio, E., Mei-Dan, O., Hackney, A. C., and Cloninger, R. (2018). Comparison of the personality traits of male and female BASE jumpers. *Front. Psychol.* 9, 1665. doi:10.3389/fpsyg.2018.01665
- Nelson, A. L., Dhimolea, E., and Reichert, J. M. (2010). Development trends for human monoclonal antibody therapeutics. *Nat. Rev. Drug Discov.* 9 (10), 767–774. doi:10.1038/nrd3229
- Nielsen, S. C. A., Yang, F., Jackson, K. J. L., Hoh, R. A., Roltgen, K., Jean, G. H., et al. (2020). Human B cell clonal expansion and convergent antibody responses to SARS-CoV-2. *Cell Host Microbe* 28 (4), 516–525. e515. doi:10.1016/j.chom.2020.09.002
- Olsen, T. H., Boyles, F., and Deane, C. M. (2022). Observed antibody Space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Sci.* 31 (1), 141–146. doi:10.1002/pro.4205
- Ortega, M. R., Spisak, N., Mora, T., and Walczak, A. M. (2021). *Modeling and predicting the overlap of B- and T-cell receptor repertoires in healthy and SARS-CoV-2 infected individuals*. One Bungtown Road, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. doi:10.1101/2021.12.17.473105
- Pan, J., Zhang, S., Chou, A., and Borchers, C. H. (2016). Higher-order structural interrogation of antibodies using middle-down hydrogen/deuterium exchange mass spectrometry. *Chem. Sci.* 7 (2), 1480–1486. doi:10.1039/c5sc03420e
- Pedrioli, A., and Oxenius, A. (2021). Single B cell technologies for monoclonal antibody discovery. *Trends Immunol.* 42 (12), 1143–1158. doi:10.1016/j.it.2021.10.008
- Peng, L., Oganiesyan, V., Damschroder, M. M., Wu, H., and Dall'Acqua, W. F. (2011). Structural and functional characterization of an agonistic anti-human EphA2 monoclonal antibody. *J. Mol. Biol.* 413 (2), 390–405. doi:10.1016/j.jmb.2011.08.018
- Picot, J., Guerin, C. L., Le Van Kim, C., and Boulanger, C. M. (2012). Flow cytometry: Retrospective, fundamentals and recent instrumentation. *Cytotechnology* 64 (2), 109–130. doi:10.1007/s10616-011-9415-0
- Pinto, D., Montani, E., Bolli, M., Garavaglia, G., Sallusto, F., Lanzavecchia, A., et al. (2013). A functional BCR in human IgA and IgM plasma cells. *Blood* 121 (20), 4110–4114. doi:10.1182/blood-2012-09-459289
- Pinto, D., Park, Y. J., Beltramello, M., Walls, A. C., Tortorici, M. A., Bianchi, S., et al. (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* 583 (7815), 290–295. doi:10.1038/s41586-020-2349-y
- Pittala, S., and Bailey-Kellogg, C. (2020). Learning context-aware structural representations to predict antigen and antibody binding interfaces. *Bioinformatics* 36 (13), 3996–4003. doi:10.1093/bioinformatics/btaa263
- Polonsky, M., Chain, B., and Friedman, N. (2016). Clonal expansion under the microscope: Studying lymphocyte activation and differentiation using live-cell imaging. *Immunol. Cell Biol.* 94 (3), 242–249. doi:10.1038/icb.2015.104
- Ponomarenko, J., Bui, H. H., Li, W., Füsseder, N., Bourne, P. E., Sette, A., et al. (2008). ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinforma.* 9, 514. doi:10.1186/1471-2105-9-514
- Pons, J., Stratton, J. R., and Kirsch, J. F. (2002). How do two unrelated antibodies, HyHEL-10 and F9.13.7, recognize the same epitope of hen egg-white lysozyme? *Protein Sci.* 11 (10), 2308–2315. doi:10.1110/ps.0209102

- Puchades, C., Kukrer, B., Diefenbach, O., Sneekes-Vriese, E., Juraszek, J., Koudstaal, W., et al. (2019). Epitope mapping of diverse influenza Hemagglutinin drug candidates using HDX-MS. *Sci. Rep.* 9 (1), 4735. doi:10.1038/s41598-019-41179-0
