Cell-free RNA reveals host and microbial correlates of broadly neutralizing antibody development against HIV

Mark Kowarsky[1], Mercedes Dalman[2,3], Mira N Moufarrej[4], Jennifer Okamoto[5], Yike Xie[2,3], Norma F Neff[5], Salim S Abdool Karim[6,7], Nigel Garrett[6,8,9], Penny L Moore[6,10,11,12,†], Joan Camunas-Soler [2,3,13,†], Stephen R Quake[4,5,†]

[1]Department of Physics, Stanford University, Stanford, CA

[2]Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, University of Gothenburg, Sweden

[3]Wallenberg Centre for Molecular and Translational Medicine, Sahlgrenska Academy, University of Gothenburg, Sweden

[4]Department of Bioengineering, Stanford University, Stanford, CA

[5]Chan Zuckerberg Biohub, San Francisco, CA

[6]Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu Natal, Durban, South Africa

[7]Department of Epidemiology, Columbia University, New York, NY 10032, USA

[8] Desmond Tutu HIV Centre, University of Cape Town, Cape Town, South Africa

[9] Discipline of Public Health Medicine, School of Nursing and Public Health, University of KwaZulu-Natal, Durban, South Africa

[10] SA MRC Antibody Immunity Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[11] Centre for HIV and STIs, National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service (NHLS), Johannesburg, South Africa.

[12] Wits Infectious Diseases and Oncology Research Institute, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[13] Science for Life Laboratory, Institute of Biomedicine, University of Gothenburg, Gothenburg, Sweden

†To whom correspondence should be addressed: steve@quake-lab.org, pennym@nicd.ac.za, joan.camunas@gu.se

Abstract:

A small number of people living with HIV (PLWH) develop broadly neutralizing antibodies (bNAbs) targeting multiple HIV strains. Although several viral and immune factors contribute to bNAb development, the genetic and environmental factors driving this response remain largely unknown. We performed combined cell-free DNA (cfDNA) and cell-free RNA (cfRNA) sequencing in 42 plasma samples from a longitudinal cohort of 14 PLWH (7 who develop bNAbs and 7 matched controls). This approach enabled us to non-invasively monitor the host transcriptome, viral genetic variation, and microbiome composition during HIV infection, and to identify molecular correlates of bNAb development. We find that development of bNAbs is associated with a transcriptomic signature of early immune activation characterized by elevated levels of MHC class I antigen presentation genes. This signature is independent of viral load or CD4 count and declines over time. In addition to host features, we recovered sufficient viral reads to reconstruct HIV consensus sequences, supporting the utility of cfRNA for viral genotyping. Finally, we also identified an enrichment of several microbial taxa in bNAb producers and increased levels of GB virus C (GBV-C), a non-pathogenic lymphotropic virus. Our findings suggest a distinct early immune activation profile in PLWH who develop bNAbs. More broadly, we show that combined cfDNA/cfRNA sequencing can reveal relationships between a protective immunogenic response to HIV infection, the host immune system, and microbiome, highlighting its potential for biomarker discovery in future vaccine and therapeutic studies.

Author Summary

A subset of people living with HIV develop broadly neutralizing antibodies (bNAbs), a phenomenon that remains poorly understood but relevant to vaccine development. We applied

combined cell-free RNA and DNA sequencing to 42 longitudinal plasma samples (7 bNAb producers, 7 controls) to monitor the host transcriptome, viral variation, and microbiome composition during bNAb development. Early in infection, bNAb producers showed an MHC class I-linked immune activation signature that declined over time, alongside changes in commensal microbial sequences and increased levels of GB virus C in cfRNA. These findings support the use of cfDNA/cfRNA sequencing for biomarker discovery in HIV research.

Introduction

HIV has infected over 36 million people worldwide, reaching epidemic proportions in Eastern and Southern Africa with an estimated adult prevalence close to 7% (1). Antiretroviral therapy (ART) has greatly improved outcomes, allowing HIV to be a manageable condition throughout life. However, global access to ART is difficult to achieve and also requires lifelong daily adherence to a pill taking regime. Moreover, despite effective treatment, transmission of HIV remains common in vulnerable populations, and people living with HIV (PLWH) are at increased risk of aging-related and chronic diseases (2).

Although virtually all PLWH develop strain-specific autologous responses (3–6), few develop broadly neutralizing antibodies (bNAbs), which can neutralize multiple strains of the virus (7), and when administered passively can prevent acquisition of susceptible HIV strains and suppress viral replication in PLWH (8). There is great interest in understanding the drivers of bNAbs, as this could inform the development of an effective vaccine. Efforts to design an effective HIV vaccine have increasingly focused on strategies to induce bNAbs by guiding B-cell affinity maturation, with studies demonstrating promising immunization approaches through the rational design of immunogens (9–11). Despite these advances, which specific cellular processes give rise to bNAbs during infection remains incompletely described. So far,

80 there is evidence that viral and host genetic factors might play a role (12–16), and that a
81 combination of high viral titres and reduced CD4 counts early in infection is associated with
82 the development of bNAbs in some individuals (17,18). Additionally, changes in frequency of
83 certain immune subpopulations, such as T follicular helper cells (Tfh) might also influence
84 affinity maturation towards bNAb production (19–23). Moreover, there is some evidence that
85 bNAbs might arise from pre-existing memory B cells initially stimulated by bacterial antigens
86 -likely from the gut microbiome- which later cross-react with the HIV glycan shield and
87 develop neutralizing capabilities (24,25). Although microbial antigen cross-reactivity has also
88 been shown to lead to non-protective immune responses in HIV vaccine trials (26), a better
89 understanding of the interplay between the host immune response, microbiome and a protective
90 bNAb response could aid in vaccine development.
91
92 Sequencing of plasma cell-free DNA (cfDNA) has become an established tool in clinical
93 medicine, with applications ranging from non-invasive prenatal testing and oncology to organ
94 transplant monitoring (27,28). In contrast, plasma cell-free RNA (cfRNA) provides an
95 emerging complementary approach to measure dynamic changes in the host transcriptome (29).
96 Because transcripts released into circulation can originate from many different cell types and
97 tissues (30), cfRNA enables non-invasive monitoring of systemic physiological and immune
98 states (31,32). To date, cfRNA sequencing has been used in diverse biomedical settings,
99 including the monitoring of placental and pregnancy complications for maternal-fetal health
100 (33–36), to follow bone marrow reconstitution after stem cell transplantation (37), and to
101 characterize host immune responses to tuberculosis infection (38). However, integrative
102 analyses that capture host, viral, and microbial nucleic acids simultaneously remain
103 uncommon, and cfRNA sequencing approaches have been rarely applied to monitor retroviral
104 infections such as HIV. Here, we conducted a pilot study to investigate whether combined

105 cfDNA and cfRNA sequencing of plasma could reveal molecular features associated with the
106 development of bNAbs in a longitudinal cohort of PLWH.
107
108 **Results**
109 ***Study design, sequencing overview and viral analysis***
110 We analyzed 42 plasma samples from 14 PLWH from the CAPRISA studies, of whom 7
111 developed bNAbs and 7 did not. For each participant, plasma was collected at approximately
112 6 months, 1 year, and 3 years post-HIV acquisition and prior to ART (Figure 1A). At the 6-
113 month time point, both groups had comparable CD4 counts and viral loads (Figure S1). The
114 extent of bNAb development was assessed by measuring plasma neutralization breadth against
115 a panel of 18 Env-pseudotyped HIV viruses, summarized as percentage breadth, a continuous
116 measure of bNAb development (Figure 1B, Table S1, Methods).
117
118 We performed combined cfDNA and cfRNA sequencing in each of the 42 samples (Methods).
119 After quality control and filtering, we obtained enough reads to sample the host transcriptome
120 (80% of total reads for cfRNA) and microbiome (0.2% and 20% reads of cfDNA and cfRNA
121 respectively) (Figure S2). Sequencing depth before and after sample QC processing (cleaning)
122 is similar for both cfDNA and cfRNA, with comparable metrics between bNAb producers and
123 non-producers (Figure S3). Similarly, cell-type deconvolution of the host transcriptome
124 showed comparable contributions of the most abundant cell types (e.g. red blood cells, immune
125 cells, platelets) across both groups (Figure S4). We detected HIV reads in the cfRNA assay for
126 all samples, with coverage sufficient to genotype the majority strain in 38 out of 42 samples
127 (Figure S5A,B). For high coverage samples, the nucleotide identity of HIV clustered within
128 individuals based on previous single-genome amplified envelope sequences (Figure 2A) except
129 for one mislabeled sample that was excluded from further analyses (Methods).

