Using High-Throughput Sequencing to Characterize the Development of the Antibody Repertoire During Infections: A Case Study of HIV-1

12

Felix Breden and Corey T. Watson

Abstract

High throughput sequencing (HTS) approaches have only recently been applied to describing the antibody/B-cell repertoire in fine detail, but these data sets have already become critical to the design of vaccines and therapeutics, and monitoring of cancer immunotherapy. As a case study, we describe the potential and present limitations of HTS studies of the Ab repertoire during infection with HIV-1. Most of the present studies restrict their analyses to lineages of specific bnAbs. We discuss future initiatives to expand this type of analysis to more complete repertoires and to improve comparing and sharing of these Ab repertoire data across studies and institutions.

Keywords

Antibody repertoire · High throughput sequencing · HIV-1 · Broadly neutralizaing antibodies

F. Breden (✉)
Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada
e-mail: breden@sfu.ca

C. T. Watson
Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, KY, USA

12.1 Introduction

B lymphocytes are key players in the adaptive immune response, and in concert with other cells of the immune system, are capable of directing the recognition of and response against a vast array of pathogens [43]. The critical functions of B cells are made possible by the expression of diverse proteins called immunoglobulins (IG), expressed either as membrane-bound cell surface B cell receptors (BCRs) or secreted antibodies (Abs), each comprised of two identical heavy and two identical light chains. Mirroring the diversity observed among the potential pathogens and related epitopes that could be encountered during a lifetime, the amino acid diversity of the Abs produced by an individual (i.e., the Ab repertoire) is astounding, with theoretical estimates exceeding 10^{15} unique Abs in a given repertoire [54]. The diversity observed in the naïve Ab repertoire is the consequence of unique molecular processes, beginning with combinatorial diversity that arises from the somatic "V(D)J" recombination of hundreds of germline heavy and light chain variable (V), diversity (D; heavy chain only), and joining (J) gene segments across three primary loci in the human genome, accompanied by additional junctional variation introduced via the addition of nucleotides at the junctions of V, D, and J gene segments, and the random pairing of heavy and light chains [43, 51, 63]. Additional diversity is later added to the repertoire following antigen stimulation via somatic hypermutation (SHM).

Effectively and comprehensively capturing the diversity of the Ab repertoire has historically been a challenging endeavor. However, the recent application of high-throughput sequencing (HTS) to Ab repertoire profiling has begun to provide great insight into the dynamic nature of the Ab response [20]. With these methods, generally referred to as "RepSeq", repertoire-wide profiling of Ab variable domains (and in some cases including constant domain isotype-level information) can be conducted in depth at either the level of rearranged genomic DNA or mRNA, and can be done across millions of cells in bulk, or at the single cell level [67]. In addition, accompanying these technologies has come a surge of computational methods for analyzing such HTS data [79], allowing for complex modeling of the Ab repertoire and in depth descriptions of a range of repertoire features (e.g., IG gene segment usage; complementarity determining region (CDR) length and diversity; SHM patterns; clonal expansion). Over the past several years, these approaches have been applied in many disease and clinical contexts and across cell subsets and time points, clearly establishing that Ab repertoires can be influenced by many factors, including age, infection history, disease status, genetic background, and treatment/prophylactics [18, 21, 23, 39, 47, 50, 61, 64, 66, 77, 81]. In summary, HTS-based analyses of immune repertoires have enormous promise for understanding repertoire dynamics in immunology, vaccinology, infectious diseases, autoimmunity, and cancer biology.

In this review, we present the results of the use of HTS of antibody repertoires in the development of the adaptive immune response to HIV-1, to illustrate the importance and utility of this method. We summarize some general conclusions from such studies, but also some significant limitations, especially related to the analysis of only a limited section of the complete Ab repertoire, typically in few individuals. Finally, we review new trends in adaptive immune repertoire analysis that may provide opportunity for overcoming these limitations.

12.2 Characterizing the Development of Ab Repertoires During HIV-1 Infection

12.2.1 General Questions and Importance of HTS Repertoire Studies

Studies of the development of Ab repertoires in health and disease have been dominated by those targeting HIV-1 Abs. This is due to several circumstances, including disproportionate funding rates and the obvious immediate application of such studies to vaccine development. These studies of Ab repertoires in HIV-1 infection provide an excellent example of the power of using HTS for studying Ab repertoires, and we therefore have chosen to focus on this set of studies in our review.

At least two related questions have substantially motivated the studies of Ab repertoires during HIV-1 infection. The first question concentrates on the characteristics of anti-HIV-1 broadly neutralizing Abs (bnAbs), which an effective HIV-1 vaccine would presumably have to elicit in naïve individuals. Much has been made of the observation that the sequences of most anti-HIV-1 bnAbs are highly somatically mutated, encode long ($\geq$20 aa) heavy chain CDR3 (CDR-H3) sequences, and are poly- or auto-reactive ([19, 40]; but see [10]). This observation led some researchers to suggest that one explanation for the rarity of bnAbs against HIV-1 is that the individual must "break tolerance" to produce these bnAbs [24]. However, all bnAbs against HIV-1 don't have these properties, and to what extent these features are required for the acquisition of neutralization breadth and potency of bnAbs remains poorly understood. The application of HTS to probe the Ab repertoire opened the possibility of describing the characteristics of bnAbs against HIV-1 in fine detail, and this remains one of the strongest motivations for these studies.

The second question motivating HTS Ab repertoire studies in HIV-1 focuses on the timing or ontogeny of the development of the broadly-neutralizing response during infection. It is well established that most HIV-1-infected individuals produce strong strain-specific bnAb responses against HIV-1 envelope (Env) soon after initial infection; yet only a small percentage of infected individuals develop broad neutralization, and then only after a year or more [22, 57, 58]. The "B-cell-lineage immunogen design" strategy for anti-HIV-1 vaccine development attempts to understand the timing and action of immunogens that led to the development of anti-HIV-1 bnAbs, and then induce such bnAbs in naïve individuals [5, 25]. This "directed evolution" approach to rational immunogen design relies on identifying clonal lineages that have led to bnAbs in order to infer their common unmutated ancestors (putative naïve B cell precursor) and intermediate sequences within the Ab clonal lineages. For a given Ab lineage, once the Ab sequences of the unmutated ancestor and intermediates along the developmental lineage have been identified, it is then possible to identify epitopes to which these Abs bind. Finally, this points to possible immunogens that might be used to drive the Ab response along specific, predetermined "evolutionary" pathways. The goal is to recapitulate these lineages in naïve individuals, and increase the chance that these naïve individuals will ultimately produce bnAbs themselves. This approach has dominated HIV-1 vaccine research for several years, and explains why many of the studies concentrate on only specific lineages, targeted to known bnAbs within the broad Ab response ("targeted" lineages). This is in contrast to studies that would attempt to capture the complete Ab milieu during infection, *i.e.*, all of the dominant Ab lineages present in the repertoire. In this review we will discuss how this emphasis on studying particular targeted lineages has shaped HIV-1 Ab repertoire research, and what advantages a broader approach examining multiple lineages might provide.

12.2.2 Survey of Studies of the Ab Response During HIV-1 Infection Using HTS

In this section, we review those studies that use HTS to characterize the anti-HIV-1 Ab response. We include only those studies that report a response to HIV-1 infection, rather than that induced by vaccination. Most of the bnAbs known to target HIV-1 bind to a small set of epitopes, and we organize the studies according to these regions (summarized in Table 12.1).

12.2.2.1 CD4 Binding Site (CD4bs)

The largest set of studies using HTS to explore antibody production and development in response to HIV-1 infection are those that target the CD4 binding site (CD4bs); one of the earliest studies of this type was by Wu et al. [76]. In this study, PBMC from 3 donors from the international AIDS-vaccine initiative (IAVI) protocol G were isolated and subjected to HTS. This set included Donor 45, from whom bnAbs VRC01, VRC02 and VRC03 had been originally isolated. These sequences, especially VRC01, were subjects of intense interest, given their ability to neutralize up to 90% of HIV-1 isolates [75].

