Virologica Sinica 37 (2022) 370–379

Contents lists available at ScienceDirect

Virologica Sinica

journal homepage: www.keaipublishing.com/en/journals/virologica-sinica
www.virosin.org

Research Article

Characterization of human IgM and IgG repertoires in individuals with chronic HIV-1 infection

Xiaolong Tian [a,1], Binbin Hong [b,*,1], Xiaoyi Zhu [a,1], Desheng Kong [c], Yumei Wen [a], Yanling Wu [a], Liying Ma [c,*], Tianlei Ying [a,d,e,*]

[a] MOE/NHC Key Laboratory of Medical Molecular Virology, School of Basic Medical Sciences, Shanghai Medical College, Fudan University, Shanghai, 200032, China
[b] Central Laboratory, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, 362000, China
[c] State Key Laboratory of Infectious Disease Prevention and Control, National Center for AIDS/STD Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, 102206, China
[d] Shanghai Engineering Research Center for Synthetic Immunology, Shanghai, 200032, China
[e] Shanghai Key Laboratory of Lung Inflammation and Injury, Shanghai, 200032, China

ARTICLE INFO

Keywords:
Ig-seq
HIV-1
Antibody repertoire
VDJ rearrangement
Junctional diversity

ABSTRACT

Advancements in high-throughput sequencing (HTS) of antibody repertoires (Ig-Seq) have unprecedentedly improved our ability to characterize the antibody repertoires on a large scale. However, currently, only a few studies explored the influence of chronic HIV-1 infection on human antibody repertoires and many of them reached contradictory conclusions, possibly limited by inadequate sequencing depth and throughput. To better understand how HIV-1 infection would impact humoral immune system, in this study, we systematically analyzed the differences between the IgM (HIV-IgM) and IgG (HIV-IgG) heavy chain repertoires of HIV-1 infected patients, as well as between antibody repertoires of HIV-1 patients and healthy donors (HH). Notably, the public unique clones accounted for only a negligible proportion between the HIV-IgM and HIV-IgG repertoires libraries, and the diversity of unique clones in HIV-IgG remarkably reduced. In aspect of somatic mutation rates of CDR1 and CDR2, the HIV-IgG repertoire was higher than HIV-IgM. Besides, the average length of CDR3 region in HIV-IgM was significant longer than that in the HH repertoire, presumably caused by the great number of novel VDJ rearrangement patterns, especially a massive use of *IGHJ6*. Moreover, some of the B cell clonotypes had numerous clones, and somatic variants were detected within the clonotype lineage in HIV-IgG, indicating HIV-1 neutralizing activities. The in-depth characterization of HIV-IgG and HIV-IgM repertoires enriches our knowledge in the profound effect of HIV-1 infection on human antibody repertoires and may have practical value for the discovery of therapeutic antibodies.

1. Introduction

Nearly 38 million people globally are living with human immunodeficiency virus (HIV) (https://www.who.int/news-room/fact-sheets/detail/hiv-aids), an integrating retrovirus that over time causes acquired immune deficiency syndrome (AIDS) by attacking human immune cells, specifically CD4 T cells. Since two of the CD4 T cell subsets, T helper 1 (Th1) and T helper 2 (Th2) cells, play an important role in B cell clonal expansion, the B cell homeostasis and normal B cell architecture are disturbed and the repertoire of antibodies would be reshaped during HIV infection (Smith et al., 2000; Zhu and Paul, 2008; Sajadi et al., 2018). Moreover, the subsequent depletion of CD4 T cells would further decrease antibody production and even lead to the disruption of the adaptive immune system, albeit with the antiretroviral therapy to control viral replication (Jacobson et al., 2002; Brenchley et al., 2006). Therefore, an in-depth analysis of the HIV-infected antibody repertoire characteristics is of great interest in illustrating the underlying mechanisms of B cell dysfunction and the molecular perturbations in the antibody repertoire.

Adaptive immunity distinguishes from innate immunity markedly in immense diversities of antibody or B cell receptor (BCR), and T cell receptor (TCR). Key determinants of the antibody diversities include the

* Corresponding authors.
E-mail addresses: bbh.0329@163.com (B. Hong), mal@chinaaids.cn (L. Ma), tlying@fudan.edu.cn (T. Ying).
[1] Xiaolong Tian, Binbin Hong and Xiaoyi Zhu contributed equally to this work.

https://doi.org/10.1016/j.virs.2022.02.010
Received 25 September 2021; Accepted 24 February 2022
Available online 3 March 2022

gene usages and rearrangement patterns of variable (*V*), diversity (*D*), and joining (*J*) genes, junctional diversity resulted from exonuclease trimmings and the random addition of palindrome (P) or non-template encoded (N) nucleotides (Li et al., 2004; Stavnezer et al., 2008). What's more, high-affinity antibodies are produced by B cells via the subsequent class switch recombination (CSR) and somatic hypermutation (SHM) to eliminate the invading pathogens. More importantly, each of the immunoglobulin heavy chain variable (IGHV) domain consists of three hypervariable loops, known as the complementarity-determining region (CDR) 1, 2, and 3, which are critical in antigen recognition and binding. Especially, CDR3 with the most variable sequences and structures is sufficient for most antibody specificities (Xu and Davis, 2000). Enormous efforts have been made to study these antibody repertoire characteristics using Ig-Seq (Weinstein et al., 2009; DeKosky et al., 2013; Turchaninova et al., 2016), providing unprecedented insight into immune responses to pathogens or vaccines, along with potential clinical applications for biomarker discovery and antibody development (Laserson et al., 2014; Wang et al., 2018; Kreer et al., 2020). Previous studies characterizing the antibody repertoire in HIV-1 infection, which accounts for 95% of all HIV infections, have uncovered the evolution of germline antibodies to broadly neutralizing antibodies (bnAbs) and the mutation pathway of HIV under the selective pressure of the immune system (Liao et al., 2013; Hoehn et al., 2015; Bonsignori et al., 2016; Setliff et al., 2018). However, none of the capacities of the unique antibody clonotypes in these studies exceeded a million to date, and controversial results were also presented in different researches (Scamurra et al., 2000; Bowers et al., 2014), demonstrating that a high-throughput and systematic investigation of HIV-1 infection antibody repertoires is still lacking.

Here, we identified approximately nine million IgM and two million IgG unique antibody clones from unsorted peripheral blood mononuclear cells (PBMCs) of ten HIV-1 infected patients by Ig-Seq. Using a comprehensive bioinformatic pipeline to analyze these two repertoires and another previously described IgM repertoire from 33 healthy donors (HH) (Hong et al., 2018), several signatures of antibody repertoires in HIV-1-infected patients have been identified. Besides, a serial of antibody lineages with potential HIV-1 neutralizing activity was predicted using bioinformatic strategies. These results provide insights into how antibody repertoires are reshaped by HIV-1 infection on a large-scale, and may contribute to the development of therapeutic antibodies against HIV-1 infection.

2. Materials and methods

2.1. Samples

Ten chronic HIV-1-infected patients (29–46 years old; 7.1 years of average infection time), with an average viral load of 22889 ± 17702.18 copies/mL and CD4 T cell count of 516.89 ± 220.90/mm^3, were enrolled in the study. A 10 mL peripheral blood sample of each HIV-1-infected patient was collected by the Chinese Center for Disease Control and Prevention. The baseline characteristics of 10 HIV-1-infected patients were summarized in Table 1 and those of 33 HH were described in our previous study (Hong et al., 2018). By the year of infection and viral load, 10 of the HIV-1-infected patients were classified into 5 groups for the further study (Table 1).

Table 1
Descriptive characteristics of HIV-1-infected patients (n = 10).

Group	Patients	Viral load (copies/mL)	CD4 (cells/mm^3)	Age (y)	Year of infection
GX03	GX2016EU03	51800	934	45	16
XC06	XC2014EU06	14100	437	41	13
XC13	XC2014EU13	42200	373	45	13
MD-05	XC2014EU16	4150	389	44	5
	GX2016EU01	14700	806	42	5
	BJ2015EU01	27100	317	46	6
	BJ2015EU19	7790	311	36	5
	XC2014EU02	4150	596	31	5
NG-02	GX2016EU06	17800	517	29	1
	GX2016EU22	45100	489	39	2

2.2. cDNA template preparation and PCR amplification

PBMCs of HIV-1 infected patients were isolated using density gradient medium (Catalog #07851, Lymphoprep™, STEMCELL, Canada), followed by the total RNA extraction with the RNeasy Mini Kit (Catalog #74104, Qiagen, Germany). The first-strand cDNA was synthesized via reverse transcription from 500 ng of total RNA using a SuperScript™ First-Strand Synthesis (Catalog #18080051, Invitrogen, USA). Two rounds of PCR amplification were performed to amplify the IgM and IgG heavy chain variable (*VH*) genes from cDNA (Supplementary Fig. S1). Concretely, the first-round amplification used forward primers set annealing to the leader sequence (L) at the beginning of each *VH* segment and a series of reverse primers specific to the first constant Ig domain of the heavy chain (CH1) of immunoglobulin heavy constant mu (*IGHM*) or immunoglobulin heavy constant gamma (*IGHG*) gene. The second-round amplification was performed to produce shorter *VH* fragments ranging from the first frame regions (FR1) of *IGHV* genes to the fourth frame regions (FR4) in the immunoglobulin Heavy Joining (*IGHJ*) gene for Ig-seq. The primers used in the two rounds PCR amplification of IgM *VH* genes as we described previously (Hong et al., 2018), were the same as the primers for IgG *VH* gene amplification, except for the reverse primers specific to the CH1 domain (Supplementary Table S1). Two-round PCR reactions were all carried out using Pfu mastermix (Catalog #CW2965F, CWBIO, China) under the following conditions: initial denaturation at 94 °C for 5 min, 30 cycles of denaturation at 94 °C for 40 s, annealing at 58 °C for 45 s, extension at 72 °C for 1 min, and final extension at 72 °C for 10 min. The PCR amplicons were then purified using the QIAquick Gel Extraction Kit (Catalog #28706X4, Qiagen, Germany).

