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, Lijing Ma b,*, 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 significantly 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).
† 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
1995-820X/© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
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; Sellif 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/mm³, 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).

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 (IGH) 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.cnbc.ac.cn/omix/releas e/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 / 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

| Table 1 | Descriptive characteristics of HIV-1-infected patients (n = 10). |
|---------|---------------------------------------------------------------|
| Group   | Patients | Viral load (copies/mL) | CD4 (cells/mm³) | Age (y) | Year of infection |
|---------|----------|------------------------|-----------------|---------|------------------|
| GX03    | GX2016EU03 | 51800 | 934 | 45 | 16 |
| HC06    | X2016EU06 | 14100 | 437 | 41 | 13 |
| GC13    | X2016EU13 | 42200 | 373 | 45 | 13 |
| MD-05   | X2016EU16 | 4150 | 389 | 44 | 5 |
| BX2016EU01 | 14700 | 806 | 42 | 5 |
| BX2016EU01 | 27100 | 317 | 46 | 6 |
| BX2015EU19 | 7790 | 311 | 36 | 5 |
| BX2015EU02 | 4150 | 596 | 31 | 5 |
| NG-02   | GX2016EU06 | 17800 | 517 | 29 | 1 |
| GX2016EU22 | 45100 | 489 | 39 | 2 |
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’Angeolo 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. 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 of 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; Lariumore 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 IGKYV1, IGKYV2, IGKYV3, and IGKYV4 were preferentially used, while the IGKYV5, IGKYV6, and IGKYV7 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 IGKYV3 and IGKYV4 in the HIV-IgG repertoire (CDR1: 34.38%; IGKYV4: 29.57%) were both lower than in the HIV-IgM repertoire (IGHV: 41.20%; IGKYV4: 42.63%), while the utilization of IGKYV1 and IGKYV2 in HIV-IgG (IGHV1: 18.92%; IGKYV2: 12.91%) was higher than in the HIV-IgM repertoire (IGHV1: 12.04%; IGKYV2: 2.98%) (Table 4). The most frequently used three IGHV genes in HIV-IgM were IGKYV4-59 > IGKYV4-34 > IGKYV4-39, and the condition in HIV-IgG was IGKYV4-59 > IGKYV2-5 > IGKYV3-30.

The VDJ gene usage in the HIV-IgM repertoire was also differentiated from HH repertoires. Although IGKYV4 and IGKYD3 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: IGKYV4-59 > IGKYV4-34 > IGKYV4-39 and IGKYD3-10 > IGKYD6-13 > IGKYD3-22; HH repertoire: IGKYV4-59 > IGKYV1-69 > IGKYV4-34 and IGKYD3-10 > IGKYD3-22 > IGKYD1-26, 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.

| CDR   | HIV-IgM     | HIV-IgG     | P     | t      | d     | 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.

| CDR   | HIV-IgM     | HH          | P     | t      | d     | 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.
when only the IGHL6 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 IGHL gene usages, especially for IGHL6.

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

| Subgroups | Repertoire | HIV-IgM (%) | HIV-IgG (%) | HH (%) | OR 95% CI * | 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   |             |                  |               |
| IGHV1D    |            | 9.20        | 8.06        | 12.07  | 0.939 (0.938, 0.939) | 0.976 (0.975, 0.977) |
| IGHV2D    |            | 18.41       | 19.76       | 19.94  |             |                  |               |
| IGHV3D    |            | 39.99       | 45.50       | 34.10  |             |                  |               |
| IGHV4D    |            | 3.94        | 3.41        | 5.89   |             |                  |               |
| IGHV5D    |            | 6.99        | 7.55        | 7.40   |             |                  |               |
| IGHV6D    |            | 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.

*Calculated by student’s t-test.

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...
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. 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.
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 antibodies specifically bind to the 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 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 cloned and 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 (Tian et al., 2016). Future studies with the light chain involved are needed to mine the native 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.cnjm.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, 81902108), and Science and Technology Commission of Shanghai Municipality (20DZ2261200, 20DZ2254600, 20DZ2261200).

