Disrupted memory T cell expansion in HIV-exposed uninfected infants is preceded by premature skewing of T cell receptor clonality.

Sonwabile Dzanibe[1], Aaron J. Wilk[2], Susan Canny[2,3], Thanmayi Ranganath[2], Berenice Alinde[1,4], Florian Rubelt[5,6], Huang Huang[5,6], Mark M. Davis[5,6], Susan Holmes[7], Heather B. Jaspan[1,8**], Catherine A. Blish[2,6,9**] and Clive M. Gray[1,4**]

Affiliations

1. Division of Immunology, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa.
2. Department of Medicine, School of Medicine, Stanford University, Stanford, CA.
3. Division of Rheumatology, Department of Pediatrics, Seattle Children's Hospital, Seattle, WA USA
4. Division of Molecular Biology and Human Genetics, Stellenbosch University, Cape Town, South Africa.
5. Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA.
6. Institute for Immunity, Transplantation and Infection, Stanford University School of Medicine, Stanford, CA, USA.
7. Department of Statistics, Stanford University, Stanford, CA, USA
8. Seattle Children's Research Institute and Department of Paediatrics and Global Health, University of Washington, Seattle, WA.
9. Chan Zuckerberg Biohub, San Francisco, CA

**Corresponding authors: Clive M. Gray cgray@sun.ac.za; Catherine A. Blish cblish@stanford.edu; Heather B. Jaspan hbjaspan@gmail.com

Keywords: HIV exposure; infant immunity; T cell receptor, NK cels and antibody responses

Word count: Abstract: 154; Main text: 4983; Methods: 2039

Abstract

With increasing success of preventing vertical HIV transmission, there are an increased number of HIV-exposed uninfected infants (iHEU), who have an elevated risk of infections throughout infancy relative to HIV-unexposed and uninfected infants (iHUU). Immune developmental differences between iHEU and iHUU remains poorly understood and here we present a longitudinal multimodal analysis of infant immune ontogeny that highlights the impact of HIV/ARV exposure. Using mass cytometry, we show alterations and differences in the emergence of NK cell populations and T cell memory differentiation between iHEU and iHUU. Specific NK cells observed at birth were also predictive of acellular pertussis and rotavirus vaccine-induced IgG and IgA responses, respectively, at 3 and 9 months of life. T cell receptor Vβ clonotypic diversity was significantly and persistently lower in iHEU preceding the expansion of T cell memory. Our findings show that HIV/ARV exposure disrupts innate and adaptive immunity from birth which may underlie relative vulnerability to infections.

43 **Introduction.**

44 Early life, especially in Africa, can be plagued by high infectious morbidity, and with
45 infectious disease accounting for 2.61 million deaths globally in under-5-year children
46 annually[1]. This high mortality rate is most likely linked to the period during which the
47 immune system adapts to extrauterine life. During this transition window, several
48 factors such as maternal morbidity, microbial exposure, and vaccination significantly
49 modulate infant immunity and consequently influences health and disease outcomes[2–
50 5]. Investigating immune ontogeny and factors that impact immune trajectory during
51 infancy can provide important insights into understanding how better to combat these
52 early life immune stressors.

53

54 Current dogma is that the *in utero* environment is sterile and that newborn infants have
55 limited antigen exposure prior to birth, with T cells being predominantly naïve with little
56 T cell receptor (TCR) engagement[6,7]. This lack of pre-existing adaptive cellular
57 memory early in life increases the vulnerability of infants to infectious agents and
58 disease, especially if there is an absence of adequate breast-feeding to provide
59 passive maternal antibody immunity[3,8,9]. Since the diversity and composition of the
60 TCR is presumably dependent on prior antigen exposure, examining changes in the
61 TCR repertoire during infancy could provide insight into immune development and how
62 specific immune modulatory factors influence susceptibility to infectious diseases.

63

64 The advent of the Option B+ vertical transmission prevention program with effective
65 antiretroviral (ARV) drugs has ensured that HIV transmission has been minimised, but
66 also introduces the possible adverse effect of HIV and ARV exposure on neonatal
67 development *in utero* and in the post-natal period. The intertwined nature of HIV and

68 ARV exposure makes it difficult to unravel, but it is known that HIV/ARV-exposed
69 uninfected infants (iHEU) have higher risk of infectious disease-related morbidity and
70 mortality compared to HIV-unexposed uninfected infants (iHUU) of a similar age,
71 suggesting disruptions to their immune maturation[10–15]. Many studies investigating
72 immunological disparities between iHEU and iHUU are limited to cross-sectional
73 analysis with few studies investigating longitudinal changes and thus there is a lack of
74 evidence on how T cells mature early in life in the context of HIV/ARV exposure. We
75 have previously shown that maternal HIV/ARV exposure alters the dynamics of the T
76 regulatory (Treg) to Th17 cell ratio resulting in a Th17/Treg imbalance associated with
77 gut damage[16]. In this paper, we extend this analysis to investigate HIV/ARV exposure
78 on T cell clonality, memory and NK cell maturation differences between iHEU and
79 iHUU.
80
81 Since early life is marked by inexperienced adaptive immunity, neonatal immune
82 defence is heavily reliant on innate immune cells such as NK cells to rapidly eliminate
83 infections. Compared to adults, however, neonatal NK cells display functional defects
84 such as reduced cytolytic activity including antibody mediated cell cytotoxicity[17,18],
85 decreased expression of adhesion molecules[19] and lower secretion of TNF-α and INF-
86 γ[20]. Earlier studies have further demonstrated that compared to iHUU, iHEU have a
87 lower proportion of NK cells measured at birth and 6 months, and having reduced IFN-
88 γ secretion and perforin expression[21,22]. In addition, the function of NK cells is
89 influenced by the combinatorial signalling through a diverse array of activating and
90 inhibitory receptors expressed on the cell surface[23]. The composition of these
91 receptors is driven by past viral exposure[20,24,25]. Given that NK cells also play a key
92 role in priming the adaptive immune system to respond to invading pathogens or to

93 tailor their immunogenicity towards specific vaccines[26], it is important to investigate
94 the maturation trajectory of NK cells and how this could shape adaptive immunity.
95 Whether early life exposure to HIV/ARV shapes the maturation and/or diversification
96 of NK cells remains unknown.
97
98 Here, we comprehensively investigated the relationship between adaptive and innate
99 immunophenotypes along with TCR diversity and the ability to mount an antibody
100 response to pertussis and rotavirus vaccination. Given the heightened infectious
101 morbidity risk in iHEU, we hypothesised that immune ontogeny is associated with a
102 narrowing TCR repertoire over time in parallel with expanded NK cell subsets in iHEU
103 relative to iHUU. We mapped the immune trajectory of NK and T cell phenotypic
104 clusters in the first 9 months of life and associated this with TCR diversity and antibody
105 titres. Using this approach, we show divergent adaptive immune maturation in iHEU
106 relative to iHUU occurring from 3 months of life with earlier phenotypic differences
107 evident in NK cells. Our data show narrowing and persistent skewing of the TCR
108 repertoire from birth in iHEU that precedes altered $CD4^+$ and $CD8^+$ T cell memory
109 development.

110 **Results**

111 ***Immune cell transition from birth to 9 months of age.***

112 We first analysed the immune ontogeny of innate and adaptive immune cell subsets
113 and marker expression profiles longitudinally during the first 9 months in our iHUU and
114 iHEU samples. Infant samples were analysed longitudinally using a mass cytometry
115 antibody panel (Table S1). This included matched PBMC at birth (n=44) and at week
116 4 (n=53), 15 (n=52) and 36 (n=53), as well as in some cord blood mononuclear cells
117 (n=5). Unsupervised cell clustering on all samples using the live singlet cell population
118 (Figure S1A) revealed 11 cell clusters, shown in Figure 1A, consisting of B cells
119 (cluster 1), T cells ($CD4^+$ T cluster 2 and $CD8^+$ T clusters 3, 4, and 9), monocytes
120 (cluster 8), NK cells (cluster 10), and NKT-like cell (cluster 11). The three $CD8^+$ T cell
121 clusters (3, 4 and 9) were delineated as naïve (13.9%, $CD45RA^+CD27^+CCR7^+$), a
122 small effector memory population (0.07%, EM: $CD45RA^+CD27^+CCR7^-$) and effector
123 cells (4.4%, Eff: $CD45RA^+CD27^-CCR7^-$). We also identified a lineage negative cluster
124 (3.2%) and one cluster expressing HLA-DR only (1.2%) that appeared most closely
125 related to the monocyte cluster (Figure 1A & S1B). Although we employed
126 computational doublet removal from our clustering analysis, a minor miscellaneous
127 population (0.65%) co-expressing typically mutually exclusive lineage markers was
128 identified (Figure 1A & S1B). This small population is shown as a central cluster 7 in
129 the dimensional reduction of the immune cell clusters (Figure 1A) and was excluded
130 from downstream analysis.