2.3. Library construction and sequencing

After the quality control of the second round *IGHV* gene fragments using ND-1000 Nanodrop and Agarose Gel Electrophoresis, library construction was executed using TruSeq® DNA HT Sample Prep Kit (Catalog #FC-121-2003, Illumina, USA) according to the manufacturer's protocol. Adapter-ligated fragments were then PCR amplified and gel purified to create the final cDNA library, following a quality control employing Agilent 2200 TapeStation and Qubit 2.0. Subsequently, high-throughput sequencing was carried out on the Illumina Hiseq 2500 platform. The sequencing data reported in this paper have been deposited in the OMIX, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix/release/OMIX604) (Chen et al., 2021).

2.4. Bioinformatic analysis

Totally, five groups of IgM and IgG sequencing data were merged respectively. For quality control, only raw reads hitting a Phred quality score of 20 over 80% of bases were reserved to exclude errors caused by PCR errors and sequencing artifacts. Filtered clean reads were then submitted to the IMGT/HighV-QUEST (version 1.5.1) (http://www.imgt.org/HighV-QUEST/search.action) for gene alignment (Ehrenmann et al., 2011). The output unique amino acid (aa) sequences were classified into productive and unproductive groups, and only productive sequences without stop codons or indels in the *V* and *J* gene segments remained. Productive sequences carrying substitution mutations in the specified conserved positions were removed to eliminate substitution mutations regularly generated in the sequencing process, and the productive sequences mentioned below all denoted the filtered ones. The remaining sequences were subsequently defined as unique clones according to unique CDR3 sequences or unique VDJ rearrangement patterns. To unbiasedly estimate the antibody repertoires' diversities with varied

numbers of input cells or sequences, an *in silico* simulation was used to generate new databases by randomly selecting the same number of sequences from each sequencing data set as previously reported (Hong et al., 2018).

2.5. Statistical analysis

All statistical analyses were performed using Perl and R. *IGHV*, immunoglobulin heavy constant delta (*IGHD*), and *IGHJ* gene usages, as well as occurrence rates of junctional modifications among antibody repertoires, were analyzed with logistic regression analyses. Length distribution of CDRs and junctional modifications within antibody repertoires were tested with Student's *t*-test. *P*-value < 0.05 was considered statistically significant. For logistic regression, the odds ratio (OR) was used as an indicator of the effect size to determine if the differences are meaningful (De Muth, 2014), with $0.9 < OR < 1.1$ considered not significantly different, $0.7 < OR < 0.8$ or $1.2 < OR < 1.4$ slightly different, $0.4 < OR < 0.6$ or $1.5 < OR < 2.9$ moderately different, and $OR < 0.3$ or $OR > 3.0$ dramatically different or strong association. For Student's *t*-test, Cohen's *d* value was used to measure the standardized difference between two groups (Cohen, 1988), and $d = 0.20$, 0.50, and 0.80 were considered as small, medium, and large differences respectively. Graphs were prepared using GraphPad Prism 7. All data represented means ± standard error from three independent experiments performed in triplicates.

3. Results

3.1. Diversities of antibody repertoires

By performing high-throughput sequencing, 46,742,356 and 59,320,054 raw reads were obtained from the IgM and IgG repertoires of HIV-1-infected patients, respectively. IMGT/HighV-QUEST analysis results showed that, among all the unique amino acid sequences, 75.7% of HIV-IgM and 72.7% of HIV-IgG sequences were productive. In this study, we define "unique clones" as sequences that had unique V(D)J gene arrangements or unique CDR3 amino acid sequences. After a series of stringent data filtering and cleaning procedures as described above, finally, a total of 9,197,007 unique clones (41.7% of productive sequences) were identified in the HIV-IgM repertoire, and 2,025,035 unique clones (16.6% of productive sequences) in the HIV-IgG repertoire (Supplementary Table S2). To evaluate the diversity of repertoires, we randomly selected sequences *in silico* and found that the proportion of unique clones from the HIV-IgM repertoire was distinctly higher than that from HIV-IgG when the number of randomly selected sequences was over 10,000 (Supplementary Fig. S2). The difference further escalated with the increased input sequences, indicating that the HIV-IgM repertoire was more diverse than the HIV-IgG repertoire.

Next, we matched unique clones from the HIV-IgM, HIV-IgG, and HH repertoires. Results showed that HIV-IgM and HIV-IgG repertoires shared 14,729 unique clones, accounting for a negligible proportion, approximately 0.2% and 0.7% of the two repertoires, respectively. For HIV-IgM and HH repertoires, 14,552 unique clones were shared, in an almost equivalent proportion of about 0.2% in both two repertoires (Fig. 1A). CDR3 sequences have also been used to measure the diversity of antibody repertoires (D'Angelo et al., 2018). In this study, there were 3,562,095 unique CDR3 sequences in the HIV-IgM repertoire and 1,252,891 in the HIV-IgG repertoire. Among them, 12,352 shared CDR3 sequences were detected, occupying 0.3% and 1.0% in HIV-IgM and HIV-IgG repertoires correspondingly (Fig. 1B). Withal, the number of shared CDR3 sequences was observed to be 43,786 within the HIV-IgM and HH repertoires (3,428,850 unique CDR3 sequences) and accounted for about 1.2% in both.

Using the IMGT/High V-QUEST tool, we further analyzed the VDJ rearrangement patterns and found 228 and 237 *IGHV* alleles in the HIV-IgM and HIV-IgG repertoires respectively, and 28 *IGHD* alleles, 13 *IGHJ* alleles in both. Interestingly, among the 49,285 VDJ rearrangements detected in the HIV-IgM and 43,992 in the HIV-IgG repertoire, 38,893 shared patterns were observed, taking up 78.9% of the total VDJ rearrangements in the HIV-IgM and 88.4% in the HIV-IgG repertoires (Fig. 1C). Meanwhile, the HIV-IgM and HH (34,688 VDJ rearrangements) repertoires shared 29,155 identical VDJ rearrangement patterns, which accounted for 84.0% of the HH repertoire but only 59.2% of the HIV-IgM repertoire (Fig. 1C). Taken together, these results highlighted that although the VDJ rearrangements in HIV-1-infected repertoires were highly conserved between the isotypes of IgM and IgG, only negligible levels of similarity of unique clones and CDR3 sequences were found among the three antibody repertoires.

3.2. Characterization of CDRs

To acquire higher antigen-binding affinity, B lymphocytes often experience a complex process of SHM, especially in the CDR regions. Thus, the CDRs are crucial to the diversity of antigen specificities and may play the role as signatures of B cells clonal expansion following antigen recognition (LeBien and Tedder, 2008). In this study, in both HIV-IgM and HIV-IgG repertoires, the length distribution of CDR1 and CDR2 ranged from 8 to 10 aa and 7 to 10 aa, respectively (Fig. 2A and B).

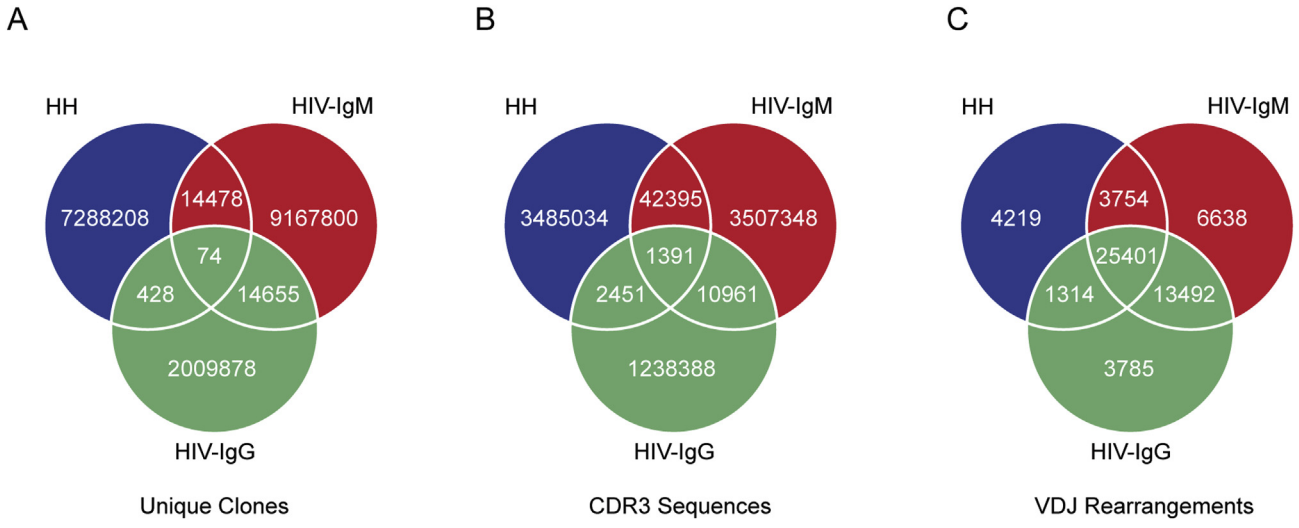

Fig. 1. Venn diagram comparing the numbers of unique clones (**A**), *VDJ* gene rearrangement (**C**), and CDR3 sequence (**B**) among the HIV-IgM, HIV-IgG, HH antibody repertoires. HH, healthy donor.

