



# cAb-Rep: A Database of Curated Antibody Repertoires for Exploring Antibody Diversity and Predicting Antibody Prevalence

Yicheng Guo<sup>1</sup>, Kevin Chen<sup>2</sup>, Peter D. Kwong<sup>3,4</sup>, Lawrence Shapiro<sup>1,3,4</sup> and Zizhang Sheng<sup>1\*</sup>

<sup>1</sup> Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, United States, <sup>2</sup> College of Arts and Science, Stony Brook University, Stony Brook, NY, United States, <sup>3</sup> Vaccine Research Center, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, MD, United States, <sup>4</sup> Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, United States

## OPEN ACCESS

### Edited by:

Deborah K. Dunn-Walters,  
University of Surrey, United Kingdom

### Reviewed by:

Duane R. Wesemann,  
Brigham and Women's Hospital and  
Harvard Medical School,  
United States

Claude-Agnes Reynaud,  
Institut National de la Santé et de la  
Recherche Médicale  
(INSERM), France

### \*Correspondence:

Zizhang Sheng  
zs2248@columbia.edu

### Specialty section:

This article was submitted to  
B Cell Biology,  
a section of the journal  
Frontiers in Immunology

**Received:** 14 August 2019

**Accepted:** 20 September 2019

**Published:** 09 October 2019

### Citation:

Guo Y, Chen K, Kwong PD, Shapiro L  
and Sheng Z (2019) cAb-Rep: A  
Database of Curated Antibody  
Repertoires for Exploring  
Antibody Diversity and Predicting  
Antibody Prevalence.  
*Front. Immunol.* 10:2365.  
doi: 10.3389/fimmu.2019.02365

The diversity of B cell receptors provides a basis for recognizing numerous pathogens. Antibody repertoire sequencing has revealed relationships between B cell receptor sequences, their diversity, and their function in infection, vaccination, and disease. However, many repertoire datasets have been deposited without annotation or quality control, limiting their utility. To accelerate investigations of B cell immunoglobulin sequence repertoires and to facilitate development of algorithms for their analysis, we constructed a comprehensive public database of curated human B cell immunoglobulin sequence repertoires, cAb-Rep (<https://cab-rep.c2b2.columbia.edu>), which currently includes 306 immunoglobulin repertoires from 121 human donors, who were healthy, vaccinated, or had autoimmune disease. The database contains a total of 267.9 million V(D)J heavy chain and 72.9 million VJ light chain transcripts. These transcripts are full-length or near full-length, have been annotated with gene origin, antibody isotype, somatic hypermutations, and other biological characteristics, and are stored in FASTA format to facilitate their direct use by most current repertoire-analysis programs. We describe a website to search cAb-Rep for similar antibodies along with methods for analysis of the prevalence of antibodies with specific genetic signatures, for estimation of reproducibility of somatic hypermutation patterns of interest, and for delineating frequencies of somatically introduced *N*-glycosylation. cAb-Rep should be useful for investigating attributes of B cell sequence repertoires, for understanding characteristics of affinity maturation, and for identifying potential barriers to the elicitation of effective neutralizing antibodies in infection or by vaccination.

**Keywords:** antibody prevalence, antibody repertoire, antibodyomics, B cell response, gene-specific substitution profile, next-generation sequencing, post-translational modification, somatic hypermutation (SHM)

## INTRODUCTION

B cells comprise a crucial component of the adaptive immune response (1). B cells recognize three-dimensional epitopes of antigens through the variable domains of the B cell receptor (BCR), or its various secreted forms of antibody. The variable domains of BCRs and antibodies are composed of immunoglobulin heavy and light chains, encoded by separate genes.

Through V(D)J gene recombination and somatic hypermutation (SHM), a high level of sequence diversity is introduced to the variable domain (1–4), allowing B cells to recognize diverse antigens. Thus, an interrogation of BCR diversity and function is key to understanding the B cell immune response. The application of next-generation sequencing (NGS) to BCR repertoires provides snapshots of BCR diversity, and such studies in the past decade have characterized numerous features of B cell responses to infection, immunization, and autoimmune disease (3, 5–18). Databases, programs, and websites have been developed to store, process, and annotate BCR repertoire data (17, 19–24) [software reviewed in Chaudhary and Wesemann (25)]. Nonetheless, most repertoire data are deposited in public databases in the format of raw NGS reads, which contain both sequencing duplicates and sequencing errors and may be difficult to annotate. A comprehensive database of curated and well-annotated BCR transcripts, should therefore accelerate repertoire studies including but not limited to characterization of B cell receptor diversity, mechanisms of clonal expansion, development of BCR repertoire analysis algorithms, estimation of frequency of antigen-specific antibodies and their precursor-like cells, and effects of SHM. In addition, such a database should assist researchers in performing repertoire-related data mining.

Designing vaccines that can elicit broadly neutralizing antibodies (bnAbs) is a long-term goal for preventing infections from fast evolving and/or highly diversified pathogens such as HIV-1 and influenza (26, 27). Studies have revealed that some epitopes could elicit bnAbs in different individuals with similar modes of recognition and shared genetic signatures [e.g., V(D)J gene origin, complementarity determining region 3 (CDR3) length, CDR3 motifs, SHM] (9, 28–33), defined as convergent, multidonor, or public bnAb classes. Such antibody classes often target conserved epitopes; those that can elicit the activation and maturation of naïve B cells with bnAb potential might provide templates for universal vaccines (30, 34–36). However, not all antibody classes appear with high frequencies in BCR repertoires; those that do not may be limited due to disfavored developmental steps or “roadblocks.” Thus, the identifications of bnAb classes with precursor-like cells prevalent in humans is a critical consideration for determining which template antibodies are good targets for elicitation, and thus for immunogen design (9, 37, 38). A comprehensive database of curated B cell repertoire transcripts, for which uncertainties and variations in B cell repertoires (e.g., sampling bias, infection history, aging, and genetic diversification) (12) have been minimized, would be helpful for predicting antibody class prevalences.

A database of curated BCR repertoires could, moreover, be used to investigate SHM preference and mechanisms of affinity maturation. Antibodies accumulate mutations with high preference determined by both intrinsic gene mutability and functional selection (1, 39). We previously developed gene-specific substitution profiles (GSSPs) to characterize positional substitution types and frequencies in 69 human V genes (14). By incorporating additional BCR transcripts, GSSPs could be built for more genes, and this could have broad application;

GSSPs or similar approaches have been applied to examine whether bnAbs mature with shared pathways (40), to identify highly frequent SHMs and common mechanisms of affinity modulation (41, 42), and to estimate whether rare SHMs (SHMs generated with very low frequency by the SHM machinery) in bnAbs could form barriers to re-elicitation by vaccination (34, 43).

In this study, we constructed a database of curated human B cell immunoglobulin sequence repertoires (cAb-Rep) from 306 high quality human repertoires, and developed methods to search cAb-Rep using either sequence or sequence signature. We used the database to construct GSSPs for 102 human antibody V genes and developed a script to identify rare SHMs in input sequences. We evaluated the robustness of query results relative to the number of repertoires, using antibody prevalence estimates as a test case. In summary, the database developed here should help to investigate B cell repertoire features, to find antibody templates for vaccine design, and overall to understand mechanisms of antibody development.

## MATERIALS AND METHODS

### BCR Repertoire Dataset

We assembled a total of 376 BCR repertoire datasets from 108 human donors deposited in the NCBI short reads archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra>). These repertoires were all sequenced using illumina MiSeq or HiSeq and were published in 11 studies (Table 1). In addition, annotated full length sequences of three donors (sequenced by AbHelix with high depth or HD repertoires,  $\sim 10^8$  high quality sequences per donor) (4) and the near full length V(D)J sequences (sequenced with framework 1 region primers) from 10 HD repertoires (2) were downloaded from the links provided by the authors.

