

Current strategies for detecting functional convergence across B-cell receptor repertoires

Matthew I. J. Raybould ^a, Anthony R. Rees ^b, and Charlotte M. Deane ^a

^aOxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford, UK; ^bRees Consulting AB, Stockholm, Sweden

ABSTRACT

Convergence across B-cell receptor (BCR) and antibody repertoires has become instrumental in prioritizing candidates in recent rapid therapeutic antibody discovery campaigns. It has also increased our understanding of the immune system, providing evidence for the preferential selection of BCRs to particular (immunodominant) epitopes post vaccination/infection. These important implications for both drug discovery and immunology mean that it is essential to consider the optimal way to combine experimental and computational technology when probing BCR repertoires for convergence signatures. Here, we discuss the theoretical basis for observing BCR repertoire functional convergence and explore factors of study design that can impact functional signal. We also review the computational arsenal available to detect antibodies with similar functional properties, highlighting opportunities enabled by recent clustering algorithms that exploit structural similarities between BCRs. Finally, we suggest future areas of development that should increase the power of BCR repertoire functional clustering.

ARTICLE HISTORY

Received 29 July 2021
Revised 10 October 2021
Accepted 12 October 2021

KEYWORDS

BCR repertoire; antibody; function; BCR-seq; Ig-seq; sequencing; clonal clustering; structural clustering; evolution; selection

Introduction

Naïve B-cell receptor (BCR) repertoires have an enormous theoretical diversity ($c. 10^{15}$) that is sampled at only a tiny rate by any individual at one time ($c. 10^9$ naïve B-cells), however we frequently observe functional convergence of the humoral immune response across populations upon particular ('immunodominant') antigen-binding sites (epitopes). This is one of the great unsolved puzzles of immunology. Studies have suggested various driving forces behind the observation of epitope immunodominance (see Rees *et al.*¹ for a recent review), but regardless of underlying mechanism it follows that overlapping epitope complementarities ought to preexist within the naïve BCR repertoires of different individuals to result in these stereotyped antibody responses (Figure 1).

The capacity to identify same-epitope specificities across the BCR or antibody repertoires of different individuals leads to exciting opportunities to both improve our understanding of immunology and ability to design effective pharmaceuticals. For instance, given a pathogenic antigen, it would aid in the prediction of epitope immunodominance, which in turn can help us to predict how quickly/effectively certain population cohorts are likely to respond to immunogen challenge. When immunodominant epitopes are also linked to humoral immunity, human monoclonal antibody cocktails can be designed to mimic these low-risk, high-reward immune strategies across broad cohorts of patients struggling to fend off a pathogen. Alternatively, antibodies could be developed that target other protective but less immunodominant epitopes. Such therapies would be expected to enjoy a more durable period of clinical benefit, since their binding sites are under lower mutation pressure by natural selection. Another advantage of

understanding the limits of a BCR repertoire's functionality is the potential to reveal epitopes that cannot be recognized by the repertoire (*i.e.* 'functional holes'), that may be exploited by pathogens or lead to increased disease susceptibility across certain sub-populations. This logic could easily be inverted, to investigate BCR functional commonalities amongst specific sub-populations with chronic autoimmune conditions that are absent from an equivalent cohort of healthy volunteers, or across diseased patients with surprisingly high resilience.

The typical pipeline to identify repertoire functional convergence involves sampling the B-cells of several individuals in a particular immune state (healthy, post-vaccination, *etc.*), obtaining the sequences of their BCRs, and comparing these sequences across individuals to find shared/convergent BCR clonal lineages that are inferred to bind to the same epitope. Many reviews have documented and benchmarked the wide variety of bioinformatics software packages that now exist to process and perform clonal analysis on BCR repertoire sequencing (BCR-seq) datasets.^{2–7} However, as yet undiscussed are how decisions made throughout this pipeline, such as experimental choices in the sourcing and stratification of B-cells or the importance and choice of reference BCR-seq datasets can significantly impact how one should interpret these functional clustering results. Moreover, only the review by Kovaltsuk *et al.* in 2017,⁴ which highlighted the potential for augmenting B-cell repertoire analysis by incorporating structural data, expands upon using BCR clonal lineage properties as an *in silico* strategy for BCR functional clustering. The field has since expanded greatly, with several novel published methods that can harness differing degrees of structure prediction to cluster together more genetically diverse antibodies with same-epitope complementarity.

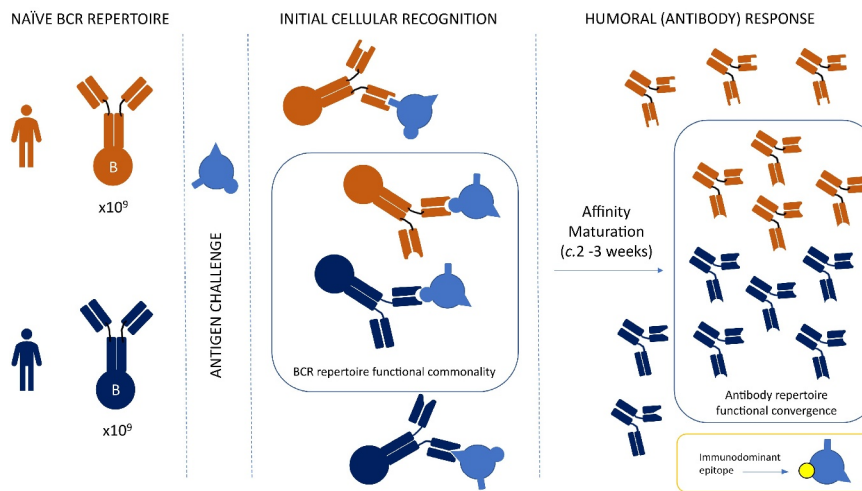


Figure 1. The concepts of functional convergence/epitope immunodominance. Each individual possesses in the region of 10^9 naïve B-cells together sampling a wide range of B-cell receptors (BCRs). Upon antigen exposure, some of these ‘baseline’ BCRs will be sufficiently complementary to various epitopes (shown as a rectangle, triangle, and circle) on the antigen surface and will initiate an immune response. In this example, BCRs exist across the two different individuals that can recognize the “circle” epitope – they have BCR repertoire functional commonality. After differentiation into plasma cells and affinity maturation, the resulting antibody repertoires can also be seen to converge around specificity to this epitope. The “circle” epitope is therefore ‘immunodominant’ over the other two epitopes, since it was the one was able to engender an immune response in both individuals.

In this review, we start by recapping the immunogenomic mechanisms that contribute to the theoretical BCR repertoire sequence diversity of different B-cell compartments and outline the case for why, even though individuals’ B-cell repertoires sample only a tiny fraction of the theoretical BCR diversity at any one time,⁸ we might still expect to observe the significant functional commonalities that are seen between different individuals. We then cover key considerations when designing a robust repertoire analysis study, and how they influence the likelihood/significance of observing functional commonality. These include donor recruitment and stratification, B-cell extraction regimen, B-cell sorting, B-cell sequencing, and finally choosing appropriate computational tools with which to functionally cluster the BCR-seq data. Finally, we discuss potential methods to further improve BCR functional clustering protocols.

Immunogenomic mechanisms contribute to different BCR repertoire diversities in different B-cell compartments

The Y-shaped BCR polypeptide is modular, with separate domains that are functionally distinct. Its cell signaling behavior is governed by the ‘fragment crystallisable (Fc) region’, a highly sequence-conserved ‘stalk’ which exists in only a small number of discrete molecular ‘isotypes’ (IgM, IgD, IgG, IgA, or IgE), each with their own unique set of effector characteristics.^{9,10} By contrast, the antigen recognition function of a BCR relies on sufficient molecular complementarity between its ‘variable (V) domain’ (at the tips of each ‘arm’) and the foreign molecule.

Sequence diversity in the V domain is initially encoded by a set of ‘germline’ gene segments (V and J, and D for certain loci) which are stitched together in different combinations (*via* a process known as ‘V(D)J recombination’¹¹). At the recombination junctions (VD, DJ, or VJ), nucleotide deletions or ‘N’/‘P’

nucleotide insertions can accentuate sequence and length diversity beyond the pre-encoded germline. Once translated and folded, the resulting polypeptide chain forms a stack of antiparallel beta sheets connected by loop regions, of which three ‘complementarity-determining region’ (CDR) loops contribute to antigen binding. These CDRs are highly diverse in sequence, length, and structure, particularly the CDR3 that is encoded across the gene recombination junction(s). Human BCRs comprise two chains encoded at different gene loci, a ‘heavy’ and ‘light’ chain (either kappa or lambda depending on the locus), adding further combinatorial diversity and extending the antigen recognition surface to a set of six neighboring CDRs. The heavy chain of the V domain (VH) and the light chain of the V domain (VL) are together termed the Fv region. VH is the most sequence and structurally diverse, owing primarily to the D-gene fragment contributing to the hypervariable CDRH3 loop (VL contains only a V- and J- gene).

The diversity described in the paragraph above explains the sequence space available to the ‘naïve BCR repertoire’ (characterized as B-cells bearing high concentrations of BCRs with the IgM and IgD isotypes, and expressing diagnostic cell surface receptors such as CD21¹²). The naïve BCR repertoire is responsible for initiating the B-cell-mediated adaptive immune response against antigens that the body has not previously encountered, typically with moderate (*i.e.* micromolar) affinity. The theoretical number of sequence-unique naïve BCRs has recently been estimated at *c.* 10^{15} ,⁸ though far fewer B-cells than this are expressed at any one time due to blood volume constraints. Recent estimates for the number of circulating naïve B-cells in humans put the figure at around 10^9 ,¹ a dynamically changing population whose sampled BCRs change as the cells die and are replaced.

