

1 **B cell receptor analysis using single cell sequencing reveals preferential antibody**  
2 **expression in a HIV candidate vaccine study.**

3 Irene Bassano<sup>1</sup>, Simon Watson<sup>2</sup>, Cinzia Malangone<sup>3</sup>, Christopher L. Pinder<sup>1</sup>, Kostas Paschos<sup>1</sup>,  
4 Hannah M. Cheeseman<sup>1</sup>, Robin J. Shattock<sup>1</sup>, Paul Kellam<sup>1,2,\*</sup>

5 <sup>1</sup> Imperial College London, Department of Infectious Disease, London W21PG, United  
6 Kingdom

7 <sup>2</sup> Kymab, Babraham Research Campus, Cambridge CB22 3AT, United Kingdom

8 <sup>3</sup> Illumina, Illumina Centre, 19 Granta Park, Great Abington, Cambridge, CB21 6DF, United  
9 Kingdom

10

11 \* Corresponding author: [i.bassano@imperial.ac.uk](mailto:i.bassano@imperial.ac.uk) and/or [p.kellam@imperial.ac.uk](mailto:p.kellam@imperial.ac.uk)

12

13

14

15 **Keywords:** single cell RNA sequencing, B cells, V(D)J recombination, 10x Chromium, cell  
16 hashing

17

## 18 **Abstract**

19 HIV (human immunodeficiency virus) is a virus that infects different cell types in the immune  
20 system. While different monoclonal antibodies have been identified, because of the antigenic  
21 diversity of HIV, defining broad neutralising antibodies (bnAbs) it has been very challenging.  
22 Using the 10x Chromium technology partitioning system, we performed single-cell sequencing  
23 of the B cell repertoire and analysed a cohort of 48 HIV-negative subjects randomised into 4  
24 groups each receiving a different HIV vaccine regime to investigate the type of antibody  
25 response to each of the proteins in the vaccine groups, namely ConM SOSIP, EDC ConM  
26 SOSIP, ConS UFO and EDC ConS UFO. We recovered over >300,000 single cells and  
27 reconstructed full-length antibody heavy- and light-chain variable regions from individual  
28 memory B cells, plasmablasts, naïve B cells and activated B cells as well as retrieved their  
29 transcriptome profile to correctly separate the different populations within each individual. We  
30 identified seven clonotypes representing functional antibodies against vaccine candidate  
31 proteins. We reconstructed their phylogenetic tree and confirmed they were novel antibodies  
32 in response to HIV. These were mainly antibodies in response to two of the four protein  
33 candidates, namely ConS UFO and EDC ConS UFO.  
34

## 35 Introduction

36 B cells of the adaptive immune system produce antibodies that are specific for an antigen, with  
37 antibody diversity achieved by immunoglobulin gene V(D)J rearrangement and somatic  
38 hypermutation (SHM). During their development, immature B cells rearrange their heavy and  
39 light chain germline immunoglobulin genes to produce functional B cell receptors and exit the  
40 bone marrow as mature antigen naïve B cells, defined by expression of the cell surface markers  
41 CD19, CD20, IgM and IgD [1]. Upon antigen encounter, naïve B cells migrate to germinal  
42 centres (GCs) where they undergo class switch recombination (CSR) and antigen affinity  
43 maturation via extensive SHM [2]. B cells proliferate within GC follicles of secondary  
44 lymphoid organs, eventually differentiating into high affinity immunoglobulin-secreting  
45 plasma cells and memory B cells.

46  
47 Memory B cells retain a “memory” of the antigen against which they matured, allowing for an  
48 anamnestic immune response upon subsequent exposures to an antigen. Pathogen genetic  
49 variation can limit the effectiveness of such an antibody response and in the case of persistent  
50 infection by HIV such variation leads to escape from antibody neutralisation. However, in  
51 some cases, HIV-positive patients develop broadly neutralising antibodies (bnAbs) [3], with  
52 high level of somatic hypermutation (SHM) and long CDR3s, which can recognize epitopes  
53 across a broad range of HIV genetic subtypes. This raises the possibility that if a vaccine can  
54 elicit such bnAbs then an effective HIV vaccine maybe achievable [4].

55  
56 Indeed, over the years several HIV vaccine candidates have been proposed based on their  
57 efficacy, safety and immunogenicity: these antibodies target the native envelope protein (Env)  
58 and include recombinant Env proteins with or without adjuvants, HIV viral like particles  
59 (VLPs), HIV DNA plasmids, or replication-competent or incompetent viral vectors[5]. The  
60 main goal is for these antibodies to block either the Env-CD4 receptor or the virion fusion with  
61 host cells [6]. BnAbs are known to target the following six epitopes on the Env protein: MPER  
62 (membrane proximal external region) of gp41, V1V2-glycan, the outer domain glycan, V3-  
63 glycan and the CD4 binding site [7].

64 Current vaccine developments commonly face many obstacles, one of which being the high  
65 diversity of the circulating virus itself. Despite this, many trials have been developed, with  
66 different degrees of success. Among those in the early phase clinical vaccine trial, it is worth  
67 mentioning the IAV G001 (NCT03547245) study which used eOD-GT8 as immunogen (a self-  
68 assembling nanoparticle with HIV Env protein) and AS01B as adjuvant. This trial ran in USA  
69 in both women and men between June 2018 and December 2020. Another clinical trial, the  
70 EAVI2020\_01 (NCT03816137) used ConM SOSIP, EDC ConM SOSIP, ConS UFO, EDC  
71 ConS UFO and mosaic SOSIPs and MPLA as adjuvant and ran in UK in both women and men  
72 between March 2019 to present. Among the HIV vaccine efficacy trials, VAX003 and 004  
73 were one of the first ones but did not prevent HIV infection in both men and women that were  
74 enrolled. On the other hand, the RV144 (NCT00223080) is one of the most successful to date,  
75 with a 31% efficacy. This trial employed a pox-protein prime-boost strategy with ALVAC-  
76 HIV (vCP1521) expressing Gag and Pro (subtype B LAI) CRF01\_AE gp120 (92TH023) linked  
77 to transmembrane anchoring portion of gp41 (LAI) AIDS VAX B/E absorbed to Aluminium  
78 hydroxide. Another notable efficacy trial was led by Johnson and Johnson between November  
79 2017 and July 2021 in south Africa. The Imbokodo HVTN 705 (NCT03060629) Phase 2b  
80 vaccine efficacy trial employed Ad26.Mos4.HIV Subtype C gp140 bound to aluminium  
81 phosphate. Current results were expected in July 2022 but the trial was halted following a  
82 25.5% efficacy result, failing to offer substantial protection against HIV [8].

83

84 The concept of designing a vaccine starting with known antibodies is known as “reverse  
85 vaccinology” [9] and can also be based on the information retrieved from deep sequencing of  
86 the BCR repertoire, which allows us to determine the type of antibodies produced by an HIV  
87 infected person or patients under clinical trials. This is different from the classical methodology  
88 of retrieving antibody repertoire by sequencing antigen sorted cells [10]. In the past decade,  
89 RNA-Seq has been widely used to assess gene expression levels within cell populations.  
90 Comparison of treated and untreated samples has given us an exhaustive overview of how the  
91 transcriptome can change under different conditions providing an insight into what an  
92 infection, antigen, or cancerous status would trigger within the cells. In addition, the more  
93 sophisticated method of single cell RNA-Seq (sc-RNA-Seq) has been developed over the last  
94 decade. Thanks to techniques such as SMARTSeq and single cell droplets methods, we are  
95 now able to sequence thousands of cells simultaneously and discover the diversity of  
96 subpopulations which otherwise would be lost in bulk sequencing. In addition, with the advent  
97 of CiteSeq and hashing methodologies, we are now for the first time able to simultaneously  
98 assess different samples in single encapsulation runs saving both costs and time towards the  
99 final libraries. These methods have been successfully used to study both B and T cell  
100 repertoires.

101  
102 Here we applied single cell sequencing to B cell populations to investigate the dynamics of  
103 antibody diversity following vaccination with novel HIV immunogens. We show that coupled  
104 B cell receptor and transcriptome sequencing allows simultaneous B cell phenotyping and BCR  
105 repertoire analysis, and that B cell lineages in peripheral blood B cells can converge in different  
106 participants following vaccination.  
107

## 108 **Material and methods**

### 109 **Clinical trial patients**

110 Samples from day 7 following the first vaccine dose for 48 HIV-negative participants of the  
111 EAVI2020\_01 (NCT03816137) clinical trial against HIV [11] designed to assess if different  
112 prime-boost combinations with novel HIV immunogens could influence serum neutralising  
113 antibody breadth, were used here. These 48 patients were randomised into 5 groups each  
114 receiving a different vaccine regime (Supplementary Table 1). As two groups (C and E) had  
115 the same vaccine regime for first injection and different vaccine regimes for the boosts, for this  
116 initial time point we combined these two groups. Here therefore, we have used samples from  
117 participants immunised with A) 100µg of an HIV trimer based on a consensus sequence of all  
118 HIV-1 group M isolates from 2004 (ConM SOSIP), with (Group B) or without (Group A) EDC  
119 (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide) (EDC ConM SOSIP), or Group D, which  
120 received 100µg of ConS UFO (Uncleaved pre-fusion optimised) HIV trimer, based on a  
121 consensus sequence from 2001, with (Group D) or without (Group C and E) addition of EDC  
122 (EDC ConS UFO) [12-16] (Supplementary Table1). The two vaccine protein candidates  
123 represent stabilised versions of the native Env protein. Briefly, the ConM SOSIP contains an  
124 extra intramolecular disulphide bond (SOS) to stabilize the gp120/gp41 interaction, in addition  
125 to a single amino acid point mutation at position 559 where an Isoleucine is replaced by a  
126 Proline (IP) [17-19]. ConS UFO forms a closed pre-fusion membrane-bound Env trimer and  
127 contains a short flexible amino acid linker to join gp120 and gp41 subunits. Therefore, ConM  
128 and ConS use two different approaches to prevent Env dissociation into monomers [13], while  
129 the addition of EDC, a crosslinking agent, preserves the nature of the bnAbs.

130

### 131 **Isolation of B cell populations**

132 Up to  $3 \times 10^7$  frozen peripheral blood mononuclear cells (PBMCs) from 48 HIV-negative donors  
133 were collected 7 days post vaccination. The immediate response to the vaccine will be led by  
134 antibody-secreting plasmablasts, while 7 days later memory B cells will proliferate with the  
135 membrane-bound form of the antibodies.

136 Memory B cell and plasmablasts were isolated using the Miltenyi Pan B and CD27 Microbead  
137 isolation kits according to manufacturer's instructions. Briefly, frozen PBMC were washed in  
138 RPMI media containing 10% FBS and twice with PBS to remove excess DMSO. Cells were  
139 subsequently incubated with a Pan-B cell biotin-antibody cocktail to remove non-B cells by  
140 negative selection. Cell populations from the eluted fraction were then enriched by positive  
141 selection using CD27 microbeads to capture all cells expressing CD27 on their surface.  
142 Although the two kits were specific for memory B cells and plasmablasts from the same pool,  
143 we were also able to simultaneously isolate naïve and activated B cells based on the expression  
144 of certain cell marker transcripts during bioinformatic analysis. Enriched cell populations were  
145 then counted and assessed for viability. Only cell populations with a viability above 85-90%  
146 were considered for single cell encapsulation.

147

### 148 **Cell hashing and 10x microencapsulation**

149 Purified B cell populations underwent the cell hashing protocol as described in [20]. The  
150 following hashing antibodies from Biolegend were used at a final concentration of 1µg/ml:  
151 TotalSeqC-251, 252, 253, 254, 255, 256, 257, 258, 259, and 260. All of the TotalSeq –C

152 antibodies use a combination of two clones that recognize CD298 and  $\beta$ 2 microglobulin,  
153 respectively. Following hashing of the single donors, cell numbers and viabilities were  
154 determined, samples pooled, and then loaded onto a 10x Chromium at 17000 cells per lane,  
155 with an expected recovery of 10000 cells. Libraries were prepared as per protocol [21]. Three  
156 types of libraries were prepared: V(D)J, whole transcriptome (GEX, gene expression) and  
157 feature barcoding (FB). The latter was used to deconvolve to each individual sample based on  
158 the unique sample barcode present. Libraries were normalised to 4nM and sequenced on  
159 Illumina platforms.