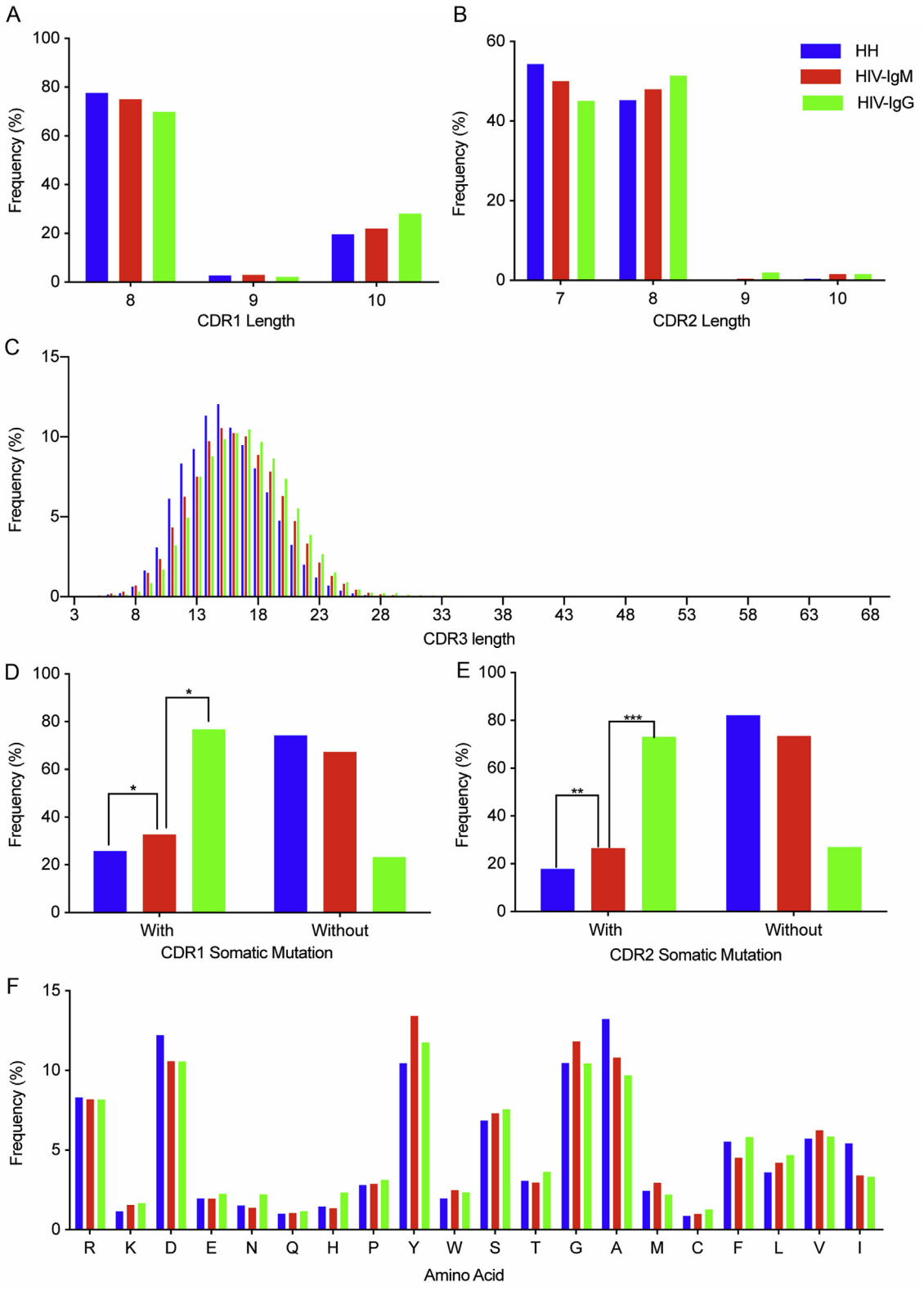

Fig. 2. The characteristics of CDRs in the HH, HIV-IgM, and HIV-IgG repertoires. (**A**) Length distribution of CDR1 region ranging from 8 to 10. (**B**) Length distribution of CDR2 region ranging from 7 to 10. (**C**) The CDR3 length distribution. (**D**) The somatic mutation frequency in the CDR1 region. "With" represented at least one amino acid change, and "Without" represented no amino acids changes compared to the germline IgH gene. (**E**) The somatic mutation frequency in the CDR2 region. (**F**) The amino acids usage of the CDR3 region. Statistical analysis was performed by R. "*" refers to statistically significant ($P < 0.05$) and "**" refers to statistically highly significant ($P < 0.01$). HH, healthy donor.

Besides, no significant difference was observed in the average length of CDR1 (HIV-IgM: 8.47, HIV-IgG: 8.58) or CDR2 (HIV-IgM: 7.54, HIV-IgG: 7.60) (Table 2). Nevertheless, the length of CDR3 was highly diverse, ranging from 3 to 70 aa, with the highest frequency of 15 aa in HIV-IgM and 17 aa in HIV-IgG (Fig. 2C). The proportion of ultra-long CDR3 sequences with a length over 30 aa, in HIV-IgM (0.1%) was lower than that in HIV-IgG (0.6%). However, interestingly, the average length of CDR3 in the two repertoires was comparable (16.31 in HIV-IgM and 16.96 in HIV-IgG, $P < 0.05$, $d = -0.17$, Table 2). Meanwhile, we also evaluated the length distribution and average length of CDR1 and CDR2 in HIV-IgM and HH repertoires and no significant difference was detected. In contrast, the average length of CDR3 in HIV-IgM was significantly longer than that of HH (16.31 in HIV-IgM, 15.48 in HH, $P < 0.05$, $d = 0.22$, Table 3), which may attribute to the 10-fold higher proportion of ultra-long CDR3 sequences in the HIV-IgM (0.1%) than HH (0.01%). These results suggested higher structural complexity of the HIV-1-experienced antibody repertoires.

As a hypervariable region, the aa changes in CDR3 cannot be compared directly due to the highly flexible structures. Instead, we analyzed the aa usage patterns and the hydrophobicity value of CDR3 by the Kyte-Doolittle scale. Approximately 75% amino acids in CDR3 were identified hydrophilic and no difference was presented in both HIV-IgM and HIV-IgG repertoires [$P < 2.2E{-}16$, OR = 1.17, 95% CI: (1.162, 1.171)]. Whereas, the proportion of hydrophilic aa in CDR3 of the HIV-IgM was slightly higher than that in the HH [75.41% vs. 69.36% , OR = 0.738 , 95% CI: (0.737, 0.740)]. In addition, the usage rates of 20 amino acids [$P < 2.2E{-}16$, OR = 0.998, 95% CI: (0.998, 0.998), Fig. 2F], average hydrophobicity (HIV-IgM: -0.39 ± 2.72, HIV-IgG: -0.44 ± 2.76, $t = 104.303$, $d = 0.02$) and charge values (HIV-IgM: 0.06 ± 1.00, HIV-IgG: 0.03 ± 1.00, $t = 92.076$, $d = 0.04$) all displayed no significant difference between HIV-IgM and HIV-IgG, as well as between HIV-IgM and HH.

3.3. SHM analysis

To investigate the characteristics of somatic mutations in different repertoires, we calculated aa changes within CDR1 and CDR2 via alignment with corresponding germline genes. The mutation frequencies, comprised in CDR1 and CDR2, in HIV-IgG were both significantly higher than that in HIV-IgM [CDR1: HIV-IgM 32.67%, HIV-IgG 76.73%, OR = 6.80, 95% CI: (6.772, 6.820); CDR2: HIV-IgM 26.56%, HIV-IgG 72.06%, OR = 8.81, 95% CI: (8.765, 8.855), Fig. 2D and E]. Meanwhile, the HIV-IgM CDR1 and CDR2 mutation frequency was statistically higher than that in HH [CDR1: 32.67% vs. 25.84%, OR = 0.718, 95% CI: (0.717, 0.720); CDR2: 26.56% vs. 17.82%, OR = 0.60, 95% CI: (0.598, 0.601), Fig. 2D and E]. Overall, these results demonstrated that the characteristics of CDRs in the HIV-IgM were similar to that in HIV-IgG, except for the lower CDR1 and CDR2 somatic mutation frequencies. However, the CDRs characteristics in HIV-IgM vastly varied from the HH repertoire, in aspects of somatic mutation rates of CDR1 and CDR2, the average CDR3 length and the proportion of hydrophilic aa in CDR3.