Once a naïve B-cell’s BCR has been activated through an antigen-binding event (and subsequent T-helper cell assistance), it begins to differentiate into a plasma cell, which is able to rapidly proliferate and secrete antibodies – serum-

soluble BCRs. Molecular modifications associated with this transition include displaying different B-cell surface markers and a process known as ‘class switching’ – where the Fc region of the BCR swaps from an IgM or IgD isotype to an IgG, IgA, or IgE isotype. Concurrently, the B-cell migrates to the ‘germinal centre’ of the lymph nodes and on arrival yet more sequence diversity can be introduced through the process of somatic hypermutation (deliberate nucleotide mutations made throughout the receptor V domain sequence of both chains, though predominantly in the CDRs). Positive selection acts to promote mutant BCRs with higher affinity toward the antigen (‘affinity maturation’). Before or after class switching, antigen-activated B-cells can differentiate into long-lived memory cells (often characterized by expression of cell-surface protein CD27¹²), which persist as a high-sensitivity, low concentration population able to be reactivated and potentially undergo further maturation upon secondary infection.¹³ Many transitional B-cell states exist, distinguished by their expression of unique combinations of cell markers and cytokines; these were recently reviewed in detail by Sanz *et al.*¹²

The apparent paradox of BCR repertoire functional commonality

As mentioned above, the size of the theoretical naïve B cell repertoire is thought to be of the order of 10^{15} different sequences, which would exceed the total number of cells of all types in the body by 100 times, the total number of B-cells in the body ($\sim 10^{11}$) by 10,000 times and more importantly, the number of circulating peripheral naïve mature B-cells ($\sim 10^9$ based on numbers of CD27-/IgD+ naïve B-cells) at any one time by a million times. While new, immature B-cells are produced at the rate of $\sim 10^9$ per day, only a small percentage of long-lived B-cells are retained in the periphery as viable mature B-cells. The majority are removed during ‘self-reactive’ depletion in the bone marrow, while those that enter the periphery but fail to successfully enter lymphoid follicles, or experience anergy by recognition of soluble self-antigens, have a much shorter half-life. The theoretical figure of 10^{15} distinct naïve BCRs is therefore likely to be unachievable for an individual, and at any rate must be sampled at an ratio of roughly $1/10^6$ due to physical constraints.

Given such dilute sampling of potential BCR diversity per individual, it might appear that searching for functional commonality in the BCRs of different people would be a fruitless endeavor. This would follow if both (a) repertoire generation is an entirely stochastic process, and (b) that each naïve B-cell clone is essentially functionally-distinct. However, there is increasing evidence to suggest that naïve BCR expression is not a random process, and furthermore that BCRs with propensity to engage the same pathogen epitope are likely to be contemporaneously sampled across individuals.^{14–17}

The basis for widespread BCR repertoire functional commonality

There is now a large body of evidence to suggest that certain germline gene segments are used preferentially across individuals during V(D)J recombination. Both Boyd *et al.* in 2010¹⁸

and Glanville *et al.* in 2011¹⁹ showed that the use frequencies of individual IGHV genes in the naïve antibody repertoire varied between 0.1% and 10%, while IGHD gene usage can vary between 1 and 15%, displaying a strong preference for one of three reading frames.^{20,21} Similar biases are observed in the usage of the IGHJ and light-chain genes, challenging the notion that “all immunoglobulin genes are made equal”.²² Biases can even be observed during the terminal deoxynucleotidyl transferase-catalyzed non-templated N-additions.^{23–25} After applying these estimates of the effects of biased gene selections, effective naïve BCR repertoire clonal diversity could be as low as 10^7 , closer to the levels frequently observed.¹

Both Briney *et al.*⁸ and Soto *et al.*²⁶ have shown that the frequency of shared clones across human naïve B-cell receptor repertoires is significantly higher than expected if clones were generated at random, estimating that typically around 1% of BCR clones are shared (‘public’) between any two individuals (though the shared clones are not common to all individuals, in fact clonal sharing drops to just $\sim 0.02\%$ across 10 people⁸). By virtue of their highly similar sequence features, the public naïve clones are *a priori* expected to have a good chance of initiating an immune response to the same epitope.

There is much evidence from studies of convergent disease-/vaccine-response antibodies to support the fact that BCRs with similar genetic characteristics often engage the same epitopes. For example, in 2013, Parameswaran *et al.*¹⁶ analyzed BCR sequences from the peripheral blood mononuclear cells (PBMCs) of 60 individuals, 48 of whom were infected with Dengue virus plus two control groups: 8 people with either a non-Dengue febrile illness or healthy individuals with no previous history of Dengue infection. They observed evidence of convergence in acute-infected dengue patients, with their BCRs displaying similar CDR3 region sequences that were entirely absent both from the control datasets and a further large reference set of patients from other clinical trials. In 2017, Robbiani *et al.* studied responses to Zika virus across 400 donors in Brazil and Mexico, finding that BCRs from infected individuals with high anti-Zika serum responses displayed common germline gene signatures.¹⁷ In the expanded memory B-cell clones of four of six individuals expressing high titer neutralizing activity, they detected the over-representation of the IGHV3-23/IGKV1-5 gene pairing, as well as strong biases in the other gene regions sampled. Even more strikingly, they found evidence of selection for a common heavy-chain residue that could only have derived from an N region addition. Ehrhardt *et al.*, by analyzing vaccinees B-cell responses to the Ebola rVSV-EBOV vaccine during a Phase I trial in 2019,²⁷ observed that not only was there a preference for response antibodies deriving from the IGHV3-15 heavy-chain germline gene but also that preferential pairing with the IGLV1-40 light-chain gene was present in 17/24 isolated neutralizing antibodies. Moreover, they detected an overlap in the hypermutated sequence positions of both the heavy and light chains from different individuals, and similarity in amino acid replacements, leading the authors to conclude that the affinity maturation process in different individuals had a ‘reproducible pattern’.

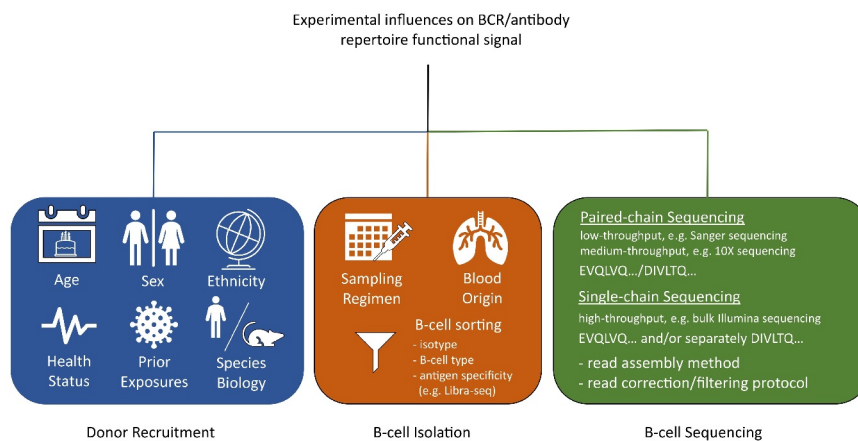


Figure 2. The many experimental parameters that can influence the functional signal in BCR-seq data, spanning the broad categories of blood donor recruitment, B-cell sampling and sorting, and B-cell sequencing and sequence processing.

More recent studies on BCR responses to SARS-CoV-2 by Nielsen *et al.*¹⁴ and Voss *et al.*¹⁵ have observed similar convergence behavior. One of the largest of such SARS-CoV-2 studies by Robbiani *et al.*²⁸ analyzed the memory BCR repertoires of 149 COVID-19 convalescent individuals. Focusing on six individuals with antibodies displaying high and medium to high neutralization activities, they found certain IGH and IGL V-genes to be over-represented, and in some cases found antibodies with almost identical amino acid sequences (up to 92% identity for a IGHV30-3/IGKV1-39 lineage and 99% identity for a IGHV1-58/IGKV3-20 lineage). The authors' conclusion was that these recurrent, clonally expanded antibody sequences derived from a memory repertoire representing a rich source of anti-Receptor Binding Domain SARS-CoV-2 antibodies, equating to < 0.01% of the circulating B cell population. It is reasonable to assume that such convergent antibodies must necessarily have been a by-product of an historical epitope-convergent naïve BCR response.

Beyond infectious diseases, genetic convergence signatures have also been observed in chronic disease contexts, such as in a recent study by DeFalco *et al.* which examined the plasmablast repertoires of patients with metastatic but non-progressing cancers.²⁹ The authors found not only a bias in the usage of heavy and light chain VJ gene combinations, but also identified several CDR sequence features common to multiple individuals across cancer categories.

There is also growing evidence that same-epitope complementarity is not limited to BCRs from the same genetic lineage.^{4,16,17,23,30-35} In these cases, convergence of markedly different sequences toward 'paratope structural signatures' that share shape and property complementarity occurs, embodying what Jackson and Boyd refer to as 'predictive features of human adaptive responses'.²³ This is supported by recent computational analyses on BCR backbone structural usages across repertoires,^{36,37} leading to the theory that 'public baseline structures' might exist across naïve repertoires that could act as broad moderate-affinity templates for somatic-hypermutation optimization against shape-complementary epitopes.^[37]

Overall, the frequency with which epitope immunodominance is observed raises the probability that considerably more than 1% of naïve BCRs across any two individuals, or 0.02% of naïve BCRs across 10 individuals, can initiate an immune response against the same epitopes.⁸ This higher probability of same epitope convergence could be achieved through a form of 'coding redundancy' in the naïve B-cell repertoire, where many different heavy and light-chain sequences can assemble to produce a sufficiently similar sequence and structural paratope to engage the same epitope. This theory has recently been further bolstered by a computational study that found a limited vocabulary of paratope-binding motifs with restricted sequence diversity across structurally solved antibodies.³⁸ Since the same paratope-binding motifs can span multiple-gene origins, the level of antibody convergence we have detected so far through clonal clustering of post-exposure BCR-seq data represents just the 'tip of the iceberg', where evolution drives similar optimizing mutations to push different individuals' epitope-reactive clones toward a consensus lineage.

Experimental factors that can influence functional signals

Many different experimental strategies can be used to isolate B-cells, with an associated impact both on the likelihood and interpretation of observing convergent signals across individuals. Before we review the growing number of computational tools available to functionally probe BCR-seq datasets, we first discuss the factors of B-cell profiling study design known to influence functional signal (Figure 2).