160

## 161 **Sequencing and data analysis**

162 Libraries were sequenced using an HiSeq 4000 or Novaseq6 from Novogene. Fastq samples  
163 were processed for quality control, adapters removed, and sequence analysis performed using  
164 CellRanger v3.1 and v5. Hashed samples were demultiplexed using CiteSeq v1.4.0 together  
165 with Seurat v.3.0 and further partitioned using Loupe Cell Browser v4.0.0 (10x Genomics)  
166 based on the sequence tag of the cell hashing antibodies. Duplets, which represent single  
167 droplets containing more than one encapsulated cell, were identified either using Scrublet [22],  
168 CiteSeq, Seurat or Loupe Cell Browser. V(D)J sequences were assembled from cells that  
169 expressed both paired heavy and light chain mRNA. Of these, only full length and productive  
170 IgH and IgL transcripts were considered for final analysis (Figure 1) using enclone  
171 (<https://10xgenomics.github.io/enclone/>). Cell type separation was performed using a list of  
172 transcription factors which define B cell subpopulations, specifically for memory B cells,  
173 plasmablasts, activated B cells and naïve B cells (ref) (Supplementary Table 2), thereby  
174 allowing IgH and IgL sequences and their respective clonotypes to be assigned using gene  
175 expression data to B cell type gene expression using Loupe Cell Browser. Additional custom  
176 scripts were used to calculate V, D, J, and C gene usage and generate tables with V(D)J gene  
177 combinations for display using Circos plots.

178

## 179 **Phylogenetic analysis**

180 Nucleic acid sequences of the  $V_H$  and  $V_L$  domain of each antibody in a clonogroup were  
181 concatenated and exported in FASTA format. The inferred germline IGHV, IGHJ, IGLV, and  
182 IGLJ genes were then added to the FASTA file. IGHD genes were not included due to the  
183 difficulty in unambiguously inferring the correct gene. The germline sequence was manually  
184 aligned to the rest of the concatenated  $V_H/V_L$  sequences from the clones in the clonogroup  
185 using AliView version 1.26 [23]. Neighbour-joining (NJ) tree was inferred from the alignment  
186 using the phangorn R library and the resultant tree was rooted on the germline taxon. Metadata  
187 such as  $V_H$  isotype, cell type, sample group and sample ID were associated to the phylogenetic  
188 tree using the ggtree R library.

189

190



## 191 **Results**

### 192 **Cell recovery and sample separation using expression profile of the hashing antibodies**

193 Human peripheral blood B cells consist mainly of naïve and memory B cells, however,  
194 immediately following antigen encounter a transient population of plasmablast B cells can be  
195 found in the peripheral blood corresponding to B cell affinity maturation in secondary  
196 lymphoid organs. Therefore, by obtaining PBMC samples on day 7 following immunisation  
197 with the different HIV immunogens we can maximise the chance of detecting vaccine antigen-  
198 responsive plasmablasts and then relate these to the subset of recently produced memory B  
199 cells to the same immunogen. We therefore enriched for plasmablasts and memory B cells from  
200 the peripheral blood at this timepoint using magnetic cell sorting for CD27 positive B cells.  
201 To simultaneously obtain B cell receptor sequences and cell type identification, we performed  
202 single cell sequencing by combining V(D)J sequencing and whole transcriptome sequencing.  
203 Furthermore, we were able to pool multiple patient samples by labelling them with hashing  
204 antibodies linked to a unique sample identifier sequence tag for each sample. Using these  
205 methods, we were able to combine up to 4 different patient samples in one 10x run (Figure 1b).  
206 Cell hashing of samples allows an estimation of the number of times ‘single cell’ sequencing  
207 occurs with a droplet containing more than one cell. Such ‘duplets’ were identified where an  
208 apparent single cell transcriptome contained two distinct cell-hashing sequences, rather than  
209 the assumed single hashing antibody sequence tag present per cell. For each multiplex hashed  
210 sample, we pairwise compared hashing sequence where single cells that were not only on the  
211 x or y axis represent potential duplets (Supplementary Figure 1a). Duplets were also identified  
212 using Seurat/CiteSeq (including for non-hashed, single sample single cell sequences) and by  
213 using Single-Cell Remover of Doublets (Scrublet)(25). Analysis of tSNE gene expression  
214 showed such duplets were not clustered in a distinct gene expression space for B cells and so  
215 could not be triaged based on a gene expression pattern (Supplementary Figure 1b). Combining  
216 these multiple duplet removal methods yielded a final single cell dataset for further analysis.  
217

### 218 **B cell populations and transcriptional profile in ConM SOSIP and ConS UFO samples**

219 Transcription factor gene expression can be used to define cell types and therefore we compiled  
220 a list of B cell transcription factors [24, 25] for the classification of B cell types from the  
221 scRNAseq data. (Supplementary Table 1). Four B cell populations were defined: memory B  
222 cells (MB), plasmablasts (PB), activated B cells (AC) and naïve B cells (N), along with gene  
223 expression markers for Leukocytes, T cells, erythrocytes, and monocytes. Cells that did not fall  
224 into these criteria were classified as undefined B cells (UB) (Figure 2b and 3a,b). These latter  
225 cells had a very low level of UMIs for the transcription factors we used to differentiate into the  
226 population of interest, but nevertheless were grouped into one cluster and analysed with the  
227 other cell types. In addition, we added to the list T cell and monocyte markers to remove any  
228 possible contamination from the magnetic separation of the B cell populations. The clustering  
229 method showed that although some cells did express transcription factors found in T cells and  
230 monocytes, it also highlighted that they had higher expression of the markers used for B cells.  
231 Consistent with the B cell enrichment methods for memory B cells and plasmablasts, all  
232 vaccine groups showed an abundance of cells that could be defined as either memory B cells  
233 or plasmablasts by transcription factor expression (Figure 2a, b). 39% of the total single cells  
234 consisted of MB, with 7% of total cell being classified as PB. 1% of cells were classified as  
235 activated B cells and 1% as Naïve B cells, perhaps reflecting low level contamination of other  
236 B cells during the cell enrichment protocol consistent with FACS sorting data [25]. By  
237 combining all sample data, it is clear that these classifications were present in all samples and  
238 vaccine groups. Due to multiple duplet identification methods, we conservatively assigned

239 47% of cells as duplets. Through combined V(D)J seq we were able to assemble complete  
240 paired IgH and IgL sequences from the majority of single cells. If paired IgH&L was not  
241 detected these were assigned as V(D)J negative cells. Interestingly most V(D)J negative cells  
242 were assigned as duplets, suggesting that as expected a duplex transcriptome would render  
243 impossible the assembly of paired IgH&L sequences.  
244

### 245 **V(D)J gene usage and CDR3 region frequency**

246 We determined the IgH and IgL gene usage irrespective of B cell type, for productive (open  
247 reading frame), full length, paired heavy and light chain containing cells (Figure 4a) per  
248 immunisation group. For all immunisation groups, IGHV3-23, IGHD3-22, IGHJ4 and the  
249 IGHM isotype were the most common. Interestingly, IGH genes IGHV3-53 and IGHV3-66  
250 commonly part of Class 1 nAbs were also commonly detected. For the lambda light chain  
251 (Figure 4b), IGLV2-14, IGLV1-44 and IGLV1-40 were the most common V segment among  
252 the groups. For the kappa light chain, IGKV3-20 was the most used segment for vaccine groups  
253 A, B and D, while IGKV4-1 was the most used in group C. For the J gene segment in the kappa  
254 light chains, we found that all the groups predominantly used the IGKJ1 gene segment (Figure  
255 4b). All groups preferentially use the kappa light chain (Figure 4d). Overall, IgHM, G2, A1 are  
256 the most common isotypes across all groups (Figure 4c). We determined the most common  
257 CDR3 sequences within each immunisation group. As shown in Figure 4e, each group had a  
258 preferential CDR3 sequence: those with the highest frequency were CARASGYYPYLDYW  
259 in group A, CASRSGRNYYGMDVW in group B, CAKWRRRDGYNFESW in group C and  
260 D (Figure 4E). IGH and IGL V(D)J recombination maps were visualised using Circos for all  
261 the 48 subjects within each vaccine group (Figure 5a) and for each individual group (Figure  
262 5a,b) but split across the cell types of interest (MB and PB) for the most frequent clonotypes  
263 (top 75). Whilst qualitative in analysis, it is clear that all patient groups and MB and PB B cell  
264 types have considerable immunoglobulin gene recombination diversity.  
265

### 266 **V(D)J clonotype analysis and isotype usage between groups and cell types**

267 By combining B cell receptor repertoires with B-cell identity, it is possible to define more  
268 precisely antigen experienced B-cells. To do this we identified BCR clonotypes (Figure 6A) from  
269 IgH&L paired single cell sequencing, assigned them their immunoglobulin isotype (Figure 6B)  
270 and assigned, by transcription factor expression, the B cell phenotype (Figure 6C). Given that  
271 peripheral blood B cells were isolated 7 days after antigen exposure, we defined antigen  
272 experience cells as those of an expanded clonotype, where the immunoglobulin isotype was  
273 switched IgG, IgA or IgE and where the cells of the clonotype were ABCs, or MBC combined  
274 with PBs. Clonotypes were used as the organising feature and these multi-level B cell  
275 clonotypes were visualised using Enclone honeycomb plots (Figure 6A). Here we show all the  
276 clones from each patient (one colour per patient) plotted onto one honeycomb plot to visualize  
277 possible shared clonotypes (represented by mixed colours combs). We represented this also by  
278 isotypes (one colour per isotype) as well as cell types (one colour or cell type). A small number  
279 of clonotypes had a single antibody isotype, namely 24 IgM clonotypes and 13 IgD clonotypes.  
280 Similarly, the majority clonotypes were made up only of plasmablasts, or contained memory  
281 B cells combined with other B cell types, especially plasmablasts.  
282

283 Within some honeycomb clonotype it is clear that different individuals from the same or  
284 different vaccine groups can share the same clonotype, (Figure 6A) where two or more different  
285 individuals are within the same clonotype cluster. Similarly, honeycomb clonotypes can have  
286 different isotypes indicative of class switching (Figure 6B). Isotype distribution among the 48



287 samples showed, the majority of the isotypes seems represented by IgM and IgG1 and IgA1,  
288 with the majority of clonotypes representing mixed IgG isotypes. Overall, IgM is the most  
289 abundant isotype for group A. Honeycomb clonotypes can also consist of different B-cell types  
290 (Figure 6C). The largest clonotypes are represented mainly by PB, while the smaller clonotypes  
291 have both MB and PB. Among all immunisation groups there were a total of 1337 clonotypes  
292 (Figure 7 - Suppl. Big table?). Of these, 206 had between 5 to 10 cells, 61 had 11 to 20 cells,  
293 52 between 21 to 7 and only 7 clonotypes had up to 100 cells, with 11 over 100 cells. The  
294 biggest clonotype had 834 cells, of these, the majority were filtered out as duplicates.

295

### 296 **Seven clonotypes represent functional antibodies against vaccine candidate proteins**

297

298 As shown in the schematic Figure 7, we filtered down the 337 clonotypes and only kept the  
299 ones where we found different patients sharing same clonotypes, between both PB and MB.  
300 We also kept only the IgM-/IgD- and the ones with high divergence from germline. This  
301 totalled up to 7 clonotypes of which two had 376 and 685 cells and the smallest with 14 cells.  
302 In total we had 1148 cells, including the duplets. These were traced back to 10 samples we  
303 named A1-A10 to maintain the blind experiment as shown in Table 2. The table shows we only  
304 have 1 cell for samples A1, A7 and A8, 4 cells for A3, 28 cells for A2, 241 cells for A4, 87  
305 cells for A5, 26 cells for A6, 5 cells for A9 and 14 cells for A10. This meant we removed 740  
306 cells as duplets from further analysis. The 408 cells are the ones we believe express functional  
307 antibodies against HIV. The majority of these were either PB, MB or undefined B cells, while  
308 we found only one being a naïve cell (sample A9). The cells were mainly from two of the four  
309 vaccine groups we tested, C and D, namely ConS UFO and EDC ConS UFO. None of the cells  
310 came from groups A or B except for one barcode which belonged to group B, EDC ConM  
311 SOSIP and it was an undefined B cell.

312

313

### 314 **Phylogenetic analysis of HIV-related antibodies**

315 We made a list of 7 clonotypes (Figure 8A-G) for further analysis via phylogeny to look into  
316 the relationships. To make the phylogenetic trees we excluded the duplets. Trees were made  
317 one for each of the 7 clonotypes selected and from these we chose the highly diversified  
318 branches for further cloning. Namely, we chose up to 3 antibodies from each tree and we  
319 matched their sequences with historical data for B cell repertoire, to exclude any duplicates  
320 antibodies which we might have found being not HIV-specific (vaccine induced). These were  
321 chosen for cloning studies currently being undertaken (data not shown).