3.4. VDJ gene usage

The *VDJ* gene usage determines CDR3 diversity and is selected and affected by pathogen infection (Chen et al., 2012; Larimore et al., 2012). Thus, the usages of *VDJ* genes in HIV-IgM and HIV-IgG repertoires were evaluated and shown in Table 4 and Fig. 3. Logistic regression analysis revealed no significant difference in *IGHD* and *IGHJ* gene usages but a minor discrepancy of *IGHV* (OR = 0.819) gene segments between HIV-IgM and HIV-IgG (Table 4). In specific, in both repertoires, families of *IGHV1*, *IGHV2*, *IGHV3*, and *IGHV4* were preferentially used, while the *IGHV5*, *IGHV6*, and *IGHV7* were less used with a total ratio of 4.81% in the HIV-IgG repertoire and 1.15% in the HIV-IgM repertoire. However, the discrepancy lied in that the usage of *IGHV3* and *IGHV4* in the HIV-IgG repertoire (*IGHV3*: 34.38%; *IGHV4*: 29.57%) were both lower than in the HIV-IgM repertoire (*IGHV3*: 41.20%; *IGHV4*: 42.63%), while the utilization of *IGHV1* and *IGHV2* in HIV-IgG (*IGHV1*: 18.92%; *IGHV2*: 12.31%) was higher than in the HIV-IgM repertoire (*IGHV1*: 12.04%; *IGHV2*: 2.98%) (Table 4). The most frequently used three *IGHV* genes in HIV-IgM were *IGHV4-59* > *IGHV4-34* > *IGHV4-39*, and the condition in HIV-IgG was *IGHV4-59* > *IGHV2-5* > *IGHV3-30*.

The *VDJ* gene usage in the HIV-IgM repertoire was also differentiated from HH repertoires. Although *IGHV4* and *IGHD3* were the most used *V*, *D* gene families in both repertoires (Fig. 3C), the top three used *V* and *D* genes were different (HIV-IgM repertoire: *IGHV4-59* > *IGHV4-34* > *IGHV4-39* and *IGHD3-10* > *IGHD6-13* > *IGHD3-22*; HH repertoire: *IGHV4-59* > *IGHV1-69* > *IGHV4-34* and *IGHD3-10* > *IGHD3-22* > *IGHD1-26*, Fig. 3D). Further logistic regression analysis revealed that there was a slight difference in the usage of the *IGHV* gene family between the two repertoires (OR = 0.789). For the *IGHJ* gene family, the most preferred gene segments in HIV-IgM and HH were *IGHJ6* and *IGHJ3*, respectively (Fig. 3C). The OR value was 0.583

Table 2
The average (aa) length of CDRs in the HIV-IgM and HIV-IgG repertoires.

	HIV-IgM	HIV-IgG	P [a]	t [a]	d [a]	95% CI [a]
CDR1	8.47 ± 0.83	8.58 ± 0.9	<2.2E−16	−172.41	−0.13	(−0.114, −0.111)
CDR2	7.54 ± 0.59	7.6 ± 0.61	<2.2E−16	−138.874	−0.11	(−0.065, −0.063)
CDR3	16.31 ± 3.82	16.96 ± 4.12	3.19E−52	−217.886	−0.17	(−0.661, −0.65)

CI, confidence interval.
[a] Calculated by student's *t*-test.

Table 3
The average (aa) length of CDRs in the HIV-IgM and HH repertoires.

	HIV-IgM	HH	P [a]	t [a]	d [a]	95% CI [a]
CDR1	8.47 ± 0.83	8.42 ± 0.8	<2.2E−16	121.561	0.06	(0.048, 0.05)
CDR2	7.54 ± 0.59	7.47 ± 0.52	<2.2E−16	252.0921	0.12	(0.07, 0.071)
CDR3	16.31 ± 3.82	15.48 ± 3.43	<2.2E−16	455.5431	0.22	(0.821, 0.828)

HH, healthy donors; *CI*, confidence interval.
[a] Calculated by student's *t*-test.

Table 4
Comparison of the VDJ gene usage frequency observed in the HIV-IgM, HIV-IgG, and HH repertoires.

Subgroups	Repertoire			OR 95% CI [a]	
	HIV-IgM (%)	HIV-IgG (%)	HH (%)	HIV-IgM vs. HIV-IgG	HIV-IgM vs. HH
IGHV1	12.04	18.92	22.05	0.819 (0.817, 0.82)	0.789 (0.788, 0.79)
IGHV2	2.98	12.31	9.19		
IGHV3	41.20	34.38	25.16		
IGHV4	42.63	29.57	43.47		
IGHV5	0.42	0.68	0.10		
IGHV6	0.44	1.93	0.02		
IGHV7	0.29	2.21	0.02		
IGHD1	9.20	8.06	12.07	0.939 (0.938, 0.939)	0.976 (0.975, 0.977)
IGHD2	18.41	19.76	19.94		
IGHD3	39.99	45.50	34.10		
IGHD4	3.94	3.41	5.89		
IGHD5	6.99	7.55	7.40		
IGHD6	20.68	15.22	19.33		
IGHD7	0.80	0.51	1.29		
IGHJ1	1.30	1.15	0.38	0.94 (0.939, 0.941)	0.851 (0.851, 0.852)
IGHJ2	14.92	11.71	8.01		
IGHJ3	28.25	33.70	57.25		
IGHJ4	12.94	18.81	4.97		
IGHJ5	2.94	3.75	1.64		
IGHJ6	39.66	30.87	27.74	0.68 (0.677, 0.682)	0.584 (0.583, 0.585)

HH, healthy donors; *OR*, odds ratio; *CI*, confidence interval.
[a] Calculated by student's *t*-test.

when only the *IGHJ6* gene family was considered, indicating a moderate difference between these two repertoires (Table 4). Taken together, these results suggested that the *VDJ* genes were selectively used among the three antibody repertoires. The *IGHV* in HIV-IgG was more evenly used than in the HIV-IgM repertoire, while between the HIV-IgM and HH repertoires differences were observed in *IGHV* and *IGHJ* gene usages, especially for *IGHJ6*.

3.5. V-D-J junctional diversity

In addition to the *VDJ* gene rearrangement, modifications of nucleotide insertion and splicing in the V-D-J junction region also contribute to the CDR3 diversity. We, therefore, compared the occurrence rates and modification lengths of palindromic nucleotides (P) and the non-template randomized nucleotides (N) insertion, as well as the nucleotides deletion caused by exonuclease trimming (T), which had been described previously (Hong et al., 2018). Generally, the N additions were observed at the region between the 3′-end of *V* gene and the 5′-end of *D* gene (N1) and the region between the 3′-end of *D* gene and the 5′-end of *J* gene (N2). The P additions and the exonuclease trimming happened at 3′-end of *V* regions (3VP and 3VT), 5′-end and 3′-end of *D* genes (5DP and 5DT, 3DP and 3DT) and 5′-end of *J* genes (5JP and 5JT). The occurrence rates of 5DT, N1, and N2 contained in HIV-IgG were significantly higher than that of HIV-IgM, and an inverse trend was detected in 5DP (Fig. 4, Supplementary Table S3). The N1 and N2 nucleotide (nt) insertions in HIV-IgG were found with longer average lengths of 1.52 and 1.86 nt than that in HIV-IgM respectively. In terms of the HIV-IgM and HH repertoires, only the occurrence rates of N1, 5JP, and 5JT presented a significant difference (Fig. 4, Supplementary Table S4). Contrary to 5JP, whose occurrence rates were less than 10% in both repertoires and with an approximate insertion length (1.47 nt vs. 1.37 nt), N1 and 5JT presented higher frequency as well as longer length (N1: 6.85 nt vs. 6.41 nt; 5JT: 6.91 nt vs. 5.78 nt) in HIV-IgM compared with HH repertoires, which altogether may also contribute to the significant longer CDR3 length in the HIV-IgM than the HH repertoire.

3.6. Potentially HIV-1 neutralizing antibodies identification

IgG, derived and differed from IgM, implies the isotype of functional secreted antibodies to build long-term immune memory in response to invading pathogens, such as the chronic HIV-1 infection (Safonova and Pevzner, 2019). In addition, clonal expansion in IgG antibody repertoires is often considered as a hallmark of therapeutic antibodies (Reddy et al., 2010). The number of unique clones shared between the HIV-IgM and HIV-IgG repertoires but not between the HIV-IgM and HH repertoires were thus then counted. Finally, 14,655 public unique clones were obtained. When evaluating the frequency variance of these shared clones between the HIV-IgM and HIV-IgG repertoires, obvious ups and downs were observed (Supplementary Fig. S3A). Specifically, 79 clones were amplified 1000-fold, and 46 clones were collapsed with 1000-fold in the HIV-IgG repertoire, while 315 clones were amplified by 100-fold, and 344 clones were collapsed by 100-fold in the HIV-IgG repertoire. Further phylogenetic analysis revealed that several antibody lineages (shown in colored lines) were gradually evolved with a great number of unique clones (Supplementary Fig. S3B).