- Qi, H., Ma, M., Hu, C., Xu, Z. W., Wu, F. L., Wang, N., et al. (2021). Antibody binding epitope mapping (AbMap) of hundred antibodies in a single run. *Mol. Cell. Proteomics* 20, 100059. doi:10.1074/mcp.RA120.002314
- Qi, H., Wang, F., and Tao, S. C. (2019). Proteome microarray technology and application: Higher, wider, and deeper. *Expert Rev. Proteomics* 16 (10), 815–827. doi:10.1080/14789450.2019.1662303
- Raybould, M. I. J., Kovaltsuk, A., Marks, C., and Deane, C. M. (2021a). CoV-AbDab: The coronavirus antibody database. *Bioinformatics* 37 (5), 734–735. doi:10.1093/bioinformatics/btaa739
- Raybould, M. I. J., Marks, C., Lewis, A. P., Shi, J., Bujotzek, A., Taddese, B., et al. (2020). Thera-SAbDab: The therapeutic structural antibody database. *Nucleic Acids Res.* 48 (D1), D383–D388. doi:10.1093/nar/gkz827
- Raybould, M. I. J., Rees, A. R., and Deane, C. M. (2021b). Current strategies for detecting functional convergence across B-cell receptor repertoires. *MABS* 13 (1), 1996732. doi:10.1080/19420862.2021.1996732
- Reddy, S. T., Ge, X., Miklos, A. E., Hughes, R. A., Kang, S. H., Hoi, K. H., et al. (2010). Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat. Biotechnol.* 28 (9), 965–969. doi:10.1038/nbt.1673
- Ripoll, D. R., Chaudhury, S., and Wallqvist, A. (2021). Using the antibody-antigen binding interface to train image-based deep neural networks for antibody-epitope classification. *PLoS Comput. Biol.* 17 (3), e1008864. doi:10.1371/journal.pcbi.1008864
- Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., et al. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl. Acad. Sci. U. S. A.* 118 (15), e2016239118. doi:10.1073/pnas.2016239118
- Robbiani, D. F., Gaebler, C., Muecksch, F., Lorenzi, J. C. C., Wang, Z., Cho, A., et al. (2020). Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* 584 (7821), 437–442. doi:10.1038/s41586-020-2456-9
- Robert, P. A., Akbar, R., Frank, R., Pavlović, M., Widrich, M., Snapkov, I., et al. (2022). *Unconstrained generation of synthetic antibody-antigen structures to guide machine learning methodology for real-world antibody specificity prediction*. One Bungtown Road, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. doi:10.1101/2021.07.06.451258
- Rosati, E., Dowds, C. M., Liaskou, E., Henriksen, E. K. K., Karlsen, T. H., and Franke, A. (2017). Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol.* 17 (1), 61. doi:10.1186/s12896-017-0379-9
- Rosenfeld, A. M., Meng, W., Luning Prak, E. T., and Hershberg, U. (2017). ImmuneDB: A system for the analysis and exploration of high-throughput adaptive immune receptor sequencing data. *Bioinformatics* 33 (2), 292–293. doi:10.1093/bioinformatics/btw593
- Roskin, K. M., Jackson, K. J. L., Lee, J. Y., Hoh, R. A., Joshi, S. A., Hwang, K. K., et al. (2020). Aberrant B cell repertoire selection associated with HIV neutralizing antibody breadth. *Nat. Immunol.* 21 (2), 199–209. doi:10.1038/s41590-019-0581-0
- Ruffolo, J. A., Sulam, J., and Gray, J. J. (2022). Antibody structure prediction using interpretable deep learning. *Patterns (N Y)* 3 (2), 100406. doi:10.1016/j.patter.2021.100406
- Saha, S., and Raghava, G. P. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 65 (1), 40–48. doi:10.1002/prot.21078
- Saravanan, V., and Gautham, N. (2015). Harnessing computational biology for exact linear B-cell epitope prediction: A novel amino acid composition-based feature descriptor. *OMICS A J. Integr. Biol.* 19 (10), 648–658. doi:10.1089/omi.2015.0095