Recruiting donors

The first step in a study to investigate BCR repertoire functional commonality is to recruit appropriate cohorts of individuals from whom to retrieve BCR sequences. A sufficient number of individuals need to be recruited to achieve meaningful statistical significance, and if possible the numbers of individuals in each cohort are balanced.

What constitutes an appropriate cohort is intimately tied to the research question posed. For example, if a study sought to investigate the functional diversity of antibodies raised in response to the first dose of a COVID-19 vaccine, blood donors might be restricted only to those who hadn't received a positive SARS-CoV-2 antigen test result at some point prior to vaccination (since the first dose of vaccine is likely to induce a secondary, and therefore distinct, immune response in pre-exposed people). Alternatively, they might still be recruited regardless but stratified into a separate cohort. In most cases, particularly when considering human BCR repertoires, such stratification is rarely perfect – in this case due to the inaccessibility of testing during the early months of the pandemic and the large degree of asymptomatic SARS-CoV-2 infection. Equally, another condition could be that no volunteer in either cohort had previously been infected with SARS-CoV (which is highly likely to bias any response BCRs through reactivation of memory cells), or had any history of immunodeficiency or chronic disease that could influence their ability to respond effectively. In this context, recruiting a cohort of healthy patients given the 'placebo' (unrelated) vaccine would provide a set of negative control samples that should not only account for antibodies that are highly expressed irrespective of health status, but also the large proportion of bystander antibodies that appear to be activated in any response and that do not possess specific antigen-reactivity (as recently seen by Horns *et al.* in their analysis of influenza vaccination³⁹).

In addition to prior infection/vaccination history, broader cohort characteristics can also be considered and either held constant, balanced, or varied, depending on the research question. Studies have found the immune response is dependent both on age^{40–42} and sex,⁴³ and that the properties of response BCRs/antibodies depend strongly on maturation state³⁶ (see below). Another important, yet often under-considered trait, is the geographical origin and/or ancestry of each volunteer. This is relevant since gene loci are under the influence of local environmental pressures, affecting allelic selection which in turn can play a role in disease susceptibility through different induced antibody responses.^{44–46} For example, IGKV2D-29*02 and IGHV3-23*03 have been found to be significantly over-represented in North American and Asian populations respectively, correlating with differential ability to engage *Haemophilus influenzae* type b (Hib),⁴⁶ while polymorphisms in IGHV1-69 and several IGHV2 loci have been associated with autoimmunity.^{47,48} A recent study has found that BCR gene recombination profiles can vary even amongst human monozygotic twins,⁴⁹ so it remains unclear to what extent individual V(D)J recombination preferences can be accounted for by controlling for volunteer genetics. The role of an individual's microbiome in shaping the nature of the baseline immune repertoire is also becoming apparent.⁵⁰

Often animal models are used to study antibody responses to a pathogen. This strategy is usually employed when sourcing human samples is difficult, to enable sampling of B-cell compartments from primary and secondary lymphoid organs,⁵¹ or to benefit from the rigor of laboratory rather than clinical experimental conditions (*i.e.* allowing for easier control of variables such as the precise time of exposure to pathogen and pathogen load). However, difference in species

biology can be significant in determining whether conclusions drawn from the model are predictive of patterns in the human antibody response to the same stimulus. For example, from the perspective of antibody function, mice have been seen to exhibit a lower frequency of amino acid mutation resulting from somatic hypermutation than humans,⁵² potentially by preferentially using different biochemical repair mechanisms.⁵³ As a result, different mice may be expected to converge on more clonally-similar affinity-matured response BCRs than humans. Mouse BCRs are likely to sample different regions of functional space, since antibodies with reactivity against human proteins will not be selected against by central/peripheral tolerance mechanisms (this is expected to be the case even in transgenic mice with human immunoglobulin loci).⁵⁴ Additionally, if the mice are laboratory-raised, their young age and sterile storage conditions may also be expected to lead to a more predictable and convergent response than would be expected by either mice or humans with an established immune memory of prior antigen exposure.⁵⁵

Extracting B-cells

Once cohorts of volunteers have been selected, an experimental regimen is chosen for the extraction of blood samples.

Typically as many clinical variables as possible are held constant so as not to confound the results. These include taking the same sample volume from each subject (as a proxy for B-cell number that can be confirmed later) and extracting samples at the same timepoint(s) post-infection/vaccination, or when all volunteers are apparently healthy if studying baseline BCR functional properties. Samples are usually processed efficiently within a consistent timeframe, to minimize the risk of sample degradation that might understate the diversity and number of functional B-cell ribonucleic acid (RNA) transcripts present at any one time.

The choice of sampling timepoint post-antigen exposure is another variable that can have a large impact on functional study conclusions. The B-cell immune response can go through many stages of selection over time, not only for affinity but, as is becoming apparent in SARS-CoV-2 vaccination, for neutralization ability and breadth.^{56,57}

The naïve component of the immune response should be strongest in the first week(s) following exposure, overtaken thereafter by class-switched plasma cells that over time benefit from increased levels of affinity maturation.⁵⁸ Antigen-complementary memory B-cells can either be detected in high concentration soon after exposure, or can develop only after antigen-encounter if the subject was originally immunologically-naïve to the pathogen.⁵⁹ Early differentiated memory cells (IgM⁺) are typically of moderate affinity but high promiscuity, able to respond to a range of homologous antigens (e.g. viral variants), while later-differentiated class-switched (IgM⁻ and IgG⁺/IgA⁺) memory B-cells are likely to have higher affinity to the initial immunogen but may exhibit a narrower functional profile.⁶⁰

Most human studies are performed on the B-cell fraction of peripheral blood mononuclear cells (PBMCs), as peripheral blood is the most practical and ethical source of B-cells. As

peripheral blood is not a lymphoid organ (and therefore not an epicenter of the developing immune response) B-cell compartments harboring true disease response BCR signal can be heavily diluted.^{40,61} To compensate for this, the decision might be made to use enrichment tactics such as cell sorting (see below).

Occasionally, the opportunity may arise to obtain samples directly from lymphoid organs by studying subjects *post mortem*.⁶² Such samples are more common in studies on animal models, where BCRs can be sourced from organs such as the bone marrow, spleen, and lymph nodes upon harvesting;⁶³ analyzing these datasets would be expected to yield purer clonal expansion signals with enhanced concentrations of plasma and activated memory B-cells.⁶²

Sorting B-cells

Different BCR compartments carry their own characteristic functional signal that reflects their role in immune surveillance or response.¹² General methods that sort B-cells from T-cells and other cell classes from the PBMC mixture result in the mixing of distinct B-cell compartments, capturing the BCR repertoire in bulk fashion. This can significantly dilute many compartment-specific functional signals, such as those present in certain memory B-cell components that exist at extremely low concentration in peripheral blood.⁶⁴ To combat this, different degrees of cell sorting (appropriate to project budget and time constraints) are sometimes performed prior to sequencing to enrich certain B-cell fractions.

This is usually achieved *via* fluorescence-activated cell sorting (FACS), a specialized form of flow cytometry. It is a form of immunophenotyping, making use of monoclonal antibody reagents (covalently bound to a fluorescent probe) that bind selectively to a particular surface marker protein or BCR isotype. As labeled and unlabeled B-cell droplets flow through the cytometer, those tagged with a fluorescent antibody are deflected into a separate vial, while those without stay on-course. Sequential FACS experiments using different monoclonal antibodies can be used to isolate B-cell compartments with complex diagnostic combinations of cell-surface markers and BCR isotypes.¹² Up to 38 distinct B-cell components can currently be isolated through careful positive and negative selection.⁴⁰ For example, activated class-switched memory B-cells, which reveal BCRs from previous infections that also have complementarity against the vaccine/pathogens, can be enriched by 'gating' for the presence (+) or absence (-) of the following B-cell surface markers: IgD-CD27+CD38-CD24-CD21-IgG/IgA+CD95+CD86+.¹² They are distinguished from resting class-switched memory B-cells *via* their absence of CD38, CD24, CD21 and their expression of CD95.

Some studies go a step further and high-throughput sort bulk B-cells (or pre-sorted B-cell compartments) directly for antigen complementarity. Fluorescence-tagged antigens can be introduced which gravitate toward complementary BCRs and become co-encapsulated within the microfluidics droplet.⁶⁵⁻⁶⁸ FACS can then be used to enrich for the B-cells more likely to be antigen-complementary before subsequent sequencing (see below). In LIBRA-seq, several different antigens can be assessed simultaneously using antigen-specific

deoxyribonucleic acid (DNA) barcodes, enabling antigen specificity to be revealed directly from the single-cell sequence reads.⁶⁶ Alternative *in vitro* and *in vivo* antigen-specificity sorting approaches, including methods of increased sensitivity to analyze rare B-cell compartments such as magnetic enrichment and recombinant phage-display library generation,⁶⁹ have recently been reviewed by Boonyaratankornkit *et al.*⁷⁰

Antigen-specificity sorting is proving to be a transformative technology in accelerating the process of rapidly homing in on high-affinity B-cell clones specific to particular functions of interest. State-of-the-art approaches such as novel epitope display technologies,^{71,72} second-generation LIBRA-seq techniques including 'ligand blocking',⁷³ or higher-throughput epitope mapping approaches⁷⁴ could soon make it possible to select for binders to particular antigen regions.

BCR sequencing

Once B-cell samples have been isolated and potentially sorted, the next procedure is to record the sequence of each B-cell's encoded BCR.

If the B-cells have been sorted for antigen-affinity in a low-throughput manner (e.g. such as cell staining microscopy⁷⁵), the number of implicated BCRs may be sufficiently low that high-fidelity sequencing techniques such as Sanger sequencing are applied. For any higher-throughput antigen-specificity study, or any study without B-cell sorting for antigen association, correspondingly high-throughput ('Next-Generation') sequencing (NGS) techniques are exploited to handle the scale of the data.