4. Discussion

HIV infection damages the human immune system and induces the reconstruction of the antibody repertoire. High-throughput Ig-Seq enables tracking of the characteristic changes in clonal populations during HIV infection, facilitates the comprehensive disclosure to HIV-derived B cell responses at the molecular level, and is also vital to evaluate the theoretical size of the antibody repertoire or identify therapeutic antibody. Here, a comprehensive antibody repertoire analysis of 10 HIV-1-infected patients, accompanied with 33 healthy individuals previously reported by our laboratory (Hong et al., 2018), was conducted to determine the changes in the antibody repertoire upon chronic HIV-1 infection and the imparities between two isotypes of IgM and IgG repertoires in HIV-infected patients. Moreover, a batch of potential HIV-1 neutralizing antibodies was specified.

Recent research suggested that the circulating immunoglobulin heavy-chain repertoire comprised about 11 million unique clones (Soto et al., 2019). However, the woefully inadequate sequencing depth hampered the characterization of the antibody repertoire upon HIV-1 infection. Moreover, the conclusions of the antibody diversity between healthy individuals and HIV-1 infected patients were controversial in previous researches (Yin et al., 2013), suggesting that an ultra-deep Ig-Seq was urgently needed to be involved. In this study, using a VH-targeted Ig-Seq strategy inexpensively, around 9 million unique

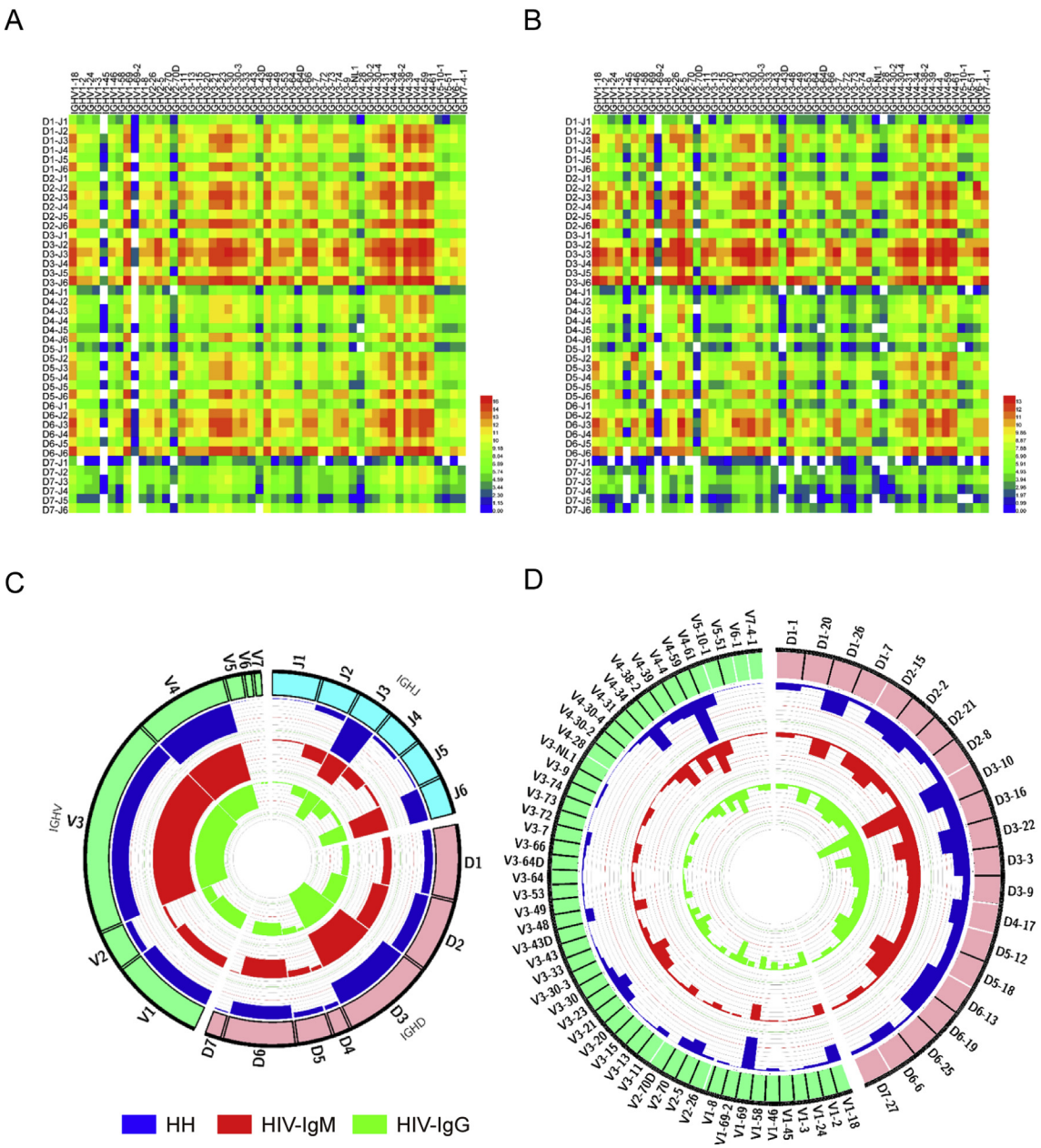

Fig. 3. The *VDJ* gene rearrangement patterns and *VDJ* gene usages in HIV-1-infected patients and healthy individuals. **A, B** Heat map of *VDJ* gene rearrangement patterns in the HIV-IgM repertoire (**A**) and the HIV-IgG repertoire (**B**). The base-2 logarithm of the count was used to present the numbers of each rearrangement pattern. Different colors represented different counts. **C, D** The usage of *VDJ* gene subgroups (**C**) and *VDJ* genes (**D**). The outer arcs of circus plots were represented the *VDJ* gene subgroups or *VDJ* genes with different colors (HH: blue; HIV-IgM: red; HIV-IgG: green), and the inner arcs were represented HH, HIV-IgM, and HIV-IgG repertoires by order. The usage rates were in proportion to the radians covered by the column. HH, healthy donor.

clones of the dominant component of BCR (Xu and Davis, 2000; Jiang et al., 2013), IgM repertoire, was obtained at a magnitude approaching the theoretical size (11 million) of an individual's circulating antibody repertoire and was superior to previous researches focusing on the antibody repertoire in HIV infection to our knowledge. Hence, this study enabled a more convincing analysis and provided a far broader view of the HIV-infected antibody repertoires. Furthermore, previous studies working on Ig-seq generally evaluated only one isotype of antibody repertoires or multiple isotypes without separate analysis of each antibody repertoire (Wu et al., 2012; Yin et al., 2013; Galson et al., 2014; Khurana et al., 2016). As founded in this study, public CDR3 sequences constituted only negligible proportions between the HIV-IgM and HIV-IgG repertoires, suggesting multiple antibody isotypes should be considered in the deep mining of antibody repertoires.

As the most variable domains of antibody, CDRs acquire somatic hypermutations that dominantly respond with specificity, duration, and strength to identify and bind to antigen epitopes (Saragovi et al., 1991; Hoet et al., 2005). We found that the preferential usages of certain *IGHV* and *IGHJ* genes did not affect the average length of CDR1 and CDR2. However, the somatic hypermutation levels of these two regions in the HIV-1-IgG repertoire were significantly higher than that in HIV-1-IgM, which in turn significantly higher than that in the HH antibody repertoire. These results, as well as previous work, indicated that antibodies underwent somatic mutations under the stimulation of HIV-1 infection to acquire high affinity, especially for the isotype of IgG (Chu et al., 1995; Jackson et al., 2014; Kitaura et al., 2017). Among the three CDRs in the heavy chain, CDR3 provided the widest range of variations in both length and structure (Chothia and Lesk, 1987; Chothia et al., 1989), and is the

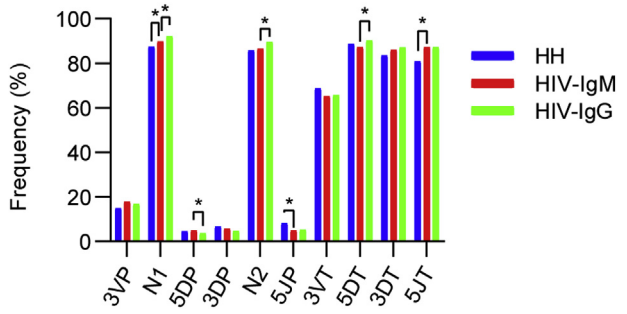

Fig. 4. Frequencies of junctional modifications. 3VP and 3VT: the palindromic nucleotides (P) additions and exonuclease trimmings in the 3′ end of *V* regions, respectively; 5DP and 5DT: the P additions and exonuclease trimmings in the 5′ end of *D* genes, respectively; 3DP and 3DT: the P additions and exonuclease trimmings in the 3′ end of *D* genes, respectively; 5JP and 5JT: the P additions and exonuclease trimmings in the 5′ end of *J* genes, respectively; N1: the non-template randomized nucleotides (N) additions between 3′ end of the *V* gene and 5′ end of the *D* gene; N2: N additions between 3′ end of the D gene and 5′ end of the *J* gene. Statistical analysis was performed by R. "*" refers to statistically significant ($P < 0.05$). HH, healthy donor.

