B cell repertoire sequencing of HIV-1 pediatric elite-neutralizers identifies multiple broadly neutralizing antibody clonotypes

Sanjeev Kumar[1,2,3,4,5,*], Prashant Bajpai[6,*], Collin Joyce[2,3,4,5], Sushil Kumar Kabra[7], Rakesh Lodha[7], Dennis R. Burton[2,3,4,5], Bryan Briney[2,3,4,8,†], and Kalpana Luthra[1,†]

[1]Department of Biochemistry, All India Institute of Medical Sciences, New Delhi 11029, India
[2]Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 92037 USA
[3]Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, The Scripps Research Institute, La Jolla, CA 92037 USA
[4]Center for Viral Systems Biology, The Scripps Research Institute, La Jolla, CA92037 USA
[5]IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, CA92037 USA
[6]ICGEB-Emory Vaccine Center, International Center for Genetic Engineering and Biotechnology, New Delhi, 110067, India
[7]Department of Pediatrics, All India Institute of Medical Sciences, New Delhi 11029, India
[8]Multi-omics Vaccine Evaluation Consortium, The Scripps Research Institute, La Jolla, CA 92037 USA
*These authors contributed equally
†Lead Contact
Correspondence may be addressed to:
Kalpana Luthra (kalpanaluthra@aiims.edu, kalpanaluthra@gmail.com)
Bryan Briney (briney@scripps.edu)

ABSTRACT

A limited subset of HIV-1 infected adult individuals typically after at least 2-3 years of chronic infection, develop broadly neutralizing antibodies (bnAbs), suggesting that highly conserved neutralizing epitopes on the HIV-1 envelope glycoprotein are difficult for B cell receptors to effectively target, during natural infection. Recent studies have shown the evolution of bnAbs in HIV-1 infected infants. We used bulk BCR sequencing (BCR-seq) to profile the B cell receptors from longitudinal samples (3 time points) collected from a rare pair of antiretroviral-naïve, HIV-1

infected pediatric monozygotic twins (AIIMS_329 and AIIMS_330) who displayed elite plasma neutralizing activity against HIV-1. BCR-seq of both twins revealed convergent antibody characteristics including V-gene use, CDRH3 lengths and somatic hypermutation (SHM). Further, antibody clonotypes with genetic features similar to highly potent bnAbs isolated from adults showed ongoing development in donor AIIMS_330 but not in AIIMS_329, corroborating our earlier findings based on plasma bnAbs responses. An increase in SHM was observed in sequences of the IgA isotype from AIIMS_330. This study is an important step toward defining specific features of the antibody repertoire and shared clonotype maturation that are associated with the development of HIV-1 elite-neutralizing activity. Some of the shared lineages identified in this pair of identical twins, who had acquired HIV-1 infection by vertical transmission, may have been elicited by exposure to common HIV-1 antigens. Defining the sequences of such shared clonotypes in a large number of children can shed light on the role of specific B cell receptor features in the response to HIV-1 infection.

KEYWORDS

HIV-1, children, elite-neutralizers, B cell repertoire sequencing, bnAbs, clonotypes

INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) infections are a global health problem that has affected an estimated 38.4 million people worldwide including children[1]. The integration of virus into the host cell genome and tremendous level of viral mutation observed in individuals living with HIV-1 are major obstacles to developing effective therapeutics and vaccines againstHIV-1[2,3]. During HIV-1 infection, neutralizing antibodies (nAbs) and non-nAbs are elicited against multiple epitopes of the HIV-1 envelope glycoprotein[4–6]. In the past 15 years, the invention of high-throughput technologies for antibody generation i.e. single B cell sorting, high-throughput antibody screening by micro-neutralization methods and single cell sequencing, have led to the discovery and molecular characterization of highly potent second-generation HIV-1 broadly neutralizing antibodies (bnAbs) both from adults and children living with HIV-1[5–7]. Presently, researchers around the globe are leveraging our understanding of the genetic, functional, and

structural properties of these bnAbs to develop HIV-1 vaccines that reliably induce broad and potently neutralizing antibody responses[8,9].

The characteristic features of HIV-1 bnAbs targeting multiple envelope epitopes that are isolated from adult donors have been extensively studied and are well-known[5,6]. However, the understanding of genetic and molecular features of HIV-1 bnAbs from children is still limited. So far, only two HIV-1 pediatric bnAbs have been identified, BF520.1 by Simonich et.al.[10] and AIIMS-P01 by us (Kumar et. al.)[11]. Further insights into the pediatric B cell repertoire are urgently needed to support next-generation vaccine design and vaccination strategies aiming to elicit protective bnAbs responses in a wide range of individuals including children. It has been reported in multiple studies that children living with HIV-1 showed broader and more potent bnAbs responses with multiple epitope specificities as compared to bnAbs from adults even within one-year of age, suggesting that bnAb responses in children are developed from different maturation mechanisms/pathways[10–21]. This is supported by the discovery of BF520.1 and AIIMS-P01:pediatric bnAbs which exhibit comparable HIV-1 neutralization breadth and potency to adult bnAbs but with limited somatic hypermutation (SHM)[10,11].

In the past 15 years, we have established a rare cohort of HIV-1 clade C chronically infected pediatric donors including infants[11,14,16,17,21–25]. Recently, we reported the identification of infant and adolescent pediatric elite-neutralizers (AIIMS_329 and AIIMS_330) from characterization of their plasma HIV-1 bnAbs responses at a single and longitudinal time points, respectively[14,21]. We observed the development of bnAbs targeting the V1/V2 apex, glycan supersite, andCD4 binding site (CD4bs) in AIIMS_330. We also noted V1/V2 apex- and glycan supersite-dependent bnAbs in AIIMS_329, but at a lower potency and breadth in comparison to AIIMS_330[21]. Herein, we performed deep sequencing of the bulk B cell repertoires (BCR-seq) of two monozygotic twin HIV-1 pediatric elite-neutralizers (AIIMS_329 and AIIMS_330). BCR-seq of both AIIMS_329 and AIIMS_330 showed convergent antibody characteristics including Variable gene use, heavy chain complementarity determining region (HCDR3) lengths, SHM frequency. Mapping BCR-seq data to knownHIV-1 bnAbs, allowed us to identify antibody clonotypes with similar features to potent HIV-1 bnAbs in AIIMS_330 but not in AIIMS_329, corroborating with our previous serological findings. This study suggests that children living with chronic HIV-1maydevelop clonotypes of

HIV-1bnAbs against multiple envelope epitopes similar to those isolated from adults, highlighting that such B cells could be steered to elicit bnAbs responses through vaccines aimed to induce bnAbs against HIV-1 in a broad range of people.

RESULTS

Deep BCR sequencing of HIV-1 pediatric elite-neutralizers

AIIMS_329 and AIIMS_330 are monozygotic twins, and both display elite HIV-1 plasma neutralization. Longitudinal samples were obtained from both twins prior to their initiation of combined antiretroviral therapy (cARTz) in 2018. The total number of sequences obtained for each subject after sequencing and annotation ranged from approximately $1.0x10^5$ to $4.0x10^5$ sequences (S**upplementary Table 1**). Immunogenetic analysis of heavy chain Variable (VH) genes identified IGHV3-21, IGHV3-23, and IGHV-34 to be predominant gene usages in both AIIMS_329 and AIIMS_330 groups in the three timepoints of their sample collection in the years 2015 (112 months p.i.), 2016 (117 months p.i.), and 2018 (138 months p.i.). Within IGHV3-21, AIIMS_330 from the three timepoints showed a trend towards higher frequency compared to AIIMS_329, and 330_2015 had the highest frequency (25.6% compared to compared to approx. 13 other groups, **Figure 1A, Supplementary Table 2**). No major differences in the VH gene frequency of the two groups-at any time point was observed (**Figure 1A, Supplementary Table 2**). Consistent with previous studies, IGHD3-22, IGHD3-10, IGHD6-13 were found to be major D genes and IGHJ4 was found to be predominant heavy J gene. No noteworthy differences were observed in the frequencies of heavy chain Diversity (D)or Joining (J) genes (**Figure 1B and IC**). In the light chain, IGKV4-1 and IGKV3-20 were the predominant gene usages. We observed a trend towards higher frequency of IGKV4-1 in AIIMS_330 compared to AIIMS_329 (14.54%, 20.35%, 14.16% in 330_2015, 330_2016 and 330_2018 compared to 8.89%, 10.27%, 12.99% in 329_2015, 329_2016 and 329_2018). We also observed frequency of IGLV2-14 to be lower in AIIMS_330 compared to AIIMS_329 (2%, 2.09%, 2.13% in 330_2015, 330_2016 and 330_2018 compared to 4.63%, 7.91%, 5.09% in 329_2015, 329_2016 and 329_2018) (**Figure 1D, Supplementary Table 3**). We also observed marked differences in the light chain J gene use between the two groups. Gene usage IGKJ2 showed a trend towards higher frequency in AIIMS_330 at three timepoints. Further, the frequency of IGLJ3 was markedly lower in AIIMS_330