This seminal study used several filters to exclude all Abs except those related to the neutralizing mAbs VRC01, VRC02, VRC03 and VRC-PG04. First, only antibodies utilizing the IGHV1 gene family were amplified and sequenced, and only those using the *IGHV1-2* gene were further analyzed. A second filter was applied that only accepted sequences that formed a cluster of sequences defined by divergence from the germline *IGHV1-2* gene in one dimension and divergence from the template antibodies VRC01 and VRC03 in the second dimension. This is a prime example of the approach used to restrict analysis only to clusters of antibodies that are related to a "target" antibody, in contrast to analysis techniques that attempt to study the entire Ab repertoire. Finally, the authors used cross-donor phylogenetic analysis, a general

Table 12.1 Survey of studies of the Ab response during HIV-1infection using HTS, organized primarily by epitope

HIV Epitope	Samples	Age	Time course	Total individuals	Total repertoires	Nucleic acid	Sequencing method	PBMC/B cell subsets	Chain/ isotypes	Lineage analysis	Link for HTS data deposition	References
CD4bs	Donors 45 & 74 IAVI Protocol G, Donor 0219 CHAVI Cohort 001	Adult	Single time point	3	3	mRNA → cDNA	454-Roche	PBMC	Heavy and light chains	Cross donor phylogenetic analysis targeted to VRC01	SRA SRP006992	[76]
CD4bs	Donor CH505 CHAVI Cohort 001	Adult	Multiple time points	1	5	mRNA → cDNA	454-Roche	PBMC	Heavy and light chains	Targeted to CH103 clonal lineage (CH103-CH106)	NA	[36]
CD4bs	Primary results from C38	Adult	Single time point	1	1	mRNA → cDNA	454-Roche	PBMC; IgG and IgM	Heavy and light chains	Cross donor phylogenetic analysis targeted to VRC01-like antibodies	SRA SRP026397	[85]
CD4bs	NIAD 43; IAV1 23, 57, 74; RU 3	Adult	Single time point	5	5	mRNA → cDNA	454-Roche	PBMC	Light chains only	Targeted to VRC01-like antibodies; gene specific amplification	SRA SRA072279	[82]
CD4bs	Donor 45 IAVI Protocol G	Adult	Multiple time points	1	10	mRNA → cDNA	454-Roche	PBMC; IgG+ B Cells	Heavy and light chains	Cross donor phylogenetic analysis targeted to VRC01	SRA SRP052625	[77]
CD4bs	RU01, RU08, 44 and C38 from NIAID protocols (new repertoires)	Adult	Single time point	4	4	mRNA → cDNA	454-Roche	PBMC	Heavy chains	Cross-donor phylogenetic analysis targeted to anti-CD4bs Abs	SRA SRP055520	[83]

CD4bs	Donor CH505 CHAVI001 acute HIV-1 infection cohort, source of CH103 class Abs	Adult	Multiple time points	1	15	gDNA	454-Roche	PBMC	Heavy chains	Cloanalyst analysis targeted to CH235 lineage	SRA SRP067168	[6]
	HIV-1 Hepatitis C negative individuals LP32647, LP08248 and LP23810											
CD4bs	Z258 from NIAID protocol	Adult	Single time point	1	1	mRNA → cDNA	Illumina MiSeq	PBMC memory B cells	Heavy and light chains	Targeted analysis to N6 lineage	SRA: SRR4417615–SRR4417632	[30]
V1 V2	1- Donor CAP256	Adult	Multiple time points	1	6	mRNA → cDNA	454-Roche/ MiSeq (paired VH/VL)	PBMC	Heavy and light chains	Intra-donor phylogenetic analysis targeted to CAP256-VRC26	SRA SRP034555 & SRP017087	[13]
V1 V2	1- Donor CAP256	Adult	Multiple time points	1	6	mRNA → cDNA	454-Roche	PBMC	Heavy chains	Targeted to CAP256-VRC26	SRA SRR2126754 SRR2126755	[4]
V1 V2	1- Donor CAP256	Adult	Multiple time points	1	6	mRNA → cDNA	454-Roche	PBMC	Heavy and light chains	Targeted to CAP256-VRC26	SRA SRR2126754 SRR2126755	[14]
MPER	2 - Donors N152 and IAVI 84	Adult	Single time point	2	2	mRNA → cDNA	454-Roche	PBMC	Heavy and light chains	Targeted bioinformatic sieving	SRA SRP018335	[85]
MPER	1-Donor N152	Adult	Single time point	1	1	mRNA → cDNA	454-Roche	PBMC	Heavy and light chains	Targeted bioinformatic sieving	SRA SRP018335	[60]

(continued)

Table 12.1 (continued)

HIV Epitope	Samples	Age	Time course	Total individuals	Total repertories	Nucleic acid	Sequencing method	PBMC/B cell subsets	Chain/isotypes	Lineage analysis	Link for HTS data deposition	References
N332 region	1 - Donor 17 IAVI Protocol G	Adult	Single time point	1	1	mRNA → cDNA	454-Roche	PBMC	Heavy and light; IgG+ memory B-cells	Gene-specific amplification targeted to PGT 121–123	NA	[59]
N332 region	Donor 17 IAVI Protocol G	Adult	Single time point	3	8	mRNA → cDNA	Ion Torrent PGM sequencing	PBMC	Heavy and light chains	Targeted to 6 PGT 121-class mAbs & full repertoire	NA	[26]
N332 region	Donor 39 IAVI Protocol G	Adult	Single time point	1	1	mRNA → cDNA	454-Roche	PBMC	Heavy and light chains	Targeted to PGT 135–137	SRA SRA055820	[84]
N332 region	Donor PC76 IAVI Protocol C	Adult	Multiple time points	1	1	mRNA → cDNA	Illumina MiSeq	PBMC or IgG+ memory B cells	Heavy and light chains	Targeted to panel of PCDN mAbs	SRA PRJNA: 304,070	[38]
Non-targeted analysis of HIV infection	HIV-1 infected patient	Adult/neonates	Multiple time points	8 (HIV+)	Pooled Libraries	mRNA → cDNA	454-Roche	IgM and IgG from PBMC and bone marrow	Heavy chains	Complete repertoire	NA	[78]
Non-targeted analysis of HIV infection	8 HIV-1 infected patients; 3 untreated, 5 treated	Adults	Up to 8 time points	8	48	mRNA → cDNA	Illumina MiSeq	PBMC	Heavy chains	Complete repertoire	ENA: ERP000572	[27]
Non-targeted analysis of HIV infection and SLE	4 controls; 4 SLE; 4 HIV+ treated; 4 HIV+ untreated	Adults	Single time point	16	16	mRNA → cDNA	454-Roche	PBMC	Heavy chains	Complete repertoire	NA	[81]

technique used in several similar studies, to further characterize the Ab repertoire. In this instance, similar antibodies from 2 other donors were added to the filtered antibodies from the primary donor, Donor 45, and a tree was constructed from these Ab sequences. A node on the tree was identified that included all VRC01 sequences from these 3 donors, and then all Ab sequences nested within that subtree emanating from that node were used in the final analysis. From studying this final set of Abs, the authors concluded that VRC01-like antibodies from all three donors exhibited similar developmental pathways, in terms of rates of diversification and other parameters. Their functional relatedness was confirmed by the observation that complementation of these identified heavy chains with standard VRC01 light chains from different donors produced neutralization.

Liao et al. [36] also examined anti-CD4bs Abs, and produced a major advance by sequencing both the Ab response to HIV-1, as well as the HIV-1 clonal lineages that were evolving in response to the Ab repertoire. They concentrated their study on one donor, CH505, the source of the VRC01-class bnAb CH103 ("VRC01-class" refers to Abs with similar properties to the canonical VRC01). By sequencing both Ab and HIV-1 lineages, Liao and colleagues were able to infer the co-evolution of the virus with the Abs, as the virus mutated to escape control by the Abs, and the Abs mutated to bind to the changing viral strains. The reconstruction of the CH103 lineage made it possible to infer the sequence of its unmutated common ancestor, and then it was shown that this ancestral sequence was able to bind transmitted/founder HIV-1 envelope glycoprotein. Furthermore, the Ab response of lineage CH103 broadened through time, following viral diversification of the CH103 binding site. This reconstruction of the evolution of the Ab lineage leading to the production of a bnAb is critical to the B-cell-lineage immunogen design strategy of vaccine development, as described in Sect. 12.2.1.

Zhou et al. [82] applied the HTS approach to multiple donors, again only targeting lineages based on VRC01-class bnAbs. They found that VRC01-class antibodies were obtained in multiple donors, characterized by similar maturation pathways and structural solutions to binding. This supported the hypothesis that the development of bnAbs such as the VRC01 class could be elicited in a large proportion of naïve individuals through the use of the proper immunogens (*i.e.*, the B-cell-lineage immunogen design strategy).

Based on several studies, VRC01-class bnAbs have been characterized by the use of the *IGHV1–2* gene for the heavy chain and a light chain with a 5-amino acid third complementary-determining region (CDR-L3). Several other recent studies have used HTS to help identify and describe the development of this class of antibodies in other donors. For example, Zhu et al. [85] discovered several Abs in this class from donor C38. Although many of the Abs from this study did not exhibit high sequence homology to other VRC01 class Abs, when reconstituted with VRC01 light chains, most of these new VRC01-class Abs showed substantial neutralization of HIV-1 isolates. Huang et al. [30] also found VRC01 class antibodies in a previously unstudied individual, Z258, from the NIAID protocol. Strikingly, the N6 Ab from this donor, in particular, neutralized 98% of HIV-1 isolates, including several strains that were not neutralized by previously discovered VRC01-class Abs

As a final example, Gao et al. [19] used HTS to discover that the donor who produced the bnAb CH103 [36] also produced a second bnAb, CH235. HTS of HIV-1 isolates and of the Ab repertoire from multiple time points during this individual's infection revealed the coevolution of the virus and the Ab response in fine detail. The authors hypothesized that CH235 developed first, and that an escape mutation on HIV-1 in response to CH235 facilitated the development of the broadly neutralizing Ab CH103; thus, it was hypothesized that cooperation of the two Ab lineages led to broad neutralization in this donor.