322

## 323 Discussion

324 B cells are an important component of the immune response against infection. In this study we  
325 analysed the antibody response to two different vaccine candidates against HIV, ConM and  
326 ConS, and we investigate the phylogenetic relationship between the antibodies produced by  
327 the different B cell subpopulations. The novelty of the study resides in that we used single cell  
328 sequencing, by applying a multiplexing protocol which together allowed us to have a broad  
329 view of the antibody response among 48 different subjects. We used different pipelines to  
330 confirm our findings, such as those offered by Seurat [24] and 10x Genomics software. For the  
331 first time we were able to report the antibody response to two HIV -proteins being investigated  
332 in the EAVI2020 clinical trials.

333  
334 Our findings highlight that among the four B cell populations, memory B cells are  
335 overrepresented, followed by plasmablasts. Notably, these two populations always clustered  
336 separately in all our samples, suggesting the diverse nature of the population. Activated B cells  
337 and naïve B cells, which instead did not cluster as much as memory B cells and plasmablasts,  
338 were also retrieved as a mixed population in memory B cells and/or plasmablasts. This is  
339 consistent with previous published data where, as infection progresses (or in our case, vaccine-  
340 induced immune response), we can see a shift of the ratios between the various B cell  
341 populations: in our experiment we did not follow the course of the infection, but by reading the  
342 cell type profile at 7 days post vaccination, where we would expect a high number of memory  
343 B cells, we found that indeed plasmablasts numbers were lower than memory B cells, as well  
344 as that at this time activated B cells can originate from memory B cells or from naïve B cells  
345 as suggested in [26]. Although we did not follow the dynamics of B cell proliferation, we were  
346 able to retrieve more populations than those manually isolated.

347  
348 We based our cell type clustering for the downstream analysis on a list of internal markers  
349 which we could use with confidence to separate the various B populations (Supplementary  
350 table 1). Our initial magnetic separation was based on a negative separation where PBMCs  
351 were incubated with antibodies which separated B cells from the rest of the cells (Pan B kit  
352 from Mylteni). We then incubated the eluted B cells with CD27 antibody to further enrich our  
353 population for memory B cells and plasmablasts. We did not expect this magnetic separation  
354 to be 100% pure, since based on our downstream analysis we found both T cells (the majority  
355 CD8+ cytotoxic cells, data not shown), as well as other B cells populations, such as naïve B  
356 cells and activated B cells. Given that these latter two populations were constantly present, we  
357 included them in our analysis. Downstream analysis clearly cannot be based on the cell surface  
358 markers we used to separate the initial populations from PBMCs, since protein expression  
359 levels do not necessarily match mRNA levels. Since the data we had comes from mRNA, we  
360 had to use internal markers known to be specific for the populations of interest. We therefore  
361 built up this list based on our 10x runs (over 50 samples) as well as from the literature. Our  
362 data shows that the separation of the various B cell sub populations is consistent with the data  
363 presented in [26]. In fact, seven days post vaccine, three transcription factors are used to  
364 identify the plasmablasts populations, namely IRF4 and IRF8. As suggested in the paper, and  
365 confirmed in our analysis, IRF4 expression is higher in plasmablasts compared to naïve or  
366 activated B cells, while IRF8 has higher expression in memory B cells, naïve and activated B  
367 cells compared to plasmablasts. This suggested to us that the B cell populations isolated from  
368 total PBMCs contained not just memory B cells and plasmablasts, but also naïve and activated  
369 B cells. This was consistent among the different vaccine regimes as outlined in our  
370 experimental plan. In addition, still consistent with the literature, we found that markers such

371 as PRDM1 and XBP1 could be used with confidence to separate plasmablasts from the rest of  
372 the populations [26], while TCL1A could be used for naive B cells separation.

373  
374 When overlapping the clonotype data generated by cellranger v<sub>dj</sub> pipeline, plasmablasts had  
375 the highest numbers of clonotypes, with some clonotypes being overrepresented more than  
376 others. Based on our analysis, in some cases cell lysis that occurred before running the 10x  
377 might have generated fragments that could have ended up being captured in background GEMs.  
378 These were eliminated from our analysis. However, the majority of the samples did not undergo  
379 lysis based on the mitochondrial gene content (Supplementary Figure 2) and any cells with  
380 >5% content of MT genes was discarded. Therefore, the high number of the same clonotypes  
381 being present in a population can be genuinely assigned to that cluster, in this case  
382 plasmablasts. We generated honeycomb plots for all clonotype that passed the filtering steps,  
383 and we showed that for each vaccine group, there are patients that have bigger clonotypes than  
384 others, but also share same clonotypes. While many clonotypes are unique to individuals,  
385 shared ('public') clonotypes are observed in different populations. We found many clonotypes  
386 that were shared between more than one patient, suggesting a strong response to the vaccine.  
387 Based on the final table we produced, the majority of the antibodies within the clonotypes and  
388 shared between the patients belong to groups C and D, and only one clonotype with a patient  
389 from group B. This suggests that ConS is more likely to induce a better antibody response  
390 compared to ConM, which we only found in one shared clonotype. In addition, we also noted  
391 that among the top most used CDR3 regions, both group C and D share the same one, namely  
392 CAKWRRRDGYNFESW. Interestingly, we also found in our final clonotype analysis a group  
393 of antibodies with this CDR3.

394  
395 It is arguable that these antibodies might not be HIV-specific since we didn't sequence the  
396 same patients before beginning the trial to investigate whether these were pre-existing  
397 clonotypes. Sorting Ag-specific cells by FACS would have been the best approach, but in our  
398 trialling of this method we found that sorting cells before running the 10x Chromium did not  
399 allow us time to successfully perform the encapsulation protocols, when considering that each  
400 sample would need to be sorted individually before hashing and loading into the 10x Chromium  
401 device, leading to an increase in cell lysis events. While this would be feasible for one or two  
402 samples, processing 12 samples for a single 10x run was not a good approach. In addition, the  
403 number of Ag-specific B cells sorted would have been very low, especially considering the cell  
404 losses during the hashing, and so even if multiple samples were combined into a single 10x run  
405 there would have still been insufficient numbers for an optimal encapsulation and thus  
406 sequencing depth. Therefore, in the absence of pre-vaccine B cells for these individuals, we  
407 used historical data from publicly available B cell repertoires. This is a common procedure  
408 when lacking baseline data as suggested in [27] where the B cell response is investigated in  
409 SARS-CoV2 patients, but for whom pre-pandemic samples are not available for the same  
410 patients.

411 In the last couple of years, the technologies and methodologies of single cell isolation and  
412 encapsulation have improved, for example the use of hashed antigen probes and of  
413 'superloading' of the 10x Chromium platform. This would allow future studies to identify  
414 antigen-specific B cells from scRNA-Seq data without prior sorting, and to load increased  
415 numbers of multiplexed cells (greater than the 17000 used in this study) with a much higher  
416 recovery rate. In this instance, we have been able to successfully isolate single cells for analysis  
417 of the B cell repertoire without the need of antigen-sorting the cells or using pre-vaccine  
418 samples. We relied on the robustness of the bioinformatic analysis to selectively remove  
419 duplets, cells with known Antibodies against other pathogens and on phylogeny studies to  
420 make a list of possible antibodies to study further binding and neutralization assays.



<b>Contributor Role</b>	<b>Role Definition</b>
<b>Conceptualisation</b>	Ideas: formulation or evolution of overarching research goals and aims: IB, PK, RS
<b>Methodology</b>	Development or design of methodology; creation of models: IB, CP
<b>Software</b>	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components: IB, SW, CM
<b>Validation</b>	Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs: IB, CP
<b>Formal Analysis</b>	Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesise study data: IB, SW, CM
<b>Investigation</b>	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection: IB, CP, HC, KP
<b>Resources</b>	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools: PK, RS
<b>Data Curation</b>	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse: IB
<b>Writing – Original Draft Preparation</b>	Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation): IB
<b>Writing – Review and Editing</b>	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages: All authors
<b>Visualisation</b>	Preparation, creation and/or presentation of the published work, specifically visualisation/data presentation: IB, CM, SW
<b>Supervision</b>	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team: PK, RS

<b>Project Administration</b>	Management and coordination responsibility for the research activity planning and execution: PK
<b>Funding</b>	Acquisition of the financial support for the project leading to this publication: This work was supported by the EAVI2020 grant 681137 to RS and PK.

422

423

424 **Conflicts of interest**

425 The authors declare no conflict of interest.

426

427 **Acknowledgements**

428 The authors thank Dr Hamish King, Walter and Eliza Hall Institute of Medical Research for  
429 useful comments and advice.

430

431

432



## 433 Bibliography

- 434 1. Melchers, F., *Checkpoints that control B cell development*. J Clin Invest, 2015. **125**(6): p.  
435 2203-10.
- 436 2. Muramatsu, M., et al., *Class switch recombination and hypermutation require*  
437 *activation-induced cytidine deaminase (AID), a potential RNA editing enzyme*. Cell,  
438 2000. **102**(5): p. 553-63.
- 439 3. Sather, D.N. and L. Stamatatos, *Epitope specificities of broadly neutralizing plasmas*  
440 *from HIV-1 infected subjects*. Vaccine, 2010. **28 Suppl 2**: p. B8-12.
- 441 4. Zhou, T. and K. Xu, *Structural Features of Broadly Neutralizing Antibodies and Rational*  
442 *Design of Vaccine*. Adv Exp Med Biol, 2018. **1075**: p. 73-95.
- 443 5. McElrath, M.J. and B.F. Haynes, *Induction of immunity to human immunodeficiency*  
444 *virus type-1 by vaccination*. Immunity, 2010. **33**(4): p. 542-54.
- 445 6. Ross, A.L., et al., *Progress towards development of an HIV vaccine: report of the AIDS*  
446 *Vaccine 2009 Conference*. Lancet Infect Dis, 2010. **10**(5): p. 305-16.
- 447 7. Burton, D.R. and J.R. Mascola, *Antibody responses to envelope glycoproteins in HIV-1*  
448 *infection*. Nat Immunol, 2015. **16**(6): p. 571-6.
- 449 8. Kim, J., et al., *Current approaches to HIV vaccine development: a narrative review*. J Int  
450 AIDS Soc, 2021. **24 Suppl 7**: p. e25793.
- 451 9. Bonsignori, M., et al., *Maturation Pathway from Germline to Broad HIV-1 Neutralizer of*  
452 *a CD4-Mimic Antibody*. Cell, 2016. **165**(2): p. 449-63.
- 453 10. Galson, J.D., D.F. Kelly, and J. Truck, *Identification of Antigen-Specific B-Cell Receptor*  
454 *Sequences from the Total B-Cell Repertoire*. Crit Rev Immunol, 2015. **35**(6): p. 463-78.
- 455 11. EAVI2020, *EAVI2020 European AIDS VAccine Initiative*. 2015.
- 456 12. Sliepen, K., et al., *Structure and immunogenicity of a stabilized HIV-1 envelope trimer*  
457 *based on a group-M consensus sequence*. Nat Commun, 2019. **10**(1): p. 2355.
- 458 13. Aldon, Y., et al., *Rational Design of DNA-Expressed Stabilized Native-Like HIV-1 Envelope*  
459 *Trimers*. Cell Rep, 2018. **24**(12): p. 3324-3338 e5.
- 460 14. Sanders, R.W., et al., *A next-generation cleaved, soluble HIV-1 Env trimer, BG505*  
461 *SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-*  
462 *neutralizing antibodies*. PLoS Pathog, 2013. **9**(9): p. e1003618.
- 463 15. Binley, J.M., et al., *A recombinant human immunodeficiency virus type 1 envelope*  
464 *glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120*  
465 *and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure*. J  
466 Virol, 2000. **74**(2): p. 627-43.
- 467 16. Liao, H.X., et al., *A group M consensus envelope glycoprotein induces antibodies that*  
468 *neutralize subsets of subtype B and C HIV-1 primary viruses*. Virology, 2006. **353**(2): p.  
469 268-82.
- 470 17. Sanders, R.W., et al., *Evolution of the HIV-1 envelope glycoproteins with a disulfide bond*  
471 *between gp120 and gp41*. Retrovirology, 2004. **1**: p. 3.
- 472 18. Sanders, R.W., et al., *Variable-loop-deleted variants of the human immunodeficiency*  
473 *virus type 1 envelope glycoprotein can be stabilized by an intermolecular disulfide bond*  
474 *between the gp120 and gp41 subunits*. J Virol, 2000. **74**(11): p. 5091-100.
- 475 19. Alshafi, N., et al., *Effects of the I559P gp41 change on the conformation and function*  
476 *of the human immunodeficiency virus (HIV-1) membrane envelope glycoprotein trimer*.  
477 PLoS One, 2015. **10**(4): p. e0122111.