130

131 We validated the accuracy of cfRNA-derived consensus sequences by comparing them to
132 previously published envelope sequences from the same CAPRISA participants (Table S2),
133 observing consistent clustering and strain concordance across time points and sequencing
134 methods (Figure 2B, S5C). By comparing differences in consensus sequences within
135 individuals (Figure S5D), we found that most mutations occur in the *env* gene, which encodes
136 the surface protein responsible for host cell binding. The Env protein is the sole target of bNAbs
137 and is known to be the most rapidly evolving part of the HIV genome (39–41). Using cfRNA
138 we estimated that the dominant viral strain in each individual accumulates approximately 10
139 mutations per year (Figure S5E), in line with other estimates from virion or PBMC-derived
140 studies (42,43). Altogether, these results show that cfRNA is a quantitative tool to perform
141 genotype analysis of viral pathogens and to track strain-specific mutational dynamics over
142 time.

143 Analysis of read lengths and mapping start and end positions indicates that HIV-derived reads
144 in plasma cfRNA (cfHIV) predominantly consist of short RNA fragments (200-500 bases) and
145 do not show systematic positional biases along the HIV genome (Figure S6). This is consistent
146 with cfHIV being largely derived from fragmented cell-free viral RNA rather than libraries
147 dominated by intact full-length viral genomes. In addition, to assess the quality of HIV genome
148 recovery from cfRNA, we performed a subsampling analysis showing that approximately
149 50,000 cfHIV reads are sufficient to recover >90% of the genome at ≥50X depth (Figure S7).
150 In this cohort, this level of coverage was achieved for most samples (83%).

151

152 ***Host transcriptome correlates of cfHIV and clinical parameters***

153 To characterize host responses to HIV infection, we measured the correlation between the host
154 transcriptome and the total abundance of cfHIV, as well as with clinical parameters such as

155 viral load and CD4 counts. Spearman correlation of cfHIV, viral load and CD4 counts to the
156 human transcriptome revealed several genes associated with these measurements (Figure 3A,
157 Figure S8A-B). Analysis of genes significantly correlated to cfHIV revealed 19 genes typically
158 associated with host immune responses to viral infection (FDR<0.05), whereas only 1 for each
159 of viral load and CD4 count, respectively (Figure 3B). CD4 count levels showed an association
160 with *IL4*, a cytokine inducing differentiation of naive T helper cells. Although the relation
161 between cytokine profile and HIV progression is controversial, increased secretion of IL4 in
162 relation to IFN-gamma is a known hallmark of switching from Th1 to Th2 lymphokine
163 secretion in PLWH (44). Our observation of IL4 being increased in low CD4 samples is
164 compatible with an increase of IL-4 levels upon disease progression and T helper cell decrease.
165 Top correlates of cfHIV include: *CCL5*, also known as RANTES, an HIV suppression factor
166 released by $CD8^+$ T cells, *TFRC* (CD71), a marker of T-cell activation, and *CCL2*, a chemokine
167 that attracts immune cells to sites of infection (Figure 3C, Figure S8C). These genes do not
168 show a differential behaviour between bNAb producers and controls, indicating that they
169 represent shared responses to HIV infection. This shows that cfRNA sequencing captures
170 shared host transcriptional responses associated with peripheral viral burden during HIV
171 infection.
172
173 ***Development of bNAbs is associated with a transcriptional phenotype of immune activation***
174 We next compared the transcriptome of bNAb producers and non-producers. We performed
175 differential gene abundance analysis finding 256 genes elevated in bNAbs producers
176 (FDR<0.05, fold-change>1, Figure 4A, Methods). This variation could not be attributed to
177 environmental factors such as site of collection, that showed little to no correlation to these
178 transcriptomic signatures. Remarkably, 52% (n=131) of elevated genes in bNAb producers are
179 immune-related (primarily associated with adaptive immunity), compared with only 14%

180 (n=39) of the 281 decreased genes. Among the elevated genes, 25 of them are associated with
181 production of RANTES (CCL5), a chemokine involved in homing and migration of effector
182 and memory T cells. Although CCL5 itself is not enriched in bNAb producers in our dataset,
183 elevated RANTES, along with other inflammatory markers, have previously been linked to the
184 development of bNAbs in spontaneous controllers of HIV (45).
185
186 A pathway analysis confirmed these results and further revealed an enrichment of molecular
187 functions related to an immunologic response in bNAb producers (Figure S9A). Terms such as
188 receptor activity, IgG and peptide antigen binding, cytokine signaling and NF-kB signaling
189 were enriched in bNAb positive samples. Analyses of the genes decreased in the bNAb
190 producers showed lower enrichment scores and did not point to the existence of other host
191 immunoregulatory effects (Figure S9B). Overall, host cfRNA sequencing provides evidence
192 for an increased immune activation in participants who developed bNAbs, despite having
193 similar CD4 counts and viral loads to non-bNAb producers.
194
195 To investigate this immune activation in more detail, we performed pseudotime analysis on the
196 two groups, using plasma neutralization breadth to define the temporal progression (Figure 4B,
197 Methods). Pseudotime is a trajectory-inference approach that orders samples by relative
198 biological state rather than chronological time, allowing comparison across individuals at
199 different stages of bNAb development. We found that the immunologic gene signature is
200 elevated only early in breadth development in bNAb producers (when only a limited number
201 of viral strains are neutralized) and later converges to levels comparable to non-bNAb
202 producers as neutralization breadth increases. This aligns with previous studies showing that
203 factors such as CD4 levels inversely correlate with bNAb development only early in infection
204 (4). However, in our study, both groups had similar CD4 counts at the first time point. Focusing

205 on the genes significantly correlated with breadth in both groups (Figure S9C-D, FDR<0.05)
206 revealed almost the entirety of the MHC class I antigen presentation pathway was
207 anticorrelated with the development of breadth (i.e. pseudotime) and sometimes differentially
208 altered or correlated with the level of cfHIV measured (Figure 4C). Hence, MHC class I antigen
209 genes are highly elevated early in infection in participants who go on to develop bNAbs, and
210 their abundance decrease as breadth develops, becoming comparable to that of participants who
211 do not develop bNAbs. MHC class I antigen presentation is present in almost all cells of the
212 body, presents antigens to be killed by cytotoxic T cells (CD8) but is not generally associated
213 with antigen presenting cells or the production of antibodies by B cells (4). Taken together,
214 these results indicate that early immune activation associated with MHC class I antigen
215 presentation, independent of viral load or CD4 count, is a relevant factor in the development
216 of bNAbs.
217
218 ***Microbiome associated with bNAbs and the development of breadth***
219 Development of bNAbs has been hypothesized to be triggered by antigen cross-reactivity
220 between HIV and bacterial components present in the gut microbiome (24,25,46,47). To test
221 whether a differentiated microbiota could be observed in participants who developed bNAbs,
222 we mapped all non-human reads (both cfDNA and cfRNA) to a database of bacterial and viral
223 genomes. After discarding reads that mapped to potential reagent contaminants (Figure S10,
224 Methods) (48), we identified 249 taxa in cfDNA and 4366 taxa in cfRNA that could be
225 confidently attributed to the participants microbiome. We combined data from all participants
226 to assemble a phylogenetic tree of the human microbiome during HIV infection as measured
227 from circulating nucleic acids (Figure S11). This accounted for ~70% of non-human reads in
228 both cfDNA and cfRNA, and spanned several areas of the tree of life including archaea,
229 bacteria and both DNA and RNA viruses. Among other features, we identified the presence of

230 several DNA and RNA viruses such as Anelloviruses, Flaviviruses, Herpesviruses indicative
231 of concomitant viral infections for several participants. In cfRNA, we found that ribosomal
232 reads encompass 37% of the nonhost microbiome reads, and we also identified a moderate
233 fraction of reads (~20%) that could not be attributed to specific species and mapped to
234 uncultured bacteria.
235
236 We then compared the microbiome of both groups (bNAb producers/controls) finding 19
237 genera that were differentially abundant between both groups (FDR<0.05, $|\log_2 FC|>1$, Figure
238 5A). Among genera enriched in bNAb producers, we identified several taxa that include
239 opportunistic pathogens and are commonly associated with mucosal surfaces and the
240 gastrointestinal tract, including Enterobacter, Serratia and Leclercia (Figure S12). By
241 performing a temporal analysis, we found that the detection of these pathogens does not differ
242 across timepoints and does not precede development of breadth (Figure S13). Overall, and
243 despite the relatively small sample size of this study, our data shows the existence of differences
244 in the microbiota compositions in participants who develop bNAbs from those who do not.
245
246 Additionally, metagenomic analysis of cfRNA revealed the presence of GB virus C (GBV-C,
247 previously known as Hepatitis G) in several participants (Fig. 5B). GBV-C is a flavivirus that
248 infects lymphocytes, has no known pathology, and has been associated with improved survival
249 in people with HIV infection (49,50). Given the phylogenetic relationship between GBV-C and
250 hepatitis C virus (HCV), we also examined reads mapping to HCV. However, we did not find
251 evidence of HCV signal, whereas reads mapping to GBV-C showed broad coverage across the
252 viral genome (Figure S13). In our study, all bNAb producers had non-zero levels of GBV-C (p
253 < 0.05, one-sided Fisher exact test) and showed higher odds of elevated GBV-C levels (Figure
254 5B). Differential gene abundance analysis on samples with high levels of GBV-C (defined as