- Satpathy, S., Wagner, S. A., Beli, P., Gupta, R., Kristiansen, T. A., Malinova, D., et al. (2015). Systems-wide analysis of BCR signalosomes and downstream phosphorylation and ubiquitylation. *Mol. Syst. Biol.* 11 (6), 810. doi:10.15252/msb.20145880
- Sela-Culang, I., Ofiran, Y., and Peters, B. (2015). Antibody specific epitope prediction-emergence of a new paradigm. *Curr. Opin. Virol.* 11, 98–102. doi:10.1016/j.coviro.2015.03.012
- Setliff, I., McDonnell, W. J., Raju, N., Bombardi, R. G., Murji, A. A., Scheepers, C., et al. (2018). Multi-donor longitudinal antibody repertoire sequencing reveals the existence of public antibody clonotypes in HIV-1 infection. *Cell Host Microbe* 23 (6), 845–854.e6. doi:10.1016/j.chom.2018.05.001
- Setliff, I., Shiakolas, A. R., Pilewski, K. A., Murji, A. A., Mapengo, R. E., Janowska, K., et al. (2019). High-throughput mapping of B cell receptor sequences to antigen specificity. *Cell* 179 (7), 1636–1646. e1615. doi:10.1016/j.cell.2019.11.003
- Seydoux, E., Homad, L. J., MacCamy, A. J., Parks, K. R., Hurlburt, N. K., Jennewein, M. F., et al. (2020). Analysis of a SARS-CoV-2-infected individual reveals development of potent neutralizing antibodies with limited somatic mutation. *Immunity* 53 (1), 98–105. e5. doi:10.1016/j.immuni.2020.06.001
- Sher, G., Zhi, D., and Zhang, S. (2017). Drrep: Deep ridge regressed epitope predictor. *BMC Genomics* 18 (6), 676. doi:10.1186/s12864-017-4024-8
- Shiakolas, A. R., Kramer, K. J., Johnson, N. V., Wall, S. C., Suryadevara, N., Wrapp, D., et al. (2022). Efficient discovery of SARS-CoV-2-neutralizing antibodies via B cell receptor sequencing and ligand blocking. *Nat. Biotechnol.* 40 (8), 1270–1275. doi:10.1038/s41587-022-01232-2
- Shiakolas, A. R., Kramer, K. J., Wrapp, D., Richardson, S. I., Schafer, A., Wall, S., et al. (2021). Cross-reactive coronavirus antibodies with diverse epitope specificities and Fc effector functions. *Cell Rep. Med.* 2 (6), 100313. doi:10.1016/j.xcrm.2021.100313
- Singh, H., Ansari, H. R., and Raghava, G. P. (2013). Improved method for linear B-cell epitope prediction using antigen's primary sequence. *PLoS One* 8 (5), e62216. doi:10.1371/journal.pone.0062216
- Singh, M., Al-Eryani, G., Carswell, S., Ferguson, J. M., Blackburn, J., Barton, K., et al. (2019). High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat. Commun.* 10 (1), 3120. doi:10.1038/s41467-019-11049-4
- Smakaj, E., Babrak, L., Ohlin, M., Shugay, M., Briney, B., Tosoni, D., et al. (2020). Benchmarking immunoinformatic tools for the analysis of antibody repertoire sequences. *Bioinformatics* 36 (6), 1731–1739. doi:10.1093/bioinformatics/btz845
- Solihah, B., Azhari, A., and Musdholifah, A. (2020). Enhancement of conformational B-cell epitope prediction using CluSMOTE. *PeerJ Comput. Sci.* 6, e275. doi:10.7717/peerj-cs.275
- Soto, C., Bombardi, R. G., Branchizio, A., Kose, N., Matta, P., Sevy, A. M., et al. (2019). High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* 566 (7744), 398–402. doi:10.1038/s41586-019-0934-8
- Sourisseau, M., Lawrence, D. J. P., Schwarz, M. C., Storrs, C. H., Veit, E. C., Bloom, J. D., et al. (2019). Deep mutational scanning comprehensively maps how Zika envelope protein mutations affect viral growth and antibody escape. *J. Virol.* 93 (23), e01291. doi:10.1128/JVI.01291-19
- Standley, D. M., Nakanishi, T., Xu, Z., Haruna, S., Li, S., Nazlica, S. A., et al. (2022). The evolution of structural genomics. *Biophys. Rev.* submitted.