- 478 20. Stoeckius, M., et al., *Cell Hashing with barcoded antibodies enables multiplexing and*  
479 *doublet detection for single cell genomics*. Genome Biol, 2018. **19**(1): p. 224.
- 480 21. Genomics, x., *Chromium Single Cell V(D)J Reagent Kits with Feature Barcode technology*  
481 *for Cell Surface Protein* 2019.
- 482 22. Wolock, S.L., R. Lopez, and A.M. Klein, *Scrublet: Computational Identification of Cell*  
483 *Doublets in Single-Cell Transcriptomic Data*. Cell Syst, 2019. **8**(4): p. 281-291 e9.
- 484 23. Larsson, A., *AliView: a fast and lightweight alignment viewer and editor for large*  
485 *datasets*. Bioinformatics, 2014. **30**(22): p. 3276-8.
- 486 24. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions,*  
487 *technologies, and species*. Nat Biotechnol, 2018. **36**(5): p. 411-420.
- 488 25. Morbach, H., et al., *Reference values for B cell subpopulations from infancy to*  
489 *adulthood*. Clin Exp Immunol, 2010. **162**(2): p. 271-9.
- 490 26. Ellebedy, A.H., et al., *Defining antigen-specific plasmablast and memory B cell subsets*  
491 *in human blood after viral infection or vaccination*. Nat Immunol, 2016. **17**(10): p. 1226-  
492 34.
- 493 27. Scheid, J.F., et al., *B cell genomics behind cross-neutralization of SARS-CoV-2 variants*  
494 *and SARS-CoV*. Cell, 2021. **184**(12): p. 3205-3221 e24.

495

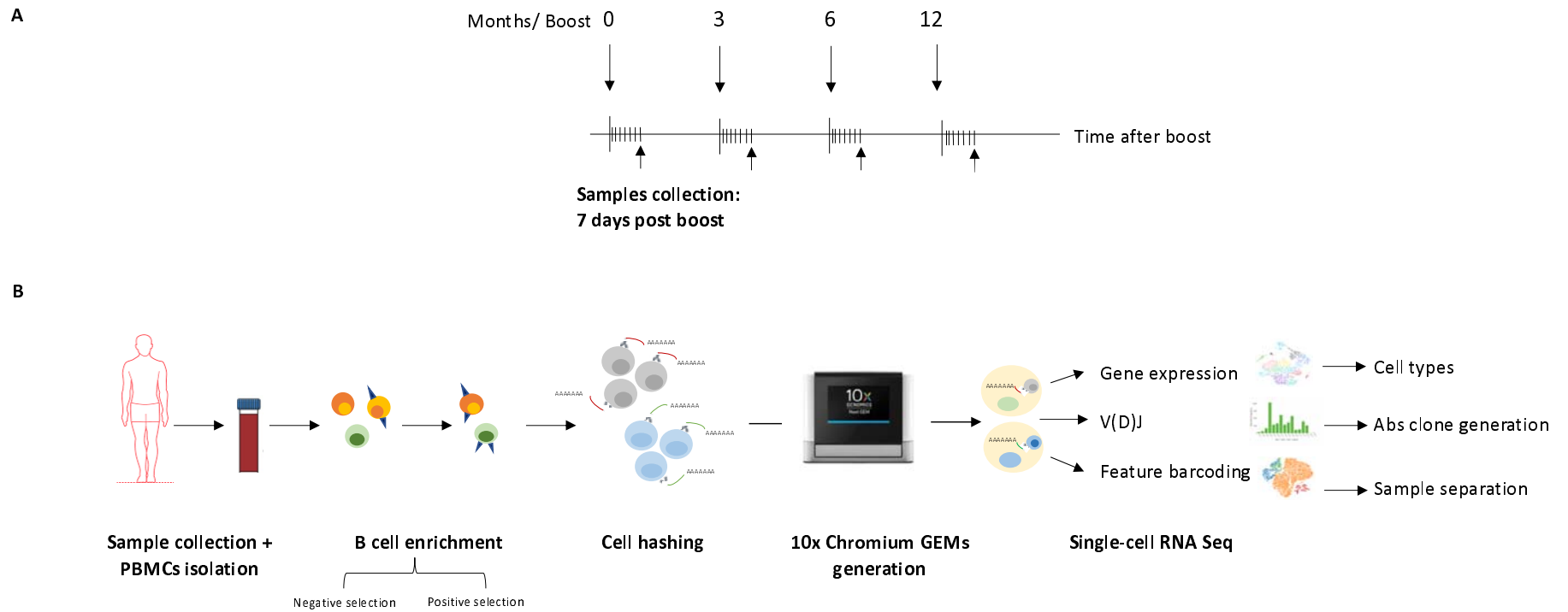


Figure 1 Schematic overview describing the process from sample collection to data analysis. A. Timeline of the vaccine program showing which samples were used for this work. B. Figure describing the steps from blood collection to encapsulation up to sequencing and data analysis.

<b>Groups</b>	<b>Time point 1</b>
<b>A</b>	ConM SOSIP = 100 µg
<b>B</b>	EDC ConM SOSIP = 100 µg
<b>C</b>	ConS UFO = 100 µg
<b>D</b>	EDC ConS UFO = 100 µg
<b>E</b>	ConS UFO = 100 µg

Table 1 Sample vaccine regime. Only one time point, representing PBMC collected 7 days following first boost, was analyzed in this study. Patients 27/29 did not participate

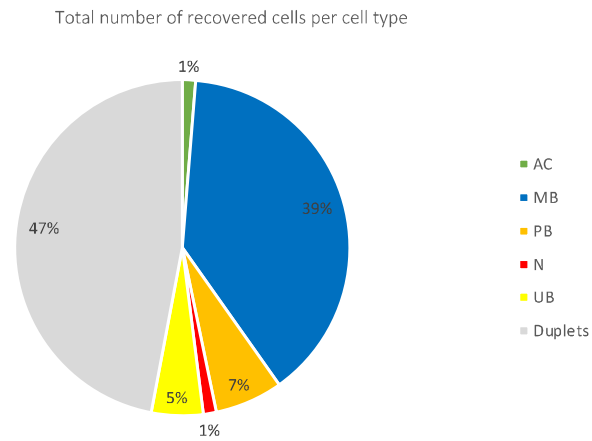
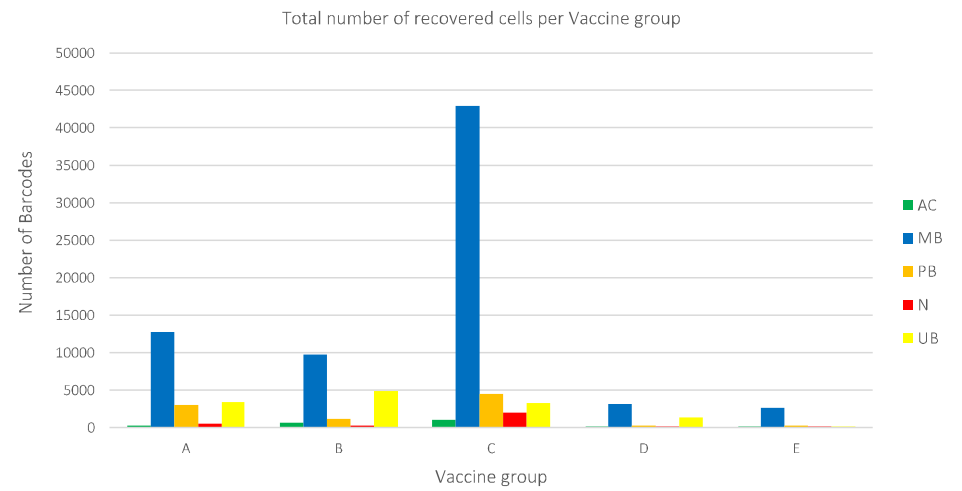
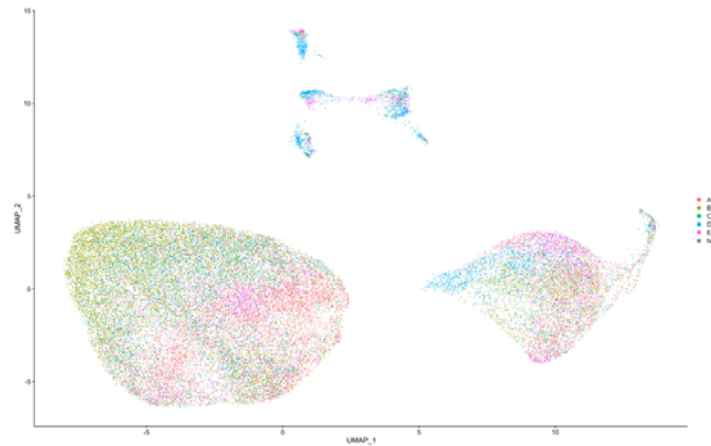
**A****B**

Figure 2 Number of recovered cell types in all the samples (A) and within each of the vaccine groups (B). On average, once we removed duplets and V(D)J negative cells, we recovered ~150000 total cells.

A



B

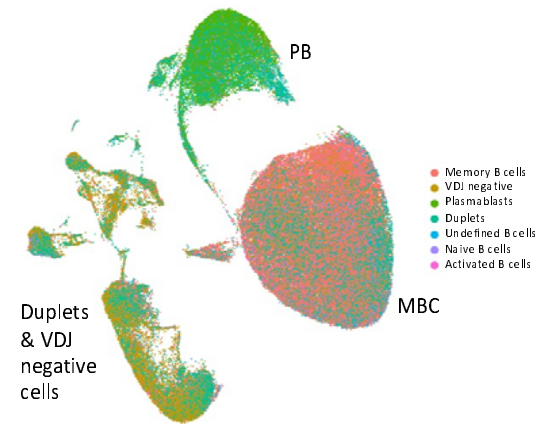


Figure 3A-B Generation of cell clusters. A. All samples were grouped together following assignment of the vaccine group they belonged to. NA represent a group of cells which were either duplets and/or V(D)J negative. B. Cell type separation based on internal transcriptional markers. The figure shows the different B cell subpopulation retrieved after separating duplets and V(D)J negative cells. Samples that showed expression of barcodes which didn't fall in any of the B cell populations we retrieved based on the transcriptional markers used were defined as undefined B cells (UB). As shown in the figure most of the population is represented by memory B cells, followed by plasmablasts, activated B cells and naïve B cells (excluding duplet barcodes).