131

132 When we examined these clusters over time, we observed an age dependent
133 segregation of cells. Proportions of clusters changed over 36 weeks of life, where
134 monocytes and CD4+ T cells diminished with age and B cells and effector memory

135 CD8+ T cells increased (Figure 1B). Monocyte (ρ=-0.4, p <0.001) and CD4$^+$ T cell (ρ=-
136 0.36, p <0.001) clusters were negatively correlated with infant age, whereas the EM
137 CD8$^+$ T cell and B cell clusters were positively correlated (ρ=0.66 and ρ=0.54
138 respectively, p value <0.001; Figure 1C). We further manually gated on CD4$^+$ and
139 CD8$^+$ T cells and targeted the NK cell population using lineage-exclusion manual
140 gating (Figure S1A). Using the median marker expression for each sample,
141 multidimensional scaling (MDS) revealed intra-individual variability associated with
142 infant age (Figure 1D). For both CD4$^+$ T cells and NK cells, the grouping of the samples
143 by marker expression displayed a converging trajectory along the MDS2-axis (Figure
144 1D). Both of these cell populations had a linear converging trajectory, although
145 stabilizing from week 15 for NK cells (Figure 1E). CD8$^+$ T cells, however, revealed a
146 divergent trajectory along the MDS1-axis (Figure 1D), with a significant expansion by
147 week 15 (Figure 1E).
148 Overall, these detailed unsupervised phenotypic changes in cell clustering illustrate
149 the archetypal progression and transition of infant immunity from neonatal to an adult-
150 like immune phenotype by 9 months of age, being characterised by a shift from innate
151 myeloid cells to an increase in lymphoid adaptive immune cells.
152
153 ***CD4 and CD8 T cell memory maturation is disrupted by HIV exposure after***
154 ***three months of life.***
155 To identify in more depth how HIV/ARV exposure disrupts T cell phenotypes, we
156 parsed manually gated CD4 and CD8 cells from all samples and time points using
157 unsupervised cell clustering which generated 13 CD4$^+$ T cell clusters and 7 CD8$^+$ T
158 cell clusters (Figures 2A and 2D). Tables 1 & 2 ascribe the phenotype for each of the
159 CD4$^+$ and CD8$^+$ T cell clusters, respectively. The marker expression patterns and

160 relationships between the cell clusters are shown using unsupervised hierarchical
161 clustering heatmaps in Figure S2A.
162
163 Across all time points, 13 clusters of CD4+ T cells existed (Figure 2A and Table 1)
164 where naïve cells (cluster 2, CD45RA^{+}CD27^{+}CCR7^{+}) made up the predominant
165 population with an overall frequency of 84.1% (Figure 2A & Table 1). The remaining
166 12 clusters for CD4+ T cells consisted of minor populations with the top three being
167 activated naïve-like cells (cluster 4, CD38^{+}HLA-DR^{+}CD45RA^{+}CD27^{+}CCR7^{+}), PD-1
168 expressing activated cells (cluster 12, PD-1+CD38+) and cytotoxic terminally
169 differentiated cells (cluster 3, CD45RA^{+}CD27^{-}CCR7^{-}CD57^{+}Per^{+}PD-1^{+}). When these
170 cell clusters were examined over infant age, there was a proportional increase in
171 cluster 1 (Th2-like, CCR4^{+}CCR6^{-}CXCR3^{-}), decrease in cluster 4 (activated cells,
172 CD38^{+}HLA-DR^{+}CD45RA^{+}CD27^{+}CCR7^{+}) and maintenance of cluster 2 (naïve,
173 CD45RA^{+}CD27^{+}CCR7^{+}) (Figure 2B). We also assessed these compositional
174 differences by performing Principal Component Analysis (PCA) on centred log-ratios
175 of the relative cluster abundance for each sample. Consistent with the observed
176 immunological trajectory shown in Figures 1D, phenotypic compositions were strongly
177 influenced by infant age, where early time points (cord, birth, and week 4) were
178 grouped together and distinguishable from week 15 and 36 time points (Figure 2C).
179 Looking at CD8^{+} T cells in the same manner, we identified 7 cell clusters, being less
180 diverse than CD4^{+} T cells, but a more distinct clustering between naïve and
181 effector/terminally differentiated memory (Figure 2D and Table 2). It was also clear
182 that naïve CD8^{+} T cells were less populous and that clusters 1 (terminally
183 differentiated, Perforin^{+}CD45RA^{-}CD27^{-}CCR7^{-}CD57^{+}), 5 (cytotoxic cells,
184 Perforin^{+}CD57^{+}) and 6 (cytotoxic effector memory, Perforin^{+}CD45RA^{-}CD27^{-}CCR7^{-})

185 made up a higher proportion of cells than in $CD4^+$ T cell compartment. When
186 separating out these clusters over time, these perforin-expressing cell clusters (1, 5
187 and 6) increased over 9 months of life (Figure 2E), with a contraction of the naïve cell
188 population. Like $CD4^+$ T cells, there was an age dependent $CD8^+$ T cell compositional
189 change (Figure 2F).
190
191 We next determined whether there were any disruptions in T cell cluster compositions
192 due to HIV exposure. There were statistically significant differences in PCA-2 at weeks
193 15 and 36 for $CD4^+$ T cells between iHEU and iHUU (Figure 2G) and a converse
194 difference at week 36 for $CD8^+$ T cells (Figure 2H). To account for these divergent T
195 cell phenotypes, we used differential abundance testing to determine which of the cell
196 clusters were responsible for these differences. CD4 clusters 4, 5, 6, 7, 9 and 11 were
197 significantly lower in iHEU compared to iHUU at week 15, with clusters 5, 7 and 9
198 remaining lower in iHEU at week 36 (Figure 2I). These less frequent of cell clusters in
199 iHEU were characterised as either being activated (cluster 4: $CD38^+$HLA-DR^+),
200 expressing perforin (clusters 7, 9 and 11) or exhibiting terminal differentiation
201 ($CD57^+$/PD-1^+; clusters 5 and 6). In contrast, the $CD4^+$ Th2-like cluster 1 was
202 significantly higher in iHEU compared to iHUU at week 15 but was no longer
203 significantly elevated by week 36 (Figure 2I). Moreover, $CD4^+$ Th2 cells (cluster 1) and
204 cytotoxic terminally differentiated cells (cluster 3: $CD45RA^+CD27^-CCR7^-$
205 $CD57^+$Perforin$^+$PD-1^+) were also higher in iHEU compared to iHUU at week 4 ($p=0.007$
206 and $p=0.013$, respectively) although these differences were not statistically significant
207 following FDR correction (Figure S2B).
208 For $CD8^+$ T cells, there were significantly lower mean frequencies for clusters 2 (NKT-
209 like cells: $CD56^+CD16^+NKp30^+NKp46^+NKG2A^+Per^+$), 3 (Naïve-like activated cell:

210 CD45RA^{+}CD27^{+}CCR7^{+}CXCR3^{+}CCR4^{+}HLA-DR^{+}) and 7 (proliferating EM: Ki67^{+}
211 CD45RA^{+}CD27^{+}CCR7^{-}Per^{+}) in iHEU compared to iHUU at week 15 (Figure 2I). By
212 week 36, only the mean frequencies of clusters 2 (NKT-like cells) and 3 (Naïve-like
213 cells) remained significantly lower in iHEU compared to iHUU (Figure 2J).
214
215 A similar analytical approach was undertaken for NK cells, and of the 10 clusters
216 identified (Figure 3A), 9 were CD56$^{dim/+}$CD16$^{dim/+}$ (Table 3). This included
217 CD56^{Hi}CD16^{-} (cluster 9), which had high NKG2A expression, six clusters that were
218 CD56$^{dim/+}$CD16$^{dim/+}$ (clusters 1, 4, 5, 6, 8 and 10) that were predominantly expressing
219 perforin, and 2 clusters that were CD56^{-}CD16^{+} (clusters 3 and 7). Only cluster 2 was
220 characterised as CD56^{-}CD16^{-} and expressed DNAM1 (Figure S2A & Table S6). The
221 NK cell phenotypic composition was also observed to be strongly associated with
222 infant age (Figure 3B and C), with cluster 1 (Per^{+}CD45RA^{+}Fc$_{\varepsilon}$RIIγ^{+}) expanding over
223 36 weeks and cluster 2 (CD56^{-}CD16–DNAM1^{+}) contracting (Figure 3B). Of note, the
224 perforin-expressing cluster 5 was expanded at weeks 15 and 36. Compositional
225 differences in NK cell clusters between iHEU and iHUU were observed at week 4,
226 earlier than for T cells, and while this trend persisted at week 15, the differences were
227 not statistically significant (Figure 3D). Differential abundance testing revealed that at
228 week 4, iHEU had higher frequencies of cluster 1 (CD56^{dim}CD16^{dim}) co-expressing
229 Per^{+}CD45RA^{+}FcERIy+ (p=0.04, p.adj=0.2) and cluster 5 (CD56^{dim}CD16^{+}) co-
230 expressing Per^{+}CD57^{+}CD45RA^{+}CD38^{+} (p=0.03, p.adj=0.2) compared to iHEU (Figure
231 S2C). This trend continued at week 15 for cluster 5 (p=0.041, p.adj=0.20). Also at
232 week 15, we observed lower frequencies of cluster 10 (CD56^{+}CD16^{+}: p=0.027,
233 p.adj=0.2) in iHEU compared to iHUU (Figure S2C). These findings show that that the
234 ontogeny of immune cells is ordered, and that HIV/ARV exposure sequentially disrupts