compared to AIIMS_329.Overall, though heavy chain V, D, and J gene frequencies were similar in AIIMS_329 and AIIMS_330, we observed marked differences in lambda and kappa light chains between the two groups at all timepoints.

Shared immunogenetic features between AIIMS_329 and AIIMS_330

The CDRH3 length distributions were similar for AIIMS_329 and AIIMS_330 at all time points, with mean CDHR3 length being lower in the AIIMS_330(12.91, 13.22 and 12.66 in 330_2015, 330_2016, and 330_2018, respectively, compared to 12.97, 13.37, and 12.97 in 329_2015, 329_2016, and 329_2018) (**Figure 2A, Supplementary Table 4**).A similar trend was observed in the distribution of light chain CDR3 lengths (**Supplementary Fig 1**). SHM frequency was significantly higher in AIIMS_330 at all three timepoints compared to AIIMS_329 group. In the 2018 timepoint, median SHM in AIIMS_330 was 2.42% compared to 0.41% in AIIMS_329 (**Figure 2B, Supplementary Table 5**). Interestingly, when the sequences were segregated into isotypes, sequences from IgA isotype were observed to have significantly higher SHM in AIIMS_330 compared to AIIMS_329. The median SHM in IgA sequences was found to be 11.74%, 6.97%, and 6.94% in 330_2015, 330_2016, and 330_2018, respectively, compared to 7.63%, 6.05%, and 4.72% in 329_2015, 329_2016, and 329_2018. No significant changes were observed in other isotypes (**Figure 2C, Supplementary Table 6**).

Shared clonotype analysis identified distinct features between AIIMS_330 and AIIMS_329

Sequences were subsampled before performing the clonotype analysis (more detail of the subsetting approach can be found in the methods section). Frequency of total shared clonotypes between any two groups was found to be less than 0.07% (77 shared clonotypes across 10918 total clonotypes, **Figure 3A bottom panel**). Of all the clonotypes identified in each group, clones in 330_2018 timepoint had the highest SHM compared to others. Most of the clonotypes in 330_2018 timepoint had mean SHM >10%. We also observed higher expansion in 2018 timepoint in AIIMS_330 compared to AIIMS_329 as observed from lower number of clonotypes and higher number of sequences in each clonotype. Interestingly, the overall clonotypes in 2018 timepoint in both groups were higher compared to 2015 and 2016 timepoints (**Figure 3A middle panels**). No notable differences were observed in the V-family distribution of shared clonotypes (**Figure 3A**

top panel). We also analyzed the clones shared between each sample and timepoints. We did not observe any clonotypes shared between all the groups or even between AIIMS_330 and AIIMS_329 at all three timepoints. Interestingly, the highest number of shared clonotypes were identified between 330_2016 and 330_2018 timepoints (39 shared clonotypes) but not between any timepoints in AIIMS_329. Of these 39 clonotypes, only 3 had mean SHM greater than 10%. However, those 3 clonotypes did not have any significant differences in their CDR3 amino acid (**Figure 3B**). We identified 12 clonotypes that were shared between AIIMS_330 and AIIMS_329 at 2016 timepoint of which 5 clonotypes had mean SHM greater than 10%. These 5 clonotypes were found to have accumulated higher mutations in CDR3 compared to the 3 clonotypes shared between 330_2016 and 330_2018 (**Figure 3C**). None of these shared clonotypes were predicted to encode sequence insertions or deletions which are frequently found in HIV bnAbs and thought to be difficult to elicit by vaccination. Such BCRs could plausibly define a more efficient path for bnAb maturation that can be exploited by rational HIV vaccine immunogens. We also observed 21 clonotypes in common between 2018 timepoint between AIIMS_330 and AIIMS_329. Of these, only one clonotype had >10% SHM. CDR3 of sequences from these clonotypes were observed to harbor higher number of mutations in 330_2018 timepoint compared to 329_2018 (**Figure 3D**).

Mapping of sequences to published adult HIV bnAbs identified 330_2018 to be capable of producing bnAbs with characteristics of adult bnAbs

To identify if either of the two children could generate known HIV-1 bnAbs at any timepoint, we mapped the sequences from each group to known HIV mAbs compiled from HIV neutralizing antibody database (CATNAP) and published studies. A total of 347 mAbs were compiled (**Supplementary Table 7**) and mapped to sequences from each group. The criteria for calling a sequence a successful hit is described in the methods. Of the total 347 mAbs, 78 mAbs mapped to at least one of the sequences in at least one of the groups. Surprisingly, majority of the sequences that mapped to these mAbs were from 330_2018 timepoint but no other timepoints. Moreover, we observed the expansion/more frequency of thefollowingHIV-1 bnAbs like sequences in AIIMS_330:VRC38.01 (V2 apex)[26], VRC34.01 (fusion peptide)[27], HGN194 (V3)[28], HK20 (gp41 HR)[29], DH270 (glycan supersite)[30], BG505.m27 (V3)[31], 8ANC131 (CD4bs)[32], 2558 (V3)[33], and VRC33.01 (CD4bs)[34](**Figure 4**). Strikingly, except DH511.1, DH511.6 (MPER)[35]and BG505.m27

HIV-1 bnAbs, no other mapped sequences from AIIMS_330 only showed SHM >10%, however, such high SHMs were not observed for any of the mapped sequences from AIIMS_329.

DISCUSSION

An understanding of pediatric HIV-1 bnAb responses elicited during chronic infection can provide critical insights for the design and development of effective universal HIV-1 vaccines for both adults and children capable of eliciting potent HIV-1 bnAb responses[36,37]. Human antibody repertoire comprises a whole group of antibody genes generated in an individual since birth, which can be specific to an antigen during an infection/disease and vaccination. BCR-seq is a powerful method that has enabled the in-depth analysis of genetic features of antibody responses and tracking of antibody evolution during an infection/vaccination[38]. Studying such antibody repertoires could reveal important information that include germline antibody genes, junctional diversity, SHM, clonotypes and allows identification of rare antibody lineage genes specific to a particular antigen.

Recently, in a study based on longitudinal characterization of plasma samples we reported the identification of adolescent pediatric elite-neutralizers who were HAART-naïve and living with chronic HIV-1 clade-C infection for more than 11 years[21]. We evaluated the V1/V2 apex, glycan supersite, and CD4bs specific responses in AIIMS_329 and AIIMS_330, using ELISA and point mutated HIV-1 pseudovirus basedneutralizationassays[21]. We showed a longitudinal development of HIV-1 bnAbs targeting V1V2, N332 and CD4bs in AIIMS_330, whereas V1V2 and N332-supersite dependent HIV-1 bnAbs in AIIMS_329, but at a lower potency and breadth in comparison to AIIMS_330[21]. Herein, we performed bulk BCR-seq of these two monozygotic twins HIV-1 pediatric elite-neutralizers (AIIMS_329 and AIIMS_330) to understand their BCR repertoire. Analysis of three time points (112, 117 and 138 months p.i.) from these pediatric elite-neutralizers showed high convergence of antibody gene-usage, CDRH3 lengths, SHM, and clonotypes. Though a higher IgA SHM was observed in AIIMS_330, no other isotype showed differential mutation patterns. The development of HIV-specific IgA responses with affinity maturation have been shown to in association with anti-gp41 IgA antibodies that occurs to a greater extent in elite controllers than in individuals on HAART[39], suggesting IgA in AIIMS_330 could be associated with lower viral load as compared to AIIMS_329.