### Next-Generation Sequencing Data Processing

The NGS data were analyzed with the SONAR pipeline, version 2.0 (<https://github.com/scharch/sonar/>) developed in our lab (23). Briefly, USEARCH was used to merge the  $2 \times 300$  or  $2 \times 250$  raw reads to single transcripts and to remove transcripts potentially containing more than 20 miscalls calculated from sequencing quality score (44). Merged transcripts shorter than 300 nucleotides were removed. BLAST (<http://www.ncbi.nlm.nih.gov/blast/>) was used to assign germline V, D, and J genes to each transcript with customized parameters (23, 45). CDR3 was identified from BLAST alignment using the conserved 2nd Cysteine in V region and WGXXG (heavy chain) or FGXXG (light chain) motifs in J region (X represents any of the 20 amino acids). To assign an isotype for each heavy chain transcript, we used BLAST with default parameters to search the 3' terminus of each transcript against a database of human heavy chain constant domain 1 region obtained from the international ImMunoGeneTics information system (IMGT) database. A BLAST E-value threshold of  $1E-6$  was used to find significant isotype assignments. Then, sequences other than the V(D)J region of a transcript were removed

**TABLE 1** | Summary of donor type, library preparation and sequencing methods, BCR repertoire statistics, and references.

Type of donors	Experimental methods	Total repertoires	Unique donors	Average no. raw reads	Average no. unique reads	Average no. clones	No. of studies
Healthy	5' RACE/FR1 primer MiSeq	59	31	1,247,695	13,432	7,142	5
High-depth healthy donor	5' RACE/FR1 primer MiSeq/HiSeq	13	13	43,984,663	25,632,108	N.D.	2
Hepatitis B prime vaccination	FR1 primer MiSeq	88	9	291,249	7,677	5,648	1
Hepatitis B Boost vaccination	FR1 primer MiSeq	18	9	338,467	26,570	21,936	1
Influenza vaccination	5' RACE/FR1 primer MiSeq	81	46	742,531	16,029	7,016	3
Multiple sclerosis brain	5' RACE MiSeq	17	4	1,654,572	34,326	11,130	1
Myasthenia gravis	5' RACE MiSeq	18	9	862,543	17,512	10,332	1
Systemic lupus erythematosus	5' RACE MiSeq	8	7	3,741,944	410,286	41,705	1
Tetanus vaccination	5' RACE MiSeq	4	4	1,999,230	58,787	17,309	1
Total	5' RACE, FR1 primer MiSeq/HiSeq	306	121	818,868,037	340,841,950	2,655,416	11

and transcripts containing frame-shift and/or stop codon were excluded.

We then applied two approaches to remove PCR duplicates and sequencing errors. For datasets sequenced with unique molecular identifier (UMI), we first grouped merged raw transcripts having identical UMI. Due to PCR crossover and UMI collisions, transcripts in a group may originate from different cells (2, 12). We therefore clustered the raw transcripts in each group using USEARCH with a 97% sequence identity. All transcripts in a USEARCH cluster were aligned using CLUSTALO (46) and a consensus sequence was generated. To further reduce redundant transcripts derived from multiple mRNA molecules of the same cell, we removed duplicates in the consensus sequences. Singletons, which have neither UMI duplicates nor consensus duplicates, were removed. SONAR was used to annotate the unique transcripts. For datasets sequenced without UMI, similar to our previous studies (14, 18, 23), we clustered transcripts of each repertoire using USEARCH with sequence identity of 0.99, and selected only one transcript with the highest sequencing depth or numbers of duplicates per cluster for later analyses. To further remove low quality reads, we excluded clusters with size <2. Finally, a unique dataset of transcripts was generated for each repertoire. To reduce artifacts and effect of small sample size on the comparisons of features of antibody repertoires, we excluded repertoires containing fewer than 1,700 unique transcripts.

For each of the 13 HD repertoires curated in previous studies, we further removed duplicates, sequences containing frameshifts or/and stop codons, and corrected annotation errors to form a

unique dataset. The gene annotation information was retrieved from the annotation files provided by the studies.

## Germline Gene Database and Prediction of New Gene/Alleles

The human germline gene database from IMGT was used to assign germline V(D)J genes. To detect new germline genes or alleles, we combined unique reads from 108 repertoires, and used IgDiscover v0.9 with default parameters to identify potential new germline genes or alleles that are observed in multiple repertoires (47). While IgDiscover prefers to identify new genes or alleles from IgM repertoire, we used repertoires containing all isotypes. Nonetheless, identical unique reads were observed for each predicted gene/allele in at least two donors, suggesting that the predictions are still reliable. The predicted genes were submitted to European Nucleotide Archive (ENA) with project accession numbers: PRJEB31020. For each of the 13 HD repertoires, we randomly selected 1 million IgM transcripts for novel gene prediction (as recommended by IgDiscover manual) and no new gene or allele was found.

## V and J Gene Usage and Antibody Position Numbering

The unique dataset for each repertoire was used to calculate the distributions of germline gene usage and SHM levels. For each transcript, we used ANARCI to assign each position according to the Kabat, Chothia, and IMGT numbering schemes (48).

## Signature Prevalence and Rarefaction Analysis

We developed a python script, `Ab_search.py`, to search the BCR repertoire database in two modes: signature motif and full V(D)J sequence. Briefly, in the signature searching mode, an amino acid signature motif from either CDR3 region or other positions of interest (defined with the Kabat, Chothia, and IMGT numbering scheme) is converted to python regular expression format and searched against all transcripts in the database. For the sequence searching mode, BLAST+ is called to find similar transcripts with  $E$ -value  $< 1E-6$  (49). Signature frequency in each repertoire is calculated by dividing the number of matched transcripts with the total number of unique transcripts originated from the same germline gene or allele.

To evaluate the effect of number of repertoires on estimation of signature frequency, we performed rarefaction analysis by sampling  $i$  subset repertoires from PBMC samples (1 to 35). For each subset  $i$ , the random sampling was performed  $N$  times (we used  $N = 20$  in the current study) and the mean signature frequency from each sampling  $\bar{f}_i$  was calculated. Then the coefficient of variance for each  $i$ ,  $CV_i$ , was calculated as:

$$CV_i = \sqrt{\frac{\sum_{j=1}^N (\bar{f}_{ij} - \frac{\sum_{j=1}^N \bar{f}_{ij}}{N})^2}{N-1}}{\frac{\sum_{j=1}^N \bar{f}_{ij}}{N}} \quad \bar{f}_i = \frac{\sum_{j=1}^N \bar{f}_{ij}}{i}$$

## Identification of Antibody Clones and Construction of GSSPs

For each repertoire, we first sorted all transcripts to groups based on identical V and J gene usages. For each group, transcripts with 90% CDR3 sequence identity and the same CDR3 length were clustered into clones using USEARCH. One representative sequence was selected in each clone and representative transcripts from all repertoires were combined to build GSSPs for V genes using mGSSP (14). We exclude GSSPs built with  $< 100$  clones because of lack of information (14). A python script, `SHM_freq.py`, was developed to identify SHMs in an input sequence, to search the GSSP of an assigned V gene, and to output the rarity of each mutation, calculated as:  $\text{Rarity} = (1 - \text{Frequency of mutation}) * 100\%$ .

## Construction of Gene-Specific N-Glycosylation Profile

Glycosylation sites for sequences having SHMs  $> 1\%$  in the unique datasets of healthy and vaccination donors were predicted using an artificial neural network method NetNGlyc v1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>)<sup>1</sup>. Predictions with high specificity (potential  $> 0.5$  and jury agreement of 9/9) and do not match Asn-Pro-Ser/Thr motifs were included. N-glycosylation sites encoded in germline genes were excluded.

<sup>1</sup> Gupta R, Jung E, Brunak S. Prediction of N-glycosylation sites in human proteins. Unpublished manuscript (2004).

## Statistical Tests

ANOVA and  $T$ -tests were performed in R.