determinant of specificity in antigen recognition (Davis et al., 1997). Our results showed that, in HIV-1-infected patients, the average CDR3 length in the IgG repertoire was 0.65 aa longer than that of the IgM repertoire. For the IgM repertoire, the average CDR3 length in HIV-1-infected individuals significantly increased 0.83 aa when compared with healthy individuals. These findings strongly suggested potentially higher structural complexity of the antibodies in HIV-1-experienced repertoires compared with naive repertoires (Wu et al., 2010; Hong et al., 2018; Hu et al., 2019). The increased length of CDR3 would be expected to expand the antibody repertoire diversity and facilitate binding to recessed epitopes of pathogens or the active sites of enzymes (Desmyter et al., 1996; Zwick et al., 2004; De Genst et al., 2006). Several reports have shown that many HIV-1 bnAbs (CAP256-VRC26, PGDM1400, and PGT141 series) contain extraordinarily long CDR3 with more than 30 aa, and even 39 aa appeared in CAP256-VRC26.08 and CAP256-VRC26.09 (Walker et al., 2011; Zhu et al., 2012; Doria-Rose et al., 2014). In our study, the proportion of ultra-long CDR3 sequences in the HIV-IgM repertoires were nearly 10 times more than the HH (IgM) repertoires, further, the HIV-IgG repertoire was 6 times more than the HIV-IgM repertoires, indicating that bnAbs may be successfully activated and expanded upon HIV-1 infection.

It is well-known that IgM provides the first line of humoral immunity defense upon antigen stimulation. After undergoing a process of class switch, responses of high-affinity IgG are initiated, which also play a key role in the long-term immune protection (Goding, 1978; Geisberger et al., 2006). In the present study, we found that *VDJ* gene rearrangements in the HIV-IgM and HIV-IgG repertoires were similar to a large extent, indicating that most of the *VDJ* gene rearrangement patterns in the HIV-IgG repertoire were derived from the IgM repertoire. However, about 40% of *VDJ* gene rearrangements in the HIV-IgM repertoire were different from the HH repertoire, demonstrating that during HIV-1 infection, a considerable number of *VDJ* genes undergo novel rearrangements. Since *VDJ* germline gene rearrangement is an important genetic basis of the diversity in immunoglobulins (Janeway et al., 2001), prior studies have found the preferential usages of some *VDJ* genes when responding to different antigens (Walker et al., 2011; Doria-Rose et al., 2014). As expected, the current study demonstrated higher *IGHV3* usage and lower *IGHV1*, *IGHV2* usage in the HIV-IgM repertoire than that of HH repertoire. However, this finding was contrary to previous research, whose results showed that the usage of *IGHV3* family decreased by 45% in HIV-1-infected patients. We presume this discrepancy was caused by the distinct throughput of antibody repertoires employed. Owing to a larger-scale antibody repertoire analysis enabled by the high-throughput sequencing technology, our study verified the significantly different propensity of *IGHV* and *IGHJ* gene usages during HIV-1 infection, especially the prominently increased utilization of *IGHJ6*, which therefore would deepen our understanding of the influence of HIV-1 infection on the human immune repertoires.

In addition to the variability in V-D-J rearrangement patterns, numerous studies have proved that junctional modifications also play an important role in expanding the diversity of CDR3 (de los Rios et al., 2015; Hong et al., 2018; Imkeller and Wardemann, 2018), despite none of them focused on the junctional diversity of antibody repertoires in HIV-1 infected patients. For the first time, our study reported the higher occurrence and length of N1 and N2 insertions, as well as the higher occurrence of 5DP and lower occurrence of 5DT in the HIV-IgG repertoire than that in the HIV-IgM repertoire. Interestingly, this feature, together with a 6-fold higher proportion of ultra-long CDR3 in HIV-IgG, did not make a significant difference between the two repertoires' CDR3 length. This result may be explained by the fact that the long *IGHV*, *IGHJ* genes were used more frequently in HIV-IgM while all the *IGHV* and *IGHJ* genes were evenly used in HIV-IgG. On the other hand, when comparing the HIV-IgM with the HH repertoire, despite relatively lower occurrence of N1 and 5JT and relatively higher occurrence of 5JP have been observed, only the average length of 5JT was statistically different, and junctional diversity had less effect on the length of CDR3, suggesting that junctional diversity contributed little to the longer CDR3 in HIV-1 infected repertoires. Hence, it could be conceivably hypothesized that the increased CDR3 length in HIV-IgM was primarily caused by the VDJ length which was determined by a great number of novel VDJ rearrangements, especially a massive use of the longest *IGHJ* gene *IGHJ6*.

Several reports have found that antibodies specific to HIV-1 or other viruses have undergone a series of somatic mutations, evolving mature antibodies with broad-spectrum neutralizing activity (Doria-Rose et al., 2014; MacLeod et al., 2016). Herein, it was found that several IgG antibodies were clonally expanded when compared with IgM, and a number of public clonotypes evolved into numerous unique clones. These amplified clones may be developed as candidate HIV-1 therapeutic antibodies through bioinformatic methods, which act as a supplement to traditionally experimental screening and overcomes the deficiency of long time-consumption and high labor-intensity (Tian et al., 2020).

5. Conclusions

In conclusion, using a series of antibody repertoire amplification primers and a comprehensive bioinformatic pipeline, we have analyzed millions of unique clones with unprecedented depth in the HIV-1-infected patients. Importantly, our study found that *VDJ* gene rearrangement patterns can be dramatically changed by HIV-1 infection. In-depth analyses of the HIV-1-infected antibody repertoires also uncover that the higher structural complexity of the HIV-1-experienced antibodies is simultaneously caused by the preferential usage of *VDJ* genes, junctional diversity, and proportion of ultra-long CDR3. Several amplified antibodies with somatic hypermutations are promising in HIV-1 neutralization. However, shortcomings like the read length limitation of 2 × 250 bp yielded out the extraordinarily long antibody sequences (Tan et al., 2016). Future studies with the light chain involved are needed to mine the natively paired human VH: VL (light chain variable gene) antibody against HIV-1 infection (Wang et al., 2018).

Data availability

The sequencing data reported in this paper have been deposited in the OMIX, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix/release/OMIX604) (Chen et al., 2021).

Ethics statement

This study was approved by the Institutional Research Ethics Community of the Chinese Center for Disease Control and Prevention, and all

subjects signed informed consent to participate in the research study prior to blood and data collection. All experiments were performed in accordance with relevant guidelines and regulations.

Author contributions

Xiaolong Tian: data curation, formal analysis, investigation, validation, visualization, writing-original draft. Binbin Hong: conceptualization, data curation, formal analysis, methodology, visualization. Xiaoyi Zhu: validation, writing-review & editing. Desheng Kong: investigation, resources, writing-review & editing. Yumei Wen: conceptualization, supervision. Yanling Wu: supervision, writing-review & editing. Liying Ma: conceptualization, investigation, resources, software, supervision, writing-review & editing. Tianlei Ying: conceptualization, funding acquisition, investigation, project administration, supervision, writing-review & editing.

Conflict of interest

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

Acknowledgments

This work was supported by grants from the National Key R&D Program of China (2019YFA0904400), National Natural Science Foundation of China (81822027, 81630090, 81902108), and Science and Technology Commission of Shanghai Municipality (20DZ2254600, 20DZ2261200).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.virs.2022.02.010.

References

Bonsignori, M., Zhou, T., Sheng, Z., Chen, L., Gao, F., Joyce, M.G., Ozorowski, G., Chuang, G.Y., Schramm, C.A., Wiehe, K., Alam, S.M., Bradley, T., Gladden, M.A., Hwang, K.K., Iyengar, S., Kumar, A., Lu, X., Luo, K., Mangiapani, M.C., Parks, R.J., Song, H., Acharya, P., Bailer, R.T., Cao, A., Druz, A., Georgiev, I.S., Kwon, Y.D., Louder, M.K., Zhang, B., Zheng, A., Hill, B.J., Kong, R., Soto, C., Mullikin, J.C., Douek, D.C., Montefiori, D.C., Moody, M.A., Shaw, G.M., Hahn, B.H., Kelsoe, G., Haber, P.T., Korber, B.T., Boyd, S.D., Fire, A.Z., Kepler, T.B., Shapiro, L., Ward, A.B., Mascola, J.R., Liao, H.X., Kwong, P.D., Haynes, B.F., 2016. Maturation pathway from germline to broad HIV-1 neutralizer of a CD4-mimic antibody. Cell 165, 449–463.

Bowers, E., Scamurra, R.W., Asrani, A., Beniguel, L., MaWhinney, S., Keays, K.M., Thurn, J.R., Janoff, E.N., 2014. Decreased mutation frequencies among immunoglobulin G variable region genes during viremic HIV-1 infection. PLoS One 9, e81913.

Brenchley, J.M., Price, D.A., Douek, D.C., 2006. HIV disease: fallout from a mucosal catastrophe? Nat. Immunol. 7, 235–239.

Chen, W., Prabakaran, P., Zhu, Z., Feng, Y., Streaker, E.D., Dimitrov, D.S., 2012. Characterization of human IgG repertoires in an acute HIV-1 infection. Exp. Mol. Pathol. 93, 399–407.