235 immune trajectory over time after birth beginning with that of innate (NK) cells and later
236 disrupting memory T cell maturation.
237
238 ***Early TCR repertoire skewing in iHEU.***
239 To relate altered T cell memory maturation trajectories to T cell repertoire changes,
240 we characterised TCR clonotypes in fractionated naïve and memory T cells from
241 matching samples used to define memory lineage in the CyTOF panel. Each PBMC
242 sample was sorted into four T cell subsets: naïve ($CD45RA^{+}CD27^{+}CCR7^{+}$) $CD4^{+}$ and
243 $CD8^{+}$ cells and total memory ($CD45RA^{-/+}CCR7^{-/+}$) T cells (Figure S3A). We subjected
244 these sorted cell fractions (n=934 samples) to bulk sequencing of the TCRβ locus,
245 which enabled identification of TCR clones in 885 samples. The number of unique
246 clones identified per T cell subset positively correlated with the total number of cells
247 sequenced (Figure S3B). There were 24 samples that had either fewer than 10 clones
248 or the number of unique clones were greater than the number of cells sequenced, and
249 these were removed from downstream analysis. Therefore, we included 861 samples
250 having a total of 238,092 reads (range: 11-5451) and an overall median number of
251 unique clones of 81 (IQR: 42-156). The CDR3 lengths of the clonotypes were evenly
252 distributed across infant age and between iHEU and iHUU (Figure S3C).
253
254 The diversity and richness of the TCR clonotypes, determined by inverse Simpson
255 diversity index and Chao1 richness scores respectively, remained unchanged with
256 infant age (Figure S3D & S3E). However, when the TCR diversity scores were
257 stratified by infant HIV exposure status, iHEU TCR clonotypes had relatively lower
258 diversity compared to iHUU starting as early as birth (Figure 4A). These differences
259 were apparent for naïve $CD4^{+}$ T cells at weeks 15 and 36 and for naïve $CD8^{+}$ T cells

260 at birth and week 36. Differences in memory T cell diversity scores were observed to
261 occur only at birth, being significantly lower in iHEU for both $CD4^+$ and $CD8^+$ T cells
262 (Figure 4A). Similarly, richness (the number of unique TCR clones) was significantly
263 lower in iHEU relative to iHUU for both $CD4^+$ and $CD8^+$ T cells at birth (Figure 4B).
264 This remained statistically significant for naïve $CD4^+$ T cells at week 36 (Figure 4B).
265 When examining memory T cells, significant differences in TCR richness were
266 restricted to earlier timepoints, being lower in iHEU compared to iHUU at birth and
267 week 4 for $CD4^+$ T cells and only at birth for $CD8^+$ T cells. We also observed differences
268 in the overall structural overlap, as measured by Jaccard indices, of the memory $CD4^+$
269 TCR repertoire between iHEU and iHUU at birth, although this difference was less
270 prominent at later time points (Figure S3F). In contrast, the TCR structural overlap for
271 naïve $CD8^+$ T cells was lower in iHEU compared to iHUU only at week 36 (Figure
272 S3F). Differences in Vβ gene usage were observed between iHEU and iHUU with
273 most genes being used more frequently in iHEU (Figure S4). Together, these results
274 show expansion of specific TCR clones among iHEU, resulting in skewing of the TCR
275 repertoire relative to iHUU.
276
277 We next wished to determine if there was a relationship between TCR diversity and
278 the clusters of $CD4^+$ and $CD8^+$ T cells identified in Figures 2A-F. By using Spearman's
279 rank correlation between T cell clusters and the inverse Simpson TCR diversity scores
280 for naïve and memory $CD4^+$ and $CD8^+$ T cells, we found that the frequencies of $CD4^+$
281 T cell clusters (4-6: CD38+HLA-$DR^+CD45RA^+CD27^+CCR7^+$,
282 $CCR4^+CCR6^+CXCR3^+CD57^-$, $CCR4^+CCR6^+CXCR3^+CD57^+$, and 9-11:
283 $Perforin^+Ki67^+CD25^+$, $Perforin^+Ki67^-HLA^-DR^+CD69^+CD39^+$, Per^+Ki67^+) from iHEU
284 were positively correlated with the TCR diversity scores derived from sorted naïve

285 CD4$^+$ T cells (Figure 4C). Similarly, frequencies of CD8$^+$ T cells clusters (2, 3 and 7:
286 CD56$^+$CD16$^+$NKp46+NKG2A$^+$, HLA-DR$^+$, Ki67$^+$CD45RA$^+$CD27$^+$CCR7$^-$) were
287 positively correlated to the TCR diversity scores derived from the sorted naïve CD8$^+$
288 T cells at either birth or week 4 (Figure 4D). No T cell clusters were significantly
289 correlated to TCR diversity derived from sorted memory T cells (data not shown).
290 Collectively, our findings show skewing of the TCR repertoire in iHEU relative to iHUU
291 from birth, in the naïve T cell compartment prior to phenotypically defined memory
292 maturation of CD4 and CD8 T cells.
293
294 ***Predicted TCR recognition of epitopes found only in iHUU.***
295 To understand the predicted antigen specificities of the TCR clonotype differences
296 between iHEU and iHUU, we used GLIPH (grouping of lymphocyte interactions by
297 paratope hotspots)[27,28]. TCR specificities from sorted naïve CD4$^+$ T cells in iHUU
298 appeared enriched with specificities for cytomegalovirus (CMV), Hepatitis C virus
299 (HCV), SARS-Cov2 and HIV-1 (Figure 4E). Enrichment of Influenzae A specific clones
300 was more apparent in memory CD4$^+$ T cells at the week 36. For memory CD4$^+$ T cells
301 at week 4, the shared specificity groups in iHUU were enriched for CMV clones.
302 Similarly, for CD8$^+$ T cells, there were more specificities in the naïve compartment at
303 both earlier and later time points with less predictions in memory CD8$^+$ T cells. TCR
304 specificities in sorted naïve CD4$^+$ and CD8$^+$ cells from iHUU appeared consistent
305 throughout 36 weeks, whereas specificities in the CD4 memory compartment
306 fluctuated and appeared predominantly at 4 and at 36 weeks (Figure 4E). Conversely
307 for iHEU sorted T cell populations, no enriched TCR specificities in either naïve or
308 memory CD4$^+$ or CD8$^+$ cells at any of the time points were predicted.
309

310 ***NK cell clusters are the strongest predictors of antibody responses.***

311 To relate how early life T and NK cell clusters may associate with vaccine-induced
312 antibody responses, we measured the response to acellular pertussis vaccination from
313 birth to 36 weeks and to rotavirus at 36 weeks. We initially hypothesised that iHEU
314 would have lower IgG and IgA responses compared with iHUU after vaccination with
315 routine administration of these vaccines, which may explain a mechanism for higher
316 infection rates of these pathogens in iHEU. To test this, IgG antibody levels against
317 pertussis were measured at birth, weeks 4, 15 and 36, with rotavirus specific IgA and
318 neutralization titres measured at week 36 (Figure 5B and C). The median anti-
319 pertussis IgG OD values were lower in iHEU compared to iHUU at birth and week 4
320 (pre-vaccination at 6 weeks; Figure 5A). However, at week 15 (1 week post 3rd
321 vaccination), iHEU showed a significantly higher IgG response relative to iHUU,
322 persisting until week 36 (Figure 5A). No significant differences were observed in anti-
323 rotaviral IgA levels or ability to neutralize rotavirus between iHEU and iHUU (Figure
324 5B, C).

325

326 We built a prediction model by grouping all infants (iHEU and iHUU) into those that
327 had a protective pertussis-specific IgG titre (≥0.469 OD_{Abs}) denoted as responders and
328 those that were non-responders (<0.469 OD_{Abs}). Most of the infants at birth and week
329 4 (89.7% and 94.1%, respectively) were determined to be non-responders as
330 expected, since this was prior to vaccination. At week 15 and 36 of age, 46.4% and
331 34.0% infants respectively were determined to be responders (Figure 5A). Multivariate
332 unbiased variable selection using MUVR[29] was performed to determine early life cell
333 clusters that could distinguish later pertussis vaccine responders from non-
334 responders, according to our classification. Certain cell cluster abundances at birth