## RESULTS

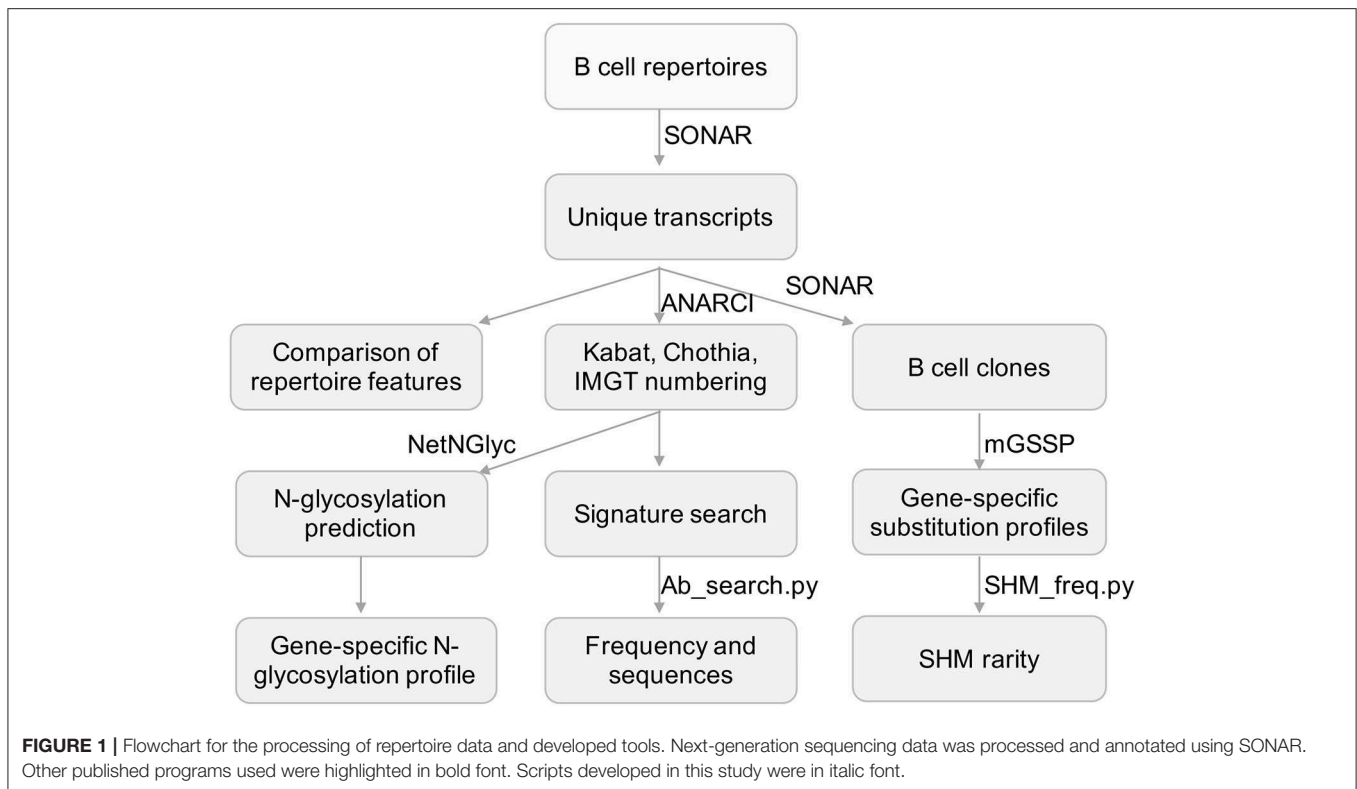
### cAb-Rep Contains a Diverse Compendium of Annotated Next-Generation Sequencing Datasets

We first assembled 376 BCR repertoire deep sequencing data sets from NCBI SRA database, each sequenced by Illumina MiSeq or HiSeq and with library preparation protocols to cover full length V(D)J region [5' primers at leader regions or 5' Rapid amplification of cDNA ends (RACE), 3' primers targeting constant region 1] or near full length (5' primers at N-terminus of framework 1 region). These datasets contain a total of 247 million raw reads (Table 1 and Table S1). We performed quality control including removing sequencing errors, PCR crossover, PCR duplicates, and annotated each sequence using SONAR (See Methods and Figure 1). A total of 293 repertoires from 108 donors were selected to construct the cAb-Rep database, with 6.6 and 0.9 million curated full length V(D)J heavy and light chain transcripts. Twenty-five thousand five hundred forty-one curated transcripts were obtained per repertoire. In addition, annotated sequences from 13 healthy donor repertoires sequenced with high depth (more than 20 million unique sequences from about  $10^7$ - $10^8$  B cells per donor) (2, 4), were curated and incorporated.

These repertoires were from studies including Hepatitis B vaccination, influenza vaccination, Tetanus vaccination, Multiple Sclerosis, Myasthenia Gravis, Systemic Lupus Erythematosus, and healthy donors [Table 1 and Table S1; (6–9, 12, 15–17, 50)]. Among all repertoires, 72 repertoires were from 43 healthy donors (35.5% of total donors). The database repertoires were obtained from three tissues: 2 from brain, 15 from cervical lymph node, and 289 from peripheral blood mononuclear cell (PBMC). Among PBMC samples, 16 were from antibody secreting B cells, 27 from Hepatitis B surface antigen (HBsAg)+ B cells, 8 from Human Leukocyte Antigen – DR isotype (HLA-DR)<sup>+</sup> plasma cells, 139 from IgG<sup>+</sup> B cells, 20 from memory B cells, 34 from IgM or naïve B cells, and 52 from whole PBMC (Table S1). Thus, cAb-Rep contains BCR repertoire data in different settings, which will enable cross-study comparisons.

For each transcript in cAb-Rep, we numbered positions using Kabat, Chothia, and IMGT schemes (Figure 1), and annotated gene origin, CDR3 sequence and length, somatic hypermutation level, isotype, donor, and clonotype. These curated datasets can be used for downstream analyses. The transcripts are stored in FASTA format with annotation information in the header line of each transcript. Such a format can be easily altered to feed into other programs for repertoire analysis.

Because the 13 HD repertoires contained over 330 million unique sequences, to avoid sampling bias for profile constructions and improve speed of searching cAb-Rep (see sections below), we randomly selected 20,000 unique sequences of each isotype of each repertoire and combined with the rest of repertoires for all analyses below. Nonetheless, in case rare



events of BCR signatures are of interests, we provided an option in our scripts and at cAb-Rep website to search the 13 HD repertoires alone.