Chen, T., Chen, X., Zhang, S., Zhu, J., Tang, B., Wang, A., Dong, L., Zhang, Z., Yu, C., Sun, Y., Chi, L., Chen, H., Zhai, S., Sun, Y., Lan, L., Zhang, X., Xiao, J., Bao, Y., Wang, Y., Zhang, Z., Zhao, W., 2021. The genome sequence archive family: toward explosive data growth and diverse data types. Dev. Reprod. Biol. S16720229, 163–167.

Chothia, C., Lesk, A.M., 1987. Canonical structures for the hypervariable regions of immunoglobulins. J. M. Biol. 196, 901–917.

Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., 1989. Conformations of immunoglobulin hypervariable regions. Nature 342, 877–883.

Chu, Y.W., Marin, E., Fuleihan, R., Ramesh, N., Rosen, F.S., Geha, R.S., Insel, R.A., 1995. Somatic mutation of human immunoglobulin V genes in the X-linked HyperIgM syndrome. J. Clin. Invest. 95, 1389–1393.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. Academic press.

Davis, M.M., Lyons, D.S., Altman, J.D., McHeyzer-Williams, M., Hampl, J., Boniface, J.J., Chien, Y., Nossal, Zinkernagel, Kirberg, 1997. T Cell Receptor Biochemistry, Repertoire Selection And General Features Of Tcr And Ig Structure, pp. 94–100.

De Genst, E., Silence, K., Decanniere, K., Conrath, K., Loris, R., Kinne, J., Muyldermans, S., Wyns, L., 2006. Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. Proc. Natl. Acad. Sci. U.S.A. 103, 4586–4591.

de los Rios, M., Criscitiello, M.F., Smider, V.V., 2015. Structural and genetic diversity in antibody repertoires from diverse species. Curr. Opin. Struct. Biol. 33, 27–41.

De Muth, J.E., 2014. Basic Statistics and Pharmaceutical Statistical Applications. CRC Press.

DeKosky, B.J., Ippolito, G.C., Deschner, R.P., Lavinder, J.J., Wine, Y., Rawlings, B.M., Varadarajan, N., Giesecke, C., Dörner, T., Andrews, S.F., Wilson, P.C., Hunicke-Smith, S.P., Willson, C.G., Ellington, A.D., Georgiou, G., 2013. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. Nat. Biotechnol. 31, 166–169.

Desmyter, A., Transue, T.R., Ghahroudi, M.A., Thi, M.-H.D., Poortmans, F., Hamers, R., Muyldermans, S., Wyns, L., 1996. Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. Nat. Struct. Biol. 3, 803.

Doria-Rose, N.A., Schramm, C.A., Gorman, J., Moore, P.L., Bhiman, J.N., DeKosky, B.J., Ernandes, M.J., Georgiev, I.S., Kim, H.J., Pancera, M., Staupe, R.P., Altae-Tran, H.R., Bailer, R.T., Crooks, E.T., Cupo, A., Druz, A., Garrett, N.J., Hoi, K.H., Kong, R., Louder, M.K., Longo, N.S., McKee, K., Nonyane, M., O'Dell, S., Roark, R.S., Rudicell, R.S., Schmidt, S.D., Sheward, D.J., Soto, C., Wibmer, C.K., Yang, Y., Zhang, Z., Mullikin, J.C., Binley, J.M., Sanders, R.W., Wilson, I.A., Moore, J.P., Ward, A.B., Georgiou, G., Williamson, C., Abdool Karim, S.S., Morris, L., Kwong, P.D., Shapiro, L., Mascola, J.R., 2014. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature 509, 55–62.

D'Angelo, S., Ferrara, F., Naranjo, L., Erasmus, M.F., Hraber, P., Bradbury, A.R.M., 2018. Many routes to an antibody heavy-chain CDR3: necessary, yet insufficient, for specific binding. Front. Immunol. 9, 395.

Ehrenmann, F., Giudicelli, V., Duroux, P., Lefranc, M.-P., 2011. IMGT/Collier de perles: IMGT standardized representation of domains (IG, TR, and IgSF variable and constant domains, MH and MhSF groove domains). Cold Spring Harb. Protoc. 726–736, 2011.

Galson, J.D., Pollard, A.J., Trück, J., Kelly, D.F., 2014. Studying the antibody repertoire after vaccination: practical applications. Trends Immunol. 35, 319–331.

Geisberger, R., Lamers, M., Achatz, G., 2006. The riddle of the dual expression of IgM and IgD. Immunology 118, 429–437.

Goding, J.W., 1978. Allotypes of IgM and IgD receptors in the mouse: a probe for lymphocyte differentiation. In: Contemporary Topics in Immunobiology. Springer, pp. 203–243.

Hoehn, K.B., Gall, A., Bashford-Rogers, R., Fidler, S.J., Kaye, S., Weber, J.N., McClure, M.O., , SPARTAC Trial Investigators, Kellam, P., Pybus, O.G., 2015. Dynamics of immunoglobulin sequence diversity in HIV-1 infected individuals. Phil. Trans. R. Soc. B. 370, 20140241.

Hoet, R.M., Cohen, E.H., Kent, R.B., Rookey, K., Schoonbroodt, S., Hogan, S., Rem, L., Frans, N., Daukandt, M., Pieters, H., van Hegelsom, R., Neer, N.C., Nastri, H.G., Rondon, I.J., Leeds, J.A., Hufton, S.E., Huang, L., Kashin, I., Devlin, M., Kuang, G., Steukers, M., Viswanathan, M., Nixon, A.E., Sexton, D.J., Hoogenboom, H.R., Ladner, R.C., 2005. Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity. Nat. Biotechnol. 23, 344–348.

Hong, B., Wu, Y., Li, W., Wang, X., Wen, Y., Jiang, S., Dimitrov, D.S., Ying, T., 2018. In-depth analysis of human neonatal and adult IgM antibody repertoires. Front. Immunol. 9, 128.

Hu, X., Zhang, J., Wang, J., Fu, J., Li, T., Zheng, X., Wang, B., Gu, S., Jiang, P., Fan, J., Ying, X., Zhang, J., Carroll, M.C., Wucherpfennig, K.W., Hacohen, N., Zhang, F., Zhang, P., Liu, J.S., Li, B., Liu, X.S., 2019. Landscape of B cell immunity and related immune evasion in human cancers. Nat. Genet. 51, 560–567.

Imkeller, K., Wardemann, H., 2018. Assessing human B cell repertoire diversity and convergence. Immunol. Rev. 284, 51–66.

Jackson, K.J.L., Wang, Y., Collins, A.M., 2014. Human immunoglobulin classes and subclasses show variability in VDJ gene mutation levels. Immunol. Cell Biol. 92, 729–733.

Jacobson, M.A., Khayam-Bashi, H., Martin, J.N., Black, D., Ng, V., 2002. Effect of long-term highly active antiretroviral therapy in restoring HIV-induced abnormal B-lymphocyte function. J. Acquir. Immune Defic. Syndr. 31, 472–477, 1999.

Janeway Jr., C.A., Travers, P., Walport, M., Shlomchik, M.J., 2001. The generation of diversity in immunoglobulins. In: Immunobiology: the Immune System in Health and Disease, fifth ed. Garland Science.

Jiang, N., He, J., Weinstein, J.A., Penland, L., Sasaki, S., He, X.-S., Dekker, C.L., Zheng, N.-Y., Huang, M., Sullivan, M., Wilson, P.C., Greenberg, H.B., Davis, M.M., Fisher, D.S., Quake, S.R., 2013. Lineage structure of the human antibody repertoire in response to influenza vaccination. Sci. Transl. Med. 5, 171ra19.

Khurana, S., Fuentes, S., Coyle, E.M., Ravichandran, S., Davey, R.T., Beigel, J.H., 2016. Human antibody repertoire after VSV-Ebola vaccination identifies novel targets and virus-neutralizing IgM antibodies. Nat. Med. 22, 1439–1447.

Kitaura, K., Yamashita, H., Ayabe, H., Shini, T., Matsutani, T., Suzuki, R., 2017. Different somatic hypermutation levels among antibody subclasses disclosed by a new next-generation sequencing-based antibody repertoire analysis. Front. Immunol. 8, 389.

Kreer, C., Gruell, H., Mora, T., Walczak, A.M., Klein, F., 2020. Exploiting B cell receptor analyses to inform on HIV-1 vaccination strategies. Vaccines 8, 13.

Larimore, K., McCormick, M.W., Robins, H.S., Greenberg, P.D., 2012. Shaping of human germline IgH repertoires revealed by deep sequencing. J. Immunol. 189, 3221–3230.

Laserson, U., Vigneault, F., Gadala-Maria, D., Yaari, G., Uduman, M., Vander Heiden, J.A., Kelton, W., Taek Jung, S., Liu, Y., Laserson, J., Chari, R., Lee, J.-H., Bachelet, I., Hickey, B., Lieberman-Aiden, E., Hanczaruk, B., Simen, B.B., Egholm, M., Koller, D., Georgiou, G., Kleinstein, S.H., Church, G.M., 2014. High-resolution antibody dynamics of vaccine-induced immune responses. Proc. Natl. Acad. Sci. U.S.A. 111, 4928–4933.