Current common strategies to generate antigen specific antibodies and evaluating humoral vaccine responses are: 1) high-throughput sorting antigen-specific B cells followed by paired heavy and light chain gene amplification and sequencing, 2) barcoded antigen approaches like LIBRA-seq which leverage single cell microfluidics to link antibody genetics and binding specificity in high throughput, and 3) activation and culture of primary B cells followed by functional supernatant screening to identify B cell clones with desirable functional profiles. Such methods are labor-and time-intensive ways of identifying antigen-specific antibodies. Mapping sequences to well-known established antibody CDR3 information can provide templates or blueprints to identify important antibody genes in a distinct set of individual populations, antibody discovery, and vaccine response evaluation without the need for antigen-specific sorting[40]. Similar methodologies have been used successfully to identify potent antibodies against Dengue, HIV-1, SARS-CoV-2, and influenza[41].

Here, we mapped BCR sequencing data against datasets of known HIV-1 bnAbs [42]. Interestingly, this analysis led to the identification of multiple HIV-1 specific antibody clonotypes isolated from both adults and children in our BCR-seq data. AIIMS_330 showed clonotypes similar to several HIV-1 bnAbs identified in adults which are known to target multiple epitopes including MPER, CD4bs, N332, FP and V1V2, suggesting that antibody data mining based on CDRH3 sequences could help in bnAbs lineage identification to a particular antigen, as recently observed by an AI-based pipeline developed by Wu et.al.[41] for COVID-19 and Flu antibody genes and DENV specific mAbs by Durham et.al.[43]Our analysis of these bnAb lineages showed the evolution/more frequency (**Figure 4**) of these antibodies in AIIMS_330 donor, which corroborates with our previous findings based on plasma characterization and evolution of multi-epitope specific antibodies including CD4bs bnAbs in AIIMS_330 than AIIMS_329[21]. Overall, the observation of some the shared lineages in this pair of identical pediatric twins, who had acquired HIV-1 infection by vertical transmission, suggest that they could evolve in response to common antigenic stimulus. Defining the sequences of such shared clonotypes, in a large number of children living with chronic HIV-1 infection worldwide, can shed light in understanding the role of specific B cell receptor repertoires in HIV-1 infection.

Our study provides insight into the BCR repertoires of pediatric twins capable of exceptionally broad and potent HIV-1 neutralization. We noted the presence of clonotypes with genetic similarity to known adults' HIV-1 bnAbs targeting multiple epitope specificities in children, which corroborated the results of our previous study which characterized the plasma neutralization breadth, potency, and epitope specificity from the same pair of donors. We anticipate these data will be useful for the design and development of effective vaccine candidates and strategies for both adults and children to combat HIV-1. Our BCR-seq findings from this unique pair of pediatric elite-neutralizers suggests that multivalent HIV-1 vaccine development strategies focused on inducing a diverse range of HIV-1bnAb specificities is likely superior to approaches that focus solely on a single epitope.

239 **METHODS**

240 **Patients Characteristics**

241 Monozygotic twin antiretroviral naïve HIV-1 clade C pediatric elite-neutralizers (AIIMS_329 and
242 AIIMS_330) were recruited from the Outdoor Patient Department of the Department of
243 Pediatrics, All India Institute of Medical Sciences (AIIMS), New Delhi, India for this study at the
244 age of 9 years and were followed for a total period of 60 months. Blood was drawn in 5-ml EDTA
245 vials, and plasma was aliquoted and stored for plasma antibody-based HIV-1 neutralization
246 assays, viral RNA isolation, and viral load determinations. The study was approved by the
247 Institute Ethics Committee (IEC/NP-295/2011 and RP-15/2011). All experiments were
248 conducted in accordance with the Institutional guidelines and protocols.

249 **Next-generation sequencing of HIV-1 pediatric B cell antibody repertoires**

250 The deep sequencing of bulk BCR was performed using primers and protocols as described
251 previously by Briney et.al.[38] Briefly, total RNA from 2 or 3 million peripheral blood mononuclear
252 cells (PBMCs) was extracted (RNeasy Maxi Kit, Qiagen) from each time point (2015 (112 months
253 p.i), 2016 (117 months p.i) and 2018 (138 months p.i)) and antibody sequences were amplified
254 using methods and primers as previously described. The PCR product sizes were verified on
255 agarose gel (E-Gel EX; Invitrogen) and quantified with fluorometry (Qubit; Life Technologies),
256 pooled at approximately equimolar concentrations and each sample pool was re-quantified
257 before sequencing on an Illumina MiSeq (MiSeq Reagent Kit v3, 600-cycle).

258 **Processing of next-generation sequencing data**

259 The Abstar analysis pipeline was used as previously described to quality trim, remove adapters
260 and merge paired sequences[38]. Sequences were then annotated with Abstar in combination with
261 UMI based error correction by AbCorrect (https://github.com/briney/abtools/). For comparison
262 of frequencies, read counts were scaled for each repertoire as previously described due to the
263 large differences in the number of reads between each group. Somatic hypermutation (SHM) was
264 calculated using the R package Shazam[44].

265 **Clonotype analysis**

Clonotype analysis was performed using Immcantation pipeline[44,45]. Sequences were grouped into clonotypes based on nucleotide hamming distance of 0.16 calculated based on bimodal distribution of distance of each sequence with its nearest neighbor. Alternatively, sequences were also clustered into clonal groups using an in-house script. The criteria used for clonal assignment was sequences having same V and J gene usage, same CDRH3 length and at least 80% CDRH3 amino acid identity. Germline V(D)J sequence was reconstructed using IMGT-gapped reference V, D and J sequences.

Deep mapping of HIV-1 mAbs to B cell repertoire sequencing data

List of HIV-1mAbs which have been reported previously was compiled from CATNAP database and literature[5–7,42]. Antibodies for which the gene usage information and CDRH3 amino acid sequence was available were selected for downstream analysis. The list of selected mAbs is given in **Supplementary Table 7**. Each mAb was mapped to all the sequences to identify the ones which have the same V and J gene usage, same CDRH3 length and at least 50% identity in CDRH3 amino acid sequence.

Alignment and phylogeny

Alignment was performed using R package MSA. Sequence logos were made using R package ggseqlogo. Distance between sequences were calculated using the neighbor joining method. Phylogenetic tree was constructed using R package ape.

Statistical analysis

All analysis was performed using R programming language (version 4.1.0). Statistical significance between groups was estimated using Wilcox rank sum test. The following packages were used for the analyses: scatterpie (0.1.7), nortest (1.0-4), scales (1.1.1), ggrepel (0.9.1), stringr (1.4.0), tigger (1.0.0), ComplexUpset (1.3.0), ggpubr (0.4.0), alakazam (1.1.0), shazam (1.1.0), patchwork (1.1.1), writexl (1.4.0), readxl (1.3.1), dplyr (1.0.6), reshape2 (1.4.4) and ggplot2 (3.3.5).

DATA AND AVAILABILITY

All bash and R scripts required for reproducing the data can be found on github repository (https://github.com/prashantbajpai/HIV_BCR_Analysis). Raw data used for generating the figures are available at the github page or in the supplementary files.