- Starr, T. N., Greaney, A. J., Addetia, A., Hannon, W. W., Choudhary, M. C., Dingsens, A. S., et al. (2021). Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* 371 (6531), 850–854. doi:10.1126/science.abf9302
- Starr, T. N., Greaney, A. J., Hannon, W. W., Loes, A. N., Hauser, K., Dillen, J. R., et al. (2022). Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science* 377 (6604), 420–424. doi:10.1126/science.abo7896
- Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingsens, A. S., et al. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 182 (5), 1295–1310. e1220. doi:10.1016/j.cell.2020.08.012
- Sulen, A., Islam, S., Wolff, A. S. B., and Oftedal, B. E. (2020). The prospects of single-cell analysis in autoimmunity. *Scand. J. Immunol.* 92 (5), e12964. doi:10.1111/sji.12964
- Suryadevara, N., Shiakolas, A. R., VanBlargan, L. A., Binshtein, E., Chen, R. E., Case, J. B., et al. (2022). An antibody targeting the N-terminal domain of SARS-CoV-2 disrupts the spike trimer. *J. Clin. Invest.* 132 (11), e159062. doi:10.1172/JCI159062
- Sutermeister, B. A., and Darling, E. M. (2019). Considerations for high-yield, high-throughput cell enrichment: Fluorescence versus magnetic sorting. *Sci. Rep.* 9 (1), 227. doi:10.1038/s41598-018-36698-1
- Sweredoski, M. J., and Baldi, P. (2009). COBEpro: A novel system for predicting continuous B-cell epitopes. *Protein Eng. Des. Sel.* 22 (3), 113–120. doi:10.1093/protein/gzn075
- Sweredoski, M. J., and Baldi, P. (2008). Pepito: Improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics* 24 (12), 1459–1460. doi:10.1093/bioinformatics/btn199
- Swindells, M. B., Porter, C. T., Couch, M., Hurst, J., Abhinandan, K. R., Nielsen, J. H., et al. (2017). abYsis: Integrated antibody sequence and structure-management, analysis, and prediction. *J. Mol. Biol.* 429 (3), 356–364. doi:10.1016/j.jmb.2016.08.019
- Tiwari, P. M., Vanover, D., Lindsay, K. E., Bawage, S. S., Kirschman, J. L., Bhosle, S., et al. (2018). Engineered mRNA-expressed antibodies prevent respiratory syncytial virus infection. *Nat. Commun.* 9 (1), 3999. doi:10.1038/s41467-018-06508-3
- Tran, M. H., Schoeder, C. T., Schey, K. L., and Meiler, J. (2022). Computational structure prediction for antibody-antigen complexes from hydrogen-deuterium exchange mass spectrometry: Challenges and outlook. *Front. Immunol.* 13, 859964. doi:10.3389/fimmu.2022.859964



- Truck, J., Ramasamy, M. N., Galson, J. D., Rance, R., Parkhill, J., Lunter, G., et al. (2015). Identification of antigen-specific B cell receptor sequences using public repertoire analysis. *J. I.* 194 (1), 252–261. doi:10.4049/jimmunol.1401405
- Tsai, K. C., Lee, Y. C., and Tseng, T. S. (2021). Comprehensive deep mutational scanning reveals the immune-escaping hotspots of SARS-CoV-2 receptor-binding domain targeting neutralizing antibodies. *Front. Microbiol.* 12, 698365. doi:10.3389/fmicb.2021.698365
- Turchaninova, M. A., Davydov, A., Britanova, O. V., Shugay, M., Bikos, V., Egorov, E. S., et al. (2016). High-quality full-length immunoglobulin profiling with unique molecular barcoding. *Nat. Protoc.* 11 (9), 1599–1616. doi:10.1038/nprot.2016.093