335 and week 4 selected using MUVR were associated with low misclassification scores
336 of a pertussis vaccine response at weeks 15 and 36 (Figure S4). We then used these
337 abundances of the MUVR selected clusters to perform a partial least squares
338 discriminant analysis (PLS-DA) with the pertussis IgG responses at week 15 and 36
339 as response variables (Figures 5D and E, respectively). The latent variables (LV) of
340 the cell clusters as determined by PLS-DA were then used to compute area under the
341 curve (AUC) of the receiver operating curve (ROC) distinguishing pertussis
342 responders from non-responders. This approach revealed that NK cell clusters (1, 3,
343 4, 6, 8 and 10) at birth and at week 4 (clusters 1, 2, 3 and 9, Table 3) were ranked the
344 highest in the list of predictors of the pertussis IgG vaccine response at week 15
345 (Figure 5D, S4A & S4B, at birth: AUC = 0.93 and p=0.005; at week 4: AUC = 0.85 and
346 p=0.012). The cell clusters derived from $CD4^+$ or $CD8^+$ T cells were not significantly
347 predictive of an IgG pertussis vaccine response at weeks 15 and 36 (Figure 5D, S4B-
348 D, for CD4 clusters at birth: AUC=0.73, p=0.063; for CD8 clusters: AUC=0.69, p=0.12;
349 at week 4: AUC=0.70, p=0.06; for CD8 clusters: AUC=0.70, p=0.053).
350
351 Using the same approach for rotavirus responses, we used an arbitrary threshold of
352 overall median concentrations for IgA responses to dichotomize infants into those that
353 had concentrations below the median (low-responders) and those above the median
354 (high-responders; Figure 5B). In iHUU, 66.7% were high rotavirus vaccine responders
355 compared to 45.5% iHEU (p=0.11). We subsequently determined which immune cell
356 clusters at birth and week 4 were predictive of IgA responses measured at week 36
357 using MUVR (Figure 5F). PLS-DA analysis using the MUVR selected cell clusters
358 revealed that NK cells at birth were highly predictive of an IgA rotavirus response at
359 week 36 (Figure 5F, AUC = 0.92, p = 0.0007). Neither $CD4^+$ or $CD8^+$ T cell clusters

360 were predictive of a rotavirus response at week 36 (Figure 5F, for CD4 clusters at
361 birth: AUC=0.7, p=0.05; for CD8 clusters: AUC=0.55, p=0.59; at week 4: AUC=0.68,
362 p=0.042; for CD8 clusters: AUC=0.6, p=0.24).
363
364 We also tested whether contemporaneous cell clusters might be related to pertussis
365 responses. When we analysed the frequencies of the cell clusters at weeks 15 and 36
366 with pertussis antibody levels at weeks 15 and 36, $CD4^+$ T cell cluster 1 (Th2 cells)
367 was significantly correlated to pertussis antibody titres at week 15, while at week 36
368 $CD4^+$ T cell clusters 6, 7, 9-11, and 13 (Table 1) were inversely correlated to pertussis
369 antibody titres (Figure 5G). Only $CD8^+$ T cell cluster 7 (Proliferating EM:
370 $Ki67^+CD45RA^+CD27^+CCR7^-$) was observed to be inversely correlated to pertussis
371 antibody titres measured at week 36 (Figure 5G). Of note, there was no significant
372 contemporaneous correlation between the NK cell clusters and pertussis antibody
373 titres (Figure 5G).
374 Collectively, these data show that various NK cell phenotypes at birth and in the first
375 1 month of life can predict vaccine-induced antibody responses to acellular pertussis
376 and rotavirus at 9 months, regardless of infants being exposed to HIV.

377 **Discussion**

378 Herein, we describe immune maturation in the first 9 months of life and how HIV/ARV
379 exposure alters this ontological trajectory. Our findings show ordered immune changes
380 during infancy, typical of the transition from innate-like and naïve neonatal immunity
381 towards memory differentiated adaptive infant immunity[30]. NK cell phenotype and
382 cluster composition differed between iHEU and iHUU at birth and week 4 and was
383 predictive of vaccine responses later in life, whereas T cell memory maturation was
384 impacted after 15 weeks. There was skewing of the TCR repertoire in naïve and
385 memory cells in iHEU from birth, relative to iHUU, and the persistent lack of TCR
386 diversity could consequently account for the increased vulnerability to infections
387 reported among iHEU[10].

388

389 Although differentiated T cells are detected as early as 12-14 weeks gestation[31,32],
390 neonatal immunity is predominantly naïve and characterised by the abundance of
391 CD45RA$^+$ T cells[7]. T cell memory expansion during infancy is therefore essential in
392 protecting against pathogens[33]. The differences in the kinetics of CD4$^+$ and CD8$^+$ T
393 cell memory maturation in our study could not be explained by changes in TCR
394 clonotypes since for both T cell subsets, TCR diversity and richness did not vary over
395 time for either iHEU or iHUU. We presume that the steady rate of CD4$^+$ T cell
396 maturation is partially explained by the abundance of CD4$^+$ T cells early in life,
397 responsible for regulating immunological responses and thus facilitating adaptation to
398 extrauterine life laden with microorganisms[6,34,35].

399 We show that the T cell compartment was significantly impacted by HIV/ARV
400 exposure. The overall trajectory of T cell phenotypic composition in iHEU diverged
401 from that of iHUU, particularly after week 15 of life. CD4$^+$ and CD8$^+$ T cell

402 compartments in iHEU had lower frequencies of activated cells and a minor cell
403 population expressing either perforin or PD-1 compared to iHUU. These findings
404 contradict earlier reports showing iHEU having heightened immune activation and/or
405 exhaustion[36,37], which may be related to mixedfeeding or non-breastmilk feeding that
406 could introduce water contaminants or alter infant microbiome[38]. Mothers in our study
407 were encouraged to exclusively breastfeed, likely accounting for lower levels of
408 activated T cells in iHEU compared to iHUU[39]. Nonetheless, reduced or impaired
409 memory T cells among iHEU compared to iHUU have been previously demonstrated
410 and associated with poor childhood vaccine responses and/or increased
411 hospitalization[16,39–41].
412
413 The 15 week period before significant expansion of T cell memory differentiation was
414 preceded by significantly reduced diversity and richness in memory T cell clonality
415 from birth in iHEU relative to iHUU. This was in agreement with others who reported
416 lower βTCR clonality in the cord blood of iHEU[42] and such reduced TCR diversity
417 found at birth infers that T cells have undergone clonal expansion *in utero*. Using
418 GLIPH[27,28], we could not detect any predicted antigen specificity enriched among
419 iHEU, in contrast to earlier studies showing high frequency of HIV-1 specific clones[42]
420 and reactivity of T cells towards HIV proteins in iHEU[43]. What was evident instead was
421 that naïve $CD4^+$ and $CD8^+$ TCR reactivities in iHUU were targeting CMV, EBV and
422 HCV antigens. There may be three possible reasons for naïve cells showing clonal
423 expansion and increased diversity in iHEU. Firstly, that our isolation of
424 CD45RA+CCR7+CD27+ naïve population was contaminated with some memory cells
425 during the sorting procedure. The contribution of sorting contamination would have
426 negligible outcomes as we obtained >95% purity after sorting and any inclusion of

427 memory cells would be of extremely low proportion. Additionally, the frequency of
428 naïve T cells was significantly higher compared to that of memory T cells throughout
429 infancy which could contribute to sampling bias resulting in enriched TCR clones
430 amongst naïve T cells compared to the memory fraction. Second, that these cell
431 populations are very early differentiated memory, and in the absence of CD95, we
432 could not account for persistence of CD4 or CD8 memory stem cell emergence at
433 birth[44–46]. The strong association between activated "naïve" CD4 and CD8 cells with
434 TCR diversity would give credibility to this possibility. We hypothesised that such an
435 outcome could be due either early differentiated cells that still expressed markers
436 characteristic of naïve cells or the inclusion of stem cell-like memory cells that co-
437 express CD27 and CD45RA[46]. Thirdly, these cells are not antigen-primed, but rather
438 akin to virtual T cells[47,48]. It has been shown that inflammatory conditions elicit virtual
439 $CD8^+$ T cells in the absence of specific antigen[49]. Further investigations are required
440 to decipher the mechanism responsible for the diverse clonality found in naïve cells
441 and whether inflammatory conditions in iHEU maybe driving this.
442
443 Unlike the somatic rearranged TCR repertoire, NK cell diversity arises through
444 combinatorial expression of germline encoded receptors that sense expression of
445 MHC class I molecules through surface bound KIR[23]. Previous viral encounters have
446 been shown to augment NKR diversity and in some instances consequently reduce
447 the cytolytic capacity of NK cells[50,51]. We observed compositional differences of the
448 NK cell clusters between iHEU and iHUU at week 4 in which iHEU had marginally
449 higher memory-like/terminally differentiated NK cells ($NKG2A^-Per^+CD57^+$) albeit not
450 significantly so after adjusting for multiple comparisons. Enhanced cytolytic activity
451 and elevated levels of the degranulation marker, CD107a, in iHEU compared to iHUU