This large set of studies examining broadly neutralizing anti-CD4bs Abs has led to several recent conclusions that may be important for guiding HIV-1 vaccine design. Zhou et al. [83] summarized these studies by concluding that bnAbs against CD4bs could be characterized by

either: (1) the use of particular IGHV genes, usually *IGHV1-2* or *IGHV1-46*, or (2) dominated by particular CDR-H3 sequences. Furthermore, Wu et al. [77] concluded that the development of multiple Ab lineages initially with high mutation rates was associated with the production of bnAbs against HIV-1, but that this mutation rate tended to slow during the development of the expressed anti-HIV-1 repertoire.

12.2.2.2 Variable regions V1 and V2 (V1V2)

HTS studies of the Ab repertoire have also been critical to our understanding of the development of bnAbs that bind to variable regions 1 and 2 (V1V2) of the HIV-1 envelope. One of the most comprehensive analyses of this sort is based on a set of 12 bnAbs that were isolated from donor CAP256 (enrolled with the Centre for the AIDS Programme of Research in South Africa (CAPRISA)), who was infected with a clade C virus; these 12 Abs are referred to as (CAP256-VRC26.01–12), where VRC26 refers to the Ab lineage [13]. This lineage is extraordinary in that it exhibits a 35–37 amino acid long CDR-H3. HTS sequencing of the Ab repertoire from this donor was conducted at eight time points between 15 and 206 weeks post-infection. Targeted lineage analysis of HTS from this donor, using these 12 bnAbs as targets, defined a broad Ab lineage whose unmutated ancestor was able to neutralize the virus that superinfected this individual 15 weeks after initial infection [13]. This result is similar to the inferred history of the lineage leading to the CH103 anti-CD4bs bnAbs, in that the inferred unmutated common ancestor bound the HIV-1 target [36]. Inference of the unmutated common ancestor heavy chain sequence supported the hypothesis that this long CDR-H3 region was present in the original B-cell whose descendants ultimately produced these VRC26 bnAbs.

Doria-Rose et al. [14] conducted further studies of the Ab response from this same individual, donor 256. They used microneutralization and single-cell sorting to isolate 21 more bnAbs from this donor, all related to the original VRC26 lineage, referred to as VRC26.13–33. All available HTS data from the Ab repertoire from this individual were used to place the complete set of 33 VRC26 Abs within one phylogenetic lineage (in this case the targeted lineage analysis was based on CDR-H3 sequences, not full Ab sequences). The lineage defined by this targeted approach bifurcated early into two distinct lineages, one of which died out and one of which developed bnAbs. HTS of the viral component in this individual [4] showed that viral escape created multiple immunotypes, some of which were able to tolerate variability at key epitope contacts and thus contribute to this neutralization breadth. This mechanism of developing breadth within one Ab lineage suggests that viral diversification may be commonly associated with broadening of neutralization, and contrasts with the multi-lineage cooperative pathway hypothesized for the development of breadth in some anti-CD4Abs (*e.g.*, [19]).

12.2.2.3 Membrane-Proximal External Region (MPER)

HTS studies of the Ab response against the MPER have concentrated on donor N152, who was the source of one of the most broadly neutralizing anti-HIV-1bnAbs so far discovered, 10^8 [29]. The heavy chain of 10^8 shows a high rate of divergence (21%) from its inferred germline gene. Zhu et al. [85] sequenced heavy and light chains of related Abs separately, and then used a targeted analysis focused on 10^8, to obtain separate pools of heavy and light chain sequences only related to 10^8. Phylogenetic trees were then inferred separately from heavy and light chain sequences. The topologies of these phylogenetic trees were similar, allowing for heavy and light chain sequences to be paired based on these topologies. Abs identified by this topology-based pairing approach had lower auto-reactivity than was exhibited by pairs of sequences drawn without respect to the tree topology. Further sequencing from one time point of the N152 donor produced more Abs related to the 10^8 lineage and allowed for the inference of the unmutated common ancestor, which bound only weakly to the original MPER target [60]. Much of the motivation for the study of 10^8 stemmed from the fact that the inferred intermediate sequences in the

reconstructed lineage might identify potential immunogens for the induction of bnAb 10⁸, following the logic of the B-cell-lineage immunogen design strategy.

12.2.2.4   N332 Glycan Supersite in V3 Loop of GP120

Several strongly neutralizing Abs have also been described that target the high mannose patch centered on glycan N332 in the V3 loop of GP120. This site is an important vaccine target, given that passive administration of Abs binding to this region have been shown to prevent infection [42] and significantly decrease the strength of an ongoing infection in non-human primates [2]. HTS was used to analyze the Ab response in donor 17 of IAVI Protocol G [59], from whom the PGT121-class bnAbs that target this region had been isolated. Within this class of Abs, they found a positive correlation between the level of SHM and the development of neutralization breadth and potency. Putative intermediates within this lineage were characterized that showed only approximately half the mutation level of PGT121-134, but were still capable of neutralizing roughly 40–80% of PGT121-134 sensitive viral isolates. Such intermediates characterized by lower SHM are attractive vaccine targets, because they may be more easily elicited by the B-cell-lineage immunogen design approach than highly mutated Abs.

He and colleagues [26] further studied the Ab repertoire from donor 17 and compared this to repertoires from two individuals that were not infected with HIV-1. The Ab repertoires were surveyed in two ways: amplified with gene specific heavy and light chain primers, matching the germline components of PG121-class Abs, and by 5′ RACE. Since 5′ RACE is not based on family- or gene-specific primers, it should produce an estimate of the complete Ab repertoire, in an unbiased manner. Much of the analysis in this study employed intra-donor phylogenetic analysis, targeted on PGT121 Abs. Results from this analysis greatly expanded the known number of Abs in this family, based on Ion-torrent PGM sequencing technology compared to the standard 454 platform. The Ab repertoire results using 5′

RACE and the bioinformatic analysis not targeted to a specific Ab family are included in the next section, which summarizes HTS studies of complete Ab repertoires during HIV-1 infection.

HTS was also used to sequence the Ab repertoire from another donor, 39, of the IAVI protocol G [84]. This individual was the source of the bnAbs PGT135-137, which also target the N332 supersite. In this study, a lineage of Abs related to PGT135-137 was identified by targeted phylogenetic analysis. This apparent lineage exhibited 15 distinct clusters of heavy chain sequences and 10 clusters of light chain sequences; sequences chosen to represent these somatic populations showed diverse neutralization characteristics.

Finally, the most recent study to use HTS data to study Ab responses against this region identified 12 somatically related bnAbs from donor PC76, 16–38 months post-infection, referred to as the PCDN lineage [38]. These Abs were not as highly mutated as previous bnAbs raised against this site, and did not include insertion–deletion variants. Again, the closer a bnAb is to its germline configuration, the better it is as a vaccine target according to the B-cell-lineage immunogen design approach, given that these Abs would be easier to elicit using a series of immunogens compared to more highly mutated Abs, which may not always follow predictable maturation pathways. Macleod et al. [38] also showed that multiple pathways led to neutralization breadth, and the authors hypothesized that early diversification followed by maturation of parallel lineages is a requirement for obtaining neutralization breadth.

12.2.2.5   Non-targeted Analysis of Ab Repertoire During HIV-1 Infection

A few studies have used HTS to examine the complete Ab repertoire during HIV-1 infection, referred to here as non-targeted approaches. These studies contrast the targeted studies in two ways: amplification of all IGHV genes (either by using primers amplifying all V genes, or 5'RACE), and the use of bioinformatic tools that examine all Abs and lineages, rather than only sequences related to a target Ab.

For example, Xiao et al. [78] used HTS to explore the Ab response in control individuals and an HIV-positive individual. They reported on V(D)J recombination frequencies, but did not perform lineage analyses. The primary findings were that the dominant IGHV family used in HIV-1+ compared to control individuals varied between IgM and IgG libraries, and from early to late infection.

The results of He et al. [26] were mentioned previously, in terms of Ab repertoires that focused on the PG121-class of Abs. They also reported on non-targeted Ab repertoires from an HIV-1+ and two control individuals. By examining the complete Ab repertoire, they observed that the dominant IGHV gene used in the repertoire from the infected individual was not the same IGHV gene associated with the PG121-class Abs. Also, one of the unbiased repertoires sequenced from a non-HIV-1 infected individual showed a strongly skewed repertoire in terms of IGHV gene usage. It is clear from this study of only a few individuals that unbiased repertoires can be highly different between individuals whether they are infected or not, consistent with what has been observed in other studies of healthy individuals (*e.g.*, [18, 9]).

Yin et al. [81] analyzed complete Ab repertoires, comparing control individuals (n = 4), systemic lupus erythematosus (SLE) patients (n = 4), and HIV-positive infected individuals either undergoing treatment (n = 4) or not (n = 4). They only analyzed the CDR-H3 region of the Ab from the IgM compartment. They observed that gene usage, CDR-H3 length distribution, and SHM did not differ between these groups. Specifically, anti-retroviral therapy (ART) did not normalize the diversity of the IgM Ab repertoire.