## Ab\_Search Identifies Similar Antibody Transcripts and Estimates Signature Frequency

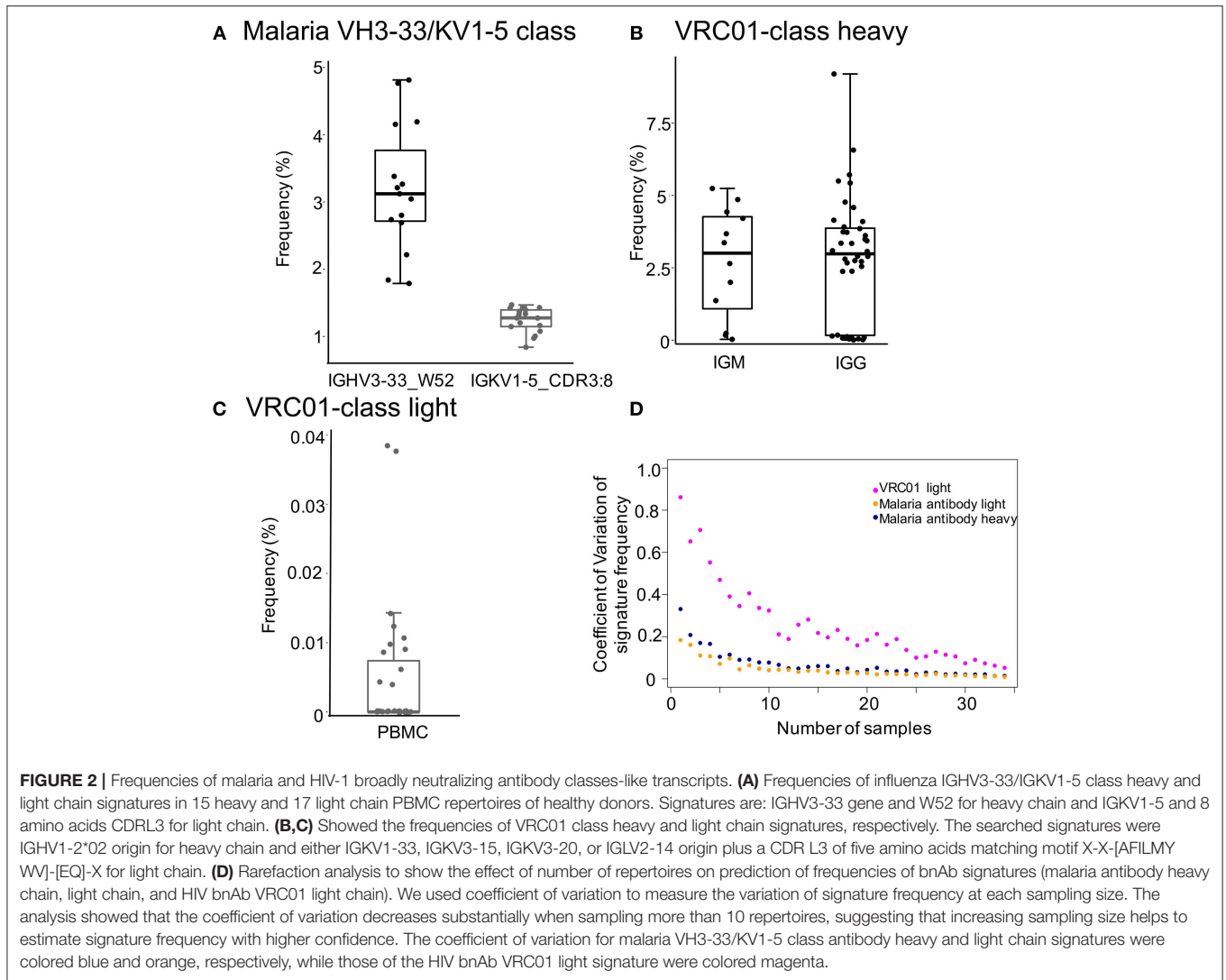
To identify antibody transcripts of interest from cAb-Rep, we developed a python script, *Ab\_search.py*. The script accepts sequences or amino acid motifs as input and output frequency of the signature and matched sequences (See Materials and Methods). Here, as examples, we searched transcripts having signature motifs of a selected malaria antibody class in PBMC BCR repertoires or of a selected HIV-1 bnAb classes in IgM and IgG repertoires (**Figure 2**).

The selected malaria antibody class contains both heavy (IGHV3-33 gene with a tryptophan at position 52) and light chain (IGKV1-5 gene with 8 amino acid CDRL3) signatures (51). The heavy chain signature was observed with a frequency of  $3.2 \pm 0.94\%$  in 15 PBMC repertoires in healthy donors (**Figure 2A**), the light chain signature was found with a frequency of  $1.24 \pm 0.18\%$  in 17 healthy donor PBMC repertoires. All healthy donors we searched contain both heavy and light sequences. By assuming random pairing (52), heavy-light paired antibodies similar to the malaria antibody class could be generated with a high frequency ( $\sim 40$  per million B cells). Thus, vaccine mediated elicitation of such antibodies are promising candidates to enable protection.

The HIV-1 VRC01 class contains both heavy and light chain signatures (31–33, 53). Its heavy chain uses IGHV1-2\*02 allele. The light chain is originated from a limited set of genes, including IGKV1-33, IGKV3-15, IGKV3-20, or IGLV2-14 and contains a CDR3 of five amino acids matching motif X-X-[AFILMYWV]-[EQ]-X. Searching cAb-Rep with these signatures showed that the heavy chain allele for VRC01 appears in IgM and IgG repertoires with similar frequencies ( $2.7 \pm 1.9\%$  and  $2.85 \pm 2.13\%$ , respectively; **Figure 2B**). Further, VRC01 light chain-like transcripts were found in  $\sim 33\%$  of donors, with a mean frequency of  $0.005 \pm 0.0098\%$  in PBMCs (**Figure 2C**). By assuming random pairing of heavy and light chains, our calculation of the mean frequency of VRC01 class-like antibodies was 1.4 per million B cells in healthy donors, close to estimates from previous studies (33, 38).

## Impact of Repertoire Diversity on Signature Prevalence

Due to diversification in BCR repertoires by antigen selection and other factors, the predicted prevalence of transcripts of interest could vary. To evaluate the effect of the number of sampled repertoires on prevalence prediction, we performed rarefaction analysis to sample a subset of repertoires (ranging from 1 to 35) with each subset randomly sampling 20 times to calculate frequencies of the malaria IGHV3-33/IGKV1-5 class and VRC01 class light chain signatures. For each sampling size, we calculated the coefficient of variation (standard deviation/mean) to measure the degree of variation among sampling repeats (**Figure 2D**). For all signatures, our analysis revealed that the coefficient



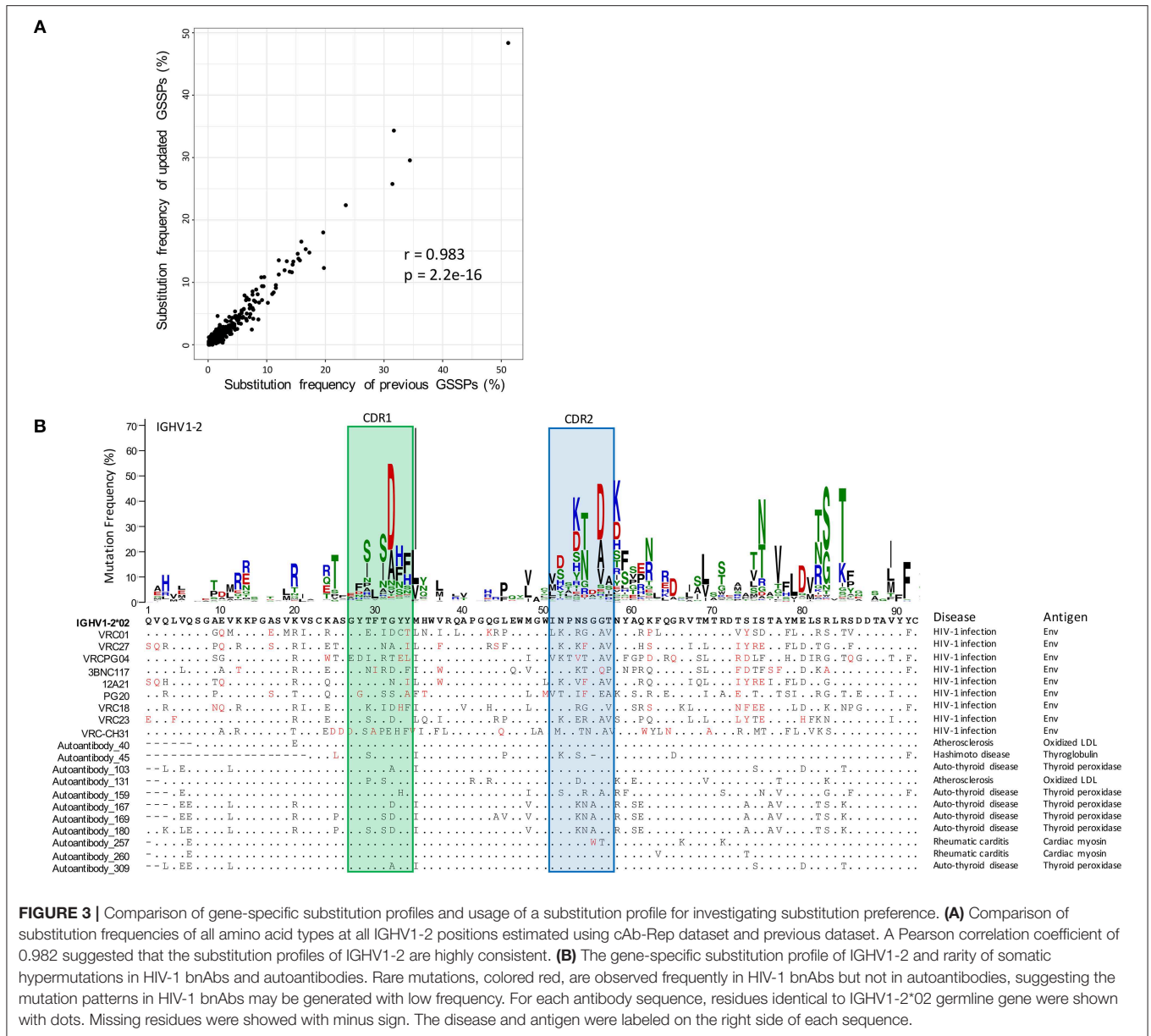
of variation decreases dramatically when the sampling size increases to 10 or more. This suggests that 10 or more repertoires will be optimal to have a consistent estimation of signature frequency.