- Upadhyay, A. A., Kauffman, R. C., Wolabaugh, A. N., Cho, A., Patel, N. B., Reiss, S. M., et al. (2018). Baldr: A computational pipeline for paired heavy and light chain immunoglobulin reconstruction in single-cell RNA-seq data. *Genome Med.* 10 (1), 20. doi:10.1186/s13073-018-0528-3
- Vander Heiden, J. A., Yaari, G., Uduman, M., Stern, J. N., O'Connor, K. C., Hafler, D. A., et al. (2014). pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* 30 (13), 1930–1932. doi:10.1093/bioinformatics/btu138
- Voss, W. N., Hou, Y. J., Johnson, N. V., Delidakis, G., Kim, J. E., Javanmardi, K., et al. (2021). Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes. *Science* 372 (6546), 1108–1112. doi:10.1126/science.abg5268
- Walker, L. M., Shiakolas, A. R., Venkat, R., Liu, Z. A., Wall, S., Raju, N., et al. (2022). High-throughput B cell epitope determination by next-generation sequencing. *Front. Immunol.* 13, 855772. doi:10.3389/fimmu.2022.855772
- Waltari, E., McGeever, A., Friedland, N., Kim, P. S., and McCutcheon, K. M. (2019). Functional enrichment and analysis of antigen-specific memory B cell antibody repertoires in PBMCs. *Front. Immunol.* 10, 1452. doi:10.3389/fimmu.2019.01452
- Wang, B., Kluwe, C. A., Lungu, O. I., DeKosky, B. J., Kerr, S. A., Johnson, E. L., et al. (2015). Facile discovery of a diverse panel of anti-ebola virus antibodies by immune repertoire mining. *Sci. Rep.* 5, 13926. doi:10.1038/srep13926
- Wang, C., Li, W., Drabek, D., Okba, N. M. A., van Haperen, R., Osterhaus, A., et al. (2020). A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat. Commun.* 11 (1), 2251. doi:10.1038/s41467-020-16256-y
- Wang, Y., Yuan, M., Lv, H., Peng, J., Wilson, I. A., and Wu, N. C. (2022). A large-scale systematic survey reveals recurring molecular features of public antibody responses to SARS-CoV-2. *Immunity* 55 (6), 1105–1117. e1104. doi:10.1016/j.immuni.2022.03.019
- Warszawski, S., Borenstein Katz, A., Lipsh, R., Khmel'nitsky, L., Ben Nissan, G., Javitt, G., et al. (2019). Optimizing antibody affinity and stability by the automated design of the variable light-heavy chain interfaces. *PLoS Comput. Biol.* 15 (8), e1007207. doi:10.1371/journal.pcbi.1007207
- Wilamowski, J., Xu, Z., Ismanto, H. S., Li, S., Teraguchi, S., Llamas-Covarrubias, M. A., et al. (2022). *InterClone: Store, search and cluster Adaptive immune receptor repertoires*. One Bungtown Road, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. doi:10.1101/2022.07.31.501809
- Wong, W. K., Leem, J., and Deane, C. M. (2019). Comparative analysis of the CDR loops of antigen receptors. *Front. Immunol.* 10, 2454. doi:10.3389/fimmu.2019.02454
- Wong, W. K., Robinson, S. A., Bujotzek, A., Georges, G., Lewis, A. P., Shi, J., et al. (2021). Ab-ligity: Identifying sequence-dissimilar antibodies that bind to the same epitope. *MAbs* 13 (1), 1873478. doi:10.1080/19420862.2021.1873478
- Woodruff, M. C., Ramonell, R. P., Nguyen, D. C., Cashman, K. S., Saini, A. S., Haddad, N. S., et al. (2020). Extrafollicular B cell responses correlate with neutralizing antibodies and morbidity in COVID-19. *Nat. Immunol.* 21 (12), 1506–1516. doi:10.1038/s41590-020-00814-z
- Wu, L., Xue, Z., Jin, S., Zhang, J., Guo, Y., Bai, Y., et al. (2022). huARdb: human Antigen Receptor database for interactive clonotype-transcriptome analysis at the single-cell level. *Nucleic Acids Res.* 50 (D1), D1244–D1254. doi:10.1093/nar/gkab857