ACKNOWLEDGEMENTS

We are thankful to the SERB, India (CRG/2021/003984), Department of Biotechnology (DBT) (BT/PR30120/MED/29/1339/2018 and BT/PR 24520/MED/29/1222/2017) for the funding provided to K.L. This study is supported in part by ViiV Healthcare and the HIV Research Trust 2019 awarded to S.K. for training and gaining expertise in Next Generation Sequencing based technologies to sequence single B cell and bulk B cell repertoire. The funders had no role in the decision to publish or the preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization: K.L., B.B., and S.K.; Methodology: B.B., C.C.J., and S.K.; Software: B.B., C.C.J., P.B.; Validation: B.B.; Formal analysis: S.K., and P.B.; Investigation: S.K., and C.C.J.; Resources: D.R.B., R.L, and S.K.K.; Data Curation: B.B., P.B., and C.C.J.; Management: S.K., P.B., and B.B.; Writing - Original Draft: S.K., and P.B.,; Writing - Review & Editing: K.L., B.B., and D.R.B.;Visualization: S.K., and P.B.; Supervision: K.L., B.B. and D.R.B.; Project administration: S.K.; Funding acquisition: K.L., B.B., and S.K.

DECLARATION OF INTERESTS

All other authors declare no competing interests.

FIGURE LEGENDS

Figure1: Gene usage of heavy and light chains. V-gene **(A)**, D-gene **(B)** and J-gene **(C)** usage of heavy chains of genes shown as percentage of total sequences and arranged by descending frequency. V-gene **(E)** and D-gene **(F)** of light chains are shown.

Figure 2: Immunogenetics characteristics of heavy chain sequences. (A) CDRH3 length distribution of heavy chains in each group. The dotted horizontal line shows the median of each group. **(B)** Violin plot shows somatic hypermutation (SHM) in each group. The dot represents the median in each group. **(C)** Violin plot shows the SHM in each group segregated by isotype. The dot represents the mean in each group.

Figure 3: Shared clonotypes between the groups. (A) Upset plot illustrating the intersection of clonotypes between the subject 329 and 330 at the three time points (2015, 2016 and 2018). Only the clonotypes which were not singlets with more than one sequences were selected for the analysis. Each vertical bar shows the number of clonotypes present in each group indicated by the intersection matrix below it (where a single dot in the matrix is a single group). The number above the bars show the number of common clonotypes between the groups of intersecting sets of clonotypes marked below the bars. For each intersection the dot plot above indicates the CDRH3 length distribution for each clonotype. Size of each dot indicates the number of sequences present in each clonotype and the color shows the somatic hypermutation (SHM) frequency. The vertical bar above it shows the distribution of VH genes for each intersection. Within the shared clonotypes, the clones with SHM >= 10% are labelled and shown. **(B)** Sequence logo representing alignment of CDRH3 region of clones common between 2016 and 2018 timepoint of subject 330 with >= 10% SHM. **(C)** Sequence logo representing alignment of CDRH3 region of clones common between 2016 and subject 329 and 330 with >= 10% SHM. **(D)** Sequence logo representing alignment of CDRH3 region of clones common between 2018 timepoint of subject 329 and 330 with >= 10% SHM.

Figure 4: Sequences shared with publicly available HIV bnAbs. Publicly available mAbs mapped to the sequences are shown. Each HIV mAb is shown on the x-axis while and each dot represents the number of sequences mapped to the respective mAb. The criteria for calling mAb a

hit was more that 50% identity in the CDHR3 amino acid and same CDRH3 length. The size of the dot represents the number of sequences mapped and color represents the group. On y-axis mean somatic hypermutation (SHM) is shown.

REFERENCES

1. UNAIDS data 2022 https://www.unaids.org/en/resources/documents/2023/2022_unaids_data.

2. Cuevas, J.M., Geller, R., Garijo, R., López-Aldeguer, J., and Sanjuán, R. (2015). Extremely High Mutation Rate of HIV-1 In Vivo. PLOS Biology *13*, e1002251. 10.1371/journal.pbio.1002251.

3. Cohn, L.B., Chomont, N., and Deeks, S.G. (2020). The Biology of the HIV-1 Latent Reservoir and Implications for Cure Strategies. Cell Host & Microbe *27*, 519–530. 10.1016/j.chom.2020.03.014.

4. Pantophlet, R., and Burton, D.R. (2006). GP120: target for neutralizing HIV-1 antibodies. Annu. Rev. Immunol. *24*, 739–769. 10.1146/annurev.immunol.24.021605.090557.

5. Sok, D., and Burton, D.R. (2018). Recent progress in broadly neutralizing antibodies to HIV. Nat. Immunol. *19*, 1179–1188. 10.1038/s41590-018-0235-7.

6. Kumar, S., Singh, S., and Luthra, K. (2023). An Overview of Human Anti-HIV-1 Neutralizing Antibodies against Diverse Epitopes of HIV-1. ACS Omega *8*, 7252–7261. 10.1021/acsomega.2c07933.

7. Burton, D.R., and Hangartner, L. (2016). Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. Annu. Rev. Immunol. *34*, 635–659. 10.1146/annurev-immunol-041015-055515.

8. del Moral-Sánchez, I., Russell, R.A., Schermer, E.E., Cottrell, C.A., Allen, J.D., Torrents de la Peña, A., LaBranche, C.C., Kumar, S., Crispin, M., Ward, A.B., et al. (2022). High thermostability improves neutralizing antibody responses induced by native-like HIV-1 envelope trimers. npj Vaccines *7*, 1–12. 10.1038/s41541-022-00446-4.

9. Sanders, R.W., Derking, R., Cupo, A., Julien, J.-P., Yasmeen, A., de Val, N., Kim, H.J., Blattner, C., de la Peña, A.T., Korzun, J., et al. (2013). A next-generation cleaved, soluble HIV-1 Env

trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. PLoS Pathog. *9*, e1003618. 10.1371/journal.ppat.1003618.

10. Simonich, C.A., Williams, K.L., Verkerke, H.P., Williams, J.A., Nduati, R., Lee, K.K., and Overbaugh, J. (2016). HIV-1 Neutralizing Antibodies with Limited Hypermutation from an Infant. Cell *166*, 77–87. 10.1016/j.cell.2016.05.055.

11. Kumar, S., Panda, H., Makhdoomi, M.A., Mishra, N., Safdari, H.A., Chawla, H., Aggarwal, H., Reddy, E.S., Lodha, R., Kumar Kabra, S., et al. (2019). An HIV-1 Broadly Neutralizing Antibody from a Clade C-Infected Pediatric Elite Neutralizer Potently Neutralizes the Contemporaneous and Autologous Evolving Viruses. J. Virol. *93*. 10.1128/JVI.01495-18.

12. Goo, L., Chohan, V., Nduati, R., and Overbaugh, J. (2014). Early development of broadly neutralizing antibodies in HIV-1-infected infants. Nat. Med. *20*, 655–658. 10.1038/nm.3565.

13. Kumar, S., Batra, H., Singh, S., Chawla, H., Singh, R., Katpara, S., Hussain, A.W., Das, B.K., Lodha, R., Kabra, S.K., et al. Effect of combination antiretroviral therapy on human immunodeficiency virus 1 specific antibody responses in subtype-C infected children. Journal of General Virology *101*, 1289–1299. 10.1099/jgv.0.001480.

14. Mishra, N., Sharma, S., Dobhal, A., Kumar, S., Chawla, H., Singh, R., Makhdoomi, M.A., Das, B.K., Lodha, R., Kabra, S.K., et al. (2020). Broadly neutralizing plasma antibodies effective against autologous circulating viruses in infants with multivariant HIV-1 infection. Nat Commun *11*, 4409. 10.1038/s41467-020-18225-x.

15. Mishra, N., Sharma, S., Dobhal, A., Kumar, S., Chawla, H., Singh, R., Das, B.K., Kabra, S.K., Lodha, R., and Luthra, K. (2020). A Rare Mutation in an Infant-Derived HIV-1 Envelope Glycoprotein Alters Interprotomer Stability and Susceptibility to Broadly Neutralizing Antibodies Targeting the Trimer Apex. Journal of Virology *94*, e00814-20. 10.1128/JVI.00814-20.