452 at birth have been previously reported[21]. Certain NK cell clusters at birth were superior
453 in predicting both pertussis and rotavirus vaccine-induced IgG and IgA at 3-9 months
454 compared to CD4+ T cells. Depending on the immunological milieu, NK cells can either
455 negatively or positively regulate adaptive immunity, including promoting
456 immunoglobulin isotype switching and enhancing antibody production[52,53]. This finding
457 was found for both iHEU and iHUU and mechanistic investigations would need to
458 determine the role of NK cells in vaccine-induced antibodies.
459
460 The novelty and power of our study lies with the longitudinal follow-up of our infants
461 with matched samples for multiple analyses allowing us to detail the ontogeny of T
462 and NK cells from birth to 36 weeks of life and their relation with vaccination. One
463 limitation of our study is the small sample volumes collected from the infants and our
464 consequent inability to investigate functional changes in the immune subsets
465 identified. A further limitation was the inability to tease out the effect of HIV from ARV
466 exposure since all pregnant women living with HIV receive ARV treatment as a
467 standard of care[54].
468
469 In conclusion, our data display transitional immunity during infancy, which is impacted
470 by HIV/ARV exposure in a sequential manner starting with alteration of NK cells
471 followed by T cell differentiation. Of particular importance, is the finding that TCR
472 clonotypic diversity is significantly lower in iHEU from birth suggesting in utero skewing
473 well before T cell memory differentiation. In addition, our data show that the
474 composition of early life NK cells could predict vaccine responses in the first 9 months
475 of life, although further mechanistic studies are still required. This study therefore
476 provides a comprehensive phenotypic view of early life immune changes in response

477 to HIV/ARV exposure. These changes may well be linked to the observed disparities
478 in co-morbidities between iHEU and iHUU.
479
480 **Methods**
481 *Study cohort*
482 We analysed infants who were delivered by mothers living with and without HIV
483 infection who were enrolled prospectively from birth and followed up until 9 months of
484 age as previously described (Table S2)[55]. Briefly, pregnant women ≥18 years who
485 recently delivered <12 hr were enrolled in the study together with their respective
486 infants following signed informed consent. All mothers living with HIV received
487 combined antiretroviral treatment during pregnancy under the Option B programme.
488 The enrolment was restricted to infants having birth weight ≥ 2.5 kg, gestation ≥ 36
489 weeks and no complication experienced during delivery. Of those delivered by
490 mothers living with HIV, viral transmission was assessed by performing HIV DNA PCR
491 test after 6 weeks of life, and those who tested positive for perinatal HIV infection were
492 excluded from further analysis. The infants participating in this study received
493 childhood vaccines according to the South African Extended Program of Immunization
494 (EPI). This included administration of acellular pertussis vaccine at weeks 6, 10 and
495 14 and rotavirus vaccine at weeks 6 and 14. Blood samples were collected at birth,
496 weeks 4, 15 and 36 for isolation of plasma and peripheral blood mononuclear cells
497 (PBMC). Demographic characteristics of the mothers and their respective infants are
498 included in the study analysis and summarised in Table S2.

499 *Plasma and peripheral blood mononuclear cell processing*

500 Infant blood samples (0.5-3 mL) were collected into sodium heparin tubes and
501 processed within 6hr. Plasma and PBMC were isolated using ficoll centrifugation.
502 Plasma samples were stored at -80°C while PBMC were cryopreserved in 90% Fetal
503 Calf serum (FCS) with 10% DMSO in liquid nitrogen.

504

505 *Intracellular and surface staining of PBMC for mass cytometry*

506 A total of 278 infant samples (including 27 duplicates) were used to assess the
507 longitudinal changes of immune cells using mass cytometry and TCR sequencing.
508 PBMC samples were retrieved from liquid nitrogen and thawed at 37°C before being
509 transferred into RPMI 1640 media supplemented with 10% FCS and 10KU
510 Benzonase. Cells were centrifuged, washed twice in PBS and counted using a TC20
511 cell counter (Biorad). We aimed to stain $2x10^6$ viable cells for mass cytometry,
512 however cell recovery varied by infant and age, with fewer cells collected at earlier
513 time points due to smaller blood volumes. PBMC samples were split into 2 aliquots
514 with $2x10^6$ (or ¾ for cells with $<2x10^6$) used for mass cytometry staining and $1x10^6$ (or
515 ¼ for cells with $<2x10^6$) used for TCR sequencing. Lyophilised surface and intracellular
516 antibody mixtures[56] stored at 4°C were reconstituted in CyFACS buffer (PBS, 2% FCS)
517 and permeabilization buffer (eBioscience permwash) respectively for cell labelling.
518 Table S1 lists the antigen targets included in the mass cytometry antibody panel. Prior
519 to antibody labelling, the cells were stained with cisplatin in PBS to determine cell
520 viability followed by staining with surface antibodies for 20 min at room temperature.
521 Cells were then fixed with 2% paraformaldehyde solution, permeabilised and stained
522 with intracellular antibodies at 4°C for 45 min. Upon which cells were washed with
523 CyFACS buffer and resuspended with 2% paraformaldehyde solution containing

524 Iridium DNA intercalator overnight at 4°C. Samples were then washed with PBS and
525 resuspended in MilliQ water containing EQ Four Element calibration beads (10% v,v,
526 Fluidigm) prior to acquisition using CyTOF 2 instrument (DVS Sciences).
527
528 *Fluorescent activated cell sorting of naïve and memory T cells.*
529 The remaining PBMC from each sample were used to sort for naïve and memory $CD4^+$
530 and $CD8^+$ T cells using fluorescent activated cell sorting. Cells were labelled by
531 surface staining using T cell markers including CD3, CD4 and CD8 and memory
532 markers CD27, CD45RA and CCR7 (Table S2 & Figure S3). Naïve cells were denoted
533 as those co-expressing CCR7, CD45RA and CD27 while the other remaining cells
534 were regarded as memory cells. A BD FACSAria Fusion (BD) was used for 4-way
535 sorting of CD4 and CD8 naïve and memory T cells from each sample and collected
536 directly into FCS. Sorted cells were centrifuged, resuspended in cell RNAProtect
537 (Qiagen) and stored at -40°C until processing for TCR sequencing.
538
539 *Bulk TCR sequencing*
540 RNA was purified from each sorted sample using the RNAeasy Plus Micro kit (Qiagen)
541 and libraries were prepared for TCR sequencing as described by Rubelt, et al.[57]. This
542 was performed at the University of Cape Town. Briefly, purified RNA was reverse
543 transcribed in a Rapid amplification of cDNA ends (RACE) reaction using
544 SMARTScribe Reverse Transcriptase using previously designed oligos (isoC-5′-
545 GTCAGATGTGTATAAGAGACAGnnnnnnnnnnCGATAGrGrGrG -3′-C3_Spacer and
546 for the Poly A tail 5′- GTGTCACGTACAGAGTCATCttttttttttttttttttttttttttttt -3′ VN) that
547 captures polyadenylated transcripts and introduces a 10bp unique molecular identifier
548 (UMI) at the 5' end of the product. cDNA was purified with AmpureXP beads (Beckman

549 Coulter) before whole transcriptome amplification using Advantage 2 Polymerase
550 (Clontech) using oligos that introduce Illumina Nextera Multiplex Identifier (MID) P5
551 Adapter sequences. The libraries were then sent to Stanford University (Department
552 of Microbiology and Immunology, Stanford University) in a blinded fashion, where
553 TCRβ-specific amplification was achieved with Q5 Hot Start Master Mix (NEB) using
554 constant region-specific oligos, simultaneously introducing P7 MID sequences. Final
555 sequencing libraries were purified using SPRISelect beads (Beckman Coulter),
556 quantified using the Agilent TapeStation and pooled for sequencing. Paired-end
557 sequencing was performed on an Illumina NovaSeq SP with 2x250 cycles, performed
558 by the Chan-Zuckerberg Biohub Initiative.
559
560 *Quantification of vaccine antibody responses*
561 Infant pertussis specific antibody responses were measured at birth, 4, 15 and 36
562 weeks of age using a commercial human IgG ELISA kit (Abcam). Plasma samples
563 were diluted 1:100 into sample diluent antibody titres measured in duplicates
564 according to manufacture instructions.
565 Rotavirus IgA titres were determined by enzyme immunoassay (EIA) and performed
566 in a blinded manner in the Division of Infectious Disease, Department of Pediatrics,
567 Cincinnati Childrens' Hospital Medical Centre, Cincinnati, Ohio. Briefly, purified rabbit
568 anti-rotavirus IgG were immobilised on microtiter plate as capture antibodies. Lysates
569 from rotavirus (strains RV3 and 8912) and mock infected cells (control) were added to
570 the immobilised capture antibodies to bind rotavirus antigens and any uncaptured
571 antigens were washed off with PBS, 0.05% Tween20. Reference standards, control
572 and test samples were diluted in PBS, 0.05% Tween20, 1% non-fat dry milk and 50
573 uL added to microtiter plate for antigen binding. Bound antibodies were detected by