## Gene-Specific Substitution Profiles and Substitution Frequency Analysis

To investigate substitution preferences in V genes, we predicted new germline genes in the database, clustered transcripts in each repertoire into clones, and selected one representative sequence per clone to build gene-specific substitution profiles (GSSPs) (see Methods). Overall, we identified 5 novel heavy chain alleles, 2 novel lambda chain alleles, and 1 kappa chain allele, each of which was found in two or more donors. By incorporating these germline gene sequences in cAb-Rep sequence annotation, we built GSSPs for 102 human V genes, compared to GSSPs built for 69 genes using three repertoires in our previous study. Our analysis further showed that the GSSPs of the

two studies were highly consistent ( $r = 0.983$  for IGHV1-2 gene, **Figure 3A**).

To facilitate exploring substitution preference, we developed a python script, SHM\_freq.py, to identify mutations in an input sequence, call the GSSP of corresponding V gene, and find the frequency of the mutation being generated by the somatic hypermutation machinery. To demonstrate how this information can be helpful, we analyzed frequencies of substitutions observed in the heavy chain of VRC01 class bnAbs (**Figure 3B**). This analysis showed that all lineages in this class contain over 30% mutations, with ~30% of the mutations being low frequency or rare mutations (frequency <0.5% in IGHV1-2 GSSP). These mutations are generated with low frequency either because they require multiple nucleotide substitutions (14) or are from single substitutions in silent SHM positions (43). Functional studies have shown that some rare mutations are essential for potency and neutralization (54). However, the likelihood of immunogens maturing antibodies to have similar mutations could be low or require longer maturation times. In contrast,



we observed that autoantibodies [e.g., collected from HIV, autoimmune thyroid disease, atherosclerosis, Hashimoto disease, and rheumatic carditis (55–60)] originated from IGHV1-2 genes contain very few rare mutations, suggesting somatic mutations may not provide a barrier to elicitation of these lineages.

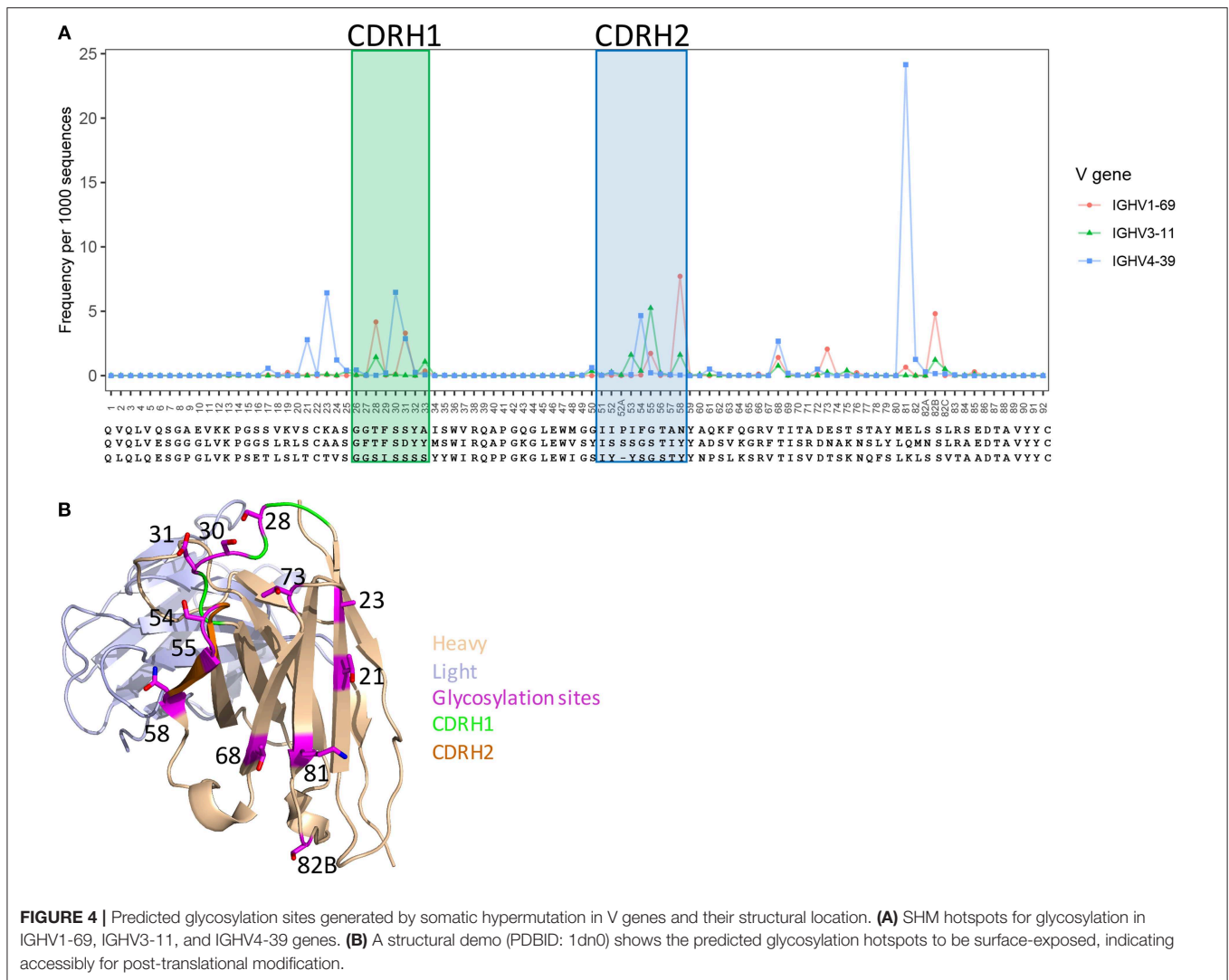
### Gene-Specific N-Glycosylation Profiles (GSNPs)

Post-translation modifications (PTM) (glycosylation, tyrosine sulfation, etc.), which affects antibody functions (42, 61), can be introduced to antibodies by V(D)J recombination and somatic hypermutation processes. To understand the frequency and preference of PTMs generated by somatic hypermutation, as an example, we predicted V-gene-specific frequency of N-glycosylation sequons at each position using healthy and

vaccination donor unique sequences that having more than 1% SHM. Overall, consistent with previous study (42), the predicted N-glycosylation sites were enriched in CDR1, CDR2, and framework 3 regions, but different genes have different hotspots for glycosylation (Figure 4A). Structural analysis showed that the side chains of these hotspot positions to be surface-exposed (Figure 4B), suggesting these sites to be spatially accessible for modification. GSNPs should thus be able to provide information for further experimental validation and investigations of impact of N-glycosylations.