Hoehn et al. [28] used HTS to probe the Ab repertoire in eight HIV-positive individuals, five of whom were being treated with ART and three who were not. These repertoires were compared to Ab repertoires of six healthy patients from a previous study. These researchers applied several measures of clonality to the entire Ab repertoire, and in general observed that the HIV-1+ individuals exhibited higher clonality. There was significant variability among HIV-positive patients treated with ART in terms of the clonality of their repertoires, such that no differences were observed between treated and untreated patients. This is similar to the conclusion of Yin et al. [81], except with larger sample sizes in two of the group subsets. Across all HIV-1+ patients, there was no association between repertoire clonality and clinical variables, such as viral load or CD4+ T cell count.

12.2.2.6 Summary of Results from HTS Studies of the Anti-HIV-1 Ab Repertoire

Many of the HTS studies of Ab repertoires during infection have focused on HIV-1 infection. As mentioned earlier, these have most often targeted particular bnAbs and their relatives within the Ab repertoire. The perfect vaccine candidate Ab would be able to neutralize a broad set of HIV-1 isolates, use common genes, not require an unusually long CDR-H3, and would be common in the naïve repertoire of most individuals. This bias toward HIV-1 studies of particular bnAbs reflects the desire to learn how to elicit such bnAbs in naïve individuals.

In our review of these studies, we have gathered a diverse set of conclusions regarding the conditions eliciting these bnAbs that could be vaccine candidates. These conditions include high rates of SHM [59], strong viral diversification [4, 36], unmutated common ancestors binding to the infecting HIV-1 strain [13, 36], multiple parallel Ab lineages [38], presence of competing Ab lineages [19], and early rates of viral diversification [36]. Despite these numerous correlations, there doesn't seem to be any conditions that are consistently associated with the development of anti-HIV-1 bnAbs across multiple studies. Several factors could plausibly explain such inconsistencies.

For one, these studies are characterized by large datasets, often in the 10s of millions of sequences, but sequenced from only a few individuals. In fact, several donors (*e.g.*, N152, CAP256, Donors 17 and 45 of IAVI Protocol G, and CH103; Table 12.1) have repeatedly been sampled across studies, because they are known to produce some of the most broadly neutralizing

Abs. This focus on a small number of donors makes sense, given the importance of the bnAbs that they produce, but this limits the generality of the results that we can draw from the set of studies undertaken so far. Also, as noted several times, the above patterns were observed mostly in studies where bioinformatic filters had been used to restrict the analysis to only Abs that are related to targeted Abs, usually the bnAbs that these donors are known to produce. This again restricts the scope of any conclusions that can be drawn from these studies. Finally, the bioinformatic approaches used to capture related Abs, such as intra- and inter-donor phylogenetic analysis, only capture sets of Abs that are functionally related (*e.g.*, utilization of the same IGHV and J gene, with similar CDR-H3), but there is no way to estimate how likely it is that these Abs may be related as ancestor/descendants which is the usual definition of a clonal lineage. In such circumstances, for example, conclusions on the necessity of binding patterns of these unmutated common ancestors or early bifurcation of B-cell lineages may not make sense if the B-cell lineages referred to are in fact multiple biological lineages. In the next section we discuss the utility of non-targeted approaches, and what is needed to make these larger scale comparisons possible (*i.e.*, sharing of large data sets across studies and institutions).

12.3 Future Directions to Enhance the Value of Ab Repertoire Studies

12.3.1 Germline Genes and Haplotypes

The Ab repertoire community is now placing a greater focus on understanding the origins of key inter-individual similarities and differences observed among repertoires responding to a given pathogen. The identification of common repertoire signatures in larger subsets of individuals is likely to improve strategies for the design of effective therapeutics and prophylactics. Given the importance of directing the development of the repertoires of naive individuals under the

B-cell-lineage immunogen design strategy, some have questioned whether all individuals have the capacity to actually respond to a given immunogen in the same way [5]. It is certainly evident in many contexts that not all vaccines elicit equivalent responses in all individuals of the population [8]. While in most scenarios, many factors are at play (*e.g.*, an individual's infection/vaccine history and age), there is a growing appreciation for the potential influence of heritable factors on the development of both the baseline naïve and antigen-stimulated Ab repertoires [18, 21, 50, 66]. This coincides with a growing body of work describing extreme levels of genetic diversity at the IG loci among human populations, including single nucleotide variants in both the coding and non-coding portions of IG V, D, J, and C genes, as well as large structural variants that can influence the number of functional IG gene segments present in a given genome (e.g., [11, 32, 35, 41, 53, 65, 69]). Importantly, there are now many studies that have directly linked IG germline polymorphisms to differences in Ab repertoire signatures and gene usage, Ab function, and disease and clinical outcomes (e.g., [1, 15, 33, 45, 46, 52, 56, 62, 73]), together suggesting that any two individuals may not necessarily be predisposed to mount identical responses. With these examples in mind, moving forward we believe the inclusion of data on germline V, D, J and C gene variants in the study of variation in the Ab repertoire in health and disease (ideally in larger cohorts) will be critically important (see [68, 70]).

A potential role for germline variation in the development of CD4-directed neutralizing Abs has been specifically observed among repertoires in HIV+ individuals. Yacoob et al. [80] recently showed that particular alleles (namely, *02, *03, and *04) at the human *IGHV1-2* gene that encode three key amino acids serve as better germline precursors for VRC01 bnAbs, with better binding affinities, noting that this is consistent with the observation that all VRC01-class bnAbs identified to date are encoded by the germline *IGHV1-2* allele *02 [72]. A critical observation was that only 8/9 subjects examined in their study carried *02 alleles, suggesting that some individuals in

the population may have a reduced capacity to produce effective VRC01-class bnAbs [80]; this may be specifically relevant to germline targeting-based vaccine priming approaches [31]. Allelic restrictions have also been noted for *IGHV1-69*-encoded bnAbs against gp41 elicited by vaccination [74], asl well as for critical binding residues of influenza hemagglutinin stem-directed bnAbs [1, 45, 73].

12.3.2 Integration of Heavy and Light Chain Data, and B-Cell and T-Cell Repertoires

Many of the studies surveyed report data only on immunoglobulin heavy chain sequences, and sometimes only for the CDR-H3 region. However, both the heavy and light chain can contribute to Ab binding properties, and including both the heavy and light chain is necessary to most accurately determine the clonal lineages within an Ab or T-cell receptor (TCR) repertoire. In fact, several labs have developed the ability to report both the heavy and the light chain sequences for B cells and T cells. The earliest approaches involved pairing heavy and light chain sequences based on the topologies and frequencies of corresponding branches of phylogenies constructed separately from heavy and light chain sequences (*e.g.*, [85]). More recent approaches are based on emulsion techniques (*e.g.*, [12]) or other sorting techniques that isolate single cells and capture paired heavy and light chain sequences from each B cell [44, 49]. Parallel analyses of Ab/B-cell and T-cell repertoires are starting to reveal common and divergent patterns between the two compartments [3, 47], and isolation of mRNA from single cells allows relating complete-length paired heavy and light chain Ab sequences to T-cell receptor repertoires in fine detail [7].

12.3.3 Integration of HTS B-Cell Repertoire Data with Other Types of Genomic-Level Data

One of the most important future developments for HTS studies of Ab repertoires will be to integrate these data with other large scale, genomic datasets. For example, Georgiou and colleagues [34] have combined HTS sequencing of the Ab repertoire with high-resolution protein mass spectrometry of the serum antibodies (Ig-seq) in order to connect soluble Ig in blood and secretions with clonally expanded peripheral B cells [34]. This can lead to identifying Ag-specific lineages both functionally and developmentally. In addition, interesting insights have come from combining HTS of gut microbiota with Ab data [16, 37].

12.3.4 Importance of Non-targeted, Complete Ab Repertoire Studies

This review shows that most of the studies of Ab repertoires in HIV-1 infected individuals have taken a "targeted" approach, in that they only examine sequences that are related to particular target bNAbs. We contrast this to non-targeted, complete repertoire studies that attempt to estimate the number and relative expansion of all of the Ab lineages in an individual's repertoire. Of course, it is practically impossible to completely characterize an individual's Ab repertoire, given that it has been estimated that an individual can have as many as 10^{11}–10^{12} lymphocytes [55], many of which express unique Ab/B-cell or T-cell receptors.

The targeted approach can answer some specific questions very well, such as describing the development of a specific lineage of Abs that led to the development of a specific bnAb. However, many questions about the development of Ab

repertoires, including understanding the conditions and broader repertoire context leading to bnAbs, require a description of the complete Ab "milieu" within which a particular Ab lineage was formed (*i.e.*, the number and estimated size of each lineage). This question comes down to: what additional insights do we obtain from unbiased, non-targeted sequencing coupled with attempts to describe the complete Ab repertoire?