574 biotyinylated goat anti-human IgA (Jackson Laboratories) and the addition of
575 peroxidase conjugated avidin:biotin (Vector Laboratories, Inc., Burlingame, CA).
576 Chromogenic signal was generated by addition of substrate O-phenylenediamine
577 (Sigma) and reaction stopped after 30 min using 1M H_2SO_4. Absorbance was
578 measured at 492_{nm} on a Molecular Devices SpectraMax Plus plate reader. The
579 reference standard was assigned 1000 arbitrary units (AU) and a four-parameter
580 logistic regression was used to extrapolate anti-rotavirus IgA tires using SoftMax
581 software.
582 Rotavirus neutralisation titres were determined as previously described[58]. In this study
583 we used Wa G1P8, 1076 G4P6 and DS-1 G2P4 virus strains obtained from the
584 National Institute of Health (NIH). Plasma samples were serially diluted and incubated
585 with the rotavirus strain for neutralization prior to adding the mixture to susceptible
586 MA104 (monkey kidney) cell line for overnight incubation. Cell lysates were used to
587 determine the level of un-neutralized rotavirus antigens using the EIA as described
588 using guinea pig anti-rotavirus antiserum to measure captured rotavirus antigens.
589 Rabbit anti-guinea pig IgG conjugated to horseradish peroxidase (Jackson
590 ImmunoResearch) was used to detect bound antibodies, and chromogenic signal
591 generated using OPD. The amount of rotavirus present in the resuspended lysate from
592 each well was inversely related to the amount of neutralizing antibody present in the
593 plasma. Each plasma dilution series as modelled using a logistic regression function.
594 For each fitted curve the dilution which corresponds to a 40% response (ED_{40}),
595 compared to the virus controls, was determined and reported as the neutralization
596 titer. The ED_{40} represents the titre of the serum against a given virus, which represents
597 a 60% reduction in amount of virus.
598

599 *Data analysis*

600 **Mass cytometry**

601 *Processing*

602 Post CyTOF acquisition, the files were normalized using Premessa R package and
603 target populations were manually gated using FlowJo software (version 10.5.3,
604 TreeStar). The target populations were exported as FCS files for downstream analysis
605 using R, an open-source statistical software[59]. All files for the target populations were
606 transformed by applying inverse hyperbolic sine with a cofactor of 5 and subjected to
607 doublet detection and removal using computeDoubletDensity from the scDblFinder
608 package[60].

609

610 *Dimensional reduction*

611 Median marker expressions for each sample were used to compute Euclidean
612 distance matrix to determine multidimensional scaling (MDS) coordinates using the
613 cmdscale function in R. MDS analysis were used to visualize immune cell trajectories.
614 Further, marker expression intensities were used to implement Uniform Manifold
615 Approximation and Projection (UMAP) while preserving local density distribution for
616 visualization using densMAP function from the denvis R package[61].

617

618 *Cell Clustering*

619 High resolution clustering of the target cell populations was performed using FlowSOM
620 algorithm and the resulting clusters grouped into metaclusters using
621 ConsensusClusterPlus available in the CATALYST R package[62]. The immune cell
622 clusters were visualised using hierarchical clustering and UMAP embedding. Centred
623 log-ratios of the relative abundance of the cell clusters per sample were determined

624 and used in Principal Component Analysis (PCA) for assessing immune cell
625 compositional differences.
626
627 **TCR sequencing**
628 *Pre-processing of TCR sequencing data*
629 Raw sequencing data was pre-processed by pRESTO[63]. Briefly, reads with mean
630 Phred quality scores less than 20 were discarded, UMIs were extracted from the first
631 10bp of read 1, and primer sequences masked. Next, a single consensus sequence
632 was constructed from reads sharing the same UMI barcode and paired-end consensus
633 sequences assembled into a full-length TCR sequence. In the case of non-overlapping
634 mate-pairs, the human TRB reference as used to properly space non-overlapping
635 reads. After removal of duplicate sequences, full-length reads were aligned and
636 clonotypes assembled using MiXCR[64,65].
637
638 *T cell immune repertoire analysis*
639 To perform quality control on our bulk TCR sequencing data, we removed samples
640 where fewer than 10 clones were detected (n = 44 samples) and where more clones
641 were detected than cells were sorted (n = 18 samples). This yielded a total of 861
642 samples for downstream analysis. The immunarch packag[66] operating in the open-
643 source statistical software R was used for all downstream immune repertoire analysis,
644 including calculation of CDR3β lengths and repertoire diversity by the Inverse Simpson
645 index and richness using Choa1.
646
647 *Identification and annotation of T cell specificity groups*

648 The GLIPH2 algorithm was used to establish T cell specificity groups, clusters of
649 CDR3 sequences that are predicted to bind to the same antigen, by discovering
650 groups of CDR3β sequences that share either global or local motifs[27,28]. To assign
651 specificity annotations to identified CDR3β sequence clusters, we adapted an
652 approach described by Chiou, et al[67]. Briefly, we collected human CDR3β sequences
653 with known specificity from VDJdb (https://vdjdb.cdr3.net/) and identified CDR3β
654 sequences within this database that could form TCR specificity groups. From 47,107
655 CDR3β sequences, this process identified 11,648 specificity groups comprised of
656 25,759 unique CDR3β sequences. We next combined these 25,759 annotated
657 sequences with 108,183 unique experimental CDR3β sequences from our bulk TCR
658 repertoire profiling cohort. This was performed separately for CD4 and CD8 T, as there
659 are differences between gene usage frequencies between CD4 and CD8 T cells that
660 can impact specificity predictions. We prioritized TCR specificity groups with at least 3
661 distinct CDR3β sequences. This process yielded 17,154 CDR3β sequences in 11,629
662 specificity groups for CD4 T cells, and 15,107 CDR3β sequences in 9,911 specificity
663 groups for CD8 T cells.
664

665 **Statistical analysis**

666 All statistical analysis and graphical visualization of the data was performed on the
667 open R software[59]. Spearman's rank correlation was used to test for the correlations
668 between the frequencies of the cell clusters with either infant age or antibody
669 responses and p-values adjusted for multiple comparisons using false discovery rates
670 (FDR). To compare differences in the abundance of cell clusters between iHEU and
671 iHUU the generalised linear model from the diffcyt package was used[68]. Pairwise
672 comparisons of medians between groups were performed using Wilcoxon-rank test or

673 Kruskal-Wallis for multiple group comparisons and the p-values adjusted for multiple
674 comparisons using Benjamin Hochberg correction. Generalised linear mixed model
675 (GLMM) with bootstrap resampling from the R package CytoGLMM[69] was used to
676 identify NK cell markers that are predictive of HIV exposure at birth. To determine cell
677 clusters that were predictive of antibody responses following vaccination, we used the
678 multivariate modelling algorithm from MUVR that incorporates recursive variable
679 selection using repeated double cross-validation within partial least squares
680 modelling[29].
681

682 **Acknowledgements**

683 We would like to acknowledge the technical staff that assisted in the completion of this
684 study. This includes Euan Johnston, Emily Tangie and Carine Kilola who assisted with
685 the CyTOF experiments, Mikayla Stabile, Michelle Leong, Kassandra Pinedo, and
686 Nicole Prins, for assistance with TCR RNA-seq processing, and Llyod Leach for
687 assistance with measuring pertussis antibody responses.
688

689 **Author contribution**

690 Conceived by CMG, CAB and HBJ, Method set-up and validation by SD, AW, SC, TR,
691 FR, HH, and MD. Study participant enrolment, sample collection and processing by
692 HBJ and BA. Experimental investigations by SD. Data and statistical analysis by SD,
693 AW and SH. Original draft by SD. Reviewed and edited by all authors.
694

695 **Declaration of Interest**

696 All authors declare no conflict of interest of any of the material included in this
697 manuscript.
698

References

1. Perin, J. *et al.* Global, regional, and national causes of under-5 mortality in 2000–19: an updated systematic analysis with implications for the Sustainable Development Goals. *The Lancet Child and Adolescent Health* **6**, 106–115 (2022).
2. Zimmermann, P. & Jones, C. E. Factors That Influence Infant Immunity and Vaccine Responses. *The Pediatric infectious disease journal* **40**, S40–S46 (2021).
3. Pasetti, M. F. *et al.* Maternal determinants of infant immunity: Implications for effective immunization and maternal-child health. *Vaccine* **38**, 4491–4494 (2020).
4. Sanidad, K. Z. & Zeng, M. Y. Neonatal gut microbiome and immunity. *Current Opinion in Microbiology* **56**, 30–37 (2020).
5. Goenka, A. & Kollmann, T. R. Development of immunity in early life. *Journal of Infection* **71**, S112–S120 (2015).
6. Aagaard, K. *et al.* The Placenta Harbors a Unique Microbiome. *ScienceTranslationalMedicine* **6**, 229–262 (2014).
7. Tosato, F. *et al.* Lymphocytes subsets reference values in childhood. *Cytometry Part A* **87**, 81–85 (2015).
8. Howie, P. W., Forsyth, J. S., Ogston, S. A., Clark, A. & Du Florey, V. C. Protective effect of breat feeding against infection. *British Medical Journal* **300**, 11–16 (1990).
9. Sankar, M. J. *et al.* Optimal breastfeeding practices and infant and child mortality: A systematic review and meta-analysis. *Acta Paediatrica, International Journal of Paediatrics* **104**, 3–13 (2015).
10. Anderson, K. *et al.* Increased infectious-cause hospitalization among infants who are HIV-exposed uninfected compared with HIV-unexposed. *AIDS* **35**, 2327–9 (2021).
11. Brennan, A. T. *et al.* A meta-analysis assessing all-cause mortality in HIVexposed uninfected compared with HIV-unexposed uninfected infants and children. *Aids* **30**, 2351–2360 (2016).
12. Cohen, C. *et al.* Epidemiology of Acute Lower Respiratory Tract Infection in HIV- Exposed Uninfected Infants. *Pediatrics* **137**, e20153272 (2016).
13. Slogrove, A. L., Goetghebuer, T., Cotton, M. F., Singer, J. & Bettinger, J. Pattern of Infectious Morbidity in HIV-Exposed Uninfected Infants and Children. *Frontiers in Immunology* **7**, 1–8 (2016).
14. Brennan, A. T. *et al.* A meta-analysis assessing diarrhea and pneumonia in HIV-exposed uninfected compared to HIV-unexposed uninfected infants and children. *JAIDS Journal of Acquired Immune Deficiency Syndromes* **82**, 1–8 (2019).
15. Slogrove, A. L., Powis, K. M., Johnson, L. F., Stover, J. & Mahy, M. Estimates of the global population of children who are HIV-exposed and uninfected, 2000–18: a modelling study. *The Lancet Global Health* **8**, e67–e75 (2020).
16. Dzanibe, S. *et al.* Stereotypic Expansion of T Regulatory and Th17 Cells during Infancy Is Disrupted by HIV Exposure and Gut Epithelial Damage. *The Journal of Immunology* **208**, 27–37 (2022).
17. Uksila, J., Lassila, O. & Hirvonen, T. Natural killer cell function of human neonatal lymphocytes. *Clinical and Experimental Immunology* **48**, 649–654 (1982).