- Wu, P. J., Kabakova, I. V., Ruberti, J. W., Sherwood, J. M., Dunlop, I. E., Paterson, C., et al. (2018). Water content, not stiffness, dominates Brillouin spectroscopy measurements in hydrated materials. *Nat. Methods* 15 (8), 561–562. doi:10.1038/s41592-018-0076-1
- Wüthrich, K. (1990). Protein structure determination in solution by NMR spectroscopy. *J. Biol. Chem.* 265 (36), 22059–22062. doi:10.1016/s0021-9258(18)45665-7
- Xu, S., Bottcher, L., and Chou, T. (2020). Diversity in biology: Definitions, quantification and models. *Phys. Biol.* 17 (3), 031001. doi:10.1088/1478-3975/ab6754
- Xu, Z., Davila, A., Wilamowski, J., Teraguchi, S., and Standley, D. M. (2022). Improved antibody-specific epitope prediction using AlphaFold and AbAdapt. *Chembiochem.* 23, e202200303. doi:10.1002/cbic.202200303
- Xu, Z., Li, S., Rozewicki, J., Yamashita, K., Teraguchi, S., Inoue, T., et al. (2019). Functional clustering of B cell receptors using sequence and structural features. *Mol. Syst. Des. Eng.* 4 (4), 769–778. doi:10.1039/c9me00021f
- Yaari, G., and Kleinstein, S. H. (2015). Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med.* 7, 121. doi:10.1186/s13073-015-0243-2
- Yao, B., Zhang, L., Liang, S., and Zhang, C. (2012). SVMTriP: A method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. *PLoS One* 7 (9), e45152. doi:10.1371/journal.pone.0045152
- Ye, J., Ma, N., Madden, T. L., and Ostell, J. M. (2013). IgBLAST: An immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* 41, W34–W40. doi:10.1093/nar/gkt382
- Yeku, O., and Frohman, M. A. (2011). Rapid amplification of cDNA ends (RACE). *Methods Mol. Biol.* 703, 107–122. doi:10.1007/978-1-59745-248-9\_8
- Yoon, H., Macke, J., West, A. P., Jr., Foley, B., Bjorkman, P. J., Korber, B., et al. (2015). Catnap: A tool to compile, analyze and tally neutralizing antibody panels. *Nucleic Acids Res.* 43 (W1), W213–W219. doi:10.1093/nar/gkv404
- Zhang, Q., Yang, J., Bautista, J., Badithe, A., Olson, W., and Liu, Y. (2018). Epitope mapping by HDX-MS elucidates the surface coverage of antigens associated with high blocking efficiency of antibodies to birch pollen allergen. *Anal. Chem.* 90 (19), 11315–11323. doi:10.1021/acs.analchem.8b01864
- Zhang, W., Wang, L., Liu, K., Wei, X., Yang, K., Du, W., et al. (2020). Pird: Pan immune repertoire database. *Bioinformatics* 36 (3), 897–903. doi:10.1093/bioinformatics/btz614
- Zheng, C., Zheng, L., Yoo, J. K., Guo, H., Zhang, Y., Guo, X., et al. (2017). Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* 169 (7), 1342–1356. e1316. doi:10.1016/j.cell.2017.05.035
- Zhou, C., Chen, Z., Zhang, L., Yan, D., Mao, T., Tang, K., et al. (2019). SEPPA 3.0-enhanced spatial epitope prediction enabling glycoprotein antigens. *Nucleic Acids Res.* 47 (W1), W388–W394. doi:10.1093/nar/gkz413
- Zhu, J., Wu, X., Zhang, B., McKee, K., O'Dell, S., Soto, C., et al. (2013). 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.* 110 (43), E4088–E4097. doi:10.1073/pnas.1306262110
- Zost, S. J., Gilchuk, P., Chen, R. E., Case, J. B., Reidy, J. X., Trivet, A., et al. (2020). Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nat. Med.* 26 (9), 1422–1427. doi:10.1038/s41591-020-0998-x