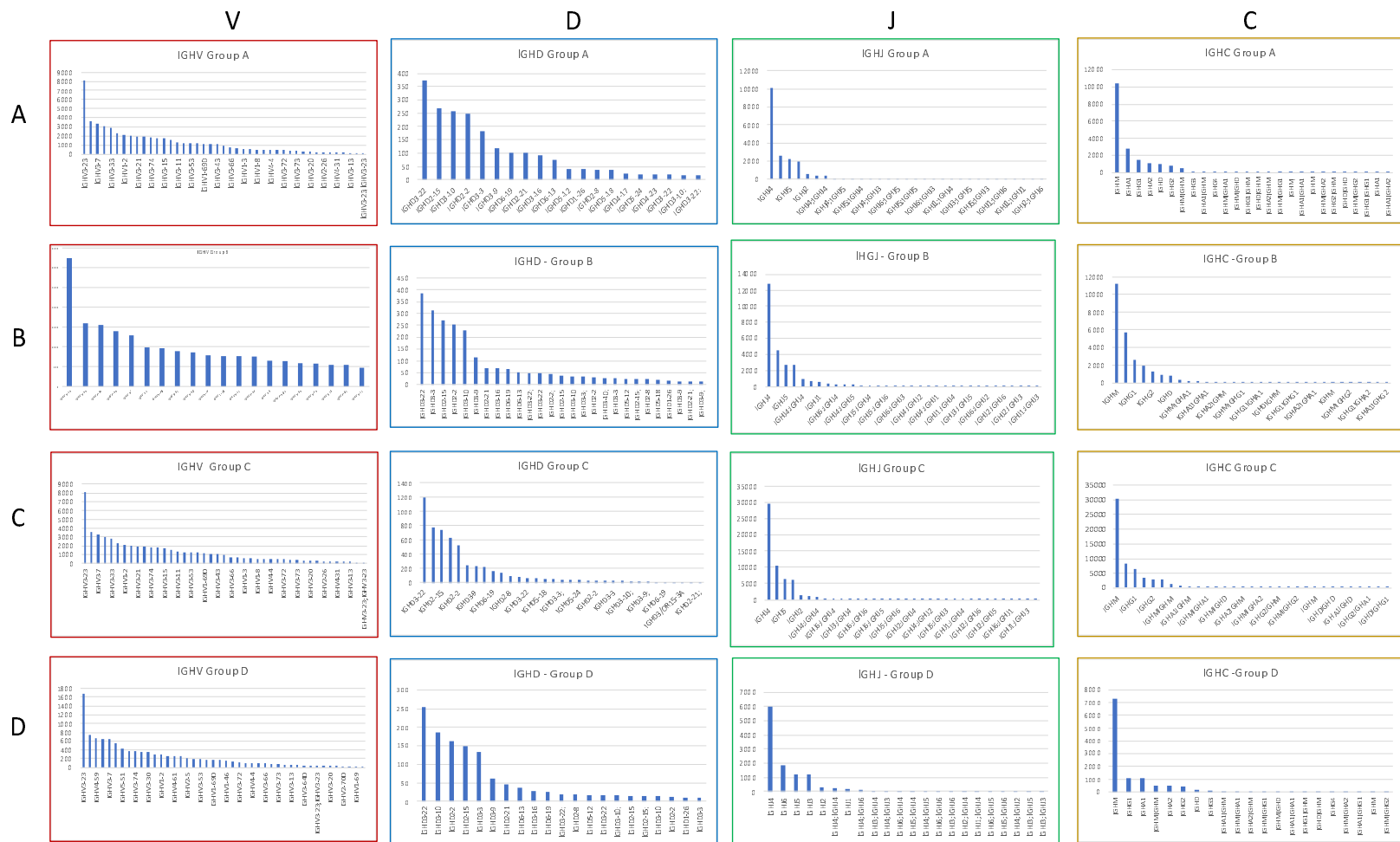


Figure 4A Plots of heavy chain V, D, J, C gene usage among the 4 vaccine groups. Only the to highly expressed genes were chosen for plotting. Columns represent the four genes and rows the four vaccine groups for which we calculated gene levels expression. For all groups, among the V genes, IGHV3-23 was the most used; for the D gene, IGHD3-22 was common to all groups; for the J gene, IGJ4 was expressed highly in all groups and the most constant gene was IGHM.

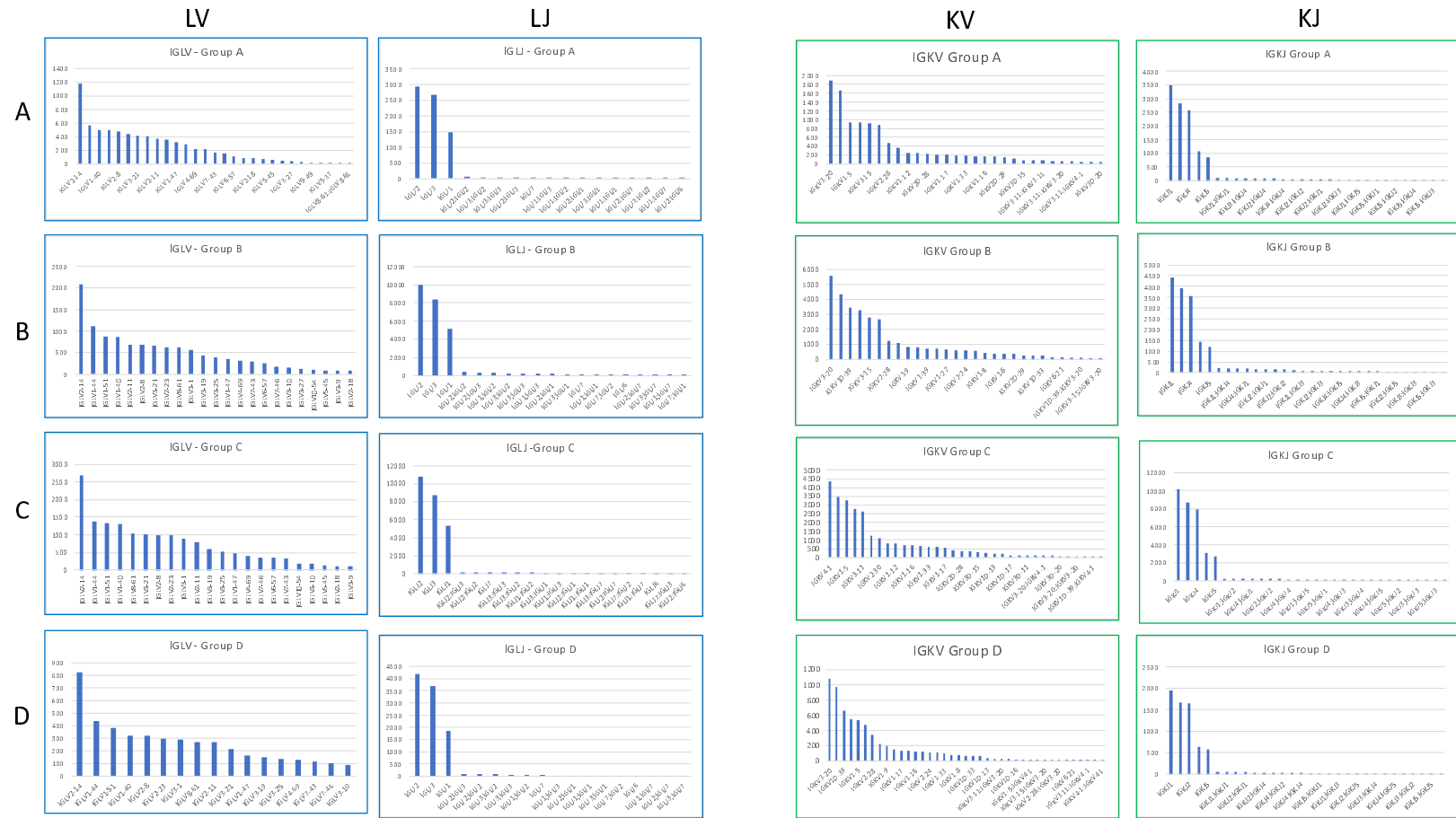


Figure 4B Plots of light chain V, J, C gene usage among the 4 vaccine groups. Left side for the Lambda light chain and right side for the Kappa light chain. The figure shows that IGLV2-14 and IGLJ2 were the top highly expressed genes for the Lambda light chain, while for the Kappa light chain IGKV3-20 was highly expressed in groups A, B, and D. Group C had IGKV4-1 as the top used gene. For the Kappa J gene, IGKJ1 was highly expressed in all four groups.

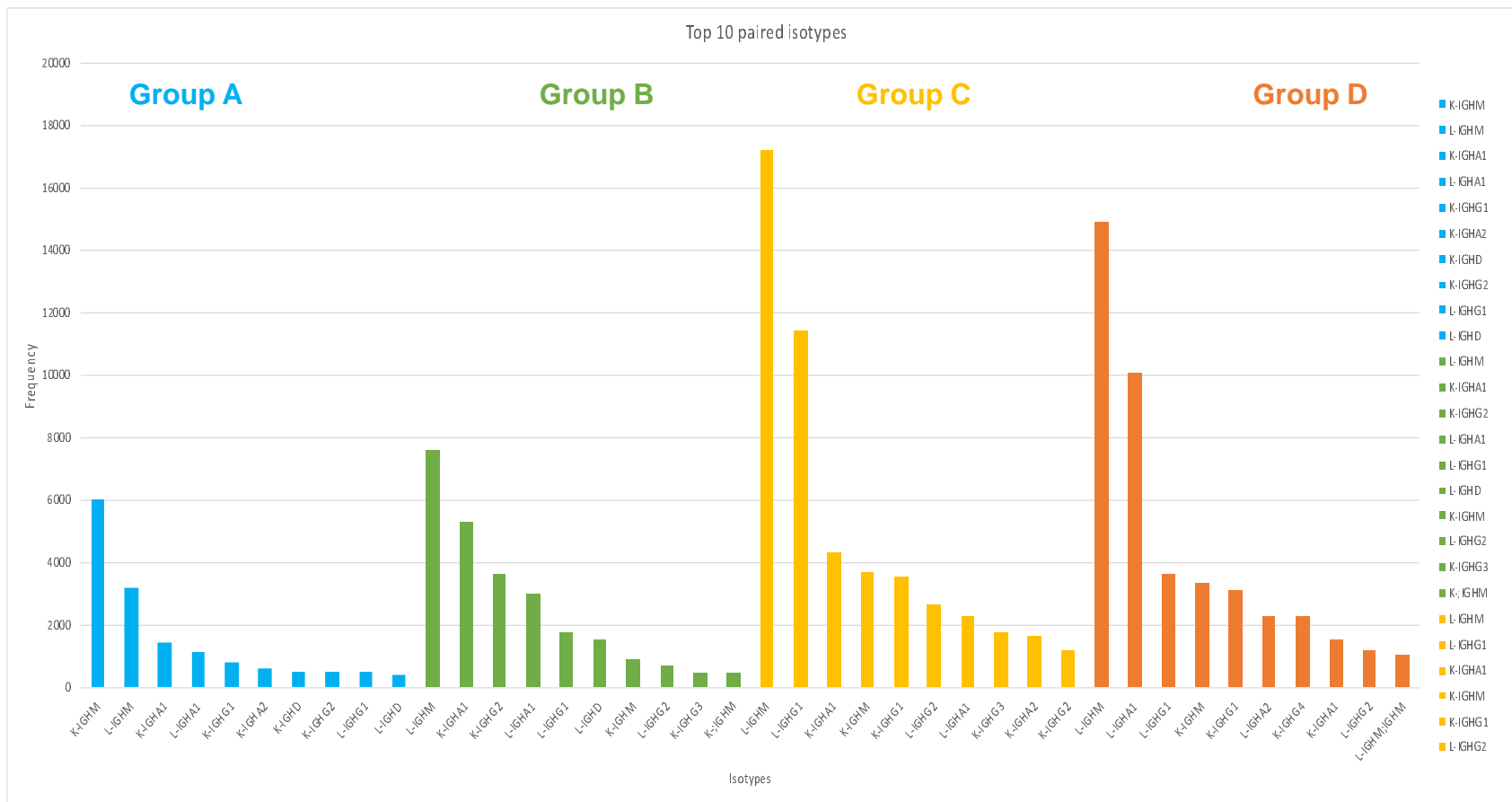


Figure 4C Frequency plots of top paired isotypes per vaccine group. The figure shows the preferential used of the paired isotypes for each vaccine group represented in four different colors: group A, blue, group B, green, group C, yellow and group D, orange. The following pairs were on top of each group: IGHM-K for group A, IGHM-L for group B, IGHG1-L for group C and IGHG1-L for group D.

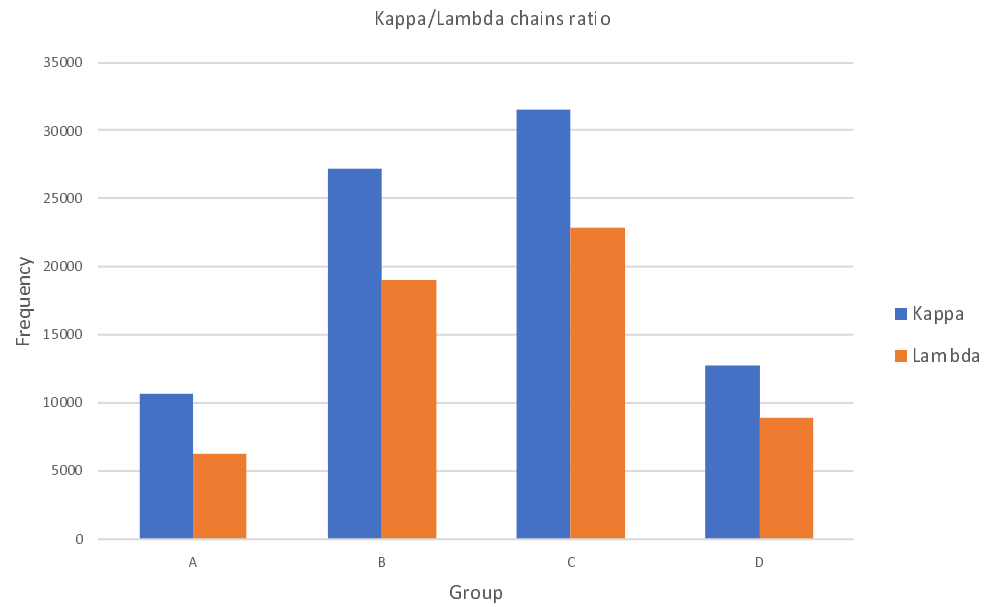


Figure 4D Frequency plot showing ratio of Kappa vs Lambda usage in each vaccine group. The figure shows the ration of Kappa vs Lambda light chain usage in each group. The frequency is higher for Kappa light chain in all groups compared to the Lambda chain.