Single-chain sequencing

In single ('unpaired') chain sequencing, B-cells are bulk lysed, V(D)J regions are extracted *via* transcription and adjoined with sequencing adaptors (and potentially DNA barcodes), and subsequent polymerase-chain reactions (PCRs) are utilized to amplify the signal and facilitate detection of the full spectrum of reads. The bulk lysis process means that native cellular VH:VL pairings are lost, but allows for a much deeper and faster analysis of the transcript repertoire since B-cells do not have to be handled individually. This represents a trade-off from a functional signal perspective: while whole-binding site resolution is lost, a much deeper sample of the repertoire offers greater statistical confidence in the signal. Sequencing libraries are produced from the amplicons and high-throughput short-read sequencing is employed to retrieve a deep sample of the VH and/or VL repertoire.

Many sequencing protocol choices can influence the resulting distribution of sequences obtained from the experiment. One such choice is whether to analyze messenger RNA (mRNA) or genomic DNA (gDNA). mRNA allows for detection of isotype information at the point of sequencing; in gDNA the switch and constant regions are too distant from the V(D)J region. However, gDNA is more stable and controls better for B-cell expression differences, ensuring read frequency better reflects B-cell abundance and does not bias toward high-expression compartments such as plasma B-cells. DNA barcodes such as Unique Molecular Identifiers (UMIs) can be helpful in correcting for amplification bias and

PCR error, since it is possible to trace multiple assembled reads with the same UMI back to the same central progenitor sequence. For full reviews of library generation protocols, see Yaari *et al.* and Chaudhary *et al.*^{2,5}

Several NGS techniques have been developed since 2009,^{76,77} each with their own advantages and disadvantages. Illumina sequencing remains the dominant platform for single-chain BCR analysis, since it benefits from relatively low cost/read and error rates, while achieving a good read depth, fast speeds through parallel sequencing, and tolerable read length. High fidelity typically lasts for a duration of *c.* 300 base pairs, so the technology must be applied in the forward and then reverse directions to span the complete VH or VL sequence. So, to obtain contiguous nucleotide reads, the Illumina forward and reverse amplicons must be ‘assembled’.

A variety of tools can achieve this, including pRESTO⁷⁸ and MIXCR.⁷⁹ Quality assessment is usually performed during the assembly pipeline to ensure that assembled reads are genuine antibody sequences and do not contain sequencing or library preparation errors. FastQ files obtained from the Illumina sequencing data can offer an indication of nucleotide assignment confidence, while a variety of tools exist to retain only assembled sequences that align tolerably to known Ig germ-lines, are in-frame, and do not contain stop codons. Several tools can also perform UMI correction (given the barcodes used) for correction of amplification bias. The many pipelines available to go from raw reads to compiled sequence datasets are comprehensively reviewed in Jacome *et al.*⁷ Due to sequencing technology biases, and the many parameters that can be tuned during read assembly, comparisons made between datasets sequenced using the same platform and processed using identical assembly pipelines will yield the most robust functional conclusions.

Paired-chain sequencing

To preserve VH:VL pairings, and therefore analyze the function of the whole BCR/antibody binding site (paratope), single-cell (‘paired-chain’) sequencing experiments are required.⁸⁰

The dominant platform for single-cell transcriptome sequencing is currently provided by 10X Genomics.⁸¹ It uses a combination of microfluidics to encapsulate individual cells, capsule-specific reverse-transcription oligonucleotides to label complementarity DNA (cDNA) libraries with their cellular origin, and short-read Illumina sequencing to rapidly capture each cell’s gene expression profile. The Chromium single-cell 5’ RNA-seq technology is designed specifically for generating immunoglobulin (Ig) sequencing libraries, while other primers can be used to create libraries for other gene transcripts. This can be helpful to draw correlates between Ig expression and B-cell state, offering finer resolution than FACS for differentiating highly similar B-cell types,⁸² and is facilitated by new algorithms such as Platypus.⁸³ Reads are assembled and deduplicated using the CellRanger package.⁸¹

Despite parallelisation of the technique, a maximum of around 10,000 cells can be analyzed in each 10X run, meaning it can be difficult, though certainly not impossible,⁸⁴ to gain whole-repertoire functional insight (e.g. clonality) from single-cell sequencing alone. Increasingly, 10X runs are run alongside standard Illumina VH-sequencing, whereby the deeper VH

repertoire samples from the NGS implicate expanded clones of interest and mapping these to the 10X runs allows for entire Fv binding sites of interest to be deduced.³⁹ Alternatively, cell sorting is run prior to single-cell analysis to enrich for a diagnostic B-cell compartment or antigen complementarity.⁸⁵

BCR-seq databases and post-processing

The field of BCR-seq functional analysis has been facilitated by efforts to aggregate BCR-seq data into repositories, including the Observed Antibody Space (OAS)⁵² and iReceptor⁸⁶ databases, containing data shared in a consistent manner according to standards established by the Adaptive Immune Receptor Repertoire Community (AIRR-C).⁸⁷ These AIRR-C standards ensure that relevant metadata (including age, gender, and isotype) is preserved alongside each BCR-seq entry. The OAS database also precompiles ‘data units’ (sequences associated with a unique combination of metadata) within each BCR-seq dataset.

While the increased availability of BCR-seq data opens the door to comparative functional analysis across studies, it may not always be appropriate. For example, finding that a dataset sequenced without UMI correction has more expanded clones than one publicly released post-UMI correction could simply reflect the fact that PCR error has not been accounted for in the first dataset. This and other significant protocol deviations between any two studies impose severe limitations on the validity of detecting differences in repertoire properties. While machine learning methods may be able to distinguish the two datasets, they are far more likely to have learnt biases associated with the different sequencing, assembly, or processing methodologies rather than informative immunogenomic features. AIRR-C standards and OAS data units enable better evaluation of the suitability of comparing two datasets, given all the potential influences on functional signal.

Additional post-processing steps can be applied to assembled BCR-seq datasets to increase repertoire read fidelity, and through the removal of data can also influence functional signal.^{3,88} For example, ABOSS⁸⁸ removes all assembled sequences that lack key conserved features and thus are very unlikely to adopt a stable immunoglobulin fold, while sequences could be filtered by clonal redundancy, with the inference that more commonly observed clones are higher-confidence.³

Computational approaches for functionally analyzing BCR-seq data

At this stage functional analysis strategies tend to diverge depending on the quantity of BCR-seq data available. If a fine-grained sorting regimen (e.g. LIBRA-seq) has been applied to yield a tractable number of sequences for classic *in vitro* profiling, one could proceed directly with antibody expression and molecular characterization. However, in most cases the scale of BCR-seq data necessitates computational analysis and/or clustering to search for features indicative of functional convergence or to home in on particular implicated clones for subsequent experimental characterization.

An array of computational approaches are used for this purpose. We will first describe whole-repertoire immunogenomic features that can be used to detect similar functional shifts between repertoires, many of which can be automatically calculated by processing pipelines.⁷ We will then detail methods that attempt to identify the particular BCRs in different individuals with propensity to engage the same epitope. In all cases where the aim is to distinguish disease-response from baseline BCR repertoires, the degree of overlap should be measured relative to a size-matched healthy or simulated repertoire data,⁸⁹ to account for serendipitous similarity between datasets deriving from innate V(D)J recombination biases.

Repertoire-wide sequence properties with functional relevance

Isotype frequencies

Biases in the usage of different isotypes across volunteers can be functionally informative, as certain isotypes are linked with characteristic functions (e.g. the role of IgA in mucosal immunity). Differential isotype usage relative to a healthy baseline could highlight certain B-cell compartments as centers of a developing immune response. It could also suggest a sampling bias that can impact functional interrogation; if a BCR-seq dataset contains 90% IgM and only 5% IgG sequences, then the functional signal from plasma cells can be expected to be significantly diluted.

CDR Lengths

CDR length usage differences from baseline can indicate a whole-repertoire functional shift driven by selection toward certain immunodominant pathogen epitopes. This is particularly evident in the more variable CDRH3 loop. Infection with the human immunodeficiency virus, for example, has been shown to shift the lengths of the CDRH3s in the IgG repertoire toward significantly longer lengths relative to healthy individuals or those with chronic cytomegalovirus infection.⁹⁰

Gene usage profiles

Each read can be aligned to a reference set of known germline genes using a variety of programs including IMGT/V-Quest⁹¹ and IgBlast⁹² (for a nucleotide sequence) and ANARCI⁹³ (for an amino acid sequence) to infer its genetic origins. Gene usages can be considered on a per-segment basis (e.g. IGHV only) or pairing frequencies can be considered between gene transcripts encoding the same chain (e.g. IGHV-IGHJ usages), or different chains (e.g. IGHV-IG[K/L]V usages). They can be used to identify similar gene expression pattern drifts common in diseased patients *vs.* healthy individuals or those with an unrelated disease. Repertoire somatic hypermutation distributions can also be derived as a byproduct of gene assignment. This can be useful for distinguishing whether an immune response is driven by naïve or affinity-matured B-cells; for example, Galson *et al.* showed that patients with lower COVID-19 disease severity tended to respond with a higher proportion of unmutated sequences.⁹⁴

Clonal clustering ('clonotyping')

Clonal lineage clustering (often termed 'clonotyping') is an extension to gene usage assignment that groups sequences with common predicted V(D)J origins and greater than a certain percentage (typically $\geq c. 80\%$) CDRH3 amino acid sequence identity.³ It can be applied on single-chain (usually VH) data or on VH:VL paired data.

Mapping BCR-seq sequences into approximate clonal lineages allows for the calculation of properties that capture repertoire diversity,^{95–97} such as the clonal diversification index (Renyi entropy, capturing unevenness in the number of V(D)J sequences per clone⁹⁷) or the proportion of BCR-seq data mapping to the ten largest clonotypes.