16. Aggarwal, H., Khan, L., Chaudhary, O., Kumar, S., Makhdoomi, M.A., Singh, R., Sharma, K., Mishra, N., Lodha, R., Srinivas, M., et al. (2017). Alterations in B Cell Compartment Correlate

394 with Poor Neutralization Response and Disease Progression in HIV-1 Infected Children. Front
395 Immunol *8*, 1697. 10.3389/fimmu.2017.01697.

396 17. Makhdoomi, M.A., Khan, L., Kumar, S., Aggarwal, H., Singh, R., Lodha, R., Singla, M., Das, B.K.,
397 Kabra, S.K., and Luthra, K. (2017). Evolution of cross-neutralizing antibodies and mapping
398 epitope specificity in plasma of chronic HIV-1-infected antiretroviral therapy-naïve children
399 from India. J. Gen. Virol. *98*, 1879–1891. 10.1099/jgv.0.000824.

400 18. Ditse, Z., Muenchhoff, M., Adland, E., Jooste, P., Goulder, P., Moore, P.L., and Morris, L.
401 (2018). HIV-1 SUBTYPE C INFECTED CHILDREN WITH EXCEPTIONAL NEUTRALIZATION
402 BREADTH EXHIBIT POLYCLONAL RESPONSES TARGETING KNOWN EPITOPES. J. Virol.
403 10.1128/JVI.00878-18.

404 19. Muenchhoff, M., Adland, E., Karimanzira, O., Crowther, C., Pace, M., Csala, A., Leitman, E.,
405 Moonsamy, A., McGregor, C., Hurst, J., et al. (2016). Nonprogressing HIV-infected children share
406 fundamental immunological features of nonpathogenic SIV infection. Sci Transl Med *8*,
407 358ra125. 10.1126/scitranslmed.aag1048.

408 20. Muenchhoff, M., Prendergast, A.J., and Goulder, P.J.R. (2014). Immunity to HIV in Early Life.
409 Front Immunol *5*, 391. 10.3389/fimmu.2014.00391.

410 21. Mishra, N., Makhdoomi, M.A., Sharma, S., Kumar, S., Dobhal, A., Kumar, D., Chawla, H.,
411 Singh, R., Kanga, U., Das, B.K., et al. (2019). Viral characteristics associated with maintenance of
412 elite neutralizing activity in chronically HIV-1 clade C infected monozygotic pediatric twins.
413 Journal of Virology, JVI.00654-19. 10.1128/JVI.00654-19.

414 22. Khan, L., Makhdoomi, M.A., Kumar, S., Nair, A., Andrabi, R., Clark, B.E., Auyeung, K.,
415 Bhattacharya, J., Vajpayee, M., Wig, N., et al. (2015). Identification of CD4-Binding Site
416 Dependent Plasma Neutralizing Antibodies in an HIV-1 Infected Indian Individual. PLoS ONE
417 *10*, e0125575. 10.1371/journal.pone.0125575.

418 23. Kumar, S., Kumar, R., Khan, L., Makhdoomi, M.A., Thiruvengadam, R., Mohata, M., Agarwal,
419 M., Lodha, R., Kabra, S.K., Sinha, S., et al. (2017). CD4-Binding Site Directed Cross-Neutralizing

scFv Monoclonals from HIV-1 Subtype C Infected Indian Children. Front Immunol *8*, 1568. 10.3389/fimmu.2017.01568.

24. Khan, L., Kumar, R., Thiruvengadam, R., Parray, H.A., Makhdoomi, M.A., Kumar, S., Aggarwal, H., Mohata, M., Hussain, A.W., Das, R., et al. (2017). Cross-neutralizing anti-HIV-1 human single chain variable fragments(scFvs) against CD4 binding site and N332 glycan identified from a recombinant phage library. Sci Rep *7*, 45163. 10.1038/srep45163.

25. Kumar, S., Kumar, R., Makhdoomi, M., Khan, L., Prakash, S., Thiruvengadam, R., Singla, M., Lodha, R., Kabra, S., Sinha, S., et al. (2014). Production of cross neutralizing single chain fragment variables (scFv) from HIV-1 infected Indian children. BMC Infectious Diseases *14*, E25. 10.1186/1471-2334-14-S3-E25.

26. Cale, E.M., Gorman, J., Radakovich, N.A., Crooks, E.T., Osawa, K., Tong, T., Li, J., Nagarajan, R., Ozorowski, G., Ambrozak, D.R., et al. (2017). Virus-like Particles Identify an HIV V1V2 Apex-Binding Neutralizing Antibody that Lacks a Protruding Loop. Immunity *46*, 777-791.e10. 10.1016/j.immuni.2017.04.011.

27. Shen, C.-H., DeKosky, B.J., Guo, Y., Xu, K., Gu, Y., Kilam, D., Ko, S.H., Kong, R., Liu, K., Louder, M.K., et al. (2020). VRC34-Antibody Lineage Development Reveals How a Required Rare Mutation Shapes the Maturation of a Broad HIV-Neutralizing Lineage. Cell Host Microbe *27*, 531-543.e6. 10.1016/j.chom.2020.01.027.

28. Watkins, J.D., Siddappa, N.B., Lakhashe, S.K., Humbert, M., Sholukh, A., Hemashettar, G., Wong, Y.L., Yoon, J.K., Wang, W., Novembre, F.J., et al. (2011). An Anti-HIV-1 V3 Loop Antibody Fully Protects Cross-Clade and Elicits T-Cell Immunity in Macaques Mucosally Challenged with an R5 Clade C SHIV. PLOS ONE *6*, e18207. 10.1371/journal.pone.0018207.

29. Sabin, C., Corti, D., Buzon, V., Seaman, M.S., Lutje Hulsik, D., Hinz, A., Vanzetta, F., Agatic, G., Silacci, C., Mainetti, L., et al. (2010). Crystal structure and size-dependent neutralization properties of HK20, a human monoclonal antibody binding to the highly conserved heptad repeat 1 of gp41. PLoS Pathog *6*, e1001195. 10.1371/journal.ppat.1001195.

30. Bonsignori, M., Kreider, E.F., Fera, D., Meyerhoff, R.R., Bradley, T., Wiehe, K., Alam, S.M., Aussedat, B., Walkowicz, W.E., Hwang, K.-K., et al. (2017). Staged induction of HIV-1 glycan-dependent broadly neutralizing antibodies. Sci Transl Med *9*. 10.1126/scitranslmed.aai7514.

31. Simonich, C., Shipley, M.M., Doepker, L., Gobillot, T., Garrett, M., Cale, E.M., Hennessy, B., Itell, H., Chohan, V., Doria-Rose, N., et al. (2021). A diverse collection of B cells responded to HIV infection in infant BG505. Cell Rep Med *2*, 100314. 10.1016/j.xcrm.2021.100314.

32. Bonsignori, M., Zhou, T., Sheng, Z., Chen, L., Gao, F., Joyce, M.G., Ozorowski, G., Chuang, G.-Y., Schramm, C.A., Wiehe, K., et al. (2016). Maturation Pathway from Germline to Broad HIV-1 Neutralizer of a CD4-Mimic Antibody. Cell *165*, 449–463. 10.1016/j.cell.2016.02.022.

33. Andrabi, R., Williams, C., Wang, X.-H., Li, L., Choudhary, A.K., Wig, N., Biswas, A., Luthra, K., Nadas, A., Seaman, M.S., et al. (2013). Cross-neutralizing activity of human anti-V3 monoclonal antibodies derived from non-B clade HIV-1 infected individuals. Virology *439*, 81–88. 10.1016/j.virol.2012.12.010.

34. Cottrell, C.A., Manne, K., Kong, R., Wang, S., Zhou, T., Chuang, G.-Y., Edwards, R.J., Henderson, R., Janowska, K., Kopp, M., et al. (2021). Structural basis of glycan276-dependent recognition by HIV-1 broadly neutralizing antibodies. Cell Rep *37*, 109922. 10.1016/j.celrep.2021.109922.