255 >1/2000 nonhost reads) revealed 42 host genes that are also elevated in these participants.
256 Among top candidates we identified several genes of the MHC class II antigen presentation
257 complex (*HLA-DQA1*, *HLA-DQB1*, *HLA-DRB1*), an HIV suppressive factor involved in
258 inflammatory signaling (*CCL3*), several proinflammatory cytokines (*IL6*, *EBI3*), and
259 immunoglobulin receptors (*KIR3DL2*). Notably, increased expression of inhibitory KIRs,
260 including KIR3DL2, has been associated with bNAb development, suggesting a potential link
261 between natural killer cell maturation and the regulation of antibody production (51). These
262 findings suggest that high levels of GBV-C might be associated with an increased activity of
263 antigen presenting cells.
264
265 **Discussion**
266 In this work, we demonstrate simultaneous monitoring of viral infection and host response
267 through sequencing of circulating nucleic acids (DNA and RNA) in plasma samples of PLWH
268 prior to ART. We show that cfRNA can be used to characterize circulating HIV viral strains
269 and, at the same time, to profile the host transcriptome, allowing characterization of immune
270 activity during HIV infection. Analysis of fragment length and genome coverage suggests that
271 HIV-derived cfRNA likely reflects fragmented viral RNA circulating in plasma rather than
272 being dominated by intact full-length viral genomes. Subsampling analyses further shows that
273 near-complete genome coverages (>90%) at moderate depths (≥50X) can be recovered from
274 cfRNA, supporting the feasibility of viral genome analyses.
275 As cfRNA contains transcripts released from diverse tissues and cell types across the body
276 (30), it offers a systemic view of immune processes that may be otherwise difficult to capture.
277 We find that overall cfHIV levels—defined as HIV-derived fragments in cfRNA—correlate
278 with abundance of host genes involved in immune responses to viral infections, including
279 CCL5, TFRC, and CCL2, consistent with responses to peripheral viral burden. For instance,

280 CCL5 encodes RANTES, a chemokine with HIV-suppressive activity that can block CCR5-
281 mediated viral entry, while TFRC (CD71) and CCL2 (MCP-1) are involved in pathways
282 commonly perturbed during infection (52–54). However, when comparing participants who
283 develop bNAbs with those who did not, we observed a distinct and orthogonal transcriptomic
284 signature suggestive of an enhanced immune activation early in infection that was specific to
285 bNAb producers. This response is independent of viral load and CD4 counts, and is enriched
286 in MHC class I antigen presentation genes, RANTES-associated pathways (though not CCL5
287 itself), and other immune activation pathways.
288
289 Previous studies have highlighted the central role of lymph nodes in shaping humoral responses
290 during HIV infection, yet their direct study remains challenging (23,55,56). Cell-free RNA
291 may provide a non-invasive window into immune activity occurring in tissue compartments,
292 including lymphoid tissues where Tfh-B cell interactions drive affinity maturation. Hence, it is
293 plausible that the immune-related cfRNA transcripts we report here reflects immune activity
294 within lymph nodes, consistent with previous evidence that cfRNA can capture tissue-resident
295 hematopoietic activity such as bone marrow progenitor transcripts (37). However, we
296 emphasize that cfRNA represents an integrated signal from multiple sources, and we cannot
297 exclude the possibility that increased cfRNA levels are due to contributions of circulating
298 immune cells or other tissues.
299
300 Analysis of non-human reads further allowed us to assess the microbiome and virome,
301 revealing a reduced set of taxa whose presence or absence is associated with the development
302 of bNAbs. Microbial sequences detected in plasma cfRNA/cfDNA most likely reflect leakage
303 of microbial products into circulation due to barrier dysfunction and dysbiosis, a process well
304 described in chronic HIV infection (57–60). While gut microbiota has been speculated to

305 increase antigen cross-reactivity, our data cannot establish a direct role in bNAb development.
306 Surprisingly, we found that participants who developed bNAbs consistently had higher levels
307 of GBV-C. GBV-C is a non-pathogenic lymphotropic flavivirus that has been associated with
308 improved outcomes in PLWH (49), with some studies suggesting this may reflect an indirect
309 rather than a causal association (50). Mechanistically, co-infection with GBV-C has been
310 reported to inhibit HIV replication and modulate cytokine release, co-receptor expression, and
311 T-cell homeostasis (50,61,62). In our study, GBV-C abundance correlated with increased
312 abundance of proinflammatory cytokines and adaptative immune genes, raising the possibility
313 that co-infection influences the host immune milieu in a way that could facilitate bNAb
314 development. While causality cannot be inferred from our dataset, understanding such
315 modulatory viral co-infections early after HIV acquisition could inform strategies to mimic
316 beneficial immune environments in the context of HIV vaccine design.
317
318 Our study represents a small pilot investigation (42 samples, 14 PLWH) in a well-characterized
319 cohort of young women from South Africa (63), followed from prior to HIV acquisition into
320 acute infection and before initiation of ART. Biobanked samples from such cohorts are an
321 invaluable resource for studies of factors associated with the development of breadth, as earlier
322 ARV treatment regimens are now standard of care. Although this provides a relatively
323 homogeneous and well-controlled population for an initial discovery study, the small sample
324 size limits generalizability and statistical power. For example, associations between viral load
325 or CD4 counts and bNAb induction reported in other studies may not be detectable here due to
326 the limit size of this pilot cohort. Expanded studies within the CAPRISA cohort or in other
327 settings, and with finer temporal resolution, will be important to validate the findings reported
328 here and their relationship to bNAb development. Despite these limitations, this work
329 illustrates the potential of cfRNA/cfDNA sequencing as a non-invasive approach to interrogate

330 host, viral, and microbial processes simultaneously from a single plasma sample. This approach
331 could be of use to support biomarker discovery in vaccine studies and help identify biological
332 contexts that favor the induction of bNAbs.
333

334 **Materials and Methods**

335 ***Ethics Statement***

336 The CAPRISA 002 Acute Infection study was reviewed and approved by the research ethics
337 committees of the University of KwaZulu-Natal (E013/04), the University of Cape Town
338 (025/2004), and the University of the Witwatersrand (MM040202). All participants provided
339 written informed consent for screening, enrolment and specimen storage.
340

341 ***Participants***

342 The study utilized biobanked plasma specimen of PLWH with known neutralization status
343 enrolled in the CAPRISA cohorts. The CAPRISA 002 Acute Infection cohort was established
344 in 2004 and has been following women from the early/acute stage of infection (63). We
345 included participants from two CAPRISA clinical sites, Ethekwini in Durban (site 1) and
346 Vulindlela (site 2), and accounted for potential site-specific effects in all downstream analyses.
347

348 ***Neutralization measurements***

349 Neutralization breadth was measured using a standardized pseudovirus assay that included 18
350 Env-pseudotyped HIV viruses (Table S1) (64,65). Briefly, HIV pseudoviruses were produced
351 by co-transfecting HEK293T cells with each HIV-1 Env plasmid and a pSG3 backbone. Serial
352 dilutions of heat-inactivated plasma were preincubated with each pseudovirus, followed by the
353 addition of TZM-bl cells. After 48h, infection was quantified by luciferase read-out, and titers