18. Choi, S. S. *et al.* Interleukin-15 enhances cytotoxicity, receptor expression, and expansion of neonatal natural killer cells in long-term culture. *Clinical and Diagnostic Laboratory Immunology* **11**, 879–888 (2004).
19. Xing, D. *et al.* Cord blood natural killer cells exhibit impaired lytic immunological synapse formation that is reversed with il-2 exvivo expansion. *Journal of Immunotherapy* **33**, 684–696 (2010).
20. Strauss-Albee, D. M., Liang, E. C., Ranganath, T., Aziz, N. & Blish, C. A. The newborn human NK cell repertoire is phenotypically formed but functionally reduced. *Cytometry Part B - Clinical Cytometry* **92**, 33–41 (2017).
21. Smith, C. *et al.* Altered natural killer cell function in HIV-exposed uninfected infants. *Frontiers in Immunology* **8**, 1–13 (2017).
22. Slyker, J. A. *et al.* The impact of HIV-1 infection and exposure on natural killer (NK) cell phenotype in kenyan infants during the first year of life. *Frontiers in Immunology* **3**, 1–16 (2012).
23. Vilches, C. & Parham, P. KIR: Diverse, rapidly evolving receptors of innate and adaptive immunity. *Annual Review of Immunology* **20**, 217–251 (2002).
24. Sun, J. C., Beilke, J. N. & Lanier, L. L. Adaptive immune features of natural killer cells. *Nature* **457**, 557–561 (2009).
25. Muntasell, A., Magri, G., Pende, D., Angulo, A. & López-Botet, M. Inhibition of NKG2D expression in NK cells by cytokines secreted in response to human cytomegalovirus infection. *Blood* **115**, 5170–5179 (2010).
26. Flommersfeld, S. *et al.* Fate mapping of single NK cells identifies a type 1 innate lymphoid-like lineage that bridges innate and adaptive recognition of viral infection. *Immunity* **54**, 2288-2304.e7 (2021).
27. Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017).
28. Huang, H., Wang, C., Rubelt, F., Scriba, T. J. & Davis, M. M. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nature Biotechnology* **38**, 1194–1202 (2020).
29. Shi, L., Westerhuis, J. A., Rosén, J., Landberg, R. & Brunius, C. Variable selection and validation in multivariate modelling. *Bioinformatics* **35**, 972–980 (2019).
30. Olin, A. *et al.* Stereotypic Immune System Development in Newborn Children. *Cell* **174**, 1277-1292.e14 (2018).
31. Haynes, B. Y. B. E., Martin, M. E., Kay, H. H. & Kurtzberg, J. Early events in human T cell ontogeny. Phenotypic characterization and immunohistologic localization of T cell precursors in early human fetal tissues. *The Journal of experimental medicine* **168**, 1061–80 (1988).
32. Haynes, B. F. & Heinly, C. S. Early human T cell development: Analysis of the human thymus at the time of initial entry of hematopoietic stem cells into the fetal thymic microenvironment. *Journal of Experimental Medicine* **181**, 1445–1458 (1995).
33. Scheible, K. M. *et al.* T cell developmental arrest in former premature infants increases risk of respiratory morbidity later in infancy. *JCI insight* **3**, (2018).
34. Semmes, E. C. *et al.* Understanding Early-Life Adaptive Immunity to Guide Interventions for Pediatric Health. *Frontiers in Immunology* **11**, 1–14 (2021).
35. Walker, R. W., Clemente, J. C., Peter, I. & Loos, R. J. F. The prenatal gut microbiome: are we colonized with bacteria in utero? *Pediatric Obesity* **12**, 3–17 (2017).

36. Rich, K. C., Siegel, J. N., Jennings, C., Rydman, R. J. & Landay, a L. Function and phenotype of immature CD4+ lymphocytes in healthy infants and early lymphocyte activation in uninfected infants of human immunodeficiency virus-infected mothers. *Clinical and diagnostic laboratory immunology* **4**, 358–361 (1997).
37. Ono, E. *et al.* Imbalance of naive and memory T lymphocytes with sustained high cellular activation during the first year of life from uninfected children born to HIV-1-infected mothers on HAART. *Brazilian Journal of Medical and Biological Research* **41**, 700–708 (2008).
38. Wood, L. F., Wood, M. P., Fisher, B. S., Jaspan, H. B. & Sodora, D. L. T cell activation in South African HIV-exposed infants correlates with ochratoxin a exposure. *Frontiers in Immunology* **8**, 1–8 (2017).
39. Garcia-Knight, M. A. *et al.* Altered memory T-cell responses to Bacillus Calmette-Guerin and Tetanus Toxoid vaccination and altered cytokine responses to polyclonal stimulation in HIV-exposed uninfected Kenyan infants. *PLoS ONE* **10**, 1–19 (2015).
40. Smith, C. *et al.* Human Immunodeficiency Virus Exposure but Not Early Cytomegalovirus Infection Is Associated With Increased Hospitalization and Decreased Memory T-Cell Responses to Tetanus Vaccine. *Journal of Infectious Diseases* **221**, 1167–1175 (2020).
41. Pérez, Y. B. *et al.* Impaired T helper cell responses in human immunodeficiency virus - exposed uninfected newborns. *Immunity Inflammation and Disease* **9**, 1541–1553 (2021).
42. Gabriel, B. *et al.* Analysis of the TCR Repertoire in HIV-Exposed but Uninfected Infants. *Scientific Reports* **9**, (2019).
43. Holditch, S. J. *et al.* Decay kinetics of HIV-1 specifict cell responses in vertically HIV-1 exposed seronegative infants. *Frontiers in Immunology* **2**, 1–6 (2012).
44. Pušnik, J. *et al.* Expansion of Stem Cell-Like CD4 + Memory T Cells during Acute HIV-1 Infection Is Linked to Rapid Disease Progression . *Journal of Virology* **93**, (2019).
45. Ribeiro, S. P. *et al.* The CD8 + Memory Stem T Cell (T SCM) Subset Is Associated with Improved Prognosis in Chronic HIV-1 Infection . *Journal of Virology* **88**, 13836–13844 (2014).
46. Gattinoni, L. *et al.* A human memory T cell subset with stem cell-like properties. *Nature Medicine* **17**, 1290–1297 (2011).
47. Savid-Frontera, C. *et al.* Exploring the immunomodulatory role of virtual memory CD8+ T cells: Role of IFN gamma in tumor growth control. *Frontiers in Immunology* **13**, 1–18 (2022).
48. Lee, J. Y., Hamilton, S. E., Akue, A. D., Hogquist, K. A. & Jameson, S. C. Virtual memory CD8 T cells display unique functional properties. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 13498–13503 (2013).
49. Rolot, M. *et al.* Helminth-induced IL-4 expands bystander memory CD8+ T cells for early control of viral infection. *Nature Communications* **9**, (2018).
50. Wilk, A. J. & Blish, C. A. Diversification of human NK cells: Lessons from deep profiling. *Journal of Leukocyte Biology* **103**, 629–641 (2018).
51. Strauss-Albee, D. M. *et al.* Human NK cell repertoire diversity reflects immune experience and correlates with viral susceptibility. *Science Translational Medicine* **7**, (2015).