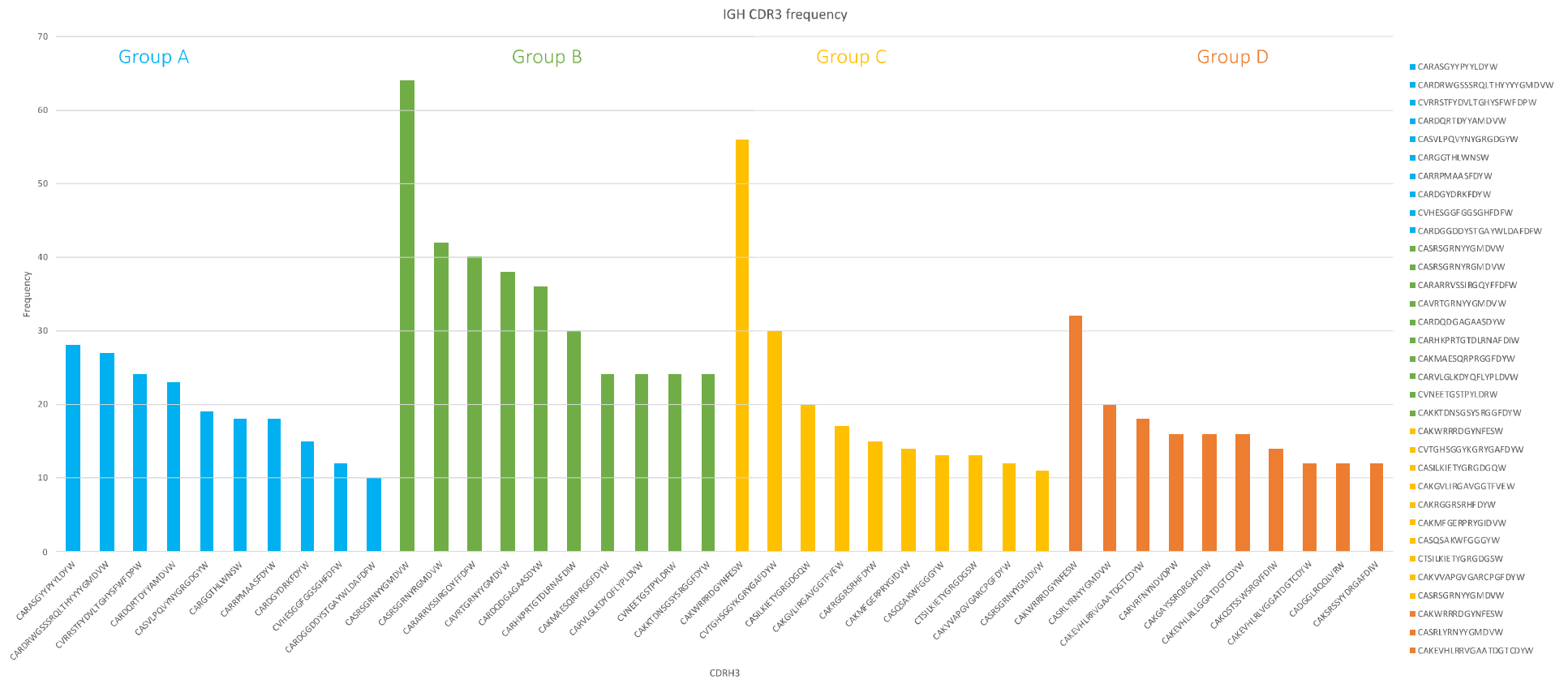
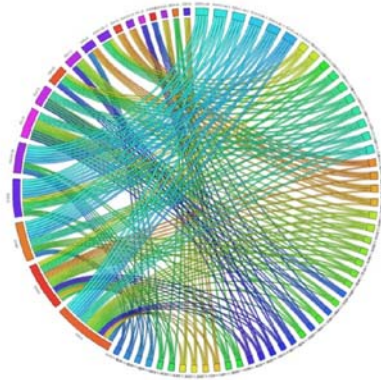
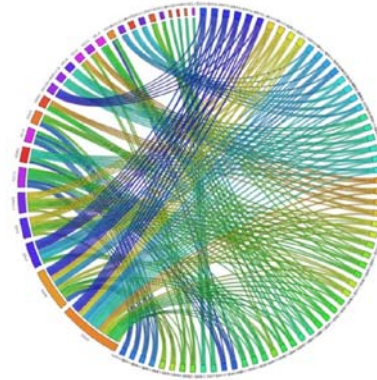


Figure 4E Frequency plot of the top 10 heavy CDR3 regions. The figure shows the top 10 CDR3 most used regions among the four vaccine groups represented in four different colors: group A, blue, group B, green, group C, yellow and group D, orange. Groups B and D share the same CDR3 region in their top usage compared to group A and C (see text for more detail about the CDR3 region sequence).

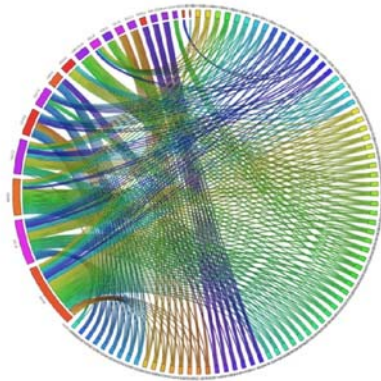
Group A – top clonotypes (75)



Group B – top clonotypes (75)



Group C – top clonotypes (75)



Group D – top clonotypes (75)

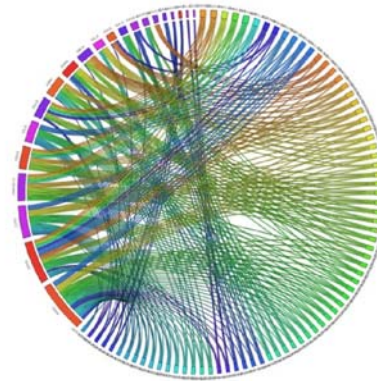


Figure 5A Circos plot showing the highest recombining events as clonotypes for the four vaccine groups. The figure shows only the top 75 recombination events. Given the high number of genes used (see text file in supplementary to read the gene names), the circos plot is only used as a visual representation of the clonotypes. All together the figure shows that there isn't a particular clonotype with a high frequency of recombination (width of curved lines), instead these events are evenly distributed among several clonotypes.



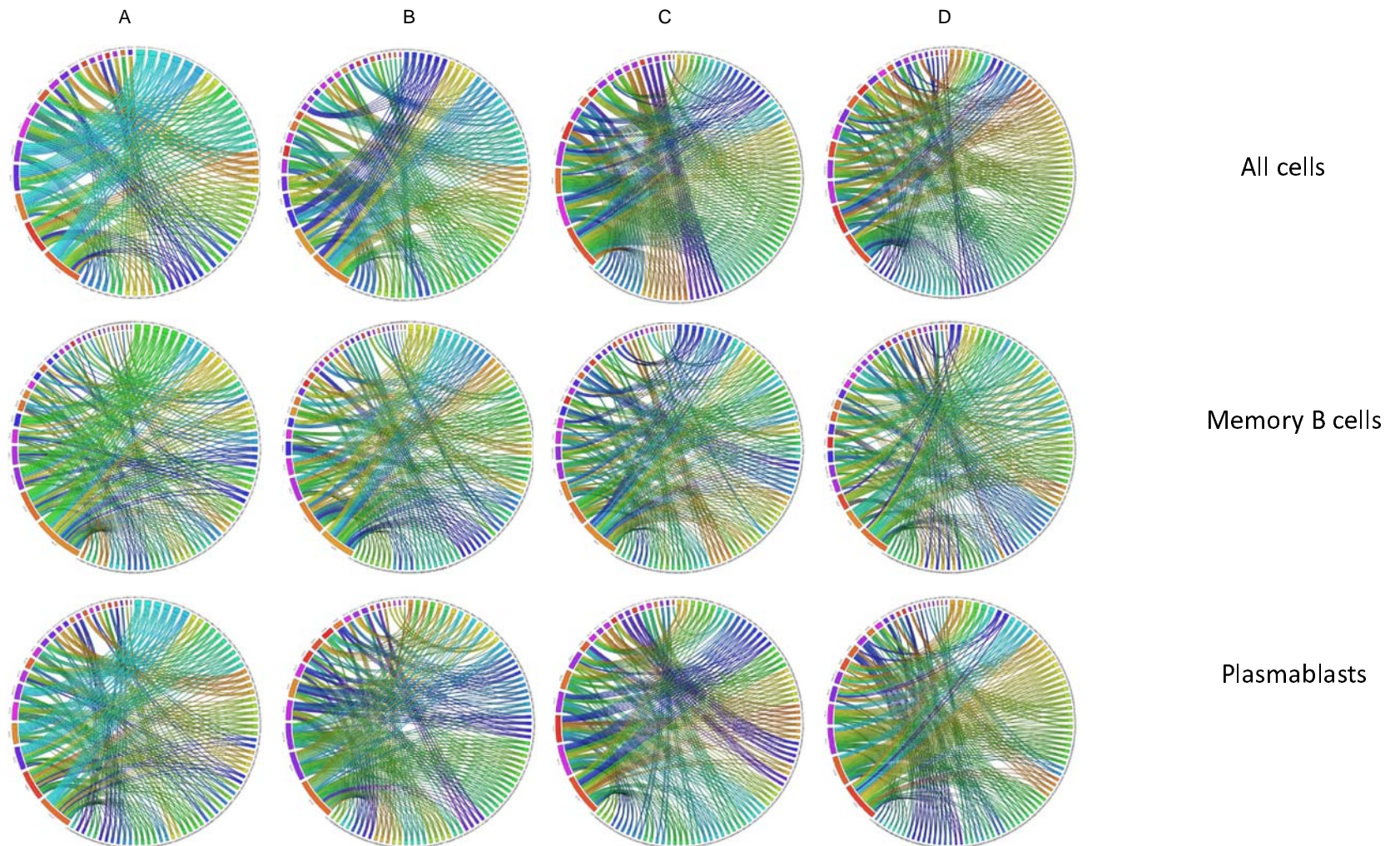


Figure 5B Circos plot of clonotypes distributions divided by cell types. Circos plot were generate specifically highlighting all cell types (top row) or focusing only onto memory B cells (middle row) and plasmablasts cells (bottom row). As shown in Figure 5A, there isn't a particular clonotype with a high frequency of recombination (width of curved lines), instead these events are evenly distributed among several clonotypes.