35. Williams, L.D., Ofek, G., Schätzle, S., McDaniel, J.R., Lu, X., Nicely, N.I., Wu, L., Lougheed, C.S., Bradley, T., Louder, M.K., et al. (2017). Potent and broad HIV-neutralizing antibodies in memory B cells and plasma. Sci Immunol *2*, eaal2200. 10.1126/sciimmunol.aal2200.

36. Andrabi, R., Bhiman, J.N., and Burton, D.R. (2018). Strategies for a multi-stage neutralizing antibody-based HIV vaccine. Curr. Opin. Immunol. *53*, 143–151. 10.1016/j.coi.2018.04.025.

37. Burton, D.R. (2017). What Are the Most Powerful Immunogen Design Vaccine Strategies? Reverse Vaccinology 2.0 Shows Great Promise. Cold Spring Harb Perspect Biol *9*. 10.1101/cshperspect.a030262.

38. Briney, B., Inderbitzin, A., Joyce, C., and Burton, D.R. (2019). Commonality despite exceptional diversity in the baseline human antibody repertoire. Nature *566*, 393–397. 10.1038/s41586-019-0879-y.

39. Nabi, R., Moldoveanu, Z., Wei, Q., Golub, E.T., Durkin, H.G., Greenblatt, R.M., Herold, B.C., Nowicki, M.J., Kassaye, S., Cho, M.W., et al. (2017). Differences in serum IgA responses to HIV-1 gp41 in elite controllers compared to viral suppressors on highly active antiretroviral therapy. PLoS One *12*, e0180245. 10.1371/journal.pone.0180245.

40. Xu, Z., Ismanto, H.S., Zhou, H., Saputri, D.S., Sugihara, F., and Standley, D.M. (2022). Advances in antibody discovery from human BCR repertoires. Frontiers in Bioinformatics *2*.

41. Wang, Y., Yuan, M., Lv, H., Peng, J., Wilson, I.A., and Wu, N.C. (2022). A large-scale systematic survey reveals recurring molecular features of public antibody responses to SARS-CoV-2. Immunity, S1074-7613(22)00142-X. 10.1016/j.immuni.2022.03.019.

42. Yoon, H., Macke, J., West, A.P., Jr, Foley, B., Bjorkman, P.J., Korber, B., and Yusim, K. (2015). CATNAP: a tool to compile, analyze and tally neutralizing antibody panels. Nucleic Acids Research *43*, W213–W219. 10.1093/nar/gkv404.

43. Durham, N.D., Agrawal, A., Waltari, E., Croote, D., Zanini, F., Fouch, M., Davidson, E., Smith, O., Carabajal, E., Pak, J.E., et al. (2019). Broadly neutralizing human antibodies against dengue virus identified by single B cell transcriptomics. Elife *8*, e52384. 10.7554/eLife.52384.

44. Gupta, N.T., Vander Heiden, J.A., Uduman, M., Gadala-Maria, D., Yaari, G., and Kleinstein, S.H. (2015). Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. Bioinformatics *31*, 3356–3358. 10.1093/bioinformatics/btv359.

45. Vander Heiden, J.A., Yaari, G., Uduman, M., Stern, J.N.H., O'Connor, K.C., Hafler, D.A., Vigneault, F., and Kleinstein, S.H. (2014). pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics *30*, 1930–1932. 10.1093/bioinformatics/btu138.