354 reported as the reciprocal dilution where 50% of the virus is neutralized (IC_{50}). Breadth was
355 defined as the percentage of pseudoviruses neutralized at an $IC_{50} \geq 1:45$.
356 Participants were classified as bNAb producers if they achieved ≥20% breadth within year of
357 infection or ≥40% breadth against the 18-pseudovirus panel at any three years post-infection.
358
359 ***Isolation of cfDNA and cfRNA and library preparation***
360 Blood samples were collected in sodium citrate tubes and plasma extracted according to the
361 standard clinical centrifugation protocol of the HIV/AIDS Network Coordination, and stored
362 at -80 C. Each plasma sample was thawed once and split into two aliquots for cfDNA/cfRNA
363 extraction (0.4-1 mL per aliquot). To minimize confounding effects from hemolysis and
364 platelet-derived RNA, plasma samples were visually inspected prior to processing for signs of
365 hemolysis (e.g. pink discoloration), which was not visible for any sample. Immediately before
366 nucleic acid extraction, plasma samples were subjected to a final centrifugation step (1 min at
367 16,000 rpm) to remove any residual debris. cfDNA was extracted using the Qiagen Circulating
368 Nucleic Acid kit. cfRNA was extracted using the Plasma/Serum Circulating RNA and
369 Exosomal Purification kit (Norgen, cat 42800), followed by DNA digestion using Baseline-
370 ZERO DNase (Epicentre) and then cleaned up using the RNA Clean and Concentrator-5 kit
371 (Zymo) (32,66). cfDNA/cfRNA libraries were prepared using the automated Ovation SP
372 Ultralow Library Systems (Mondrian ST) and the SMARTer Stranded Total RNA Pico Input
373 Mammalian kit (Clontech) respectively (32,66). A negative control (1 mL PBS, Gibco) was
374 processed in parallel for every batch to control for potential contaminants (67).
375
376 ***Sequencing methods***

377 Paired end Illumina sequencing (2×75 bp) was performed on the Nextseq 500 for cfDNA and
378 on the Novaseq 6000 for cfRNA. This resulted in a median of 69 million (cfDNA) and 109
379 million (cfRNA) reads per sample (Figure S3).
380
381 ***Sequencing Analysis***
382 Sequencing reads from cfRNA and cfDNA were processed using a shared initial pipeline. First,
383 low quality bases and adapter sequences were trimmed by Trimmomatic v0.38 (68), and
384 overlapping read pairs were merged using FLASH v1.2.11 (69). Merged reads were then
385 aligned to the UniVec database using bowtie2 to remove reads mapping to common vectors,
386 linkers and control sequences (70). These processing QC steps removed only a fraction of reads
387 in cfRNA and cfDNA libraries from most plasma samples (Figure S2-S3). Higher cleaning
388 rates were observed in cfDNA negative controls (~50%) and in three cfRNA plasma samples
389 (~25%), that may reflect lower sample input or quality. Overall, no differences in QC
390 processing were observed between bNAb producers and non-producers.
391
392 The remaining reads were first aligned to the human genome (GRCh38, Ensembl) using
393 bowtie2, retaining unmapped reads, which were subsequently passed through a second
394 homology filtering by BLASTN against the NCBI nt database restricted to taxid 9606, to obtain
395 non-human reads for microbial and viral analysis. Taxonomic classification of non-human
396 reads was performed by BLASTN (NCBI BLAST+ v2.8) against the NCBI nt database, using
397 curated taxid lists limited to viral, bacterial and fungal references (66,67). BLAST hits were
398 filtered by requiring ≥80% nucleotide identity and ≥80% per-read query coverage, also de-
399 duplicating counts to the best-supported accession to avoid counting the same read twice across
400 hits. Subsequently, per-accession counts and coverage profiles were generated and aggregated
401 at taxonomic ranks for downstream analysis.

402

403 For cfRNA, reads were additionally aligned to the human genome using STAR v.25.4a (71)
404 and duplicate-filtered using Picard v2.18.2,and gene abundance was quantified using HTSeq
405 (72). Downstream analysis of gene abundance data is detailed in the following section.

406

407 For HIV-specific analysis, non-human cfRNA reads were re-aligned to the HIV-1 group M
408 consensus reference (CON-C.fa) using bowtie2. Consensus sequences were generated with
409 samtools mpileup and bcftools consensus (73), followed by multiple sequence alignment using
410 MAFFT v7 (74). A maximum-likelihood phylogenetic tree was constructed using IQ-TREE
411 (75), and visualized with Bio.Phylo (76), with metadata-based label coloring and shaded
412 backgrounds highlighting predefined sample clusters. Previously published HIV reference
413 sequences from the same CAPRISA participants were obtained from the Los Alamos National
414 Laboratory (LANL) HIV Sequence Database (Table S2). Sample identity was primarily
415 validated by assessing HIV sequence consistency across all the three time-points per participant
416 (Fig. 2 and S3). As an orthogonal check, we performed HLA-B typing on cfRNA using
417 seq2HLA (v2.3, 4-digit resolution) (77), which confirmed exclusion of one sample (CAP200,
418 time point 1), where a mismatch was detected in both the HIV sequence analysis and HLA-B
419 typing.

420

421 To control for potential environmental contamination in both cfDNA and cfRNA data, six
422 negative controls were used to create a database of background environmental taxa that were
423 blacklisted in downstream analysis. This filtering preserved 72% of microbial reads in cfRNA
424 and 17% in cfDNA. In addition, non-template control barcodes (n=8) were spiked into the
425 cfRNA libraries to estimate sample cross-contamination, which was found to be <0.5% and

426 compatible with index-hopping during sequencing. This threshold was used to define the
427 detection limit for each individual microbial taxon.
428
429 ***Differential Gene Abundance Analysis***
430 Differential Gene Abundance Analysis was performed using edgeR with a generalized linear
431 model and quasi-likelihood F-tests. Site of sample collection was included as a covariate in the
432 model to account for potential confounding effects. All p-values were corrected for multiple
433 hypothesis testing using the Benjamini-Hochberg method, and results are reported as false
434 discovery rate (FDR). For downstream analysis, we selected genes with FDR < 0.05 and
435 absolute fold change > 1. Associations between gene abundance and clinical metadata (e.g.
436 viral load, CD4+ counts) were measured using nonparametric Spearman correlation unless
437 otherwise stated (SciPy package, Python). Differential abundance of microbial taxa was also
438 assessed in edgeR using generalized linear models, with the same thresholds and covariate
439 adjustments. Overall, no significant site-associated differences were observed in either the
440 transcriptomic or microbial enrichment analyses.
441
442 ***Pseudotime analysis***
443 Pseudotime analysis was used as an alternative approach to investigate factors associated with
444 the development of neutralization breadth, given the limited number of longitudinal samples
445 and heterogeneity in bNAb development across individuals. Samples were ordered according
446 to the number of pseudoviruses neutralized by each plasma sample, which was used as a proxy
447 for progression along the bNAb development continuum. This trajectory-inference approach
448 allows us to account for differences in the pace of bNAb development across individuals.
449 Spearman's rank correlation analyses were then performed against this pseudotime variable to

evaluate changes both between bNAb producers and non-producers and along the continuous trajectory of breadth progression.

GBV-C analysis

GBV-C reads were normalized to the total number of non-human reads per sample and scaled by a factor of 1,000 (GBV-C reads per 1,000 non-human reads), analogous to the normalization used for cfHIV. Differences in GBV-C abundance between bNAb producers and non-producers were assessed using a logistic mixed-effects model (statsmodels package, Python), where we model GBV-C abundance as a binary outcome and included participant as a random effect to account for repeated sampling across time points.

Acknowledgments

This work was supported by the funding from the Chan Zuckerberg Biohub to S.R.Q, the Bill & Melinda Gates Foundation (OPP1113682) and the Global Health Vaccine Accelerator Platform (GH-VAP-SI-ID-14) to S.R.Q and P.L.M. We thank the Centre for the AIDS Programme of Research In South Africa (CAPRISA) for sample procurement, Dr Mashudu Madzivhandila for neutralization data, and MD Gerlinde Obermoser (Stanford Center for Human Systems Immunology) for project coordination. P.L.M. is supported by South African Medical Research Council (MRC) SHIP program and the South African Research Chairs Initiative of the Department of Science and Technology and the NRF (Grant No 98341). J.C.-S. is supported by the Knut and Alice Wallenberg Foundation (Wallenberg Molecular Medicine Fellow), the Swedish Research Council (2021-05109) and the Erling Perssons Stiftelse (Swedish Foundations' Starting Grant). M.N.M. was supported by the Stanford Bio-X Bowes Fellowship. Finally, we are indebted to the study participants for their generous support of scientific research.

475

476 **Data availability**

477 The datasets generated and analyzed in the study are available in the NCBI Gene Expression
478 Omnibus (GEO) and Sequence Read Archive (SRA) with accession number GSE313190. The
479 code to reproduce the analyses in this study is publicly available at
480 https://github.com/CamunasLab/cfHIV

481

482 **Bibliography**

483

484 1. UNAIDS Report 2017. Ending AIDS. Progress towards the 90-90-90 targets. (Global Aids
485 Update 2017.). Joint United Nations Programme on HIV/AIDS (UNAIDS); 2017.

486 2. Deeks SG, Overbaugh J, Phillips A, Buchbinder S. HIV infection. Nature reviews Disease
487 primers. 2015;1:15035.