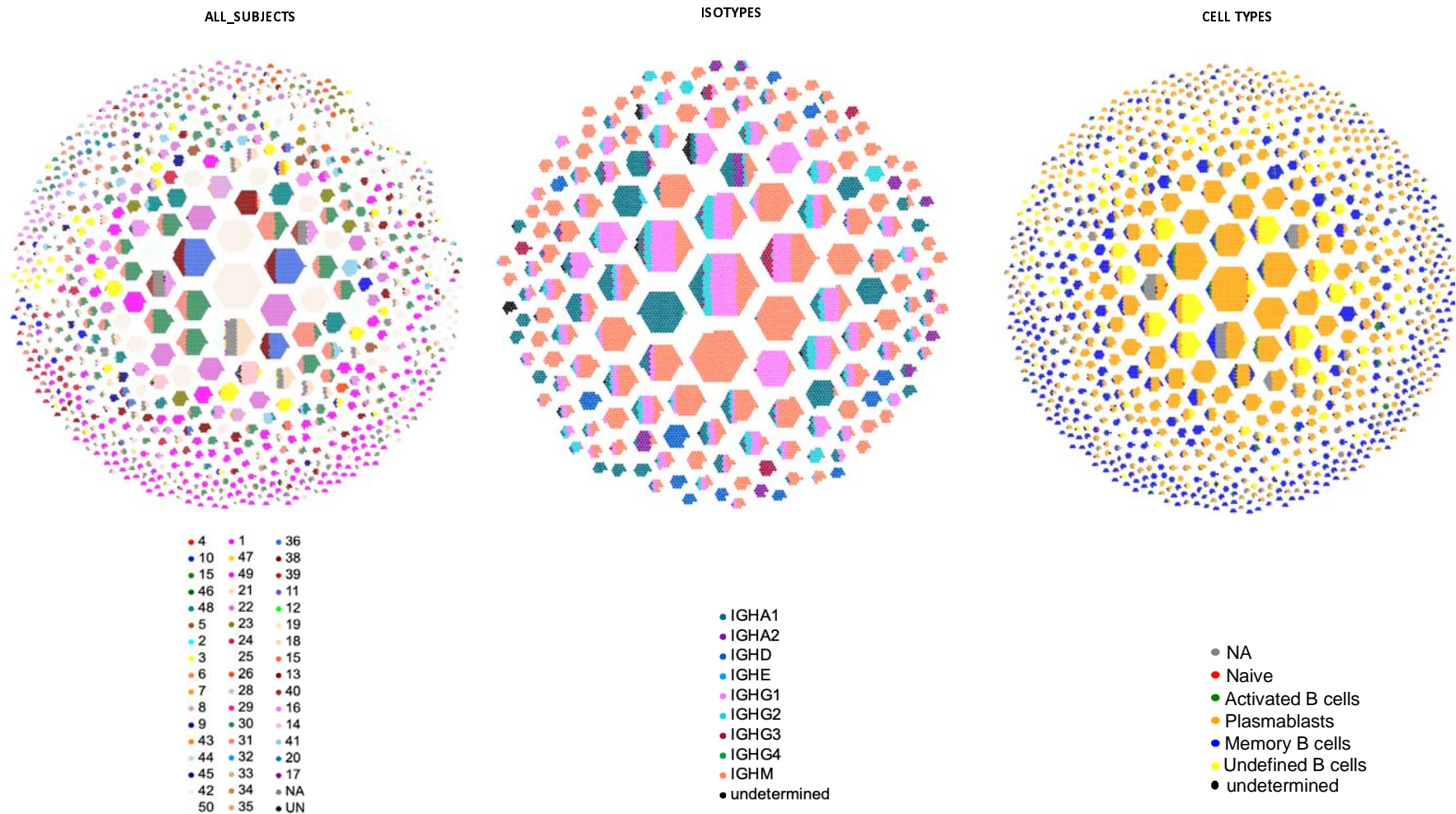


Figure 6A Honeycomb plots of shared clonotypes. The figures shows the expression of clonotype by taking all subjects (left), highlighting the expression of isotypes (middle) or cell types (right). Colors indicate either the subjects from all the 4 vaccine groups, the type of isotypes used or the type cells. Take al together, each clonotype is shown in the same shape in each plot. Each comb where more than one color is shared, represents, e.g. , a subject with more than one isotypes across different cell types. Sometimes, a clonotype is shared among two or more subjects (left honeycomb where one comb has more than one color, e.g. more than one subject).



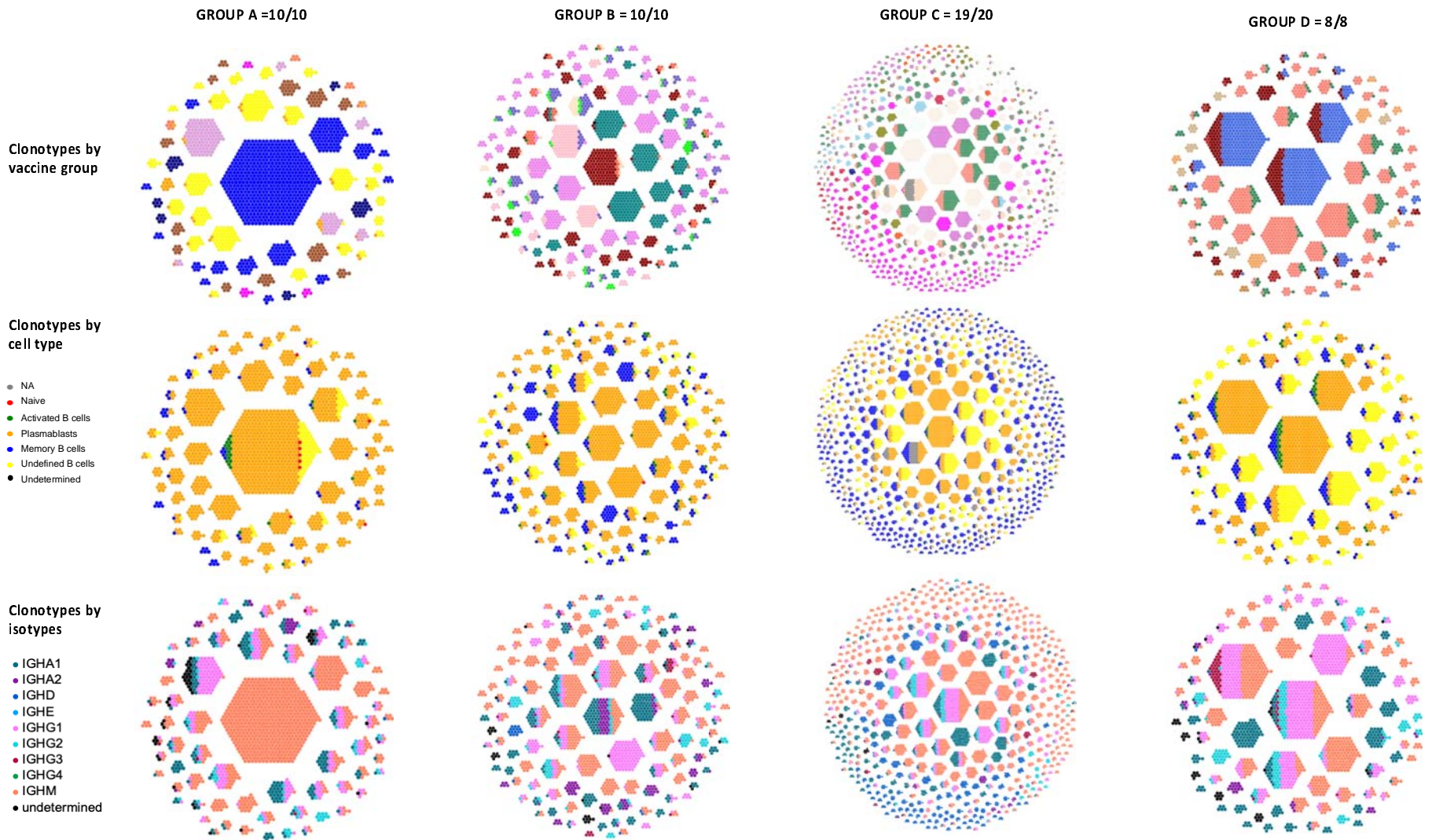


Figure 6B Honeycomb plots of shared clonotypes. In this figure the honeycombs are made by highlighting the vaccine groups (columns) vs number of clonotypes in total in a vaccine group (top row), type of cell (middle row) or isotypes (bottom row). Group C has the highest number of clonotypes as it had more subjects compared to groups A, B and D. when loo

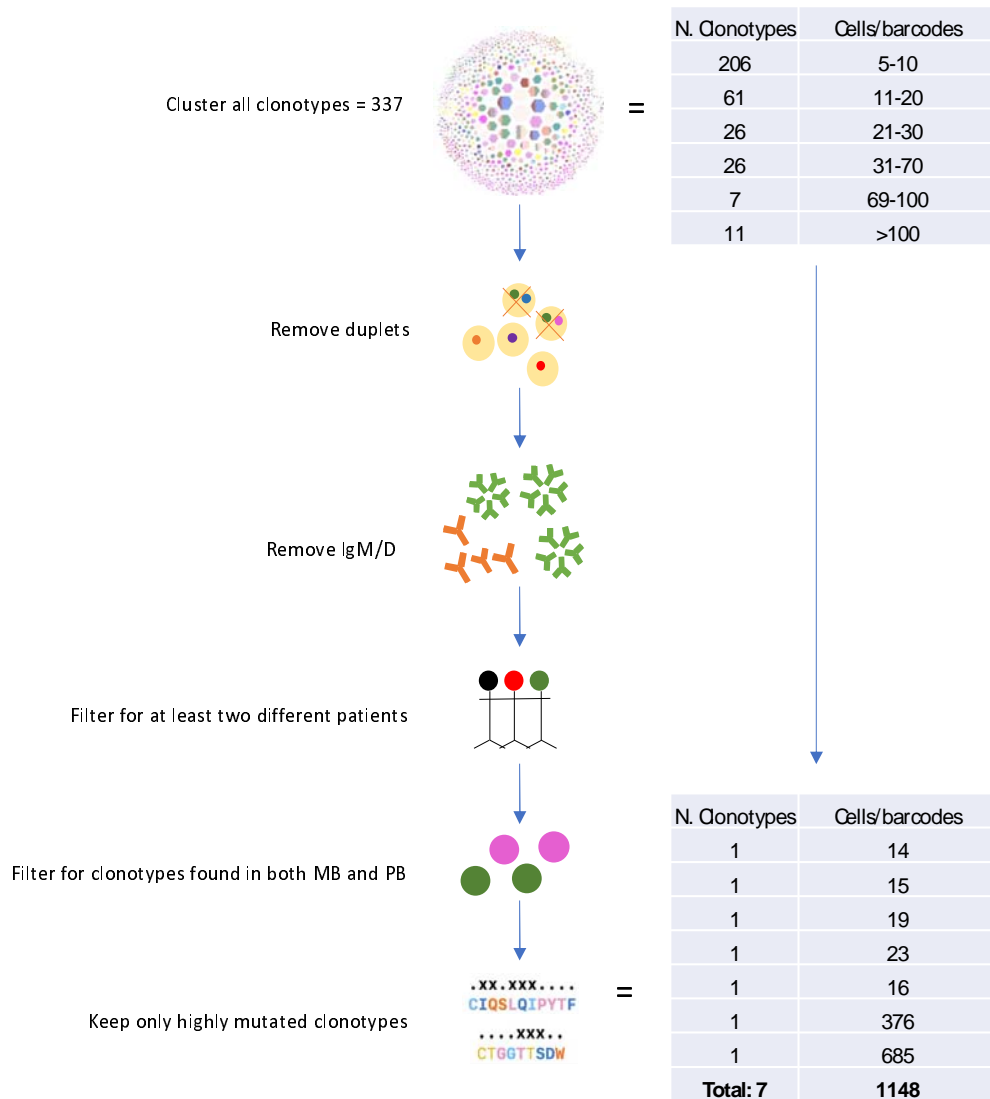


Figure 7 Schematic representation of the filtering process. Following data analysis, 337 clonotypes were found, divided into 6 groups and associated with a number of barcodes (cells). From these, all duplets were removed, as well as IgM/D isotypes and clonotypes were filtered to only keep those having at least two different patients, as these will be more unique than those not shared. Further filtering included for cell types (memory B cells, MB, and plasmablasts, PB) and highly mutated clonotypes. This filtering narrowed down the number to 7 clonotypes for a total of 1148 cells that we chose for further cloning experiments.

Barcode count	Cell type	Samples	Vaccine group
1	TB	A1	B
28	MB/PB	A2	C
4	PB	A3	C
241	MB/PB/TB	A4	D
87	MB/PB/TB	A5	D
26	MB/PB/TB	A6	D
1	PB	A7	D
1	PB	A8	C
5	MB/N	A9	C
14	MB/PB	A10	C
408			

Table 2 Summary of final cell types, vaccine groups and barcode(cell) count found in the clonotypes of interest.