Proliferation and expansion of a clone is indicative of immune selection and subsequent hypermutation, so can reveal dominant antibodies raised in response to antigen stimulation. In some patients the signal can be striking. For example, the peripheral blood BCR repertoire samples of some recipients of an influenza vaccine in 2016–2017 contained up to 22% of sequences mapping to the same clonal lineage.³⁰ The direction and degree of evolution within members of a clonotype can be estimated by constructing lineage trees from the precursor germline sequence.⁹⁸

Structural Annotation of Antibodies (SAAB+)

SAAB+,³⁶ developed from the original SAAB program,⁹⁹ offers a way to annotate BCR repertoires with estimated structure usage distributions. Typically applied to VH BCR-seq datasets, it uses SCALOP¹⁰⁰ to predict the approximate structure as given by the canonical class of CDRH1 and CDRH2, and FREAD^{101,102} to homology model the structure of the CDRH3 loop, which is binned into the nearest CDRH3 structural cluster. Repertoire-wide properties such as canonical class usage, deviation of canonical class from germline, and CDRH3 structural cluster usage can be calculated. Recent studies have shown structural usage pattern changes along the B-cell development and maturation axis,³⁶ as humans age,⁴² and during a developing immune response.¹⁰³

Repertoire clustering to find BCRs with similar epitope complementaries

Clonotyping

Clonotyping remains the dominant sequence-based methodology for functionally grouping BCRs. It is applied across BCR-seq datasets in a similar manner to intra-repertoire clonal clustering, by aggregating a pair of datasets into a single input file (each sequence labeled with its origin) and 'public' clonotypes are detected as those that contain at least one sequence from both datasets. Clonotyping can also be applied between a BCR-seq dataset and other reference datasets of antibodies to search for convergent signals.^{94,104}

When used for functional annotation, clonotyping assumes certain properties of the BCRs/antibodies that are able to engage the same epitope: (a) that the heavy and (if available) light-chain gene origins must be identical to achieve epitope-complementarity, and (b) that high CDRH3 sequence identity

is required between any two antibodies capable of binding the same epitope. These assumptions cluster a large proportion of binders to some epitopes, but do not hold for many other epitopes that can be engaged by a broad range of lineages.^{4,32–35}

For example, some IGHV genes (such as IGHV3-53 and IGHV3-66) are much closer in sequence identity to one another than many other gene pairs. They are correspondingly more likely to be misclassified for one another and, regardless of classification, to be co-complementary to the same epitope, as seen for in the context of a neutralizing SARS-CoV-2 epitope.⁵⁷ However, clonotyping does not consider gene transcript sequence similarity when assigning lineages, so would exaggerate the difference between these lineages. In several other epitopes, only a handful of contacts to the CDRH3 loop are observed.^{34,105} In these cases CDRH3 identity would not be a good indicator of common function, since the interaction motif necessary for binding can be achieved with a markedly lower net sequence identity than typical clonotyping thresholds.³⁴

Leniency can be introduced to the clonotyping procedure, either by reducing the CDRH3 sequence identity threshold,^{26,33,34} by requiring only a V-gene match instead of V and J,³⁴ or simply focusing on CDRH3 properties alone.¹⁶ However, this comes with the disadvantage that clusters become increasingly likely to contain antibodies that don't bind to the same site as the thresholds are relaxed.

Paratyping

Rather than imposing a sequence identity restriction across the whole CDRH3 loop, paratyping³³ seeks to differentiate between CDR loop residues whose identity can vary without consequence and those whose identity is crucial for same-epitope complementarity. This builds on recent improvements in epitope-agnostic paratope prediction, to the point where heavy chain (VH) paratope residues can now be predicted with a Receiver Operating Characteristic Area Under the Curve (ROC AUC) of over 87%.¹⁰⁶ Paratyping clusters over those antibody residues predicted to be in the paratope using a threshold of 75% sequence identity. While paratyping is sequence-based like clonotyping, some general features of antibody structure may be captured due to the use of predicted paratope residues. Experimental validation on antigen-sorted and bulk BCR-seq data from repertoires responding to pertussis toxoid demonstrated that paratyping can successfully identify a higher proportion of antigen-complementary VH sequences than clonotyping.³³

Ab-Ligity

Ab-Ligity³² also uses predicted paratope residues,¹⁰⁶ but more explicitly considers their relative spatial orientation by building a homology model of the antibody using ABodyBuilder.¹⁰⁷ Once built, the relative disposition of paratope residues is converted into a hash table (using the algorithm originally developed for the small-molecule tool, Ligity¹⁰⁸), that can be directly compared to other converted paratope hash tables to determine binding site similarity. This method was shown to selectively identify anti-lysozyme antibodies that engage the same epitope despite exhibiting as little as 43% CDRH3

sequence identity to one another, while ignoring an antibody with comparable CDRH3 sequence identity that binds a different protein.

Repertoire Structural Profiling

Repertoire Structural Profiling³⁷ (RSP) is an approach that attempts to capture the maximal (modellable) structural diversity in a BCR-seq repertoire. By first predicting homologous templates for each CDR sequence^{101,102} and Fv orientation in a VH:VL pair (artificially created or natural, depending on sequencing methodology), the algorithm then performs rapid greedy structural clustering on the selected templates to map sequence pairings onto representative 'distinct structures'. The analysis can be performed for a single individual or for many individuals. If multiple people are profiled, a set of 'public' model structures can be derived representing topologies likely to be common to every repertoire. One or more sequences from each representative structure, or the full set of sequences assigned to a representative structure of interest, could then be selected for *in vitro* or *in silico* screening library generation,¹⁰⁷ the latter of which can then be functionally probed using methods such as high-throughput docking into an epitope of interest.^{109,110} By holding predicted structure constant rather than sequence, libraries designed by RSP could result in the identification of disparate clones with sufficient structural conservation to engage the same epitope.

A related structural clustering technique has recently been applied to the first virus family-specific antibody database, the Coronavirus Antibody Database (CoV-AbDab),¹⁰⁴ pooling together antibodies from different lineages but with similar model structures that are predicted to engage the same coronavirus epitope.³⁴ By clustering in this way, the binding domain-consistency of multiple-occupancy structural clusters was over 90%.

Comparing BCR-seq data to reference antigen-specific datasets

Any of the above tools can be used to compare the properties of the repertoire sequences against molecules shown experimentally to bind the target of interest in antigen-specific datasets, such as CoV-AbDab³⁷ or an antigen-filtered version of the Structural Antibody Database (SAbDab).¹⁰⁵ Detecting hits to these databases can be used to make a case for the functional 'publicness' of a response BCR/antibody, if the same functional signal was observed multiple times across independent investigations.

For example, Galson *et al.*⁹⁴ compared the convergent VH clonotypes from SARS-CoV-2 patients with CoV-AbDab,¹⁰⁴ finding 10 clonal matches to molecules known to engage SARS-CoV-2 or a related coronavirus. As each CoV-AbDab entry comes with an array of metadata, this knowledge facilitated the prioritization of antibody clones with heightened neutralization potential for further pre-clinical development. Several other papers have since performed clonotype comparison between BCR-seq data and CoV-AbDab^{111–114} to learn from the growing body of knowledge (over 4,150 sequences as of 26th October 2021) on antibodies able to engage coronaviruses.

Discussion

Our current understanding of the extent of BCR repertoire commonality has derived from repertoire-wide property analysis, or clonal lineage/CDRH3 sequence comparisons between antibodies raised across different individuals. However, with the growing evidence that markedly different loop sequences can result in co-complementary paratopes to a given epitope,^{4,32–35} the possibility arises that the already strong convergent signatures we identify in many disease response repertoires represent only a fraction of same-epitope complementarity across individuals' BCR repertoires. Several new BCR-seq analysis tools have been released that attempt to capture this missing signal.

In all cases, these methods relax the conventional genetic constraints of clonotyping to detect molecules with sufficient chemical and/or structural similarity that they may be able to recognise the same epitope. This is likely to be necessary to spot functional commonality between naïve BCRs, which only engage antigens with moderate (micromolar) affinity and typically belong to different clonal definitions from their affinity-matured counterparts. It should also result in the identification of more diverse affinity-matured antibodies able to engage the same target. This will be highly valuable to early-stage antibody drug discovery, where 'hopping' to a new clonal lineage may achieve a common function but with improved developability.³³ Further improvements to these algorithms are likely to come from identifying more optimal balances between the interaction and structural conservation necessary for same-epitope engagement. Current methods will be facilitated by redoubled efforts to solve structures of more diverse antibodies in complex with a variety of pharmacologically relevant antigens (we note also that the recent uptick in data on antibody binders to coronavirus antigens is beginning to lead to significant biases in databases such as SAbDab,¹⁰⁵ that will need to be considered when training/evaluating future algorithms). Furthermore, recent dramatic improvements in our ability to accurately model general protein single-domain structures^{115,116} may soon be translated to improved antibody structure prediction, and eventually to accurate antibody-antigen complex structure prediction, yielding orders of magnitude more training data and concomitant new computational methods.

An assumption in all (non antigen-sorted) BCR-seq clustering methodologies is that selection by the immune system is sufficient to implicate antigen-complementarity. This is not always the case, as shown recently by Horns *et al.*³⁹ Further accuracy improvements in BCR-seq clustering may therefore come through harnessing prior understanding of the epitopes of interest on an antigen. Recently, Akbar *et al.* have shown proof of principle that knowing a solved epitope's structural interaction motif one can reasonably accurately predict the corresponding paratope interaction motifs, each of which tend to have restricted sequence identity.³⁸ This may in the future lead to improved "epitope-aware" paratope prediction, building on earlier work¹¹⁷ that focuses more selectively on

necessary conserved features to engage the epitope and cluster over just these regions. Along these lines, Dumet *et al.* recently released data suggesting that their commercial MAbCluster and MAbTope artificial intelligence methods are able to group together RBD-binding antibodies from CoV-AbDab in a manner consistent with our understanding of their function.¹¹⁸ At this stage, such strategies are only possible and interpretable with some prior knowledge of antibodies able to engage the antigen of interest (e.g. solved structures, or binding assay data), as epitope prediction on the antigen structure alone remains too inaccurate.¹¹⁰ For application in vaccinology, however, it is likely that some initial information would be known about antibodies that can engage a conserved epitope, and therefore epitope-aware methods could offer insight into how likely such an immunogen is to elicit a robust and specific antibody response across a population.