LeBien, T.W., Tedder, T.F., 2008. B lymphocytes: how they develop and function. Blood 112, 1570–1580.
Li, Z., Woo, C.J., Iglesias-Ussel, M.D., Ronai, D., Scharff, M.D., 2004. The generation of antibody diversity through somatic hypermutation and class switch recombination. Genes Dev. 18, 1–11.
Liao, H.-X., Lynch, R., Zhou, T., Gao, F., Alam, S.M., Boyd, S.D., Fire, A.Z., Roskin, K.M., Schramm, C.A., Zhang, Z., Zhu, J., Shapiro, L., Mullikin, J.C., Gnanakaran, S., Hraber, P., Wiehe, K., Kelsoe, G., Yang, G., Xia, S.-M., Montefiori, D.C., Parks, R., Lloyd, K.E., Scearce, R.M., Soderberg, K.A., Cohen, M., Kamanga, G., Louder, M.K., Tran, L.M., Chen, Y., Cai, F., Chen, S., Moquin, S., Du, X., Joyce, M.G., Srivatsan, S., Zhang, B., Zheng, A., Shaw, G.M., Hahn, B.H., Kepler, T.B., Korber, B.T.M., Kwong, P.D., Mascola, J.R., Haynes, B.F., 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature 496, 469–476.
MacLeod, D.T., Choi, N.M., Briney, B., Garces, F., Ver, L.S., Landais, E., Murrell, B., Wrin, T., Kilembe, W., Liang, C.-H., Ramos, A., Bian, C.B., Wickramasinghe, L., Kong, L., Eren, K., Wu, C.-Y., Wong, C.-H., Kosakovsky Pond, S.L., Wilson, I.A., Burton, D.R., Poignard, P., Price, M.A., Gilmour, J., Fast, P., Kamali, A., Sanders, E.J., Anzala, O., Allen, S., Hunter, E., Karita, E., Kilembe, W., Lakhi, S., Inambao, M., Edward, V., Bekker, L.-G., 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.
Reddy, S.T., Ge, X., Miklos, A.E., Hughes, R.A., Kang, S.H., Hoi, K.H., Chrysostomou, C., Hunicke-Smith, S.P., Iverson, B.L., Tucker, P.W., Ellington, A.D., Georgiou, G., 2010. Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. Nat. Biotechnol. 28, 965–969.
Safonova, Y., Pevzner, P.A., 2019. De novo inference of diversity genes and analysis of non-canonical V(DD)J recombination in immunoglobulins. Front. Immunol. 10, 987.
Sajadi, M.M., Dashti, A., Rikhtegaran Tehrani, Z., Tolbert, W.D., Seaman, M.S., Ouyang, X., Gohain, N., Pazgier, M., Kim, D., Cavet, G., Yared, J., Redfield, R.R., Lewis, G.K., DeVico, A.L., 2018. Identification of near-pan-neutralizing antibodies against HIV-1 by deconvolution of plasma humoral responses. Cell 173, 1783–1795 e14.
Saragovi, H.U., Fitzpatrick, D., Raktabutr, A., Nakanishi, H., Kahn, M., Greene, M.I., 1991. Design and synthesis of a mimetic from an antibody complementarity-deterinLing region. Science 253, 792–795.
Scamurra, R.W., Miller, D.J., Dahl, L., Abrahamsen, M., Kapur, V., Wahl, S.M., Milner, E.C.B., Janoff, E.N., 2000. Impact of HIV-1 Infection on VH3 gene repertoire of naive human B cells. J. Immunol. 164, 5482–5491.
Setliff, I., McDonnell, W.J., Raju, N., Bombardi, R.G., Murji, A.A., Scheepers, C., Ziki, R., Mynhardt, C., Shepherd, B.E., Mamchak, A.A., Garrett, N., Karim, S.A., Mallal, S.A., Crowe, J.E., Morris, L., Georgiev, I.S., 2018. Multi-donor longitudinal antibody repertoire sequencing reveals the existence of public antibody clonotypes in HIV-1 infection. Cell Host Microbe 23, 845–854 e6.
Smith, K.M., Pottage, L., Thomas, E.R., Leishman, A.J., Doig, T.N., Xu, D., Liew, F.Y., Garside, P., 2000. Th1 and Th2 CD4+ T cells provide help for B cell clonal expansion and antibody synthesis in a similar manner in vivo. J. Immunol. 165, 3136–3144.
Soto, C., Bombardi, R.G., Branchizio, A., Kose, N., Matta, P., Sevy, A.M., Sinkovits, R.S., Gilchuk, P., Finn, J.A., Crowe, J.E., 2019. High frequency of shared clonotypes in human B cell receptor repertoires. Nature 566, 398–402.
Stavnezer, J., Guikema, J.E.J., Schrader, C.E., 2008. Mechanism and regulation of class switch recombination. Annu. Rev. Immunol. 26, 261–292.
Tan, J., Pieper, K., Piccoli, L., Abdi, A., Foglierini, M., Geiger, R., Maria Tully, C., Jarrossay, D., Maina Ndungu, F., Wambua, J., Bejon, P., Silacci Fregni, C., Fernandez-Rodriguez, B., Barbieri, S., Bianchi, S., Marsh, K., Thathy, V., Corti, D., Sallusto, F., Bull, P., Lanzavecchia, A., 2016. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. Nature 529, 105–109.
Tian, X., Li, C., Wu, Y., Ying, T., 2020. Deep mining of human antibody repertoires: concepts, methodologies, and applications. Small Methods 4, 2000451.
Turchaninova, M.A., Davydov, A., Britanova, O.V., Shugay, M., Bikos, V., Egorov, E.S., Kirgizova, V.I., Merzlyak, E.M., Staroverov, D.B., Bolotin, D.A., Mamedov, I.Z., Izraelson, M., Logacheva, M.D., Kladova, O., Plevova, K., Pospisilova, S., Chudakov, D.M., 2016. High-quality full-length immunoglobulin profiling with unique molecular barcoding. Nat. Protoc. 11, 1599–1616.
Walker, L.M., Huber, M., Doores, K.J., Falkowska, E., Pejchal, R., Julien, J.-P., Wang, S.-K., Ramos, A., Chan-Hui, P.-Y., Moyle, M., Mitcham, J.L., Hammond, P.W., Olsen, O.A., Phung, P., Fling, S., Wong, C.-H., Phogat, S., Wrin, T., Simek, M.D., Principal Investigators, P.G., Koff, W.C., Wilson, I.A., Burton, D.R., Poignard, P., 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. Nature 477, 466–470.
Wang, B., DeKosky, B.J., Timm, M.R., Lee, J., Normandin, E., Misasi, J., Kong, R., McDaniel, J.R., Delidakis, G., Leigh, K.E., Niezold, T., Choi, C.W., Viox, E.G., Fahad, A., Cagigi, A., Ploquin, A., Leung, K., Yang, E.S., Kong, W.-P., Voss, W.N., Schmidt, A.G., Moody, M.A., Ambrozak, D.R., Henry, A.R., Laboune, F., Ledgerwood, J.E., Graham, B.S., Connors, M., Douek, D.C., Sullivan, N.J., Ellington, A.D., Mascola, J.R., Georgiou, G., 2018. Functional interrogation and mining of natively paired human VH:VL antibody repertoires. Nat. Biotechnol. 36, 152–155.
Weinstein, J.A., Jiang, N., White, R.A., Fisher, D.S., Quake, S.R., 2009. High-throughput sequencing of the zebrafish antibody repertoire. Science 324, 807–810.
Wu, Y.-C., Kipling, D., Leong, H.S., Martin, V., Ademokun, A.A., Dunn-Walters, D.K., 2010. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. Blood 116, 1070–1078.
Wu, Y.-C.B., Kipling, D., Dunn-Walters, D.K., 2012. Age-related changes in human peripheral blood IGH repertoire following vaccination. Front. Immunol. 3, 193.
Xu, J.L., Davis, M.M., 2000. Diversity in the CDR3 region of VH is sufficient for most antibody specificities. Immunity 13, 37–45.
Yin, L., Hou, W., Liu, L., Cai, Y., Wallet, M., Gardner, B., Chang, K., Lowe, A., Rodriguez, C., Sriaroon, P., Farmerie, W., Sleasman, J., Goodenow, M., 2013. IgM repertoire biodiversity is reduced in HIV-1 infection and systemic lupus erythematosus. Front. Immunol. 4, 373.
Zhu, J., Paul, W.E., 2008. CD4 T cells: fates, functions, and faults. Blood 112, 1557–1569.
Zhu, J., O'Dell, S., Ofek, G., Pancera, M., Wu, X., Zhang, B., Zhang, Z., 2012. Somatic populations of PGT135–137 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinformatics. Front. Microbiol. 3, 315.
Zwick, M.B., Komori, H.K., Stanfield, R.L., Church, S., Wang, M., Parren, P.W.H.I., Kunert, R., Katinger, H., Wilson, I.A., Burton, D.R., 2004. The long third complementarity-determining region of the heavy chain is important in the activity of the broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2F5. J. Virol. 78, 3155–3161.