488 3. Moore PL, Gray ES, Choge IA, Ranchobe N, Mlisana K, Abdool Karim SS, et al. The C3-V4
489 Region Is a Major Target of Autologous Neutralizing Antibodies in Human Immunodeficiency
490 Virus Type 1 Subtype C Infection. J Virol. 2008;82:1860–9.

491 4. Gray ES, Moore PL, Choge IA, Decker JM, Bibollet-Ruche F, Li H, et al. Neutralizing
492 Antibody Responses in Acute Human Immunodeficiency Virus Type 1 Subtype C Infection. J
493 Virol. 2007;81:6187–96.

494 5. Li B, Decker JM, Johnson RW, Bibollet-Ruche F, Wei X, Mulenga J, et al. Evidence for Potent
495 Autologous Neutralizing Antibody Titers and Compact Envelopes in Early Infection with
496 Subtype C Human Immunodeficiency Virus Type 1. J Virol. 2006;80:5211–8.

497 6. Richman DD, Wrin T, Little SJ, Petropoulos CJ. Rapid evolution of the neutralizing antibody
498 response to HIV type 1 infection. Proc Natl Acad Sci USA. 2003;100:4144–9.

499 7. Landais E, Huang X, Havenar-Daughton C, Murrell B, Price MA, Wickramasinghe L, et al.
500 Broadly Neutralizing Antibody Responses in a Large Longitudinal Sub-Saharan HIV Primary
501 Infection Cohort. PLoS Pathog. 2016;12:e1005369.

502 8. Corey L, Gilbert PB, Juraska M, Montefiori DC, Morris L, Karuna ST, et al. Two Randomized
503 Trials of Neutralizing Antibodies to Prevent HIV-1 Acquisition. N Engl J Med. 2021;384:1003–
504 14.

505 9. Caniels TG, Prabhakaran M, Ozorowski G, MacPhee KJ, Wu W, Van Der Straten K, et al.
506 Precise targeting of HIV broadly neutralizing antibody precursors in humans. Science.
507 2025;eadv5572.

508 10. Willis JR, Prabhakaran M, Muthui M, Naidoo A, Sincomb T, Wu W, et al. Vaccination with
509 mRNA-encoded nanoparticles drives early maturation of HIV bnAb precursors in humans.
510 Science. 2025;eadr8382.

511 11. Leggat DJ, Cohen KW, Willis JR, Fulp WJ, deCamp AC, Kalyuzhniy O, et al. Vaccination
512 induces HIV broadly neutralizing antibody precursors in humans. Science. 2022;378:eadd6502.

513 12. The Swiss HIV Cohort Study, Rusert P, Kouyos RD, Kadelka C, Ebner H, Schanz M, et al.
514 Determinants of HIV-1 broadly neutralizing antibody induction. Nat Med. 2016;22:1260–7.

515 13. Moody MA, Pedroza-Pacheco I, Vandergrift NA, Chui C, Lloyd KE, Parks R, et al. Immune
516 perturbations in HIV-1-infected individuals who make broadly neutralizing antibodies. Science
517 Immunology. 2016;1:aag0851.

518 14. Haynes BF, Shaw GM, Korber B, Kelsoe G, Sodroski J, Hahn BH, et al. HIV-Host Interactions:
519 Implications for Vaccine Design. Cell Host & Microbe. 2016;19:292–303.

520 15. Roskin KM, Jackson KJL, Lee J-Y, Hoh RA, Joshi SA, Hwang K-K, et al. Aberrant B cell
521 repertoire selection associated with HIV neutralizing antibody breadth. Nature Immunology.
522 2020;21:199–209.

523 16. Landais E, Moore PL. Development of broadly neutralizing antibodies in HIV-1 infected elite
524 neutralizers. Retrovirology. 2018;15:61.

525 17. Euler Z, van Gils MJ, Bunnik EM, Phung P, Schweighardt B, Wrin T, et al. Cross-reactive
526 neutralizing humoral immunity does not protect from HIV type 1 disease progression. J Infect
527 Dis. 2010;201:1045–53.

528 18. Gray ES, Madiga MC, Hermanus T, Moore PL, Wibmer CK, Tumba NL, et al. The
529 Neutralization Breadth of HIV-1 Develops Incrementally over Four Years and Is Associated
530 with $CD4^+$ T Cell Decline and High Viral Load during Acute Infection. J Virol. 2011;85:4828–
531 40.

532 19. Cottrell CA, Hu X, Lee JH, Skog P, Luo S, Flynn CT, et al. Heterologous prime-boost
533 vaccination drives early maturation of HIV broadly neutralizing antibody precursors in
534 humanized mice. Sci Transl Med. 2024;16:eadn0223.

535 20. Cohen K, Altfeld M, Alter G, Stamatatos L. Early preservation of CXCR5+ PD-1+ helper T
536 cells and B cell activation predict the breadth of neutralizing antibody responses in chronic HIV-
537 1 infection. J Virol. 2014;88:13310–21.

538 21. Locci M, Havenar-Daughton C, Landais E, Wu J, Kroenke MA, Arlehamn CL, et al. Human
539 Circulating PD-1+CXCR3−CXCR5+ Memory Tfh Cells Are Highly Functional and Correlate
540 with Broadly Neutralizing HIV Antibody Responses. Immunity. 2013;39:758–69.

541 22. Perreau M, Savoye A-L, De Crignis E, Corpataux J-M, Cubas R, Haddad EK, et al. Follicular
542 helper T cells serve as the major CD4 T cell compartment for HIV-1 infection, replication, and
543 production. J Exp Med. 2013;210:143–56.

544 23. Crotty S. T Follicular Helper Cell Biology: A Decade of Discovery and Diseases. Immunity.
545 2019;50:1132–48.

546 24. Liao H-X, Chen X, Munshaw S, Zhang R, Marshall DJ, Vandergrift N, et al. Initial antibodies
547 binding to HIV-1 gp41 in acutely infected subjects are polyreactive and highly mutated. Journal
548 of Experimental Medicine. 2011;208:2237–49.

549 25. Trama AM, Moody MA, Alam SM, Jaeger FH, Lockwood B, Parks R, et al. HIV-1 Envelope
550 gp41 Antibodies Can Originate from Terminal Ileum B Cells that Share Cross-Reactivity with
551 Commensal Bacteria. Cell Host & Microbe. 2014;16:215–26.

552 26. Williams WB, Liao H-X, Moody MA, Kepler TB, Alam SM, Gao F, et al. Diversion of HIV-1
553 vaccine–induced immunity by gp41-microbiota cross-reactive antibodies. Science.
554 2015;349:aab1253.

555 27. Loy C, Ahmann L, De Vlaminck I, Gu W. Liquid Biopsy Based on Cell-Free DNA and RNA.
556 Annual Review of Biomedical Engineering. 2024;26:169–95.

557 28. Moufarrej MN, Bianchi DW, Shaw GM, Stevenson DK, Quake SR. Noninvasive Prenatal
558 Testing Using Circulating DNA and RNA: Advances, Challenges, and Possibilities. Annu Rev
559 Biomed Data Sci. 2023;6:397–418.

560 29. Koh W, Pan W, Gawad C, Fan HC, Kerchner GA, Wyss-Coray T, et al. Noninvasive in vivo
561 monitoring of tissue-specific global gene expression in humans. Proc Natl Acad Sci USA.
562 2014;111:7361–6.

563 30. Vorperian SK, Moufarrej MN, Tabula Sapiens Consortium, Overall Project Direction and
564 Coordination, Jones RC, Karkanias J, et al. Cell types of origin of the cell-free transcriptome.
565 Nat Biotechnol. 2022;40:855–61.

566 31. Loy CJ, Servellita V, Sotomayor-Gonzalez A, Bliss A, Lenz JS, Belcher E, et al. Plasma cell-
567 free RNA signatures of inflammatory syndromes in children. Proc Natl Acad Sci USA.
568 2024;121:e2403897121.

569 32. Pan W, Ngo TTM, Camunas-Soler J, Song C-X, Kowarsky M, Blumenfeld YJ, et al.
570 Simultaneously Monitoring Immune Response and Microbial Infections during Pregnancy
571 through Plasma cfRNA Sequencing. Clinical Chemistry. 2017;63:1695–704.

572 33. Camunas-Soler J, Gee EPS, Reddy M, Mi JD, Thao M, Brundage T, et al. Predictive RNA
573 profiles for early and very early spontaneous preterm birth. American Journal of Obstetrics and
574 Gynecology. 2022;227:72.e1-72.e16.

575 34. Ngo TTM, Moufarrej MN, Rasmussen M-LH, Camunas-Soler J, Pan W, Okamoto J, et al.
576 Noninvasive blood tests for fetal development predict gestational age and preterm delivery.
577 Science. 2018;360:1133–6.