Figure 1

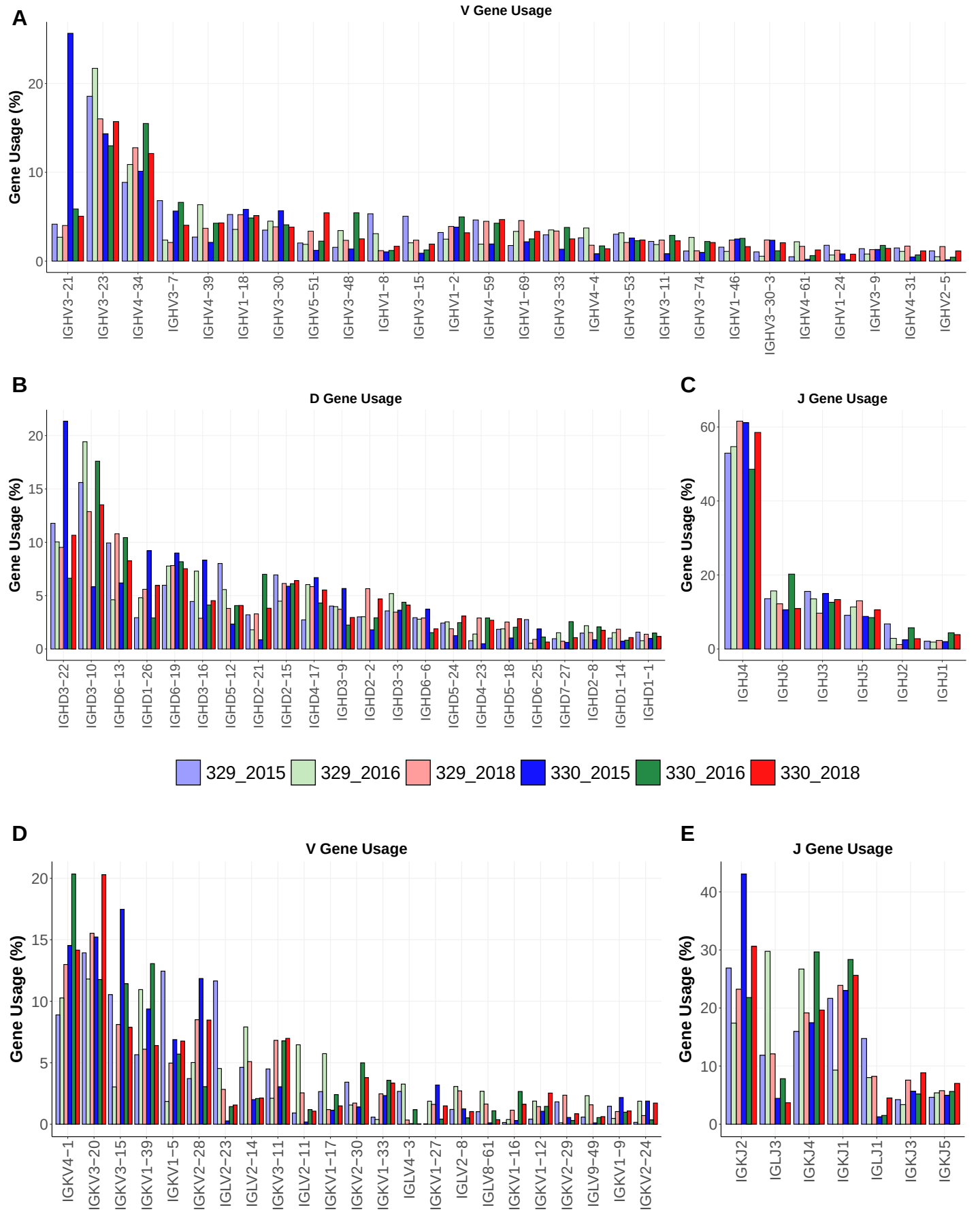

A
V Gene Usage
Gene Usage (%)
0
10
20
IGHV3−21
IGHV3−23
IGHV4−34
IGHV3−7
IGHV4−39
IGHV1−18
IGHV3−30
IGHV5−51
IGHV3−48
IGHV1−8
IGHV3−15
IGHV1−2
IGHV4−59
IGHV1−69
IGHV3−33
IGHV4−4
IGHV3−53
IGHV3−11
IGHV3−74
IGHV1−46
IGHV3−30−3
IGHV4−61
IGHV1−24
IGHV3−9
IGHV4−31
IGHV2−5
B
D Gene Usage
Gene Usage (%)
0
5
10
15
20
IGHD3−22
IGHD3−10
IGHD6−13
IGHD1−26
IGHD6−19
IGHD3−16
IGHD5−12
IGHD2−21
IGHD2−15
IGHD4−17
IGHD3−9
IGHD2−2
IGHD3−3
IGHD6−6
IGHD5−24
IGHD4−23
IGHD5−18
IGHD6−25
IGHD7−27
IGHD2−8
IGHD1−14
IGHD1−1
C
J Gene Usage
Gene Usage (%)
0
20
40
60
IGHJ4
IGHJ6
IGHJ3
IGHJ5
IGHJ2
IGHJ1
329_2015
329_2016
329_2018
330_2015
330_2016
330_2018
D
V Gene Usage
Gene Usage (%)
0
5
10
15
20
IGKV4−1
IGKV3−20
IGKV3−15
IGKV1−39
IGKV1−5
IGKV2−28
IGLV2−23
IGLV2−14
IGKV3−11
IGLV2−11
IGKV1−17
IGKV2−30
IGKV1−33
IGLV4−3
IGKV1−27
IGLV2−8
IGLV8−61
IGKV1−16
IGKV1−12
IGKV2−29
IGLV9−49
IGKV1−9
IGKV2−24
E
J Gene Usage
Gene Usage (%)
0
10
20
30
40
IGKJ2
IGLJ3
IGKJ4
IGKJ1
IGLJ1
IGKJ3
IGKJ5

Figure 2

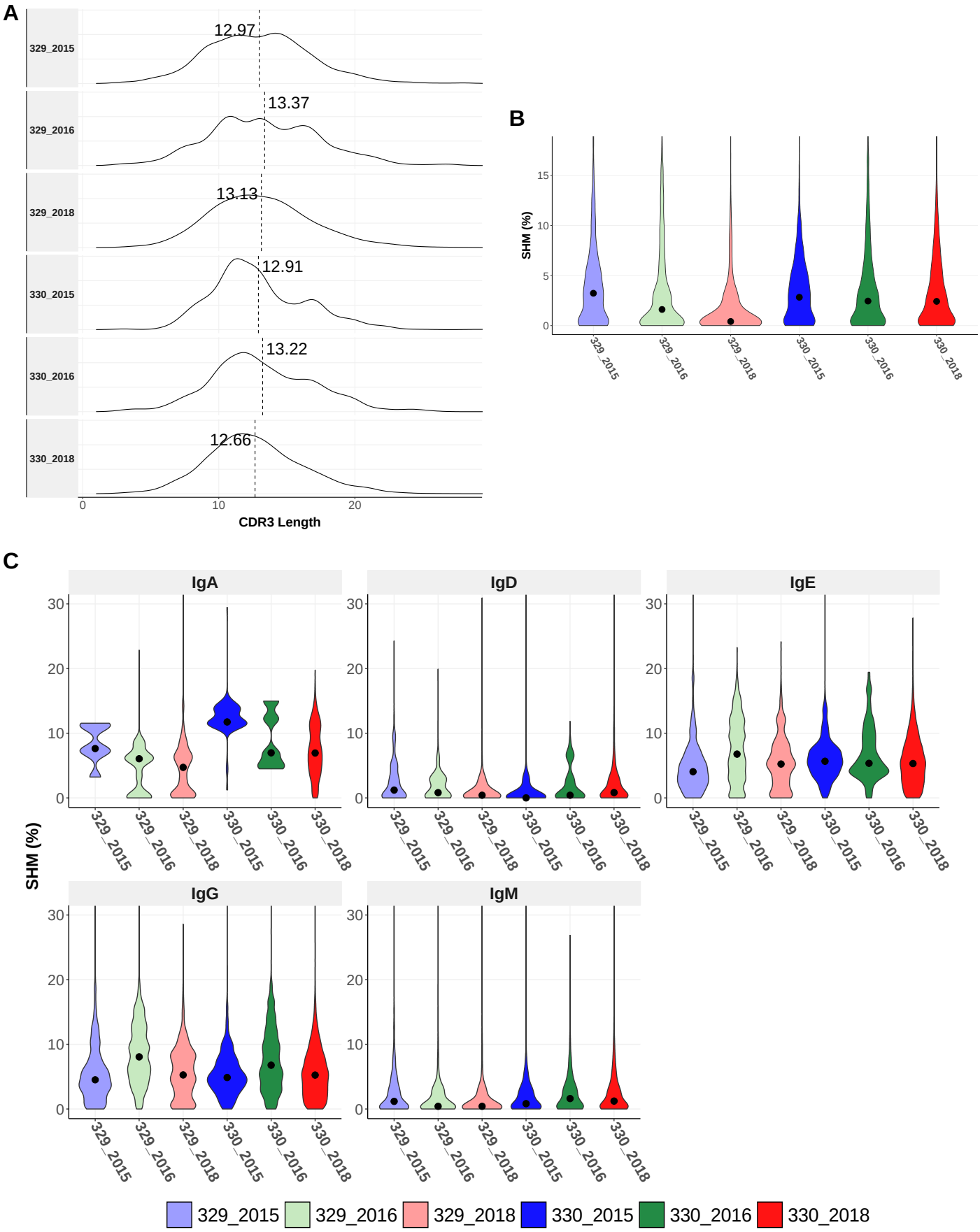

A
329_2015
12.97
329_2016
13.37
329_2018
13.13
330_2015
12.91
330_2016
13.22
330_2018
12.66
0
10
20
CDR3 Length
B
SHM (%)
0
5
10
15
329_2015
329_2016
329_2018
330_2015
330_2016
330_2018
C
IgA
IgD
IgE
IgG
IgM
SHM (%)
0
10
20
30
329_2015
329_2016
329_2018
330_2015
330_2016
330_2018