A

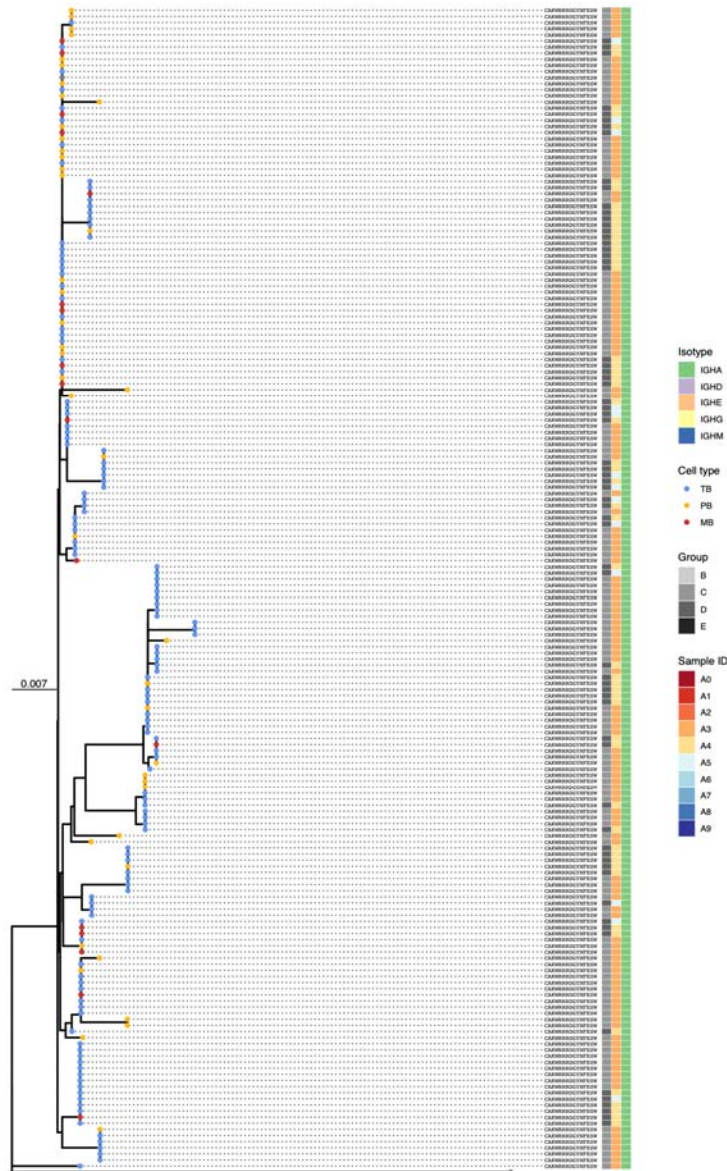
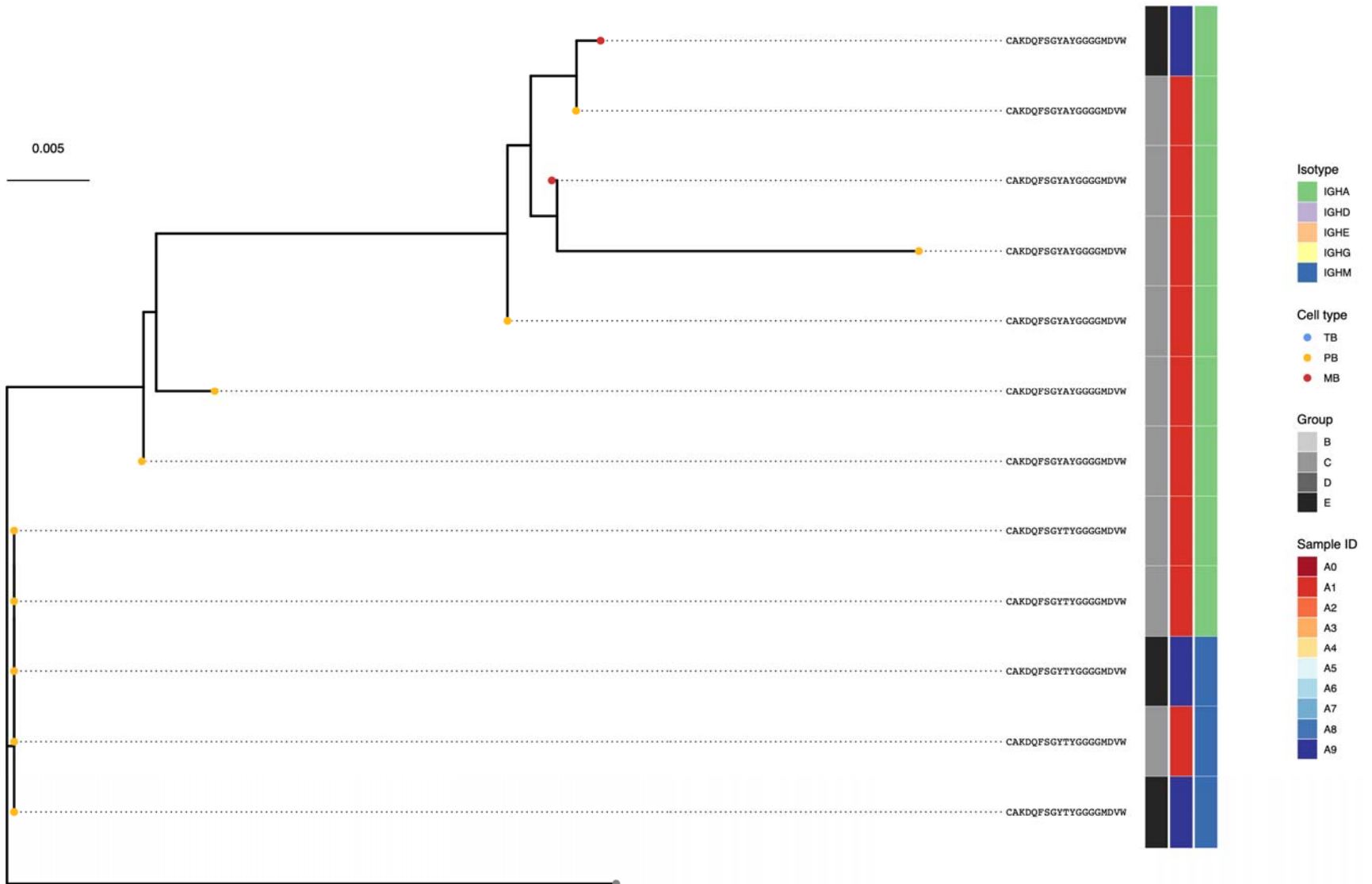


Figure 8A-G Phylogenetic trees for each for the clonotypes from Figure 7.



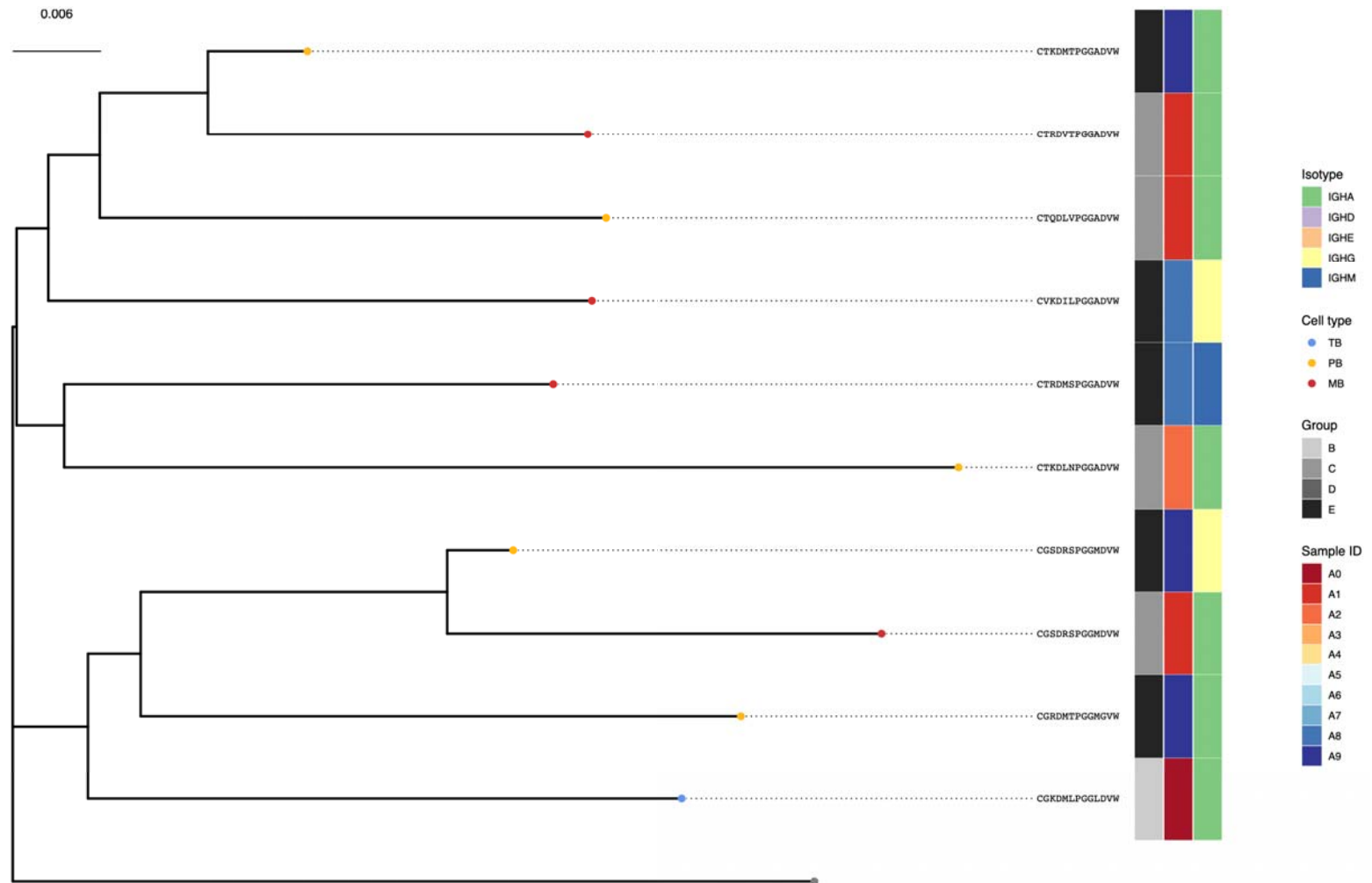


C



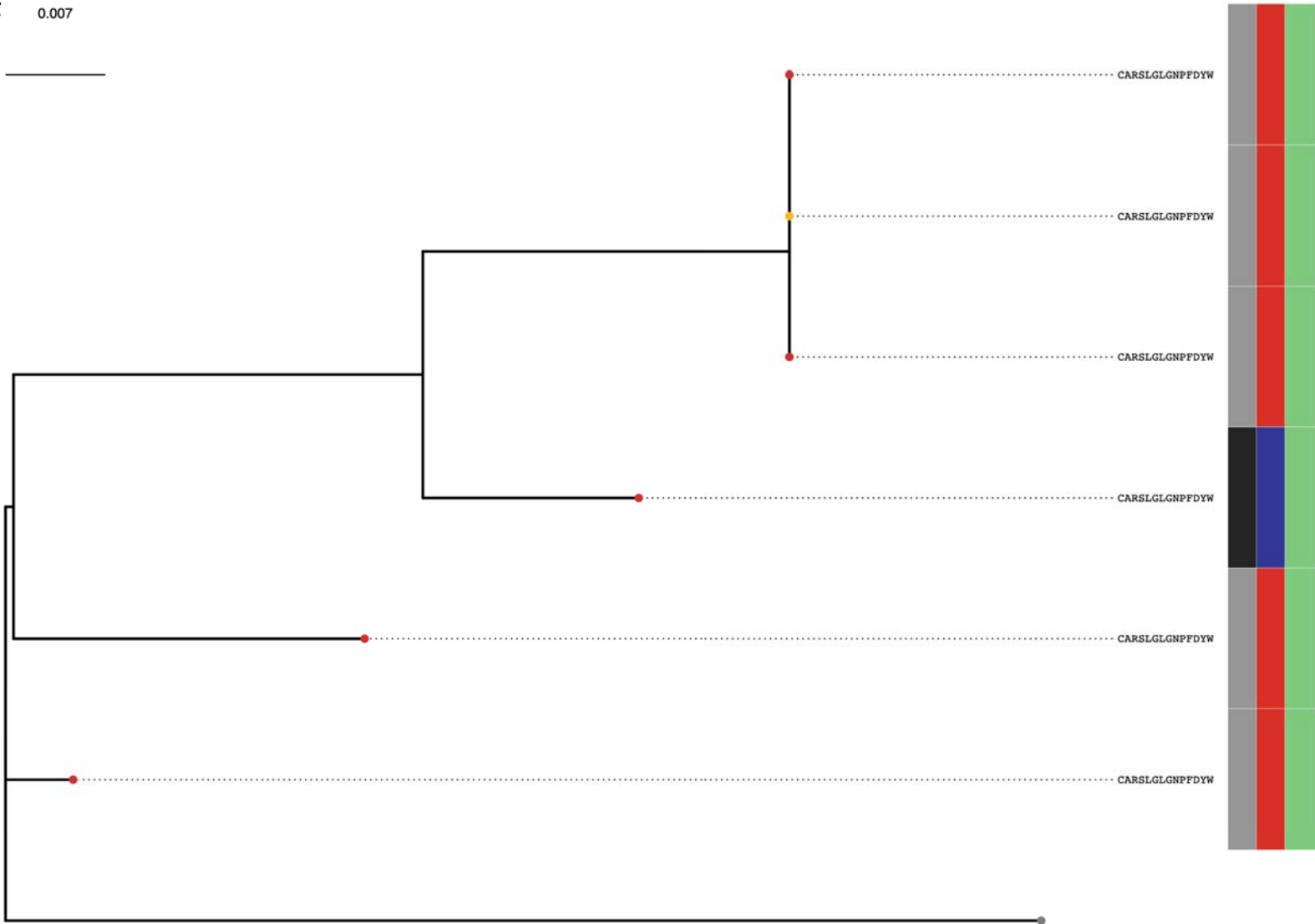


D





F 0.007



Isotype

- IGHA
- IGHD
- IGHE
- IGHG
- IGHM

Cell type

- TB
- PB
- MB

Group

- B
- C
- D
- E

Sample ID

- A0
- A1
- A2
- A3
- A4
- A5
- A6
- A7
- A8
- A9