Appendix A Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vjs.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., Wiebe, K., Alam, S.M., Bradley, T., Ghadgen, M.A., Hwang, K.K., Iyengar, S., Kumar, A., Lu, X., Luo, K., Mangiapiani, M.C., Parks, R.J., Song, H., Acharya, P., Bailey, R.T., Cao, A., Druz, A., Georgiev, I., 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., Keloce, G., Hraber, P.T., Korber, B.T., Boyd, S.D., Fire, A.Z., Kepler, T.B., Shapiro, L., Ward, A.B., Mascola, J.R., Luo, X., Kwong, P.D., Haynes, B.F., 2016. Maturation pathway features a germ line to broaden HIV-1 neutralizer a CD4+ mimic antibody. Cell 165, 449–463.

Bowers, E., Scamurra, R.W., Arzani, A., Beniguel, I., MWhinney, S., Kenys, 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 immunodeficiency syndrome. J. Clin. Invest. 95, 1389–1393.

Brenchley, J.M., Price, D.A., Douek, D.C., 2006. Shaping of human immunoglobulin V genes in the X-linked HyperIgM syndrome. J. Clin. Invest. 95, 1389–1393.

Brenchley, J.M., Price, D.A., Douek, D.C., 2006. Somatic mutation of human immunoglobulin V genes in the X-linked HyperIgM syndrome. Nature 342, 877–883.

Brenchley, J.M., Price, D.A., Douek, D.C., 2006. Exploding data growth and diverse data types. Dev. Reprod. Biol. S16720229, –167.

Brenchley, J.M., Price, D.A., Douek, D.C., 2006. Syndr. 31, 472–477, 1999.

Janeway Jr., C.A., Travers, P., Walport, M., 2001. The generation of antibody responses. Immunol. 9, 128.

Janeway Jr., C.A., Travers, P., Walport, M., 2001. The generation of antibody responses. Immunol. 9, 128.

Janeway Jr., C.A., Travers, P., Walport, M., Shlomchik, M.J., 2001. The generation of antibody responses. Immunol. 9, 128.

Janeway Jr., C.A., Travers, P., Walport, M., Shlomchik, M.J., 2001. The generation of antibody responses. Immunol. 9, 128.

Janeway Jr., C.A., Travers, P., Walport, M., Shlomchik, M.J., 2001. The generation of antibody responses. Immunol. 9, 128.

Janeway Jr., C.A., Travers, P., Walport, M., Shlomchik, M.J., 2001. The generation of antibody responses. Immunol. 9, 128.
X. Tian et al.

Virologica Sinica 37 (2022) 370–379

LeLenn, T.W., Tedder, T.F., 2008. B lymphocytes: how they develop and function. Blood 112, 1570–1580.

Li, Z., Woo, C.J., Iglesias-Usiel, 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., Katinger, H., Zhao, X., Pieper, K., Abdi, A., Fogliani, M., Geiger, R., Maria Tully, C., Jarrossay, D., Maina Ndungu, F., Wambua, J., Bejon, P., Bi, S., Fregni, P., Fernandez-Rodriguez, B., Barbieri, S., Bianchi, S., Marsh, K., Tshayi, V., Corti, D., Sallustio, F., Bull, P., Lanzavecchia, A., 2016. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. Nature 529, 105–109.

Tan, X., Li, C., Wu, Y., Ying, T., 2020. Deep mining of human antibody repertoires: concepts, methodologies, and applications. Small Methods 4, 2000045.

Turchaninova, M.A., Davydov, A., Britney, N.M., Briney, B., Garces, F., Ver, L.S., Landais, E., Murrell, B., Chrysostomou, C., Hunicke-Lano, H.-X., Lynch, R., Zhou, T., Gao, F., Alam, S.M., Boyd, S.D., Fire, A.Z., Roskin, K.M., 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. Annu. Rev. Immunol. 26, 261–292.

Stavnezer, J., Guikema, J.E.J., Schrader, C.E., 2008. Mechanism and regulation of class switch recombination. Annu. Rev. Immunol. 26, 261–292.