One example of a question that demands a complete repertoire analysis, rather than one targeted to specific pre-defined lineages, would be how often are certain lineages produced in response to HIV-1, such as those related to the important bnAb, VRC01? In one such study, Zhu et al. [85] reported 15 clusters of sequences all related to PGT135-137, and these clusters exhibited highly variable levels of neutralization. From this observation, the authors concluded that the VRC01 lineages could be produced by a large proportion of individuals. However, without comparing the frequency of this type of lineage within a broader representation of related Ab lineages, it is difficult to estimate the naturally occurring frequency of this lineage among individuals.

Also, several studies have investigated the relationship between the expansion of clones early in HIV-1 infection and the production of bnAbs. Again, one would have to have a complete description of the clonal structure of each individual to determine which individuals are characterized by clonal expansion, even of related clones, and when those clones expanded. Any comparison of V, D, or J gene usage, or the unevenness of repertoires in individuals who do or do not produce bnAbs, would also only be possible based on an unbiased view of the repertoire. Given that these signals might discriminate individuals producing or not producing bnAbs, it is obvious that we must strive to describe the entire Ab repertoire as often as possible.

Furthermore, it is worth noting that the necessity of understanding the complete Ab repertoire is not just restricted to studies of the development of anti-HIV-1 bnAbs. It is likely that the Ab repertoire will be completely different between individuals depending on many different conditions, from infection to autoimmune disease to cancer, and understanding the effects of these conditions will demand describing the complete Ab responses in many individuals. "Public" Abs, those produced by many, including presumably healthy, individuals, and using commonly expressed V, D, and J genes, would be obvious targets for most vaccine attempts. Targeted approaches to Ab repertoires demand large sequence datasets, but attempts to describe the complete Ab repertoire will demand even larger datasets, and answering the above questions will likely involve comparing and sharing these huge data sets across studies, disease states and institutions. The next section addresses some of the challenges in reporting these data to improve reproducibility and in sharing these data in order to understand what drives similarities and differences in Ab repertoires.

12.3.5 Standardizing Data Generation, Analysis and Sharing Protocols

HTS of Ab/B-cell and T-cell receptor repertoires has increased dramatically since the technique was first applied to immune receptor repertoires in 2009 [9, 17, 48, 71]. This experimental approach allows us to explore the development of the adaptive immune system in exquisite detail, and holds significant translational promise as diagnostic and prognostic tools. However, as discussed in this review, there are several limitations to the present generation and analysis of these type of data. Efforts to standardize protocols and

facilitate the sharing of data could help to answer many of these limitations.

A primary issue that arises when trying to draw conclusions based on analyses across large, complex and shared data sets is reproducibility. This largely concerns the development of standards centered around the reporting of key details related to the production and analysis of data that ensure another researcher can reproduce the data and recapitulate the analysis, and ultimately be confident that the data are of sufficient quality to be shared and compared.

Facilitating comparing and sharing of these data sets within and across labs, disease states and institutions also demands new protocols for data deposition and sharing in public repositories, as well as bioinformatic solutions to analyze these complex data. In some cases, these protocols will also need to take into account the protection of intellectual property (IP), data security, and donor confidentiality. The Adaptive Immune Receptor Repertoire (AIRR) Community is a group of immunologists, immunogeneticists, clinicians, bioinformaticians, and experts in legal, IP and security aspects of genomic data sharing, formed in 2015 to address these issues for this rapidly expanding area of immunological research. The progress and recommendations of the working groups of the AIRR Community are summarized at airr-community.org; anyone interested in joining the Community can also do so at that site.

References

1. Avnir Y, Watson CT, Glanville J, Peterson EC, Tallarico AS, Bennett AS, Qin K, Fu Y, Huang C-Y, Beigel JH, Breden F, Quan Z, Marasco WA (2016) IGHV1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with IGHV utilization shifts and varies by ethnicity. Sci Rep 6:20842. https://doi.org/10.1038/srep20842
2. Barouch DH, Whitney JB, Moldt B, Klein F, Oliveira TY, Liu J, Stephenson KE, Chang H-W, Shekhar K, Gupta S, Nkolola JP, Seaman MS, Smith KM, Borducchi EN, Cabral C, Smith JY, Blackmore S, Sanisetty S, Perry JR, Beck M, Lewis MG, Rinaldi W, Chakraborty AK, Poignard P, Nussenzweig MC, Burton DR (2013) Therapeutic efficacy of potent neutralizing HIV-1-specific monoclonal antibodies in SHIV-infected rhesus monkeys. Nature 503:224–228. https://doi.org/10.1038/nature12744
3. Benichou JIC, van Heijst JWJ, Glanville J, Louzoun Y (2017) Converging evolution leads to near maximal junction diversity through parallel mechanisms in B and T cell receptors. Phys Biol 14:45003. https://doi.org/10.1088/1478-3975/aa7366
4. Bhiman JN, Anthony C, Doria-Rose NA, Karimanzira O, Schramm CA, Khoza T, Kitchin D, Botha G, Gorman J, Garrett NJ, Abdool Karim SS, Shapiro L, Williamson C, Kwong PD, Mascola JR, Morris L, Moore PL (2015) Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. Nat Med 21:1332–1336. https://doi.org/10.1038/nm.3963
5. Bonsignori M, Alam SM, Liao HX, Verkoczy L, Tomaras GD, Haynes BF, Moody MA (2012) HIV-1 antibodies from infection and vaccination: insights for guiding vaccine design. Trends Microbiol. 20:532–539. https://doi.org/10.1016/j.tim.2012.08.011
6. Bonsignori M, Zhou T, Sheng Z, Chen L, Gao F, Joyce MG, Ozorowski G, Chuang GY, Schramm CA, Wiehe K, Alam SM, Bradley T, Gladden MA, Hwang KK, Iyengar S, Kumar A, Lu X, Luo K, Mangiapani MC, Parks RJ, Song H, Acharya P, Bailer RT, Cao A, Druz A, Georgiev IS, Kwon YD, Louder MK, Zhang B, Zheng A, Hill BJ, Kong R, Soto C; NISC Comparative Sequencing Program, Mullikin JC, Douek DC, Montefiori DC, Moody MA, Shaw GM, Hahn BH, Kelsoe G, Hraber PT, Korber BT, Boyd SD, Fire AZ, Kepler TB, Shapiro L, Ward AB, Mascola JR, Liao HX, Kwong PD, Haynes BF. (2016) Maturation Pathway from Germline to Broad HIV-1 Neutralizer of a CD4-Mimic Antibody. Cell 165(2):449–463. doi:10.1016/j.cell.2016.02.022
7. Bouquet J, Gardy JL, Brown S, Pfeil J, Miller RR, Morshed M, Avina-Zubieta A, Shojania K, McCabe M, Parker S, Uyaguari M, Federman S, Tang P, Steiner T, Otterstater M, Holt R, Moore R, Chiu CY, Patrick DM (2017) RNA-Seq analysis of gene expression, viral pathogen, and B-Cell/T-Cell receptor signatures in complex chronic disease. Clin Infect Dis 64:476–481. https://doi.org/10.1093/cid/ciw767
8. Boyd SD, Jackson KJL (2015) Predicting vaccine responsiveness. Cell Host Microbe 17:301–307. https://doi.org/10.1016/j.chom.2015.02.015
9. Boyd SD, Collins AM, Nadeau KC, Egholm M, Miklos DB, Birgitte J, Simen B, Hanczaruk B, Nguyen KD, Zhang LN, Sahaf B, Jones CD, Fire AZ, Marshall EL, Merker JD, Scott J, Boyd D, Gaëta BA, Jackson KJ, Maniar JM, Simen BB, Zehnder JL (2010) Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. J Immunol 184:6986–6992. https://doi.org/10.4049/jimmunol.1000445
10. Breden F, Lepik C, Longo NS, Montero M, Lipsky PE, Scott JK (2011) Comparison of antibody repertoires produced by HIV-1 infection, other chronic and acute infections, and systemic autoimmune disease.