### cAb-Rep Website to Search Frequencies of Signature Motif and SHM

While we developed scripts to search cAb-Rep, these may be of limited utility to users not familiar with programming. Therefore,



we developed a website for searching cAb-Rep (<https://cab-rep.c2b2.columbia.edu/>). The website implements all functions of the scripts we developed above, including querying cAb-Rep using the three signature modes (CDR3, position, BLAST) with specified isotype, numbering scheme, and VJ recombinations, identifying rare SHMs for an input sequence, and showing the GSSP of a V gene (Figure 5). Users can also query GSNP of an input gene as well as download all the datasets in this study.

## DISCUSSION

To understand the mechanisms of BCR diversity generated by V(D)J recombination and somatic hypermutation, high quality BCR repertoire datasets are critical. In this study, we have constructed a database of curated BCR transcripts from previous deep sequencing studies of B cell immunoglobulin sequence repertoires and have demonstrated how this database can provide helpful information for repertoire studies both locally and online. Currently, heavy-light pairing information is not obtained in

cAb-Rep datasets. As new technologies are applied to repertoire sequencing, we believe more paired heavy-light chain transcripts will be incorporated, which will greatly advance functional characterization of BCRs. We will continue to update this database to include more disease conditions as well as those from animal models.

In cAb-Rep, we incorporated BCR datasets sequenced both with and without UMI. Besides filtering out transcripts with low sequencing quality, different approaches were used to identify high quality transcripts. For repertoires sequenced using UMI, we used the UMI information to generate consensus sequences, which have proven effective at removing sequencing errors (62). For BCR repertoires sequenced without UMI, we used a clustering approach, which we first sorted transcripts based on redundancy and selected transcripts with the most redundancy as seeds to cluster transcripts at a given identity cutoff and only used the seeds as high quality transcripts. The assumption was that each B cell contains multiple BCR mRNA molecules and the original BCR could be PCR amplified



**A Estimate prevalence of antibody with signature of interest<sup>2</sup>:**Search Mode:  Chain:  Isotype:  Scheme:  Only for high-depth data set: 

Sequence signature (or fasta sequence for Blast mode):

Signature string or fasta format nucleotide sequences

CDR3 length:  V gene:  J gene:  **Estimate rarity of somatic hypermutation<sup>2</sup>:**Germline V gene:  Rarity for identifying rare SHM(%): 

I. Input a sequence and estimate the rarity of the substitution:

Sequence type:  Select a fasta file:  

II. Input amino acids and find the substitution rarity at a V gene position (Kabat scheme):

Position:  Amino Acid:  **Gene-specific frequency of glycosylation<sup>2</sup>:**Germline V gene: 

Please input a V gene position (Kabat scheme):

Position:  **B Job succeeded!**

The frequency of the input signature is shown below.

Click to download the statistics and BCR transcripts: [Download](#)

Donor_ID	Repertoire_ID	Signature_Frequency_(%)	B_Cell_Type	Donor_condition
95	SRR4026018	0.0099	Memory B cell	Healthy
93	SRR4026019	0.0	Naïve B cell	Healthy
95	SRR4026020	0.0	Naïve B cell	Healthy
95	SRR4026021	0.0096	Memory B cell	Healthy
96	SRR4026022	0.0	Naïve B cell	Myasthenia gravis
96	SRR4026023	0.0646	Memory B cell	Myasthenia gravis
97	SRR4026025	0.0109	Memory B cell	Myasthenia gravis
98	SRR4026026	0.0091	Memory B cell	Myasthenia gravis

**FIGURE 5 |** Querying frequencies of signature and somatic hypermutation at cAb-Rep website. **(A)** cAb-Rep interfaces to estimate frequency of an input signature, gene-specific rarity of somatic hypermutation, and gene-specific glycosylation frequency. **(B)** Result for search of VRC01 bnAb-like light chain sequences using the CDR3 mode. The full results including sequences and donor information can be obtained from the download link, and the frequency of the motif in each donor is shown online. The parameters used are: IGKV3-20 gene and a five amino acid CDR L3 matching [A-Z]{2}[AFILMYWV][EQ][A-Z] motif.

and sequenced with many copies. Sequencing errors and PCR crossover are rare and close to random events (63). Sequences that differ from the seed due to PCR crossover are likely to appear as singletons after clustering, whereas differences arising from sequencing errors are likely to be only a few hamming distances away from the seed transcripts, and cluster together. Thus, removing transcripts highly similar to the seed transcripts and singletons will remove many sequencing errors, although we may lose a portion of biological

transcripts with low sequencing coverage. This approach is proven effective at finding functional transcripts in previous studies (64, 65).

Compared to other BCR repertoire databases, cAb-Rep provides more flexibility to genetic signature search. Another public database, iReceptor, supported by the adaptive immune receptor repertoire community, has been developed to store, query, and analyze immune receptor repertoire data (66). While iReceptor includes BCR repertoires amplified with various library

preparation protocols and sequenced in multiple platforms (Illumina, 454, ion torrent, etc.), cAb-Rep is limited to include full-length or near-full length sequences with enough sample size for repertoire related analyses. For repertoire analysis, iReceptor allows query with V(D)J recombination and a CDR3 peptide (no regular expression grammars allowed), while cAb-Rep provides BLAST mode and a more flexible CDR3 searching mode. cAb-Rep also allows to search motifs across the V(D)J region in the position mode. Although iReceptor includes curated repertoire data deposited by researchers, these curated datasets are processed using different bioinformatics pipeline and parameters (germline gene database, exclusion of redundancy, etc.), which may introduce bias when performing statistical analysis across studies. Another B cell repertoire database is available, but only includes repertoires from three donors (20). Moreover, the 13 HD curated BCR repertoires, which provide high depth BCR diversity information, have not been incorporated in other databases. These datasets increase the statistical power to study features as well as rare events of BCR diversity.

A further goal of cAb-Rep is to advance the investigation of impact of somatic hypermutation, which, as far as we know, is not available in other B cell repertoire databases. We provide a query portal and GSSPs for human antibody V genes to understand gene- and positional-specific substitution preferences. We also predicted gene-specific frequencies of N-glycosylation in human antibody V genes. Such sequence-derived information, together with functional study, is critical to identify common mechanism of modulation of affinity and understand similarities in pathways of antibody affinity maturation (40–42). Nonetheless, our current knowledge on impact of SHM is still very limited. In the future, we will collect literatures to characterize how SHM affects antibody structure and function as well as develop new tools to predict effects of SHM. A portal in cAb-Rep will be created to query studied SHMs, which will elucidate the relations

of antibody sequence, structure, and function and provide knowledge for antibody design.

In summary, we believe cAb-Rep and the tools developed in this study to be complimentary to other B cell repertoire databases and to be helpful to researchers not familiar with repertoire annotation for exploring features of repertoires and compare datasets across studies and diseases.

## DATA AVAILABILITY STATEMENT

The raw next-generation sequencing data were downloaded from NCBI SRA database (<https://www.ncbi.nlm.nih.gov/sra>). The project accession numbers were listed in the ProjectID column of **Table S1**. The curated data can be found at our website: <https://cab-rep.c2b2.columbia.edu/tools/>.

## AUTHOR CONTRIBUTIONS

PK, LS, and ZS designed the research. YG, KC, and ZS analyzed data. YG, PK, LS, and ZS wrote the paper. All authors reviewed, commented on, and approved the manuscript.