Continued symbiotic advances in both BCR-seq experimental design (such as deep paired-sequencing) and computational BCR functional clustering algorithms could usher in an era where epitope complementarities (and therefore 'functional holes') are identified solely through analysis of patient BCR-seq data. This would not only lead to step-change improvements in drug discovery, but would also contribute to the continued wider debate concerning the driving forces behind repertoire selection.^{1,25,119}

Abbreviations

AIRR-C: Adaptive Immune Receptor Repertoire Community
 BCR: B-cell receptor
 BCR-seq: B-cell receptor repertoire sequencing
 (c)DNA: (complementary) deoxyribonucleic acid
 CDR: Complementarity-determining region
 CoV-AbDab: Coronavirus Antibody Database
 COVID-19: Coronavirus Disease 2019
 FACS: fluorescence-activated cell sorting
 Fc region: Fragment crystallizable region
 Fv region: Fragment variable region
 Hib: Haemophilus influenzae type b
 (m)RNA: (messenger) ribonucleic acid
 NGS: Next-generation sequencing
 OAS: Observed Antibody Space (database)
 PBMCs: peripheral blood mononuclear cells
 PCR: polymerase chain reaction
 RBD: Receptor-Binding Domain
 ROC AUC: Receiver Operator Characteristic Area Under the Curve
 SAbDab: Structural Antibody Database
 SARS-CoV(-2): Severe acute respiratory syndrome coronavirus (-2)
 UMIs: Unique molecular identifiers
 V domain: Variable domain
 VH: heavy chain variable region
 VL: light chain variable region

Disclosure statement


No potential conflict of interest was reported by the author(s).

Funding

This review was supported by funding from Boehringer Ingelheim, awarded to MR.

ORCID

Matthew I. J. Raybould  <http://orcid.org/0000-0002-5663-5297>

Charlotte M. Deane  <http://orcid.org/0000-0003-1388-2252>

References

- Rees AR. Understanding the human antibody repertoire. *mAbs*. 2020;12(1):1729683. doi:10.1080/19420862.2020.1729683.
- Yaari G, Kleinstein SH. Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med*. 2015;7:121.
- Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires. *Trends Immunol*. 2015;36:738–49.
- Kovaltsuk A, Krawczyk K, Galson JD, Kelly DF, Deane CM, Trück J. How B-Cell receptor repertoire sequencing can be enriched with structural antibody data. *Front Immunol*. 2017;8:1753.
- Chaudhary N, Wesemann DR. Analyzing immunoglobulin repertoires. *Front Immunol*. 2018;9:462.
- Miho E, Yermanos A, Weber C, Berger CT, Reddy ST, Greiff V. Computational strategies for dissecting the High-Dimensional Complexity of Adaptive Immune Repertoires. *Front Immunol*. 2018;9:224.
- López-Santibáñez Jácome L, Avendaño Vázquez SE, Flores-Jasso CF. The Pipeline Repertoire for Ig-Seq Analysis. *Front Immunol*. 2019;10:899.
- Briney B, Inderbitzin A, Joyce C, Burton DR. Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature*. 2019;566:393–97.
- Murphy K, Travers P, Walport M, Janeway CJ. *Janeway's Immunobiology*. 8th Ed. New York, USA: Garland Science; 2011.
- Bruce A, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology of the Cell*. 6th Ed. New York, USA: Garland Science; 2015.
- Chapman DD. Overview of the immune response. *J Allergy Clin Immunol*. 2010;125:S3–S23.
- Sanz I, Wei C, Jenks SA, Cashman KS, Tipton C, Woodruff MC, Hom J, Lee FEH. Challenges and opportunities for consistent classification of human B cell and plasma cell populations. *Front Immunol*. 2019;10:2458.
- Sokal A, Barba-Spaeth G, Fernández I, Broketa M, Azzaoui I, Ad LS, Vandenberghe A, Fourati S, Roeser A, Meola A, et al. Memory B cells control SARS-CoV-2 variants upon mRNA vaccination of naive and COVID-19 recovered individuals. *bioRxiv*. 2021. doi:10.1101/2021.06.17.448459.
- Nielsen SCA, Yang F, Jackson KJL, Hoh RA, Röltgen K, Jean GH, Stevens BA, Lee JY, Rustagi A, Rogers AJ, et al. Human B cell clonal expansion and convergent antibody responses to SARS-CoV-2. *Cell Host Microbe*. 2020;28(4):516–25.
- Voss WN, Hou YJ, Johnson NV, Delidakis G, Kim JE, Javanmardi K, Horton AP, Bartzoka F, Paresi CJ, Tanno Y, et al. Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes. *Science*. 2021;372(6546):1108–12.
- Parameswaran P, Liu Y, Roskin KM, Jackson KKL, Dixit VP, Lee JY, Artiles KL, Zompi S, Vargas MJ, Simen BB, et al. Convergent Antibody Signatures in Human Dengue. *Cell Host Microbe*. 2013;13(6):691–700.
- Robbiani DF, Bozzacco L, Keeffe JR, Khouri R, Olsen PC, Gazumyan A, Schaefer-Babajew D, Avila-Rios S, Nogueira L, Patel R, et al. Recurrent potent human neutralizing antibodies to Zika virus in Brazil and Mexico. *Cell*. 2017;169(4):597–609.
- Boyd SD, Gaëta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, Jones CD, et al. Individual Variation in the Germline Ig Gene Repertoire Inferred from Variable Region Gene Rearrangements. *J Immunol*. 2010;184(12):6986–92.
- Glanville J, Kuo TC, von Büdingen HC, Guey L, Berka J, Sundar PD, Huerta G, Mehta GR, Oksenberg JR, Hauser SL, et al. Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. *Proc Natl Acad Sci*. 2011;108(50):20066–71.
- Zemlin M, Schelonka RL, Ippolito GC, Zemlin C, Zhuang Y, Gartland GL, Nitschke L, Pelkonen J, Rajewsky K, Schroeder HW. Regulation of Repertoire Development through Genetic Control of DH Reading Frame Preference. *J Immunol*. 2008;181:8416–24.
- Larimore K, McCormick MW, Robins HS, Greenberg PD. Shaping of Human Germline IgH Repertoires Revealed by Deep Sequencing. *J Immunol*. 2012;189:3221–30.
- Shi B, Dong X, Ma Q, Sun S, Ma L, Yu J, Wang X, Pan J, He X, Su D, et al. The Usage of Human IGHJ Genes Follows a Particular Non-random Selection: the Recombination Signal Sequence May Affect the Usage of Human IGHJ Genes. *Front Genet*. 2020;11:1505.
- Jackson KJL, Boyd SD. Broadening Horizons: new Antibodies Against Influenza. *Cell*. 2016;166:532–33.
- DeKosky BJ, Lungu OI, Park D, Johnson EL, Charab W, Chrysostomou C, Kuroda D, Ellington AD, Ippolito GC, Gray JJ, et al. Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. *Proc Natl Acad Sci*. 2016;113(19):E2636–E2645.
- Mroczek ES, Ippolito G, Rogosch T, Hoi KH, Hwangpo T, Brand M, Zhuang Y, Liu CR, Schneider D, Zemlin M, et al. Differences in the Composition of the Human Antibody Repertoire by B Cell Subsets in the Blood. *Front Immunol*. 2014;5:96.
- Soto C, Bombardi RG, Branchizio A, Kose N, Matta P, Sevy AM, Sinkovits RS, Gilchuk P, Finn JA, Crowe JJE. High frequency of shared clonotypes in human B cell receptor repertoires. *Nature*. 2019;566:398–402.
- Ehrhardt SA, Zehner M, Krähling V, Coven-Dvashi H, Kreer C, Elad N, Gruell H, Ercanoglu MS, Schommers P, Gieselmann L, et al. Polyclonal and convergent antibody response to Ebola virus vaccine rVSV-ZEBOV. *Nat Med*. 2019;25:1589–600.
- Robbiani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z, Cho A, Agudelo M, Barnes CO, Gazumyan A, Finkin S, et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature*. 2020;584:437–42.
- DeFalco J, Harbell M, Manning-Bog A, Baia G, Scholz A, Millare B, Sumi M, Zhang D, Chu F, Dowd C, et al. Non-progressing cancer patients have persistent B cell responses expressing shared antibody paratopes that target public tumor antigens. *Clinical Immunology*. 2018;187:37–45.
- Forgacs D, Abreu RB, Sautto GA, Kirchenbaum GA, Drabek E, Williamson KS, Kim D, Emerling DE, Ross TM. Convergent antibody evolution and clonotype expansion following influenza virus vaccination. *PLoS ONE*. 2021;16:e0247253.
- Scheid JF, Mouquet H, Ueberheide B, Diskin R, Klein F, Oliveira TYK, Pietzsch J, Fenyo D, Abadir A, Velinzon K, et al. Sequence and Structural Convergence of Broad and Potent HIV Antibodies That Mimic CD4 Binding. *Science*. 2011;333(6049):1633–37.
- Wong WK, Robinson SA, Bujotzek A, Georges G, Lewis AP, Shi J, Snowden J, Taddese B, Deane CM. Ab-Ligity: identifying sequence-dissimilar antibodies that bind to the same epitope. *mAbs*. 2021;13:1873478.
- Richardson E, Galson JD, Kellam P, Kelly DF, Smith SE, Palser A, Watson S, Deane CM. A computational method for immune repertoire mining that identifies novel binders from different clonotypes, demonstrated by identifying anti-pertussis toxoid antibodies. *mAbs*. 2021;13:1869406.
- Robinson SA, Raybould MIJ, Schneider C, Wong WK, Marks C, Deane CM. Epitope profiling of coronavirus-binding antibodies using computational structural modelling. *bioRxiv*. 2021. doi:10.1101/2021.04.12.439478.