PLoS One 6:e16857. https://doi.org/10.1371/journal.pone.0016857

11. Chimge N-O, Pramanik S, Hu G, Lin Y, Gao R, Shen L, Li H (2005) Determination of gene organization in the human IGHV region on single chromosomes. Genes Immun 6:186–193. https://doi.org/10.1038/sj.gene.6364176

12. DeKosky BJ, Kojima T, Rodin A, Charab W, Ippolito GC, Ellington AD, Georgiou G (2015) In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. Nat Med 21:86–89. https://doi.org/10.1038/nm.3743

13. Doria-Rose NA, Schramm CA, Gorman J, Moore PL, Bhiman JN, DeKosky BJ, Ernandes MJ, Georgiev IS, Kim HJ, Pancera M, Staupe RP, Altae-Tran HR, Bailer RT, Crooks ET, Cupo A, Druz A, Garrett NJ, Hoi KH, Kong R, Louder MK, Longo NS, McKee K, Nonyane M, O'Dell S, Roark RS, Rudicell RS, Schmidt SD, Sheward DJ, Soto C, Wibmer CK, Yang Y, Zhang Z, Mullikin JC, Binley JM, Sanders RW, Wilson IA, Moore JP, Ward AB, Georgiou G, Williamson C, Abdool Karim SS, Morris L, Kwong PD, Shapiro L, Mascola JR (2014) Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature 509:55–62. https://doi.org/10.1038/nature13036

14. Doria-Rose NA, Bhiman JN, Roark RS, Schramm CA, Gorman J, Chuang G-Y, Pancera M, Cale EM, Ernandes MJ, Louder MK, Asokan M, Bailer RT, Druz A, Fraschilla IR, Garrett NJ, Jarosinski M, Lynch RM, McKee K, O'Dell S, Pegu A, Schmidt SD, Staupe RP, Sutton MS, Wang K, Wibmer CK, Haynes BF, Abdool-Karim S, Shapiro L, Kwong PD, Moore PL, Morris L, Mascola JR (2015) New member of the V1V2-directed CAP256-VRC26 lineage that shows increased breadth and exceptional potency. J Virol 90:76–91. https://doi.org/10.1128/JVI.01791-15

15. Feeney AJ, Atkinson MJ, Cowan MJ, Escuro G, Lugo G (1996) A defective Vkappa A2 allele in Navajos which may play a role in increased susceptibility to haemophilus influenzae type b disease. J Clin Invest 97:2277–2282. https://doi.org/10.1172/JCI118669

16. Fransen F, Zagato E, Mazzini E, Fosso B, Manzari C, El Aidy S, Chiavelli A, D'Erchia AM, Sethi MK, Pabst O, Marzano M, Moretti S, Romani L, Penna G, Pesole G, Rescigno M (2015) BALB/c and C57BL/6 mice differ in polyreactive IgA abundance, which impacts the generation of antigen-specific IgA and microbiota diversity. Immunity 43:527–540. https://doi.org/10.1016/j.immuni.2015.08.01

17. Freeman JD, Warren RL, Webb JR, Warren L, Nelson BH, Holt RA (2009) Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. Genome Res. 10:1817–1824. https://doi.org/10.1101/gr.092924.109

18. Galson JD, Trück J, Fowler A, Münz M, Cerundolo V, Pollard AJ, Lunter G, Kelly DF (2015) In-depth assessment of within-individual and inter-individual variation in the B cell receptor repertoire. Front Immunol 6:531. https://doi.org/10.3389/fimmu.2015.00531

19. Gao F, Bonsignori M, Liao HX, Kumar A, Xia SM, Lu X, Cai F, Hwang KK, Song H, Zhou T, Lynch RM, Alam SM, Moody MA, Ferrari G, Berrong M, Kelsoe G, Shaw GM, Hahn BH, Montefiori DC, Kamanga G, Cohen MS, Hraber P, Kwong PD, Korber BT, Mascola JR, Kepler TB, Haynes BF (2014) Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. Cell 158:481–491. https://doi.org/10.1016/j.cell.2014.06.022

20. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR (2014) The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat Biotechnol 32:158–168. https://doi.org/10.1038/nbt.2782

21. Glanville J, Kuo TC, von Büdingen HC, Guey L, Berka J, Sundar PD, Huerta G, Mehta GR, Oksenberg JR, Hauser SL, Cox DR, Rajpal A, Pons J (2011) Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. Proc Natl Acad Sci 108:20066–20071. https://doi.org/10.1073/pnas.1107498108

22. Gray ES, Madiga MC, Hermanus T, Moore PL, Wibmer CK, Tumba NL, Werner L, Mlisana K, Sibeko S, Williamson C, Abdool Karim SS, Morris L (2011) The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. J Virol 85:4828–4840. https://doi.org/10.1128/JVI.00198-11

23. Greiff V, Menzel U, Miho E, Weber C, Riedel R, Cook S, Valai A, Lopes T, Radbruch A, Winkler TH, Reddy ST (2017) Systems analysis reveals high genetic and antigen-driven predetermination of antibody repertoires throughout B cell development. Cell Rep 19:1467–1478. https://doi.org/10.1016/j.celrep.2017.04.054

24. Haynes BF (2005) Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. Science 308:1906–1908. https://doi.org/10.1126/science.1111781

25. Haynes BF, Kelsoe G, Harrison SC, Kepler TB (2012) B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. Nat Biotechnol 30:423–433. https://doi.org/10.1038/nbt.2197

26. He L, Sok D, Azadnia P, Hsueh J, Landais E, Simek M, Koff WC, Poignard P, Burton DR, Zhu J (2014) Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. Sci Rep 4:6778. https://doi.org/10.1038/srep06778

27. Hoehn KB, Gall A, Bashford-Rogers R, Fidler SJ, Kaye S, Weber JN, McClure MO; SPARTAC Trial Investigators, Kellam P, Pybus OG (2015) Dynamics of immunoglobulin sequence diversity in HIV-1 infected individuals. Philos Trans R Soc Lond B Biol Sci 370(1676). pii: 20140241. https://doi: 10.1098/rstb.2014.0241

28. Hoehn KB, Fowler A, Lunter G, Pybus OG (2016) The diversity and molecular evolution of B-cell recep-

tors during infection. Mol Biol Evol 33:1147–1157. https://doi.org/10.1093/molbev/msw015

29. Huang J, Ofek G, Laub L, Louder MK, Doria-Rose NA, Longo NS, Imamichi H, Bailer RT, Chakrabarti B, Sharma SK, Alam SM, Wang T, Yang Y, Zhang B, Migueles SA, Wyatt R, Haynes BF, Kwong PD, Mascola JR, Connors M (2012) Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. Nature 491:406–412. https://doi.org/10.1038/nature11544

30. Huang J, Kang BH, Ishida E, Zhou T, Griesman T, Sheng Z, Wu F, Doria-Rose NA, Zhang B, McKee K, O'Dell S, Chuang GY, Druz A, Georgiev IS, Schramm CA, Zheng A, Joyce MG, Asokan M, Ransier A, Darko S, Migueles SA, Bailer RT, Louder MK, Alam SM, Parks R, Kelsoe G, Von Holle T, Haynes BF, Douek DC, Hirsch V, Seaman MS, Shapiro L, Mascola JR, Kwong PD, Connors M (2016) Identification of a CD4-Binding-Site antibody to HIV that evolved near-pan neutralization breadth. Immunity 45(5):1108–1121. https://doi: 10.1016/j.immuni.2016.10.027

31. Jardine JG, Kulp DW, Havenar-daughton C, Briney B, Sok D, Sesterhenn F, Ereño-orbea J, Kalyuzhniy O, Deresa I, Hu X, Spencer S, Georgeson E, Adachi Y, Kubitz M, Allan C, Julien J, Wilson IA, Burton DR (2016) HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeted immunogen. Science 351:1458–1463. https://doi.org/10.1126/science.aad9195.HIV-1

32. Kidd MJ, Chen Z, Wang Y, Jackson KJ, Zhang L, Boyd SD, Fire AZ, Tanaka MM, Gaëta BA, Collins AM (2012) The inference of phased haplotypes for the immunoglobulin H chain V region gene loci by analysis of VDJ gene rearrangements. J Immunol 188:1333–1340. https://doi.org/10.4049/jimmunol.1102097

33. Kidd MJ, Jackson KJL, Boyd SD, Collins AM (2016) DJ pairing during VDJ recombination shows positional biases that vary among individuals with differing IGHD locus immunogenotypes. J Immunol 196:1158–1164. https://doi.org/10.4049/jimmunol.1501401

34. Lavinder JJ, Wine Y, Giesecke C, Ippolito GC, Horton AP, Lungu OI, Hoi KH, DeKosky BJ, Murrin EM, Wirth MM, Ellington AD, Dörner T, Marcotte EM, Boutz DR, Georgiou G (2014) Identification and characterization of the constituent human serum antibodies elicited by vaccination. Proc Natl Acad Sci U S A 111:2259–2264. https://doi.org/10.1073/pnas.1317793111

35. Lefranc MP, Hammarstrom L, Smith CI, Lefranc G (1991) Gene deletions in the human immunoglobulin heavy chain constant region locus: molecular and immunological analysis. Immunodefic Rev 2:265–281

36. Liao H-X, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, Fire AZ, Roskin KM, Schramm CA, Zhang Z, Zhu J, Shapiro L, Mullikin JC, Gnanakaran S, Hraber P, Wiehe K, Kelsoe G, Yang G, Xia S-M, Montefiori DC, Parks R, Lloyd KE, Scearce RM, Soderberg KA, Cohen M, Kamanga G, Louder MK, Tran LM, Chen Y, Cai F, Chen S, Moquin S, Du X, Joyce MG, Srivatsan S, Zhang B, Zheng A, Shaw GM, Hahn BH, Kepler TB, Korber BTM, Kwong PD, Mascola JR, Haynes BF (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature 496:469–476. https://doi.org/10.1038/nature12053