## FUNDING

This work was supported by National Institute of Allergy and Infectious Disease, National Institute of Health [1R21AI138024-01A1] to ZS.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02365/full#supplementary-material>

## REFERENCES

- Murphy K. *Janeway's Immunobiology*. New York, NY: Garland Science (2014).
- Briney B, Inderbitzin A, Joyce C, Burton DR. Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature*. (2019) 566:393–7. doi: 10.1038/s41586-019-0879-y
- Elhanati Y, Sethna Z, Marcou Q, Callan CG, Mora T, Walczak AM. Inferring processes underlying B-cell repertoire diversity. *Philos Trans R Soc Lond B Biol Sci*. (2015) 370:20140243. doi: 10.1098/rstb.2014.0243
- Soto C, Bombardi RG, Branchizio A, Kose N, Matta P, Sevy AM, et al. High frequency of shared clonotypes in human B cell receptor repertoires. *Nature*. (2019) 566:398–402. doi: 10.1038/s41586-019-0934-8
- Francica JR, Sheng Z, Zhang Z, Nishimura Y, Shingai M, Ramesh A, et al. Analysis of immunoglobulin transcripts and hypermutation following SHIV(AD8) infection and protein-plus-adjuvant immunization. *Nat Commun*. (2015) 6:6565. doi: 10.1038/ncomms7565
- Galson JD, Truck J, Clutterbuck EA, Fowler A, Cerundolo V, Pollard AJ, et al. B-cell repertoire dynamics after sequential hepatitis B vaccination and evidence for cross-reactive B-cell activation. *Genome Med*. (2016) 8:68. doi: 10.1186/s13073-016-0337-5
- Galson JD, Truck J, Fowler A, Clutterbuck EA, Munz M, Cerundolo V, et al. Analysis of B cell repertoire dynamics following hepatitis B vaccination in humans, and enrichment of vaccine-specific antibody sequences. *EBioMed*. (2015) 2:2070–9. doi: 10.1016/j.ebiom.2015.11.034
- Galson JD, Truck J, Kelly DF, van der Most R. Investigating the effect of AS03 adjuvant on the plasma cell repertoire following pH1N1 influenza vaccination. *Sci Rep*. (2016) 6:37229. doi: 10.1038/srep37229
- Joyce MG, Wheatley AK, Thomas PV, Chuang GY, Soto C, Bailer RT, et al. Vaccine-induced antibodies that neutralize group 1 and group 2 influenza A viruses. *Cell*. (2016) 166:609–23. doi: 10.1016/j.cell.2016.06.043
- Kwong PD, Chuang GY, DeKosky BJ, Gindin T, Georgiev IS, Lemmin T, et al. Antibodyomics: bioinformatics technologies for understanding B-cell immunity to HIV-1. *Immunol. Rev*. (2017) 275:108–28. doi: 10.1111/imr.12480
- Nielsen SCA, Boyd SD. Human adaptive immune receptor repertoire analysis-past, present, and future. *Immunol. Rev*. (2018) 284:9–23. doi: 10.1111/imr.12667
- Rubelt F, Bolen CR, McGuire HM, Vander Heiden JA, Gadala-Maria D, Levin M, et al. Individual heritable differences result in unique cell lymphocyte receptor repertoires of naive and antigen-experienced cells. *Nat. Commun*. (2016) 7:11112. doi: 10.1038/ncomms11112

13. Sheng Z, Schramm CA, Connors M, Morris L, Mascola JR, Kwong PD, et al. Effects of darwinian selection and mutability on rate of broadly neutralizing antibody evolution during HIV-1 infection. *PLoS Comput Biol.* (2016) 12:e1004940. doi: 10.1371/journal.pcbi.104940
14. Sheng Z, Schramm CA, Kong R, Program NCS, Mullikin JC, Mascola JR, et al. Gene-specific substitution profiles describe the types and frequencies of amino acid changes during antibody somatic hypermutation. *Front Immunol.* (2017) 8:537. doi: 10.3389/fimmu.2017.00537
15. Stern JN, Yaari G, Vander Heiden JA, Church G, Donahue WF, Hintzen RQ, et al. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci Transl Med.* (2014) 6:248ra107. doi: 10.1126/scitranslmed.3008879
16. Tipton CM, Fucile CF, Darce J, Chida A, Ichikawa T, Gregoret I, et al. Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus. *Nat Immunol.* (2015) 16:755–65. doi: 10.1038/ni.3175
17. Vander Heiden JA, Stathopoulos P, Zhou JQ, Chen L, Gilbert TJ, Bolen CR, et al. Dysregulation of B cell repertoire formation in myasthenia gravis patients revealed through deep sequencing. *J Immunol.* (2017) 198:1460–73. doi: 10.4049/jimmunol.1601415
18. Wu X, Zhang Z, Schramm CA, Joyce MG, Kwon YD, Zhou T, et al. Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. *Cell.* (2015) 161:470–85. doi: 10.1016/j.cell.2015.03.004
19. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.* (2008) 36:W503–8. doi: 10.1093/nar/gkn316
20. DeWitt WS, Lindau P, Snyder TM, Sherwood AM, Vignali M, Carlson CS, et al. A public database of memory and naive B-cell receptor sequences. *PLoS ONE.* (2016) 11:e0160853. doi: 10.1371/journal.pone.0160853
21. Margreitter C, Lu HC, Townsend C, Stewart A, Dunn-Walters DK, Fraternali F. BR repertoire: a user-friendly web server for analysing antibody repertoire data. *Nucleic Acids Res.* (2018) 46:W264–70. doi: 10.1093/nar/gky276
22. Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V. Computational strategies for dissecting the high-dimensional complexity of adaptive immune repertoires. *Front Immunol.* (2018) 9:224. doi: 10.3389/fimmu.2018.00224
23. Schramm CA, Sheng Z, Zhang Z, Mascola JR, Kwong PD, Shapiro L. SONAR: a high-throughput pipeline for inferring antibody ontogenies from longitudinal sequencing of B cell transcripts. *Front Immunol.* (2016) 7:372. doi: 10.3389/fimmu.2016.00372
24. Vander Heiden JA, Marquez S, Marthandan N, Bukhari SAC, Busse CE, Corrie B, et al. AIRR community standardized representations for annotated immune repertoires. *Front Immunol.* (2018) 9:2206. doi: 10.3389/fimmu.2018.02206
25. Chaudhary N, Wesemann DR. Analyzing immunoglobulin repertoires. *Front Immunol.* (2018) 9:462. doi: 10.3389/fimmu.2018.00462
26. Burton DR, Mascola JR. Antibody responses to envelope glycoproteins in HIV-1 infection. *Nat Immunol.* (2015) 16:571–6. doi: 10.1038/ni.3158
27. Corti D, Lanzavecchia A. Broadly neutralizing antiviral antibodies. *Ann Rev Immunol.* (2013) 31:705–42. doi: 10.1146/annurev-immunol-032712-095916
28. Andrabi R, Voss JE, Liang CH, Briney B, McCoy LE, Wu CY, et al. Identification of common features in prototype broadly neutralizing antibodies to HIV envelope V2 apex to facilitate vaccine design. *Immunity.* (2015) 43:959–73. doi: 10.1016/j.immuni.2015.10.014
29. Ekiert DC, Bhabha G, Elsliger MA, Friesen RH, Jongeneelen M, Throsby M, et al. Antibody recognition of a highly conserved influenza virus epitope. *Science.* (2009) 324:246–51. doi: 10.1126/science.1171491
30. Kwong PD, Mascola JR. HIV-1 vaccines based on antibody identification, B cell ontogeny, and epitope structure. *Immunity.* (2018) 48:855–71. doi: 10.1016/j.immuni.2018.04.029
31. Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, et al. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science.* (2011) 333:1593–602. doi: 10.1126/science.1207532
32. Zhou T, Lynch RM, Chen L, Acharya P, Wu X, Doria-Rose NA, et al. Structural repertoire of HIV-1-neutralizing antibodies targeting the CD4 supersite in 14 donors. *Cell.* (2015) 161:1280–92. doi: 10.1016/j.cell.2015.05.007
33. Zhou T, Zhu J, Wu X, Moquin S, Zhang B, Acharya P, et al. Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. *Immunity.* (2013) 39:245–58. doi: 10.1016/j.immuni.2013.04.012
34. Chuang GY, Zhou J, Acharya P, Rawi R, Shen CH, Sheng ZZ, et al. Structural survey of broadly neutralizing antibodies targeting the HIV-1 Env trimer delineates epitope categories and characteristics of recognition. *Structure.* (2019) 27:196–206.e6. doi: 10.1016/j.str.2018.10.007
35. Kwong PD, Wilson IA. HIV-1 and influenza antibodies: seeing antigens in new ways. *Nat Immunol.* (2009) 10:573–8. doi: 10.1038/ni.1746
36. Pappas L, Foglierini M, Piccoli L, Kallewaard NL, Turrini F, Silacci C, et al. Rapid development of broadly influenza neutralizing antibodies through redundant mutations. *Nature.* (2014) 516:418–22. doi: 10.1038/nature13764
37. Dosenovic P, Kara EE, Pettersson AK, McGuire AT, Gray M, Hartweiger H, et al. Anti-HIV-1 B cell responses are dependent on B cell precursor frequency and antigen-binding affinity. *Proc Natl Acad Sci USA.* (2018) 115:4743–8. doi: 10.1073/pnas.1803457115
38. Jardine JG, Kulp DW, Havenar-Daughton C, Sarkar A, Briney B, Sok D, et al. HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. *Science.* (2016) 351:1458–63. doi: 10.1126/science.aad9195
39. Odegard VH, Schatz DG. Targeting of somatic hypermutation. *Nat Rev Immunol.* (2006) 6:573–83. doi: 10.1038/nri1896
40. Bonsignori M, Zhou T, Sheng Z, Chen L, Gao F, Joyce MG, et al. Maturation pathway from germline to broad HIV-1 neutralizer of a CD4-mimic antibody. *Cell.* (2016) 165:449–63. doi: 10.1016/j.cell.2016.02.022
41. Koenig P, Lee CV, Walters BT, Janakiraman V, Stinson J, Patapoff TW, et al. Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding. *Proc Natl Acad Sci USA.* (2017) 114:E486–95. doi: 10.1073/pnas.1613231114
42. van de Bovenkamp FS, Derksen NIL, Ooijevaar-de Heer P, van Schie KA, Kruijthof S, Berkowska MA, et al. Adaptive antibody diversification through N-linked glycosylation of the immunoglobulin variable region. *Proc Natl Acad Sci USA.* (2018) 115:1901–6. doi: 10.1073/pnas.1711720115
43. Hwang JK, Wang C, Du Z, Meyers RM, Kepler TB, Neuberger D, et al. Sequence intrinsic somatic mutation mechanisms contribute to affinity maturation of VRC01-class HIV-1 broadly neutralizing antibodies. *Proc Natl Acad Sci USA.* (2017) 114:8614–9. doi: 10.1073/pnas.1709203114
44. Edgar RC, Flyvbjerg H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics.* (2015) 31:3476–82. doi: 10.1093/bioinformatics/btv401
45. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* (1997) 25:3389–402. doi: 10.1093/nar/25.17.3389
46. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* (2011) 7:539. doi: 10.1038/msb.2011.75
47. Corcoran MM, Phad GE, Vazquez Bernat N, Stahl-Hennig C, Sumida N, Persson MA, et al. Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. *Nat. Commun.* (2016) 7:13642. doi: 10.1038/ncomms13642
48. Dunbar J, Deane CM. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics.* (2016) 32:298–300. doi: 10.1093/bioinformatics/btv552
49. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* (2009) 10:421. doi: 10.1186/1471-2105-10-421
50. Gupta NT, Adams KD, Briggs AW, Timberlake SC, Vigneault F, Kleinstein SH. Hierarchical clustering can identify B cell clones with high confidence in Ig repertoire sequencing data. *J Immunol.* (2017) 198:2489–99. doi: 10.4049/jimmunol.1601850
51. Imkeller K, Scally SW, Bosch A, Marti GP, Costa G, Triller G, et al. Antihomotypic affinity maturation improves human B cell responses against a repetitive epitope. *Science.* (2018) 360:1358–62. doi: 10.1126/science.aar5304
52. Jayaram N, Bhowmick P, Martin AC. Germline VH/VL pairing in antibodies. *Protein Eng Des Sel.* (2012) 25:523–9. doi: 10.1093/protein/gzs043