35. McCarthy KR, Lee J, Watanabe A, Kuraoka M, Robinson-mccarthy LR, Georgiou G, Kelsoe G, Harrison SC, Prevalent Focused A. Human Antibody Response to the Influenza Virus Hemagglutinin Head Interface. *MBio*. 2021;12:e01144–21.
36. Kovaltsuk A, Raybould MIJ, Wong WK, Marks C, Kelm S, Snowden J, Trück J, Deane CM. Structural diversity of B-cell receptor repertoires along the B-cell differentiation axis in humans and mice. *PLoS Computat Biol*. 2020;16:e1007636.
37. Raybould MIJ, Marks C, Kovaltsuk A, Lewis AP, Shi J, Deane CM. Public Baseline and shared response structures support the theory of antibody repertoire functional commonality. *PLoS Comput Biol*. 2021;17:e1008781
38. Akbar R, Robert PA, Pavlović M, Jeliazkov JR, Snapkov I, Slabodkin A, Weber CR, Scheffer L, Miho E, Haff IH, et al. A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding. *Cell Rep*. 2021;34(11):108856.
39. Horns F, Dekker CL, Quake SR. Memory B Cell Activation, Broad Anti-influenza Antibodies, and Bystander Activation Revealed by Single-Cell Transcriptomics. *Cell Rep*. 2020;30:905–13.
40. Blanco E, Péres-Andrés M, Arriba-Méndez S, Contreras-Sanfeliciano T, Criado I, Pelak O, Serra-Caetano A, Romero A, Puig N, Remesal A, et al. Age-associated distribution of normal B-cell and plasma cell subsets in peripheral blood. *J Allergy Clin Immunol*. 2018;141(6):2208–P2219.
41. Alpert A, Pickman Y, Leipold M, Rosenberg-Hasson Y, Ji X, Gaujoux R, Rabani H, Starosvetsky E, Kveler K, Schaffert S, et al. A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nat Med*. 2019;25:487–95.
42. Ghraichy M, Galson JD, Kovaltsuk A, von Niederhäusern V, Pachlopnik Schmid J, Recher M, Jauch AJ, Miho E, Kelly DF, Deane CM, et al. Maturation of the Human Immunoglobulin Heavy Chain Repertoire With Age. *Front Immunol*. 2020;11:1734
43. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol*. 2016;16:626–38.
44. Kerui P, Safonova Y, Shugar M, Popejoy AB, Rodriguez OL, Breden F, Brodin P, Burkhardt AM, Bustamante C, Cao-Lormeau VM, et al. Diversity in immunogenomics: the value and the challenge. *Nat Methods*. 2021. doi:10.1038/s41592-021-01169-5.
45. Mikocziova I, Greiff V, Sollid LM. Immunoglobulin germline gene variation and its impact on human disease. *Genes Immun*. 2021. doi:10.1038/s41435-021-00145-5.
46. Watson CT, Breden F. The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes Immun*. 2012;13:363–73.
47. Vencovský J, Zďárský E, Moyes SP, Hajeer A; Ruzicková u, Cimburek Z, Ollier WE, Maini RN, Mageed RA. Polymorphism in the immunoglobulin VH gene V1-69 affects susceptibility to rheumatoid arthritis in subjects lacking the HLA-DRB1 shared epitope. *Rheumatology*. 2002;41(4):401–10.
48. Hashimoto LL, Walter MA, Cox DW, Ebers GC. Immunoglobulin heavy chain variable region polymorphisms and multiple sclerosis susceptibility. *J Neuroimmunol*. 1993;44:77–83.
49. Slabodkin A, Chernigovskaya M, Mikocziova I, Akbar R, Scheffer L, Pavlović M, Bashour H, Snapkov I, Mehta BB, Weber CR, et al. Individualized VDJ recombination predisposes the available Ig sequence space. *bioRxiv*. 2021. doi:10.1101/2021.04.19.440409.
50. Vogl T, Klompus S, Leviatan S, Kalka IN, Weinberger A, Wijmenga C, Fu J, Zhernakova A, Weersma RK, Segal E. Population-wide diversity and stability of serum antibody epitope repertoires against human microbiota. *Nat Med*. 2021. doi:10.1038/s41591-021-01409-3.
51. Aizik L, Dror Y, Taussig D, Barzel A, Carmi Y, Wine Y. Antibody Repertoire Analysis of Tumor-Infiltrating B Cells Reveals Distinct Signatures and Distributions Across Tissues. *Front Immunol*. 2021;12:2820.
52. Kovaltsuk A, Leem J, Kelm S, Snowden J, Deane CM, Krawczyk K. Observed Antibody Space: a Resource for Data Mining Next-Generation Sequencing of Antibody Repertoires. *J Immunol*. 2018;201:2502–09.
53. Cui A, Di Niro R, Vander Heiden JA, Briggs AW, Adams K, Gilbert T, O'Connor KC, Vigneault F, Shlomchik MJ, Kleinstein SH. A Model of Somatic Hypermutation Targeting in Mice Based on High-Throughput Ig Sequencing Data. *J Immunol*. 2016;197:3566–74.
54. Nemazee D. Mechanisms of central tolerance for B cells. *Nat Rev Immunol*. 2017;17:281–94.
55. Masopust D, Sivula CP, Jameson SC. Of Mice, Dirty Mice, and Men: using Mice To Understand Human Immunology. *J Immunol*. 2017;199:383–88.
56. Wang Z, Muecksch F, Schaefer-Babajew D, Finkin S, Viant C, Gaebler C, Barnes C, Cipolla M, Ramos V, Oliveira TY, et al. Vaccination boosts naturally enhanced neutralizing breadth to SARS-CoV-2 one year after infection. *bioRxiv*. 2021. doi:10.1101/2021.05.07.443175.
57. Clark SA, Clark LE, Pan J, Coscia A, Lga M, Shankar S, Johnson RI, Brusica V, Choudhary MC, Regan J, et al. SARS-CoV-2 evolution in an immunocompromised host reveals shared neutralization escape mechanisms. *Cell*. 2021;184(10):2605–17.
58. Galson JD, Trück J, Fowler A, Clutterbuck EA, Münz M, Cerundolo V, Reinhard C, Van Der Most R, Pollard AJ, Lunter G, et al. Analysis of B Cell Repertoire Dynamics Following Hepatitis B Vaccination in Humans, and Enrichment of Vaccine-specific Antibody Sequences. *EBioMedicine*. 2015;2(12):2070–79.
59. Akkaya M, Kwak K, Pierce SK. B cell memory: building two walls of protection against pathogens. *Nat Rev Immunol*. 2020;20:229–38.
60. Palm AKE, Henry C. Remembrance of Things Past: long-Term B Cell Memory After Infection and Vaccination. *Front Immunol*. 2019;10:1787.
61. Briney BS, Willis JR, Finn JA, McKinney BA, Crowe JE Jr. Tissue-Expressed Antibody Variable Gene Repertoires. *PLoS ONE*. 2014;9:6.
62. Yang F, Nielsen SCA, Hoh RA, Röltgen K, Wirz OF, Haraguchi E, Jean GH, Lee JY, Pham TD, Jackson KJL, et al. Shared B cell memory to coronaviruses and other pathogens varies in human age groups and tissues. *Science*. 2021;372(6543):738–41.
63. Asensio MA, Lim YW, Wayham NW, Stadtmiller K, Edgar RC, Leong J, Leong R, Mizrahi RA, Adams MS, Simons JF, et al. Antibody repertoire analysis of mouse immunization protocols using microfluidics and molecular genomics. *mAbs*. 2019;11(5):870–83.
64. Franz B, May KF, Dranoff G, Wucherpfennig K. Ex vivo characterization and isolation of rare memory B cells with antigen tetramers. *Blood*. 2011;118:348–57.
65. Ouisse LH, Gautreau-Rolland L, Devilder MC, Osborn M, Moyon M, Visentin J, Halary F, Bruggemann M, Buelow R, Anegón I, et al. Antigen-specific single B cell sorting and expression-cloning from immunoglobulin humanized rats: a rapid and versatile method for the generation of high affinity and discriminative human monoclonal antibodies. *BMC Biotechnol*. 2017;17:3.
66. Setliff I, Shiakolas AR, Pilewski KA, Murji AA, Mapengo RE, Janowska K, Richardson S, Oosthuysen C, Raju N, Ronsard L, et al. High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity. *Cell*. 2019;179:1636–46.
67. Lei L, Tran K, Wang Y, Steinhardt JJ, Xiao Y, Chiang CI, Wyatt RT, Li Y. Antigen-Specific Single B Cell Sorting and Monoclonal Antibody Cloning in Guinea Pigs. *Front Microbiol*. 2019;10:672.
68. Gerard A, Woolfe A, Mottet G, Reichen M, Castrillon C, Menrath V, Ellouze S, Poitou A, Doineau R, Briseno-Roa L, et al. High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics. *Nat Biotechnol*. 2020;38:715–21.