578 35. Rasmussen M, Reddy M, Nolan R, Camunas-Soler J, Khodursky A, Scheller NM, et al. RNA
579 profiles reveal signatures of future health and disease in pregnancy. Nature. 2022;601:422–7.

580 36. Moufarrej MN, Vorperian SK, Wong RJ, Campos AA, Quaintance CC, Sit RV, et al. Early
581 prediction of preeclampsia in pregnancy with cell-free RNA. Nature. 2022;602:689–94.

582 37. Ibarra A, Zhuang J, Zhao Y, Salathia NS, Huang V, Acosta AD, et al. Non-invasive
583 characterization of human bone marrow stimulation and reconstitution by cell-free messenger
584 RNA sequencing. Nat Commun. 2020;11:400.

585 38. Chang A, Loy CJ, Eweis-LaBolle D, Lenz JS, Steadman A, Andgrama A, et al. Circulating cell-
586 free RNA in blood as a host response biomarker for detection of tuberculosis. Nat Commun.
587 2024;15:4949.

588 39. Burton DR, Stanfield RL, Wilson IA. Antibody vs. HIV in a clash of evolutionary titans.
589 Proceedings of the National Academy of Sciences. 2005;102:14943–8.

590 40. Haddox HK, Dingens AS, Bloom JD. Experimental Estimation of the Effects of All Amino-
591 Acid Mutations to HIV's Envelope Protein on Viral Replication in Cell Culture. Swanstrom R,
592 editor. PLoS Pathog. 2016;12:e1006114.

593 41. Burton DR, Mascola JR. Antibody responses to envelope glycoproteins in HIV-1 infection. Nat
594 Immunol. 2015;16:571–6.

595 42. Cuevas JM, Geller R, Garijo R, López-Aldeguer J, Sanjuán R. Extremely High Mutation Rate of
596 HIV-1 In Vivo. Rowland-Jones SL, editor. PLoS Biol. 2015;13:e1002251.

597 43. Zanini F, Puller V, Brodin J, Albert J, Neher RA. In vivo mutation rates and the landscape of
598 fitness costs of HIV-1. Virus Evolution. 2017;3.

599 44. Meyaard L, Hovenkamp E, Keet IP, Hooibrink B, de Jong IH, Otto SA, et al. Single cell
600 analysis of IL-4 and IFN-gamma production by T cells from HIV-infected individuals:
601 decreased IFN-gamma in the presence of preserved IL-4 production. Journal of Immunology.
602 1996;157:2712–8.

603 45. Dugast A-S, Arnold K, Lofano G, Moore S, Hoffner M, Simek M, et al. Virus-driven
604 Inflammation Is Associated With the Development of bNAbs in Spontaneous Controllers of
605 HIV. Clinical Infectious Diseases. 2017;64:1098–104.

606 46. Mascola JR, Haynes BF. HIV-1 neutralizing antibodies: understanding nature's pathways.
607 Immunological Reviews. 2013;254:225–44.

608 47. Moir S, Fauci AS. B-cell responses to HIV infection. Immunological Reviews. 2017;275:33–48.

609 48. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and
610 laboratory contamination can critically impact sequence-based microbiome analyses. BMC
611 Biology. 2014;12:87.

612 49. Xiang J, Wünschmann S, Diekema DJ, Klinzman D, Patrick KD, George SL, et al. Effect of
613 Coinfection with GB Virus C on Survival among Patients with HIV Infection. N Engl J Med.
614 2001;345:707–14.

615 50. Schwarze-Zander C, Blackard JT, Rockstroh JK. Role of GB virus C in modulating HIV
616 disease. Expert Review of Anti-Infective Therapy. 2012;10:563–72.

617 51. Bradley T, Peppa D, Pedroza-Pacheco I, Li D, Cain DW, Henao R, et al. RAB11FIP5
618 Expression and Altered Natural Killer Cell Function Are Associated with Induction of HIV
619 Broadly Neutralizing Antibody Responses. Cell. 2018;175:387-399.e17.

620 52. Mengozzi M, De Filippi C, Transidico P, Biswas P, Cota M, Ghezzi S, et al. Human
621 Immunodeficiency Virus Replication Induces Monocyte Chemotactic Protein-1 in Human
622 Macrophages and U937 Promonocytic Cells. Blood. 1999;93:1851–7.

623 53. Packard TA, Schwarzer R, Herzig E, Rao D, Luo X, Egedal JH, et al. CCL2: a Chemokine
624 Potentially Promoting Early Seeding of the Latent HIV Reservoir. Moscona A, editor. mBio.
625 2022;13:e01891-22.

626 54. Madrid R, Janvier K, Hitchin D, Day J, Coleman S, Noviello C, et al. Nef-induced Alteration of
627 the Early/Recycling Endosomal Compartment Correlates with Enhancement of HIV-1
628 Infectivity. Journal of Biological Chemistry. 2005;280:5032–44.

629 55. Cirelli KM, Carnathan DG, Nogal B, Martin JT, Rodriguez OL, Upadhyay AA, et al. Slow
630 Delivery Immunization Enhances HIV Neutralizing Antibody and Germinal Center Responses
631 via Modulation of Immunodominance. Cell. 2019;177:1153-1171.e28.

632 56. Lindqvist M, Van Lunzen J, Soghoian DZ, Kuhl BD, Ranasinghe S, Kranias G, et al. Expansion
633 of HIV-specific T follicular helper cells in chronic HIV infection. J Clin Invest. 2012;122:3271–
634 80.

635 57. Alzahrani J, Hussain T, Simar D, Palchaudhuri R, Abdel-Mohsen M, Crowe SM, et al.
636 Inflammatory and immunometabolic consequences of gut dysfunction in HIV: Parallels with
637 IBD and implications for reservoir persistence and non-AIDS comorbidities. eBioMedicine.
638 2019;46:522–31.

639 58. Das Adhikari U, Froehle LM, Pipkin AN, Baharlou H, Linder AH, Shah P, et al.
640 Immunometabolic defects of CD8+ T cells disrupt gut barrier integrity in people with HIV. Cell.
641 2025;S0092867425009791.

642 59. Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, Rao S, et al. Microbial
643 translocation is a cause of systemic immune activation in chronic HIV infection. Nat Med.
644 2006;12:1365–71.

645 60. Nazli A, Chan O, Dobson-Belaire WN, Ouellet M, Tremblay MJ, Gray-Owen SD, et al.
646 Exposure to HIV-1 Directly Impairs Mucosal Epithelial Barrier Integrity Allowing Microbial
647 Translocation. Hope TJ, editor. PLoS Pathog. 2010;6:e1000852.

648 61. Stapleton JT. Human Pegivirus Type 1: A Common Human Virus That Is Beneficial in
649 Immune-Mediated Disease? Front Immunol. 2022;13:887760.

650 62. Acchioni C, Sandini S, Acchioni M, Sgarbanti M. Co-Infections and Superinfections between
651 HIV-1 and Other Human Viruses at the Cellular Level. Pathogens. 2024;13:349.

652 63. Van Loggerenberg F, Mlisana K, Williamson C, Auld SC, Morris L, Gray CM, et al.
653 Establishing a Cohort at High Risk of HIV Infection in South Africa: Challenges and
654 Experiences of the CAPRISA 002 Acute Infection Study. Ugarte-Gil CA, editor. PLoS ONE.
655 2008;3:e1954.

656 64. Montefiori DC. Evaluating Neutralizing Antibodies Against HIV, SIV, and SHIV in Luciferase
657 Reporter Gene Assays. CP in Immunology . 2005;64: Chapter 12:12.11.1-12.11.17.

658 65. Moyo-Gwete T, Scheepers C, Makhado Z, Kgagudi P, Mzindle NB, Ziki R, et al. Enhanced
659 neutralization potency of an identical HIV neutralizing antibody expressed as different isotypes
660 is achieved through genetically distinct mechanisms. Sci Rep. 2022;12:16473.

661 66. De Vlaminck I, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, et al. Temporal response of
662 the human virome to immunosuppression and antiviral therapy. Cell. 2013;155:1178–87.

663 67. Kowarsky M, Camunas-Soler J, Kertesz M, De Vlaminck I, Koh W, Pan W, et al. Numerous
664 uncharacterized and highly divergent microbes which colonize humans are revealed by
665 circulating cell-free DNA. Proc Natl Acad Sci USA. 2017;114:9623–8.

666 68. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
667 Bioinformatics. 2014;30:2114–20.

668 69. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome
669 assemblies. Bioinformatics. 2011;27:2957–63.

670 70. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
671 2012;9:357–9.

672 71. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal
673 RNA-seq aligner. Bioinformatics. 2013;29:15–21.