53. West AP Jr, Diskin R, Nussenzweig MC, Bjorkman PJ. Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proc Natl Acad Sci USA*. (2012) 109:E2083–90. doi: 10.1073/pnas.1208984109
54. Jardine JG, Sok D, Julien JP, Briney B, Sarkar A, Liang CH, et al. Minimally mutated HIV-1 broadly neutralizing antibodies to guide reductionist vaccine design. *PLoS pathogens*. (2016) 12:e1005815. doi: 10.1371/journal.ppat.1005815
55. Chazenbalk GD, Portolano S, Russo D, Hutchison JS, Rapoport B, McLachlan S. Human organ-specific autoimmune disease. Molecular cloning and expression of an autoantibody gene repertoire for a major autoantigen reveals an antigenic immunodominant region and restricted immunoglobulin gene usage in the target organ. *J Clin Invest*. (1993) 92:62–74. doi: 10.1172/JCI116600
56. Hexham JM, Furmaniak J, Pegg C, Burton DR, Smith BR. Cloning of a human autoimmune response: preparation and sequencing of a human anti-thyroglobulin autoantibody using a combinatorial approach. *Autoimmunity*. (1992) 12:135–41. doi: 10.3109/08916939209150320
57. Jeon YE, Seo CW, Yu ES, Lee CJ, Park SG, Jang YJ. Characterization of human monoclonal autoantibody Fab fragments against oxidized LDL. *Mol Immunol*. (2007) 44:827–36. doi: 10.1016/j.molimm.2006.04.005
58. Pichurin P, Guo J, Yan X, Rapoport B, McLachlan SM. Human monoclonal autoantibodies to B-cell epitopes outside the thyroid peroxidase autoantibody immunodominant region. *Thyroid*. (2001) 11:301–13. doi: 10.1089/10507250152039037
59. Pichurin PN, Guo J, Estienne V, Carayon P, Ruf J, Rapoport B, et al. Evidence that the complement control protein-epidermal growth factor-like domain of thyroid peroxidase lies on the fringe of the immunodominant region recognized by autoantibodies. *Thyroid*. (2002) 12:1085–95. doi: 10.1089/105072502321085180
60. Wu X, Liu B, Van der Merwe PL, Kalis NN, Berney SM, Young DC. Myosin-reactive autoantibodies in rheumatic carditis and normal fetus. *Clin Immunol Immunopathol*. (1998) 87:184–92. doi: 10.1006/clin.1998.4531
61. Doria-Rose NA, Schramm CA, Gorman J, Moore PL, Bhiman JN, DeKosky BJ, et al. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature*. (2014) 509:55–62. doi: 10.1038/nature13036
62. Khan TA, Friedensohn S, de Vries AR, Straszewski J, Ruscheweyh HJ, Reddy ST. Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Sci Adv*. (2016) 2:e1501371. doi: 10.1126/sciadv.1501371
63. Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res*. (2015) 43:e37. doi: 10.1093/nar/gku1341
64. Bhiman JN, Anthony C, Doria-Rose NA, Karimanzira O, Schramm CA, Khoza T, et al. Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. *Nat Med*. (2015) 21:1332–6. doi: 10.1038/nm.3963
65. Krebs SJ, Kwon YD, Schramm CA, Law WH, Donofrio G, Zhou KH, et al. Longitudinal analysis reveals early development of three MPER-directed neutralizing antibody lineages from an HIV-1-infected individual. *Immunity*. (2019) 50:677–91.e613. doi: 10.1016/j.immuni.2019.02.008
66. Corrie BD, Marthandan N, Zimonja B, Jaglale J, Zhou Y, Barr E, et al. iReceptor: a platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol Rev*. (2018) 284:24–41. doi: 10.1111/imr.12666

**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.

Copyright © 2019 Guo, Chen, Kwong, Shapiro and Sheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.