69. Rajan S, Kierny MR, Mercer A, Wu J, Tovchigrechko A, Wu H, Dall'Acqua WF, Xiao X, Chowdhury PS. Recombinant human B cell repertoires enable screening for rare, specific, and natively paired antibodies. *Commun Biol.* 2018;1:5.
70. Boonyaratanakornkit J, Taylor JJ. Techniques to Study Antigen-Specific B Cell Responses. *Front Immunol.* 2019;10:1694.
71. Rossmann M, Greive J, S, Moschetti T, Dinan M, Hyvönen M. Development of a multipurpose scaffold for the display of peptide loops. *Protein Eng Des Sel.* 2017;30(6):419–30. 4.
72. Fühner V, Heine PA, Zilkens KJC, Meier D, Roth KDR, Moreira GMSG, Hust M, Russo G. Epitope Mapping via Phage Display from Single-Gene Libraries in Human Monoclonal Antibodies: methods and Protocols. New York, USA: Springer; 2019.
73. Shiakolas AR, Johnson N, Kramer KJ, Suryadevara N, Wrapp D, Periasamy S, Pilewski KA, Raju N, Nargi R, Sutton RE, et al. Efficient discovery of potentially neutralizing SARS-CoV-2 antibodies using LIBRA-seq with ligand blocking. *bioRxiv.* 2021. doi:10.1101/2021.06.02.446813.
74. Dejnirattisai W, Zhou D, Ginn HM, Duyvesteyn HME, Supasa P, Case JB, Zhao Y, Walter TS, Mentzer AJ, Liu C, et al. The antigenic anatomy of SARS-CoV-2 receptor binding domain. *Cell.* 2021;184(8):2183–200.
75. Krautler NJ, Suan D, Butt D, Bourne K, Hermes JR, Chan TD, Sundling C, Kaplan W, Schofield P, Jackson J, et al. Differentiation of germinal center B cells into plasma cells is initiated by high-affinity antigen and completed by Tfh cells. *J Exp Med.* 2017;214:1259–67.
76. Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. Direct Comparisons of Illumina vs. Roche 454 Sequencing Technologies on the Same Microbial Community DNA Sample. *PLoS ONE.* 2012;7:e30087.
77. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. A tale of three next generation sequencing platforms: comparison of ion torrent, pacific biosciences and illumina miseq sequencers. *BMC Genomics.* 2012;13:341.
78. Vander Heiden JA, Yaari G, Uduman M, Stern JN, O'Connor KC, Hafner DA, Vigneault F, Kleinstein SH. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics.* 2014;30:1930–32.
79. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, Chudakov DM. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods.* 2015;12:380–81.
80. Neu KE, Tang Q, Wilson PC, Khan AA. Single-Cell Genomics: approaches and Utility in Immunology. *Trends Immunol.* 2017;38:140–49.
81. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017;8:14049.
82. He B, Liu S, Wang Y, Xu M, Cai W, Liu J, Bai W, Ye S, Ma Y, Hu H, et al. Rapid isolation and immune profiling of SARS-CoV-2 specific memory B cell in convalescent COVID-19 patients via LIBRA-seq. *Sig Transduct Target Ther.* 2021;6:195.
83. Yermanos A, Agrafiotis A, Kuhn R, Robbiani D, Yates J, Papadopoulou C, Han J, Sandu I, Weber C, Bieberich F, et al. Platypus: an open-access software for integrating lymphocyte single-cell immune repertoires with transcriptomes. *NAR Genom Bioinform.* 2021;3:2.
84. Sureshchandra S, Lewis SA, Doratt B, Jankeel A, Ibraim I, Messaoudi I. Single cell profiling of T and B cell repertoires following SARS-CoV-2 mRNA vaccine. *bioRxiv.* 2021. doi:10.1101/2021.07.14.452381.
85. Andreano E, Nicastrì E, Paciello I, Pileri P, Manganaro N, Piccini G, Manenti A, Pantano E, Kabanova A, Troisi M, et al. Extremely potent human monoclonal antibodies from COVID-19 convalescent patients. *Cell.* 2021;184(7):1821–35.
86. Corrie BD, Marthandan N, Zimonja B, Jaglale J, Zhou Y, Barr E, Knoetze N, Breden FMW, Christley S, Scott JK, et al. iReceptor: a platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol Rev.* 2018;284(1):24–41.
87. Scott JK, Breden F. The adaptive immune receptor repertoire community as a model for FAIR stewardship of big immunology data. *Curr Opin Syst Biol.* 2020;24:71–77.
88. Kovaltsuk A, Krawczyk K, Kelm S, Snowden J, Deane CM. Filtering Next-Generation Sequencing of the Ig Gene Repertoire Data Using Antibody Structural Information. *J Immunol.* 2018;201:3694–704.
89. Mora QM, T, Walczak AM. High-throughput immune repertoire analysis with IGoR. *Nat Commun.* 2018;9:561.
90. Roskin KM, Jackson KJL, Lee JY, Hoh RA, Joshi SA, Hwang KK, Bonsignori M, Pedroza-Pacheco I, Liao HX, Moody MA, et al. Aberrant B cell repertoire selection associated with HIV neutralizing antibody breadth. *Nat Immunol.* 2020;21:199–209.
91. Brochet X, Lefranc MP, Giudicelli V. IGMT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.* 2008;36:W503–W508.
92. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* 2013;41:W34–W40.
93. Dunbar J, Deane CM. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics.* 2015;32:298–300.
94. Galson JD, Schaetzle S, Rjm B-R, Raybould MIJ, Kovaltsuk A, Kilpatrick GJ, Minter R, Finch DK, Dias J, James LK, et al. Deep Sequencing of B Cell Receptor Repertoires From COVID-19 Patients Reveals Strong Convergent Immune Signatures. *Front Immunol.* 2020;11:3283.
95. Hill MO. Diversity and Evenness: a Unifying Notation and Its Consequences. *Ecology.* 1973;54:427–32.
96. Bashford-Rogers RJM, Palser AL, Huntly BJ, Rance R, Vassiliou GS, Follows GA, Kellam P. Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. *Genome Res.* 2013;23:1874–84.
97. Bashford-Rogers RJM, Bergamaschi L, McKinney EF, Pombal DC, Mescia F, Lee JC, Thomas DC, Flint SM, Kellam P, Jayne DRW, et al. Analysis of the B cell receptor repertoire in six immune-mediated diseases. *Nature.* 2019;574:122–26.
98. Yermanos AD, Dounas AK, Stadler T, Oxenius A, Reddy ST. Tracing Antibody Repertoire Evolution by Systems Phylogeny. *Front Immunol.* 2018;9:2149.
99. Krawczyk K, Kelm S, Kovaltsuk A, Galson JD, Kelly D, Trück J, Regep C, Leem J, Wong WK, Nowak J, et al. Structurally Mapping Antibody Repertoires. *Front Immunol.* 2018;9:1698.
100. Wong WK, Georges G, Ros F, Kelm S, Lewis AP, Taddese B, Leem J, Deane CM. SCALOP: sequence-based antibody canonical loop structure annotation. *Bioinformatics.* 2018;35:1774–76.
101. Choi Y, Deane CM. FREAD revisited: accurate loop structure prediction using a database search algorithm. *Proteins.* 2010;78:1431–40.
102. Choi Y, Deane CM. Predicting antibody complementarity determining region structures without classification. *Mol BioSyst.* 2011;7:3327–34.
103. Cao Y, Su B, Guo X, Sun W, Deng Y, Bao L, Zhu Q, Zhang X, Zheng Y, Geng C, et al. Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell.* 2020;182(1):73–84.
104. Raybould MIJ, Kovaltsuk A, Marks C, Deane CM. CoV-AbDab: the Coronavirus Antibody Database. *Bioinformatics.* 2021;37:734–35.
105. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, Shi J, Deane CM. SAbDab: the structural antibody database. *Nucleic Acids Res.* 2013;42:D1140–D1146.
106. Liberis E, Veličković P, Sormanni P, Vendruscolo M, Liò P. Parapred: antibody paratope prediction using convolutional and recurrent neural networks. *Bioinformatics.* 2018;34:2944–50.

107. Leem J, Dunbar J, Georges G, Shi J, Deane CM. ABodyBuilder: automated antibody structure prediction with data-driven accuracy estimation. *mAbs*. 2016;8:1259–68.
108. Ebejer JP, Finn PW, Wong WK, Deane CM, Morris GM. Lidity: a Non-Superpositional, Knowledge-Based Approach to Virtual Screening. *J Chem Inf Model*. 2019;59:2600–16.
109. Schneider C, Buchanan A, Taddese B, Deane CMDLAB. Deep learning methods for structure-based virtual screening of antibodies. *bioRxiv*. 2021;10.1101/2021.02.12.430941.
110. Raybould MIJ, Wong WK, Deane CM. Antibody-antigen complex modelling in the era of immunoglobulin repertoire sequencing. *Mol Syst Des Eng*. 2019;4:679–88.
111. Paschold L, Simmica D, Willscher E, Vehreschild MJGT, Dutzmann J, Sedding DG, Schultheis C, Binder M. SARS-CoV-2-specific antibody rearrangements in prepandemic immune repertoires of risk cohorts and patients with COVID-19. *J Clin Invest*. 2021;131:e142996.
112. Bullen G, Galson JD, Hall G, Villar P, Moreels L, Ledsgaard L, Mattiuzzo G, Bentley EM, Masters EW, Tang D, et al. Cross-reactive sars-cov-2 neutralizing antibodies from deep mining of early patient responses. *Front Immunol*. 2021;12:2049.
113. Ehling RA, Weber CR, Mason DM, Friedensohn S, Wagner B, Bieberich F, Kapetanovic E, Vazquez-Lombardi R, Di Roberto RB, Hong KL, et al. Single-cell sequencing of plasma cells from COVID-19 patients reveals highly expanded clonal lineages produce specific and neutralizing antibodies to SARS-CoV-2. *bioRxiv*. 2021. doi:10.1101/2021.02.12.430940.
114. Collier DA, Ferreira IATM, Kotagiri P, Datir R, Lim E, Touizer E, Meng B, Abdullahi A, Collaboration TCNBC, Elmer A, et al. Age-related immune response heterogeneity to SARS-CoV-2 vaccine BNT162b2. *Nature*. 2021. doi:10.1038/s41586-021-03739-1.
115. Jumper J, Evans R, Prizel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021. doi:10.1038/s41586-021-03819-2.
116. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021. doi:10.1126/science.abj8754.
117. Krawczyk K, Baker T, Shi J, Deane CM. Antibody i-Patch prediction of the antibody binding site improves rigid local antibody-antigen docking. *Protein Eng Des Sel*. 2013;26:621–29.
118. Dumet C, Jullian Y, Musnier A, Rivière P, Poirier N, Watier H, Bourquard T, Poupon A. Exploring epitope and functional diversity of anti-SARS-CoV2 antibodies using AI-based methods. *bioRxiv*. 2020;10.1101/2020.12.23.424199.
119. Jackson K, Kidd M, Wang Y, Collins A. The shape of the lymphocyte receptor repertoire: lessons from the B Cell Receptor. *Front Immunol*. 2013;4:263.