674 72. Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. Analysing high-throughput sequencing data
675 in Python with HTSeq 2.0. Boeva V, editor. Bioinformatics. 2022;38:2943–5.

676 73. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of
677 SAMtools and BCFtools. GigaScience. 2021;10:giab008.

678 74. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
679 Improvements in Performance and Usability. Molecular Biology and Evolution. 2013;30:772–
680 80.

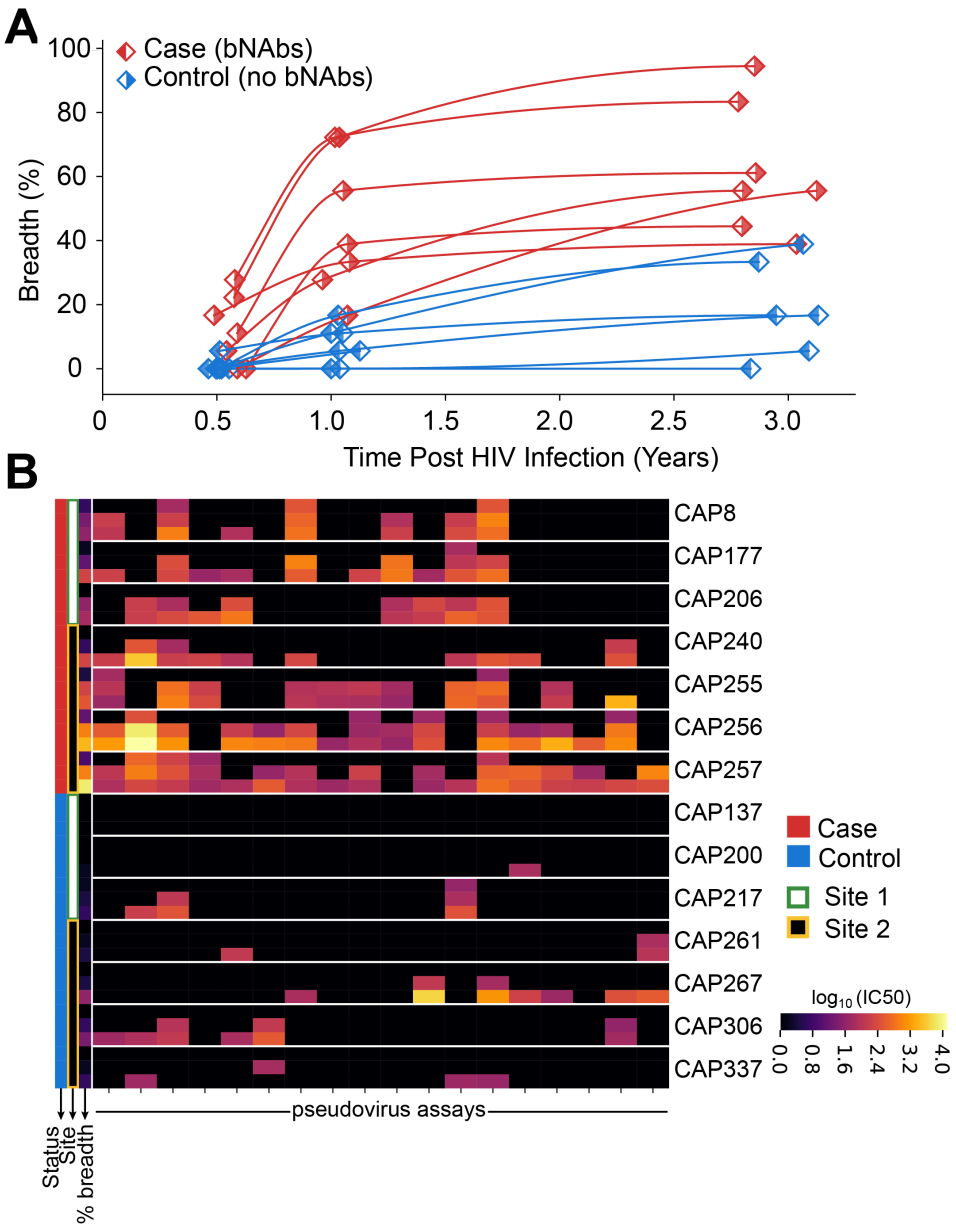
681 75. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al.
682 Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in
683 the Genomic Era. Molecular Biology and Evolution. 2020;37:2461–2461.

684 76. Talevich E, Invergo BM, Cock PJ, Chapman BA. Bio.Phylo: A unified toolkit for processing,
685 analyzing and visualizing phylogenetic trees in Biopython. BMC Bioinformatics. 2012;13:209.

686 77. Boegel S, Löwer M, Schäfer M, Bukur T, De Graaf J, Boisguérin V, et al. HLA typing from
687 RNA-Seq sequence reads. Genome Med. 2012;4:102.