37. Lindner C, Thomsen I, Wahl B, Ugur M, Sethi MK, Friedrichsen M, Smoczek A, Ott S, Baumann U, Suerbaum S, Schreiber S, Bleich A, Gaboriau-Routhiau V, Cerf-Bensussan N, Hazanov H, Mehr R, Boysen P, Rosenstiel P, Pabst O (2015) Diversification of memory B cells drives the continuous adaptation of secretory antibodies to gut microbiota. Nat Immunol 16:880–888. https://doi.org/10.1038/ni.3213

38. MacLeod DT, Choi NM, Briney B, Garces F, Ver LS, Landais E, Murrell B, Wrin T, Kilembe W, Liang CH, Ramos A, Bian CB, Wickramasinghe L, Kong L, Eren K, Wu CY, Wong CH, Kosakovsky Pond SL, Wilson IA, Burton DR, Poignard P, The IAVI Protocol C Investigators (2016) Early antibody lineage diversification and independent limb maturation lead to broad HIV-1 neutralization targeting the Env high-mannose patch. Immunity 44:1215–1226. https://doi.org/10.1016/j.immuni.2016.04.016

39. Martin VG, Wu Y-CB, Townsend CL, Lu GHC, O'Hare JS, Mozeika A, Coolen ACC, Kipling D, Fraternali F, Dunn-Walters DK (2016) Transitional B cells in early human B cell development – time to revisit the paradigm? Front Immunol 7:546. https://doi.org/10.3389/fimmu.2016.00546

40. Mascola JR, Haynes BF (2013) HIV-1 neutralizing antibodies: understanding nature's pathways. Immunol Rev 254:225–244. https://doi.org/10.1111/imr.12075

41. Milner EC, Hufnagle WO, Glas AM, Suzuki I, Alexander C (1995) Polymorphism and utilization of human VH genes. Ann N Y Acad Sci 764:50–61

42. Moldt B, Le K, Carnathan DG, Whitney JB, Schultz N, Lewis MG, Borducchi E, Smith K, Mackel JJ, Sweat SL, Hodges AP, Godzik A, Parren PWHI, Silvestri G, Barouch DH, Burton DR (2016) Neutralizing antibody affords comparable protection against vaginal and rectal SHIV challenge in macaques. AIDS 30:1543–1551. https://doi.org/10.1097/QAD.0000000000001102

43. Murphy KP, Travers P, Walport M (eds) (2008) Janeway's immunobiology. Garland Science, London

44. Murugan R, Imkeller K, Busse CE, Wardemann H (2015) Direct high-throughput amplification and sequencing of immunoglobulin genes from single human B cells. Eur J Immunol 45:2698–2700. https://doi.org/10.1002/eji.201545526

45. Pappas L, Foglierini M, Piccoli L, Kallewaard NL, Turrini F, Silacci C, Fernandez-Rodriguez B, Agatic G, Giacchetto-Sasselli I, Pellicciotta G, Sallusto F, Zhu Q, Vicenzi E, Corti D, Lanzavecchia A (2014) Rapid development of broadly influenza neutralizing antibodies through redundant mutations. Nature 516:418–422. https://doi.org/10.1038/nature13764

46. Parks T, Mirabel MM, Kado J, Auckland K, Nowak J, Rautanen A, Mentzer AJ, Marijon E, Jouven X, Perman ML, Cua T, Kauwe JK, Allen JB, Taylor H, Robson KJ, Deane CM, Steer AC, Hill AVS (2017) Association between a common immunoglobulin heavy chain allele and rheumatic heart disease risk in Oceania. Nat Commun 8:14946. https://doi.org/10.1038/ncomms14946

47. Rechavi E, Lev A, Lee YN, Simon AJ, Yinon Y, Lipitz S, Amariglio N, Weisz B, Notarangelo LD, Somech R (2015) Timely and spatially regulated maturation of B and T cell repertoire during human fetal development. Sci Transl Med 7:276ra25. https://doi.org/10.1126/scitranslmed.aaa0072

48. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, Riddell SR, Warren EH, Carlson CS (2010) Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. Blood 114:4099–4107. https://doi.org/10.1182/blood-2009-04-217604

49. Roy B, Neumann RS, Snir O, Iversen R, Sandve GK, Lundin KEA, Sollid LM (2017) High-throughput single-cell analysis of B cell receptor usage among autoantigen-specific plasma cells in celiac disease. J Immunol 199:782–791. https://doi.org/10.4049/jimmunol.1700169

50. Rubelt F, Bolen CR, McGuire HM, Vander HJA, Gadala-Maria D, Levin M, Euskirchen GM, Mamedov MR, Swan GE, Dekker CL, Cowell LG, Kleinstein SH, Davis MM (2016) Individual heritable differences result in unique cell lymphocyte receptor repertoires of naïve and antigen-experienced cells. Nat Commun 7:11112. https://doi.org/10.1038/ncomms11112

51. Sakano H, Maki R, Kurosawa Y, Roeder W, Tonegawa S (1980) Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. Nature 286:676–683. https://doi.org/10.1038/286676a0

52. Sasso EH, Johnson T, Kipps TJ (1996) Expression of the immunoglobulin VH gene 51p1 is proportional to its germline gene copy number. J Clin Invest 97:2074–2080. https://doi.org/10.1111/j.1749-6632.1997.tb52108.x

53. Scheepers C, Shrestha RK, Lambson BE, Jackson KJL, Wright IA, Naicker D, Goosen M, Berrie L, Ismail A, Garrett N, Abdool Karim Q, Abdool Karim SS, Moore PL, Travers SA, Morris L (2015) Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline Ig gene repertoire. J Immunol 9:4371. https://doi.org/10.4049/jimmunol.1500118

54. Schroeder HW (2006) Similarity and divergence in the development and expression of the mouse and human antibody repertoires. Dev Comp Immunol 30:119–135

55. Sender R, Fuchs S, Milo R (2016) Revised estimates for the number of human and bacteria cells in the body. PLoS Biology 14:1–14. https://doi.org/10.1371/journal.pbio.1002533

56. Sharon E, Sibener LV, Battle A, Fraser HB, Garcia KC, Pritchard JK (2016) Genetic variation in MHC proteins is associated with T cell receptor expression biases. Nat Genet 48:995–1002. https://doi.org/10.1038/ng.3625

57. Shen X, Parks RJ, Montefiori DC, Kirchherr JL, Keele BF, Decker JM, Blattner WA, Gao F, Weinhold KJ, Hicks CB, Greenberg ML, Hahn BH, Shaw GM, Haynes BF, Tomaras GD (2009) In vivo gp41 antibodies targeting the 2F5 mAb epitope mediate HIV-1 neutralization breadth. J Virol 83:3617–3625. https://doi.org/10.1128/JVI.02631-08

58. Simek MD, Rida W, Priddy FH, Pung P, Carrow E, Laufer DS, Lehrman JK, Boaz M, Tarragona-Fiol T, Miiro G, Birungi J, Pozniak A, McPhee DA, Manigart O, Karita E, Inwoley A, Jaoko W, Dehovitz J, Bekker L-G, Pitisuttithum P, Paris R, Walker LM, Poignard P, Wrin T, Fast PE, Burton DR, Koff WC (2009) Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. J Virol 83:7337–7348. https://doi.org/10.1128/JVI.00110-09

59. Sok D, Laserson U, Laserson J, Liu Y, Vigneault F, Julien JP, Briney B, Ramos A, Saye KF, Le K, Mahan A, Wang S, Kardar M, Yaari G, Walker LM, Simen BB, St. John EP, Chan-Hui PY, Swiderek K, Kleinstein SH, Alter G, Seaman MS, Chakraborty AK, Koller D, Wilson IA, Church GM, Burton DR, Poignard P (2013) The effects of somatic hypermutation on neutralization and binding in the PGT121 family of broadly neutralizing HIV antibodies. PLoS Pathog 9:e1003754. https://doi.org/10.1371/journal.ppat.1003754

60. Soto C, Ofek G, Joyce MG, Zhang B, McKee K, Longo NS, Yang Y, Huang J, Parks R, Eudailey J, Lloyd KE, Alam SM, Haynes BF, Mullikin JC, Connors M, Mascola JR, Shapiro L, Kwong PD, NISC Comparative Sequencing Program (2016) Developmental pathway of the MPER-directed HIV-1-neutralizing antibody 10E8. PLoS One 11:e0157409. https://doi.org/10.1371/journal.pone.0157

61. Stern JNH, Yaari G, Heiden JA Vander, Church G, William F, Hintzen RQ, Huttner AJ, Laman JD, Nagra RM, Pitt D, Ramanan S, Siddiqui BA, Vigneault F, Kleinstein H, Hafler DA, Connor KCO (2015) B cells populating the multiple sclerosis brain develop in the draining cervical lymph nodes. https://doi.org/10.1126/scitranslmed.3008879.B409

62. Sui J, Hwang WC, Perez S, Wei G, Aird D, Chen L, Santelli E, Stec B, Cadwell G, Ali M, Wan H, Murakami A, Yammanuru A, Han T, Cox NJ, Bankston LA, Donis RO, Liddington RC, Marasco WA (2009) Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. Nat Struct Mol Biol 16:265–273. https://doi.org/10.1038/nsmb.1566

63. Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302:575–581. https://doi.org/10.1038/302575a0

64. Vander Heiden JA, Stathopoulos P, Zhou JQ, Chen L, Gilbert TJ, Bolen CR, Barohn RJ, Dimachkie MM, Ciafaloni E, Broering TJ, Vigneault F, Nowak RJ, Kleinstein SH, O'Connor KC (2017) Dysregulation of B cell repertoire formation in myasthenia gravis patients revealed through deep sequencing. J Immunol 198:1460–1473. https://doi.org/10.4049/jimmunol.1601415

65. Wang Y, Jackson KJ, Gäeta B, Pomat W, Siba P, Sewell WA, Collins AM (2011) Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants. Immunogenetics 63:259–265. https://doi.org/10.1007/s00251-010-0510-8

66. Wang C, Liu Y, Cavanagh MM, Le Saux S, Qi Q, Roskin KM, Looney TJ, Lee J-Y, Dixit V, Dekker CL, Swan GE, Goronzy JJ, Boyd SD (2015) B-cell repertoire to varicella-zoster vaccination in human identical twins. Proc Natl Acad Sci 112:500–505. https://doi.org/10.1073/pnas.1415875112

67. Wardemann H, Busse CE (2017) Novel approaches to analyze immunoglobulin repertoires. Trends Immunol 38:471–482. https://doi.org/10.1016/j.it.2017.05.003

68. Watson C, Breden F (2012) The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. Genes Immun 1312:363–373. https://doi.org/10.1038/gene.2012.12

69. Watson CT, Steinberg KM, Huddleston J, Warren RL, Malig M, Schein J, Willsey AJ, Joy JB, Scott JK, Graves TA, Wilson RK, Holt RA, Eichler EE, Breden F (2013) Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. Am J Hum Genet 92:530–546. https://doi.org/10.1016/j.ajhg.2013.03.004

70. Watson CT, Glanville J, Marasco WA (2017) The individual and population genetics of antibody immunity. Trends Immunol 38:459–470. https://doi.org/10.1016/j.it.2017.04.003

71. Weinstein JA, Jiang N, White RA, Fisher DS, Quake SR (2009) High-throughput sequencing of the zebrafish antibody repertoire. Science 324:807–810. https://doi.org/10.1126/science.1170020

72. West APJ, Diskin R, Nussenzweig MC, Bjorkman PJ (2012) Structural basis for germ line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. Proc Natl Acad Sci 109:2083–2090. https://doi.org/10.1073/pnas.1208984109

73. Wheatley AK, Whittle JRR, Lingwood D, Kanekiyo M, Yassine HM, Ma SS, Narpala SR, Prabhakaran MS, Matus-Nicodemos RA, Bailer RT, Nabel GJ, Graham BS, Ledgerwood JE, Koup RA, McDermott AB (2015) H5N1 Vaccine-elicited memory B Cells are genetically constrained by the IGHV locus in the recognition of a neutralizing epitope in the hemagglutinin stem. J Immunol 195:602. https://doi.org/10.4049/jimmunol.1402835

74. Williams WB, Liao H-X, Moody MA, Kepler TB, Alam SM, Gao F, Wiehe K, Trama AM, Jones K, Zhang R, Song H, Marshall DJ, Whitesides JF, Sawatzki K, Hua A, Liu P, Tay MZ, Seaton KE, Shen X, Foulger A, Lloyd KE, Parks R, Pollara J, Ferrari G, Yu J-S, Vandergrift N, Montefiori DC, Sobieszczyk ME, Hammer S, Karuna S, Gilbert P, Grove D, Grunenberg N, McElrath MJ, Mascola JR, Koup RA, Corey L, Nabel GJ, Morgan C, Churchyard G, Maenza J, Keefer M, Graham BS, Baden LR, Tomaras GD, Haynes BF (2015) HIV-1 VACCINES. Diversion of HIV-1 vaccine-induced immunity by gp41-microbiota cross-reactive antibodies. Science 349:1253. https://doi.org/10.1126/science.aab1253

75. Wu YC, Kipling D, Leong HS, Martin V, Ademokun AA, Dunn-Walters DK (2010) High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. Blood 116:1070–1078. https://doi.org/10.1182/blood-2010-03-275859

76. Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, Chen X, Longo NS, Louder M, McKee K, O'Dell S, Perfetto S, Schmidt SD, Shi W, Wu L, Yang Y, Yang Z-YZ, Yang Z-YZ, Zhang Z, Bonsignori M, Crump JA, Kapiga SH, Sam NE, Haynes BF, Simek M, Burton DR, Koff WC, Doria-Rose NA, Connors M, Mullikin JC, Nabel GJ, Roederer M, Shapiro L, Kwong PD, Mascola JR (2011) Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. Science 333:1593–1602. https://doi.org/10.1126/science.1207532

77. Wu X, Zhang Z, Schramm CA, Joyce MG, Do Kwon Y, Zhou T, Sheng Z, Zhang B, O'Dell S, McKee K, Georgiev IS, Chuang GY, Longo NS, Lynch RM, Saunders KO, Soto C, Srivatsan S, Yang Y, Bailer RT, Louder MK, Mullikin JC, Connors M, Kwong PD, Mascola JR, Shapiro L (2015) Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. Cell 161:480–485. https://doi.org/10.1016/j.cell.2015.03.004

78. Xiao M, Prabakaran P, Chen W, Kessing B, Dimitrov DS (2013) Deep sequencing and Circos analyses of antibody libraries reveal antigen-driven selection of Ig VH genes during HIV-1 infection. Exp Mol Pathol 95:357–363. https://doi.org/10.1016/j.yexmp.2013.10.004

79. Yaari G, Kleinstein SH (2015) Practical guidelines for B-cell receptor repertoire sequencing analysis. Genome Med 7:121. https://doi.org/10.1186/s13073-015-0243-2

80. Yacoob C, Pancera M, Vigdorovich V, Oliver BG, Glenn JA, Feng J, Sather DN, McGuire AT, Stamatatos L (2016) Differences in allelic frequency and CDRH3 region limit the engagement of HIV Env immunogens by putative VRC01 neutralizing antibody precursors. Cell Rep 17:1560–1570. https://doi.org/10.1016/j.celrep.2016.10.017

81. Yin L, Hou W, Liu L, Cai Y, Wallet MA, Gardner BP, Chang K, Lowe AC, Rodriguez CA, Sriaroon P, Farmerie WG, Sleasman JW, Goodenow MM (2013)

IgM repertoire biodiversity is reduced in HIV-1 infection and systemic lupus erythematosus. Front Immunol 4:373. https://doi.org/10.3389/fimmu.2013.00373

82. Zhou T, Zhu J, Wu X, Moquin S, Zhang B, Acharya P, Georgiev IS, Altae-Tran HR, Chuang GY, Joyce MG, DoKwon Y, Longo NS, Louder MK, Luongo T, McKee K, Schramm CA, Skinner J, Yang Y, Yang Z, Zhang Z, Zheng A, Bonsignori M, Haynes BF, Scheid JF, Nussenzweig MC, Simek M, Burton DR, Koff WC, Mullikin JC, Connors M, Shapiro L, Nabel GJ, Mascola JR, Kwong PD (2013) Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. Immunity 39:245–258. https://doi.org/10.1016/j.immuni.2013.04.012

83. Zhou T, Lynch RM, Chen L, Acharya P, Wu X, Doria-Rose NA, Joyce MG, Lingwood D, Soto C, Bailer RT, Ernandes MJ, Kong R, Longo NS, Louder MK, McKee K, O'Dell S, Schmidt SD, Tran L, Yang Z, Druz A, Luongo TS, Moquin S, Srivatsan S, Yang Y, Zhang B, Zheng A, Pancera M, Kirys T, Georgiev IS, Gindin T, Peng HP, Yang AS, Mullikin JC, Gray MD, Stamatatos L, Burton DR, Koff WC, Cohen MS, Haynes BF, Casazza JP, Connors M, Corti D, Lanzavecchia A, Sattentau QJ, Weiss RA, West AP, Bjorkman PJ, Scheid JF, Nussenzweig MC, Shapiro L, Mascola JR, Kwong PD (2015) Structural repertoire of HIV-1-neutralizing antibodies targeting the CD4 supersite in 14 donors. Cell 161:1280–1292. https://doi.org/10.1016/j.cell.2015.05.007

84. Zhu J, O'Dell S, Ofek G, Pancera M, Wu X, Zhang B, Zhang Z, Mullikin JC, Simek M, Burton DR, Koff WC, Shapiro L, Mascola JR, Kwong PD (2012) Somatic populations of PGT135-137 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinformatics. Front Microbiol 3:315. https://doi.org/10.3389/fmicb.2012.00315

85. Zhu J, Wu X, Zhang B, Mckee K, Dell SO, Soto C, Zhou T (2013) De novo identification of VRC01 class HIV-1 – neutralizing antibodies by next-generation sequencing of B-cell transcripts. Proc Natl Acad Sci 110:4088–4097. doi.org/10.1073/pnas.1306262110/