Figure 3

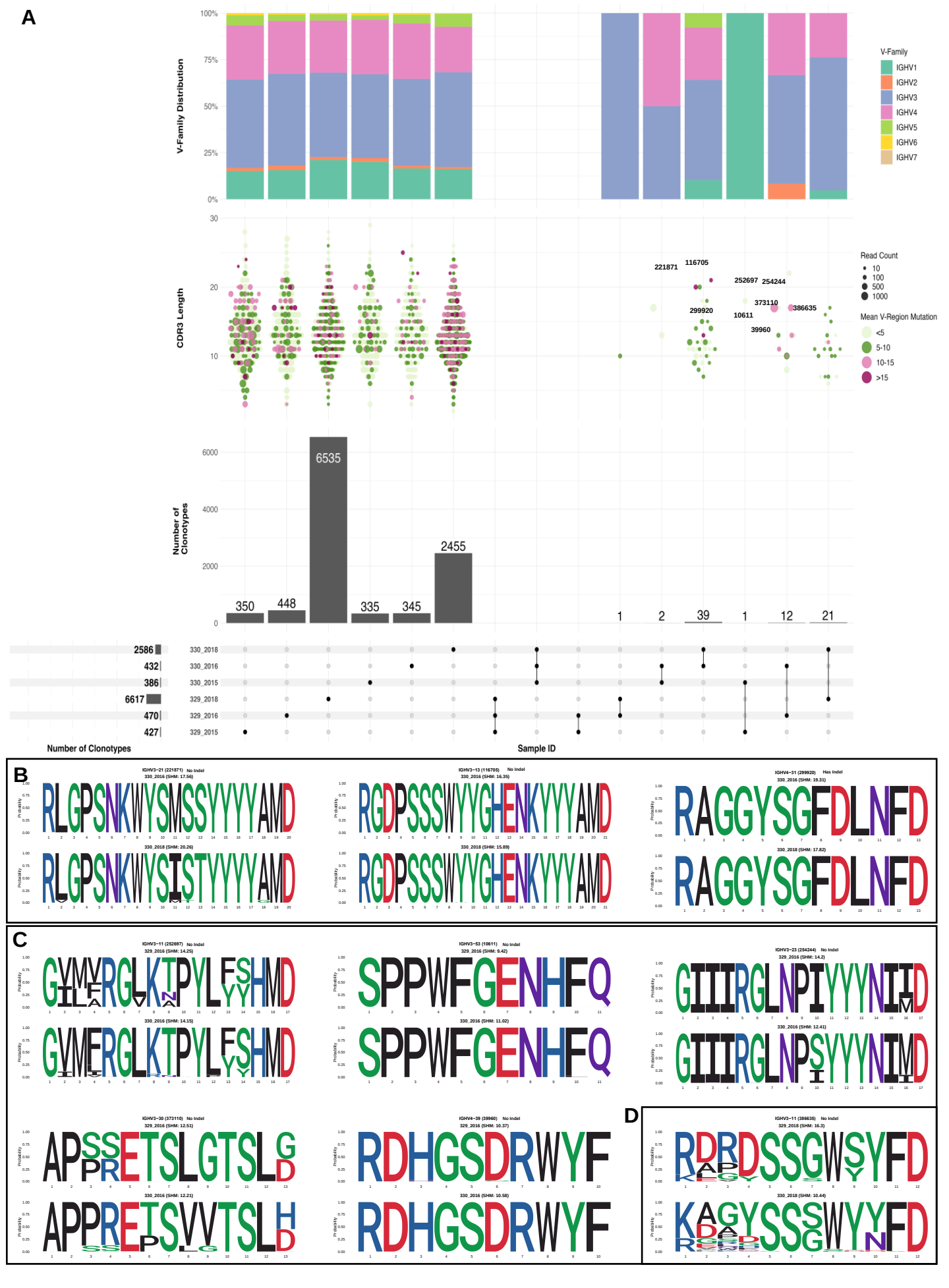
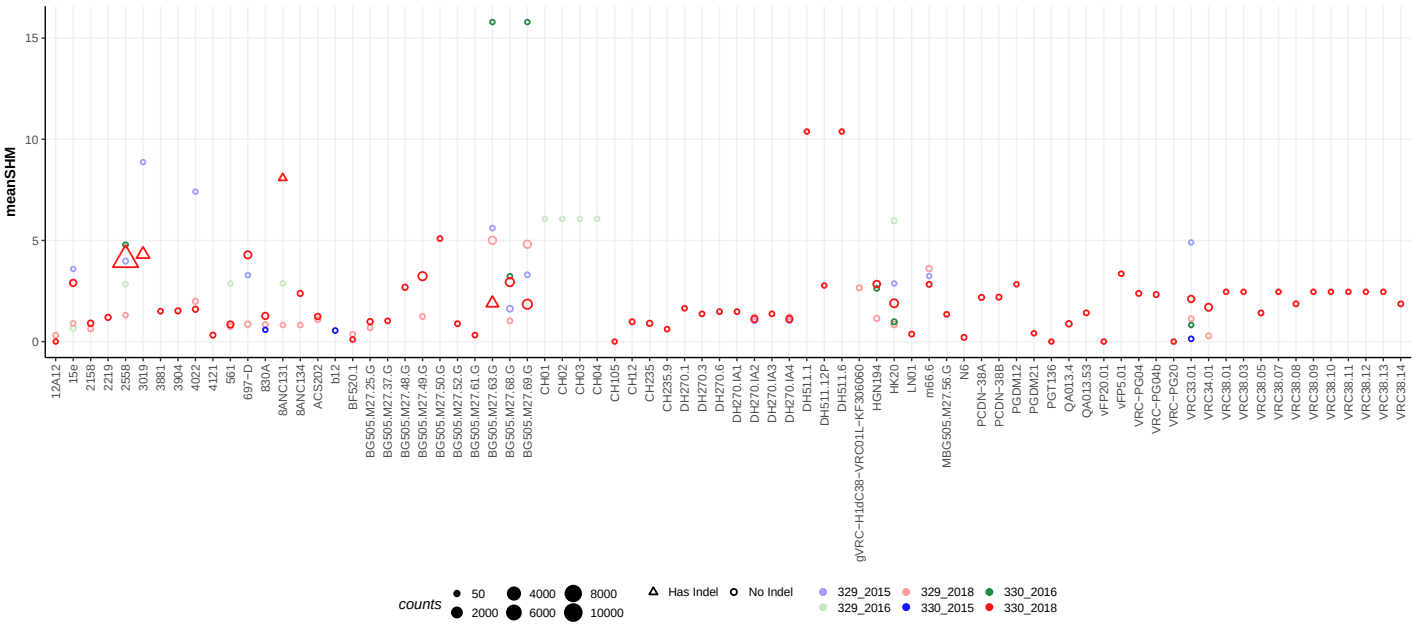

A
V-Family Distribution
100%
75%
50%
25%
0%
V-Family
IGHV1
IGHV2
IGHV3
IGHV4
IGHV5
IGHV6
IGHV7
CDR3 Length
30
20
10
Read Count
10
100
500
1000
Mean V-Region Mutation
<5
5-10
10-15
>15
221871
116705
252697
254244
373110
386635
299920
10611
39960
Number of Clonotypes
6000
4000
2000
0
6535
2455
350
448
335
345
1
2
39
1
12
21
2586
432
386
6617
470
427
330_2018
330_2016
330_2015
329_2018
329_2016
329_2015
Number of Clonotypes
Sample ID
B
IGHV3-21 (221871) No Indel
330_2016 (SHM: 17.56)
330_2018 (SHM: 20.26)
IGHV3-13 (116705) No Indel
330_2016 (SHM: 16.35)
330_2018 (SHM: 15.89)
IGHV4-31 (299920) Has Indel
330_2016 (SHM: 19.31)
330_2018 (SHM: 17.82)
C
IGHV3-11 (252697) No Indel
329_2016 (SHM: 14.25)
330_2016 (SHM: 14.15)
IGHV3-53 (10611) No Indel
329_2016 (SHM: 9.42)
330_2016 (SHM: 11.02)
IGHV3-23 (254244) No Indel
329_2016 (SHM: 14.2)
330_2016 (SHM: 12.41)
IGHV3-30 (373110) No Indel
329_2016 (SHM: 12.51)
330_2016 (SHM: 12.21)
IGHV4-39 (39960) No Indel
329_2016 (SHM: 10.37)
330_2016 (SHM: 10.58)
D
IGHV3-11 (386635) No Indel
329_2018 (SHM: 16.3)
330_2018 (SHM: 10.44)
Probability

Figure 4


meanSHM
15
10
5
0
12A12
15e
2158
2219
2558
3019
3881
3904
4022
4121
561
697−D
830A
8ANC131
8ANC134
ACS202
b12
BF520.1
BG505.M27.25.G
BG505.M27.37.G
BG505.M27.48.G
BG505.M27.49.G
BG505.M27.50.G
BG505.M27.52.G
BG505.M27.61.G
BG505.M27.63.G
BG505.M27.68.G
BG505.M27.69.G
CH01
CH02
CH03
CH04
CH105
CH12
CH235
CH235.9
DH270.1
DH270.3
DH270.6
DH270.IA1
DH270.IA2
DH270.IA3
DH270.IA4
DH511.1
DH511.12P
DH511.6
gVRC−H1dC38−VRC01L−KF306060
HGN194
HK20
LN01
m66.6
MBG505.M27.56.G
N6
PCDN−38A
PCDN−38B
PGDM12
PGDM21
PGT136
QA013.4
QA013.53
vFP20.01
vFP5.01
VRC−PG04
VRC−PG04b
VRC−PG20
VRC33.01
VRC34.01
VRC38.01
VRC38.03
VRC38.05
VRC38.07
VRC38.08
VRC38.09
VRC38.10
VRC38.11
VRC38.12
VRC38.13
VRC38.14
counts
50
2000
4000
6000
8000
10000
Has Indel
No Indel
329_2015
329_2016
329_2018
330_2015
330_2016
330_2018