688

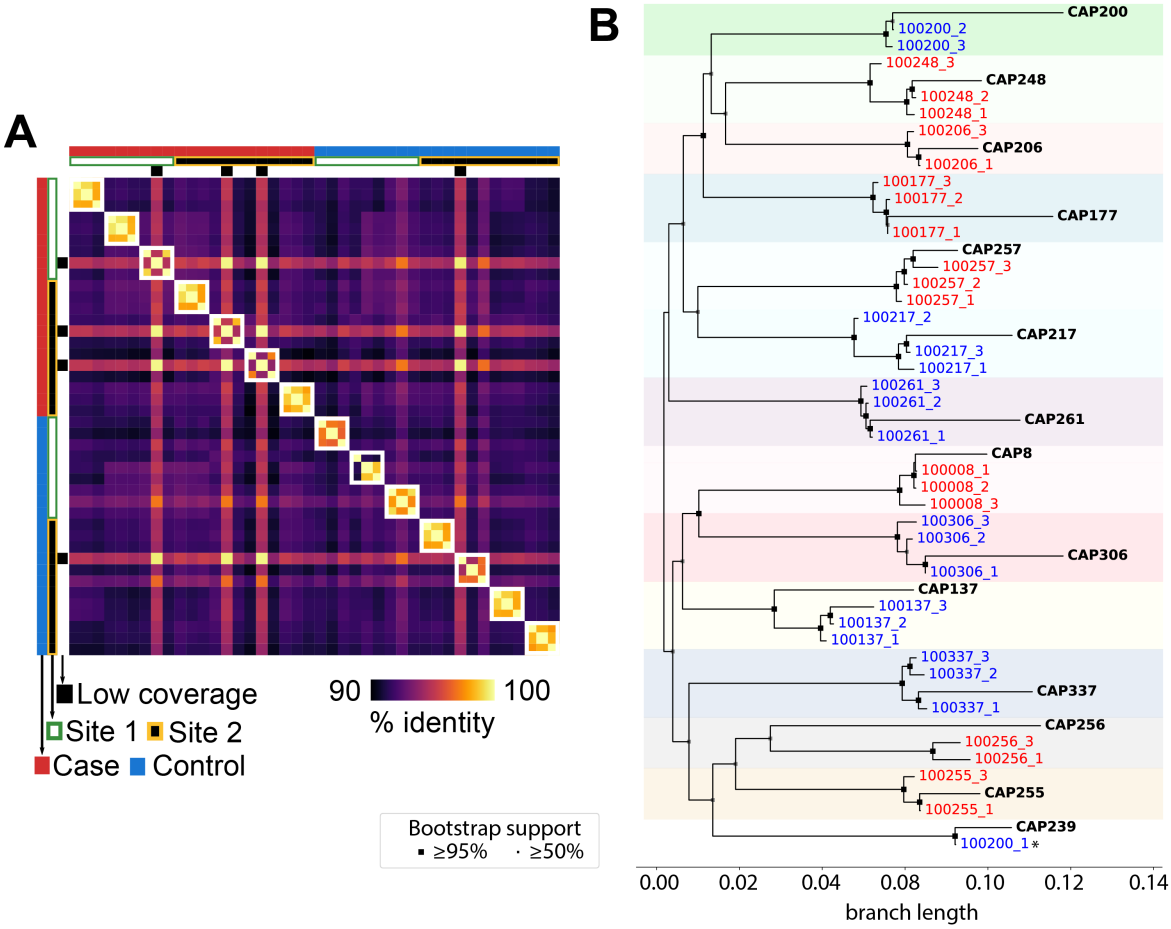
689 **Figure Legends**



690

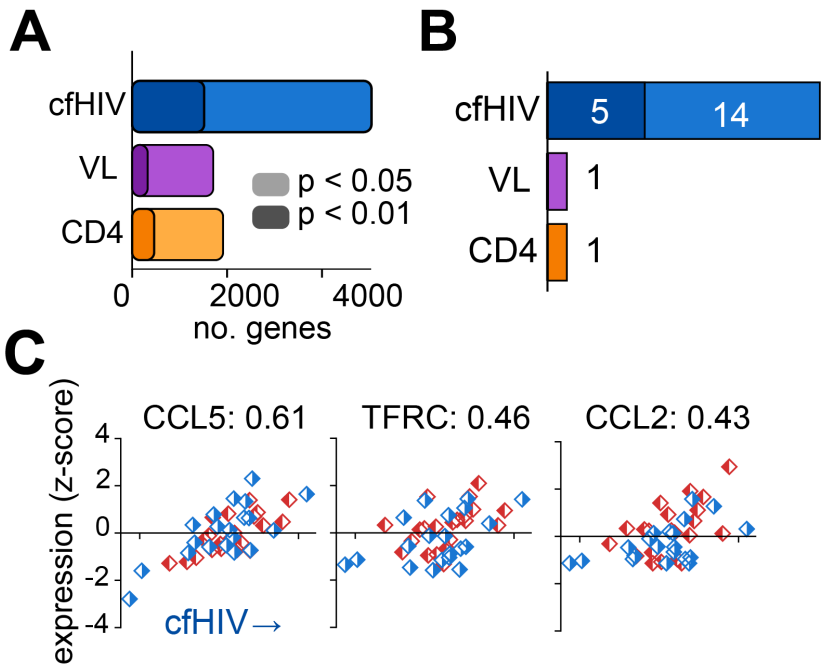
691 **Figure 1: Study design and development of neutralization breadth.** A) Longitudinal
692 trajectories of neutralization breadth in bNAb producers (red) and controls (blue). Time zero
693 corresponds to the estimated HIV infection point, with breadth defined as the ability of plasma
694 antibodies to neutralize a cross-clade panel of 18 env-pseudotyped viruses (Table S2). B)
695 Heatmap of serum neutralization titers (IC50) against the 18-virus panel, grouped by
696 individual, with time points sorted from first (top) to last (bottom). Colors represent log10-

697 transformed IC50 neutralization titers, with higher values (yellow) indicating stronger
698 neutralization potency and black indicating values below the limit of detection.



699

700 **Figure 2: HIV consensus sequences from cfRNA.** A) Genomic similarity matrix of consensus
701 HIV sequences obtained from cfRNA, showing clustering within individuals. B) Maximum-
702 likelihood phylogeny of cfRNA-derived HIV consensus sequences (colored labels) with
703 previously published envelope sequences (bold, CAPRISA). Shaded boxes highlight
704 individual-level clustering. With the exception of one sample, cfRNA consensus sequences
705 consistently cluster with the corresponding envelope sequences.



706

707 **Figure 3: Correlation between cfHIV levels and host transcriptome.** A) Number of genes

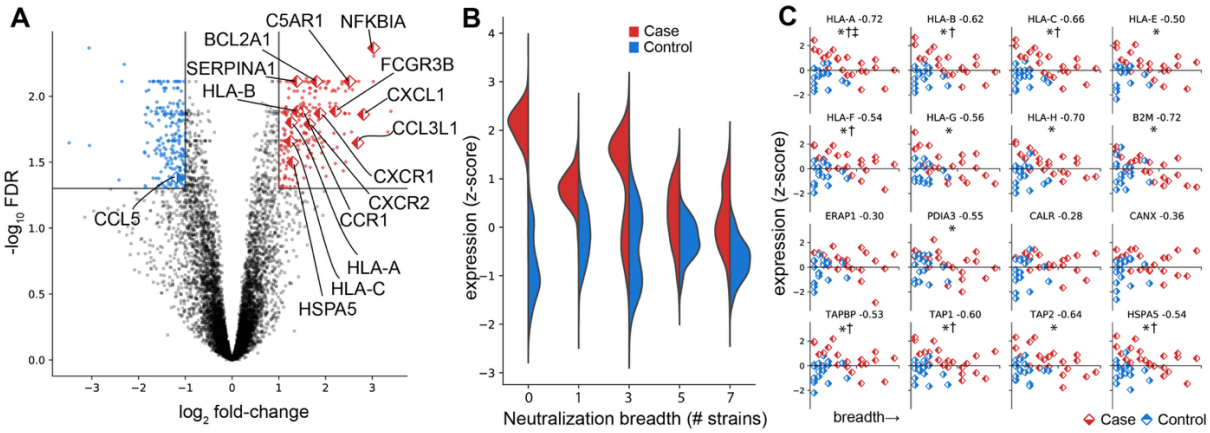
708 significantly correlated with cfHIV counts, viral load (VL) and CD4 counts (Spearman's rank

709 correlation, see Methods). B) Subset of genes from panel A previously described as involved

710 in host immune responses to viral infection. C) Example correlations between cfHIV counts

711 and three HIV-associated genes. Spearman correlation coefficients indicated. No partitioning

712 is observed between bNAb producers (red) and controls (blue).



713

714 **Figure 4: Gene level differences in cfRNA between bNAb producers and controls.** A)

715 Volcano plot of differentially abundant genes. Genes with a fold change > 2 and FDR < 0.05

716 are highlighted (red = increased, blue = decreased in bNAb producers). Labelled points are

genes involved in host immune responses to viral infection. B) Violin plots of the distribution of elevated genes in bNAb producers, stratified by group and neutralization breadth (number of viral strains neutralized), used here as a proxy for progression along bNAb development. C) Genes in the MHC class I antigen presentation pathway. Most show negative correlations with breadth, and some are also differentially abundant between bNAb producers and controls. Correlations with breadth were assessed using Spearman's rank correlation (Spearman's ρ indicated next to each gene, see Methods). * correlated with breadth at FDR <0.05, † differentially abundant between bNAb producers and controls at FDR<0.05, ‡ correlated with cfHIV at FDR <0.05.

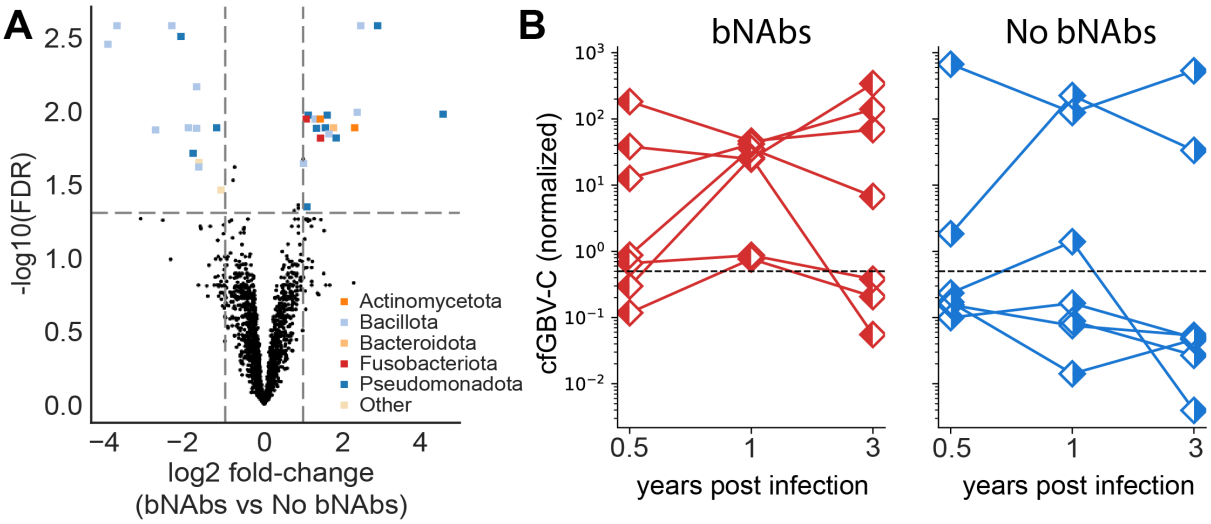


Figure 5: Microbial differences and GBV-C enrichment in bNAb producers. A) Volcano plot of differentially abundant microbial genera between bNAb producers and non-producers, colored by phylum (FDR < 0.05, Benjamini-Hochberg). Positive log2 fold change indicates enrichment in bNAb producers. B) Scatterplots showing cfGBV-C abundance in samples partitioned by time and study group (left, bNAbs producers, right, controls). The dashed line indicates the threshold used to define elevated cfGBV-C levels. Using a logistic mixed-effects model with participant as a random effect, bNAb producers showed higher odds of elevated

736 GBV-C levels (odds ratio [OR]: 6.4, 95% CI: 2.3-17.9). In some participants, GBV-C read

737 counts exceeded those of HIV up to 100-fold.