848 52. Rydyznski, C. E. & Waggoner, S. N. Boosting vaccine efficacy the natural
849 (killer) way. *Trends in Immunology* **36**, 536–546 (2015).
850 53. Rydyznski, C. E. *et al.* Affinity Maturation Is Impaired by Natural Killer Cell
851 Suppression of Germinal Centers. *Cell Reports* **24**, 3367-3373.e4 (2018).
852 54. Wessels, J. *et al.* The updated South African National Guideline for the
853 Prevention of Mother to Child Transmission of Communicable Infections
854 (2019). *Southern African Journal of HIV Medicine* **21**, 1–8 (2020).
855 55. Tchakoute, C. T. *et al.* Breastfeeding mitigates the effects of maternal HIV on
856 infant infectious morbidity in the Option B+ era. *AIDS (London, England)* **32**,
857 2383–2391 (2018).
858 56. Vendrame, E. *et al.* Profiling of the Human Natural Killer Cell Receptor-Ligand
859 Repertoire. *Journal of Vizualized Experiments* **165**, 139–148 (2020).
860 57. Rubelt, F. *et al.* Individual heritable differences result in unique cell lymphocyte
861 receptor repertoires of naïve and antigen-experienced cells. *Nature*
862 *Communications* **7**, 1–12 (2016).
863 58. Knowlton, D. R., Spector, D. M. & Ward, R. L. Development of an improved
864 method for measuring neutralizing antibody to rotavirus. *Journal of Virological*
865 *Methods* **33**, 127–134 (1991).
866 59. RCoreTeam. R: A language and environment for statistical computing. Preprint
867 at (2021).
868 60. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D.
869 Doublet identification in single-cell sequencing data using scDblFinder.
870 *F1000Research* **10**, 979 (2022).
871 61. Narayan, A., Berger, B. & Cho, H. Assessing single-cell transcriptomic
872 variability through density-preserving data visualization. *Nature Biotechnology*
873 **39**, 765–774 (2021).
874 62. Crowell, H. L., Zanotelli, V. R. T., Chevrier, S. & Robinson, M. D. CATALYST:
875 Cytometry dATa anALYSis Tools. Preprint at (2021).
876 63. Vander Heiden, J. A. *et al.* PRESTO: A toolkit for processing high-throughput
877 sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**,
878 1930–1932 (2014).
879 64. Bolotin, D. A. *et al.* MiXCR: Software for comprehensive adaptive immunity
880 profiling. *Nature Methods* **12**, 380–381 (2015).
881 65. Bolotin, D. A. *et al.* Antigen receptor repertoire profiling from RNA-seq data.
882 *Nature biotechnology* **35**, 908–911 (2017).
883 66. Nazarov, V. I. *et al.* immunarch: Bioinformatics Analysis of T-Cell and B-Cell
884 Immune Repertoires. Preprint at (2022).
885 67. Chiou, S. H. *et al.* Global analysis of shared T cell specificities in human non-
886 small cell lung cancer enables HLA inference and antigen discovery. *Immunity*
887 **54**, 586-602.e8 (2021).
888 68. Weber, L. M. diffcyt: Differential discovery in high-dimensional cytometry via
889 high-resolution clustering. Preprint at (2021).
890 69. Seiler, C. *et al.* CytoGLMM: conditional differential analysis for flow and mass
891 cytometry experiments. *BMC Bioinformatics* **22**, 1–14 (2021).
892

893 Tables

894 **Table 1**: Phenotypic description of $CD4^+$ T cell clusters idenfied by unsupervised cell
895 clustering.

896 **Table 2**: Phenotypic description of $CD8^+$ T cell clusters idenfied by unsupervised cell
897 clustering.

898 **Table 3**: Phenotypic description of NK cell clusters idenfied by unsupervised cell
899 clustering.

900

901 Figures

902 **Figure 1**: Immunophenotypic trajectory of infant peripheral blood mononuclear cells
903 (PBMC) in the first 9 months of life. **A)** Dimensional reduction of infant PBMC Immune
904 lineage clusters embedded on uniform manifold approximation and proximation
905 (UMAP) with density preservation and table summarising immune cell cluster
906 phenotype. **B)** Relative abundance of immune lineage clusters measured in cord blood
907 (CB) and infant peripheral blood at birth (BTH), weeks (WK) 4, 15 and 36. **C)**
908 Spearman's rank correlation between proportion of immune lineage cell clusters and
909 infants age from birth until week 36. **D)** Age related maturation trajectory of CD4+ and
910 CD8+ T cells and NK cells depicted using multidimensional scaling coordinates
911 derived from median marker expressions for each infant samples. **E)** Boxplots
912 summarizing MDS coordinate associated with infant age for CD4+ and CD8+ T cells
913 and NK cells.

914

915 **Figure 2**: Divergent T cell memory differentiation in HIV-exposed uninfected infants
916 (iHEU) compared to HIV-unexposed uninfected infants (iHUU). **A & D)** Uniform
917 manifold approximation and proximation (UMAP) with density preservation showing

918 dimensional reduction of FlowSOM clusters of CD4+ and CD8 T cells **B &E)** Relative
919 abundances of CD4+ and CD8 + T cell clusters measured in cord blood (CB) and
920 infant peripheral blood at birth (BTH), weeks (WK) 4, 15 and 36. **C &F)** Age related
921 phenotypic composition of CD4+ and CD8+ T cells depicted as principal component
922 (PC) coordinates of centred log-odd ratios of FlowSOM clusters for each infant
923 samples and arrows indicating contribution of each cell cluster in scatter of PC
924 components. **G &H)** Boxplots comparing PC coordinates of CD4+ and CD8+ T cells
925 between iHEU and iHUU longitudinally. **I & J)** Generalized linear mixed model (GLMM)
926 comparing the abundances of CD4+ and CD8+ T cell clusters between iHEU and
927 iHUU at weeks 15 and 36.
928
929 **Figure 3**: Early life immunophenotypic alteration in NK cells in HIV-exposed uninfected
930 infants (iHEU) compared to HIV-unexposed uninfected infants (iHUU). **A)** Uniform
931 manifold approximation and proximation (UMAP) with density preservation showing
932 dimensional reduction of FlowSOM clusters on NK cells and table summarising
933 immune cell cluster phenotype. **B)** Relative abundances of NK cell clusters measured
934 in cord blood (CB) and infant peripheral blood at birth (BTH), weeks (WK) 4, 15 and
935 36. **C)** Age related phenotypic composition of NK cells depicted as principal
936 component (PC) coordinates of centred log-odd ratios of FlowSOM clusters for each
937 infant samples and arrows indicating contribution of each cell cluster in scatter of PC
938 components. **D)** Boxplots comparing PC coordinates of NK cells between iHEU and
939 iHUU
940
941 **Figure 4:** Premature CD4+ and CD8+ T cell receptor (TCR) repertoire skewing in
942 HIV-exposed uninfected infants (iHEU) relative to HIV-unexposed uninfected infants

943 (iHUU). **A)** Boxplots comparing Inverse Simpson TCR diversity scores between
944 iHUU and iHEU at birth and weeks 4, 15 and 36. **B)** Boxplots comparing Chao1 TCR
945 clonotype richness between iHUU and iHEU at birth and weeks 4, 15 and 36. **C &D**)
946 Spearman's rank correlation between naive CD4+ and CD8+ T cell Inverse Simpson
947 scores and frequencies of FlowSOM clusters for CD4+ and CD8+ T cell clusters
948 respectively. **E**) . GLIPH analysis showing antigen specificity groups of the top 100
949 TCR clones that are significantly different between iHEU and iHUU and between
950 naïve and memory CD4+ and CD8+ T cell subsets.
951
952 **Figure 5** Association of vaccine antibody responses to immune cell phenotypes. **A)**
953 Comparing IgG levels against pertussis between HIV-exposed uninfected infants
954 (iHEU) and HIV-unexposed uninfected infants (iHUU), grey shaded area indicate
955 threshold IgG levels for protective pertussis vaccine response. **B & C**) Boxplots
956 comparing rotavirus specific IgA and neutralization titres between iHEU and iHUU at
957 week 36. **D, E & F)** Summary of ROC analysis using the latent variable axis-1 derived
958 from partial least square discriminate analysis (PLS-DA) of NK, CD4 and CD8 T cell
959 clusters determined to be best predictors at birth and week 4 of pertussis antibody
960 responses at weeks 15 and 36 and rotavirus antibody response at week 37. **G)**
961 Spearman's correlation between abundances FlowSOM clusters of CD4+ and CD8+
962 T cells and NK cells and anti-pertussis IgG titres measured at week 15 and 36.

Table 1: Phenotypic description of $CD4^+$ T cell clusters idenfied by unsupervised cell clustering.

Cluster	Phenotype	(%)
(1)	CCR4+CCR6-CXCR3- (Th2)	5.07
(2)	CD45RA+CD27+CCR7+ (Naïve)	84.1
(3)	CD45RA+CD27-CCR7-CD57+Per+PD-1+ (Cytotoxic TD)	1.3
(4)	CD38+HLA-DR+CD45RA+CD27+CCR7+ (Activated cells)	4.7
(5)	CCR4+CCR6+CXCR3+CD57-	0.7
(6)	CCR4+CCR6+CXCR3+CD57+	0.2
(7)	Per+Ki67+	0.3
(8)	CD127-CD25+CD39+TIGIT+ (Treg cells)	0.9
(9)	Per+Ki67+CD25+	0.3
(10)	Per+Ki67-HLA-DR+CD69+CD39+	0.1
(11)	Per+Ki67+CD27+$Fc_{\varepsilon}RI_{\gamma}$+	0.2
(12)	PD-1+CD38+	2.07
(13)	CD39+CD45RA+CCR7-CD27-	0.1

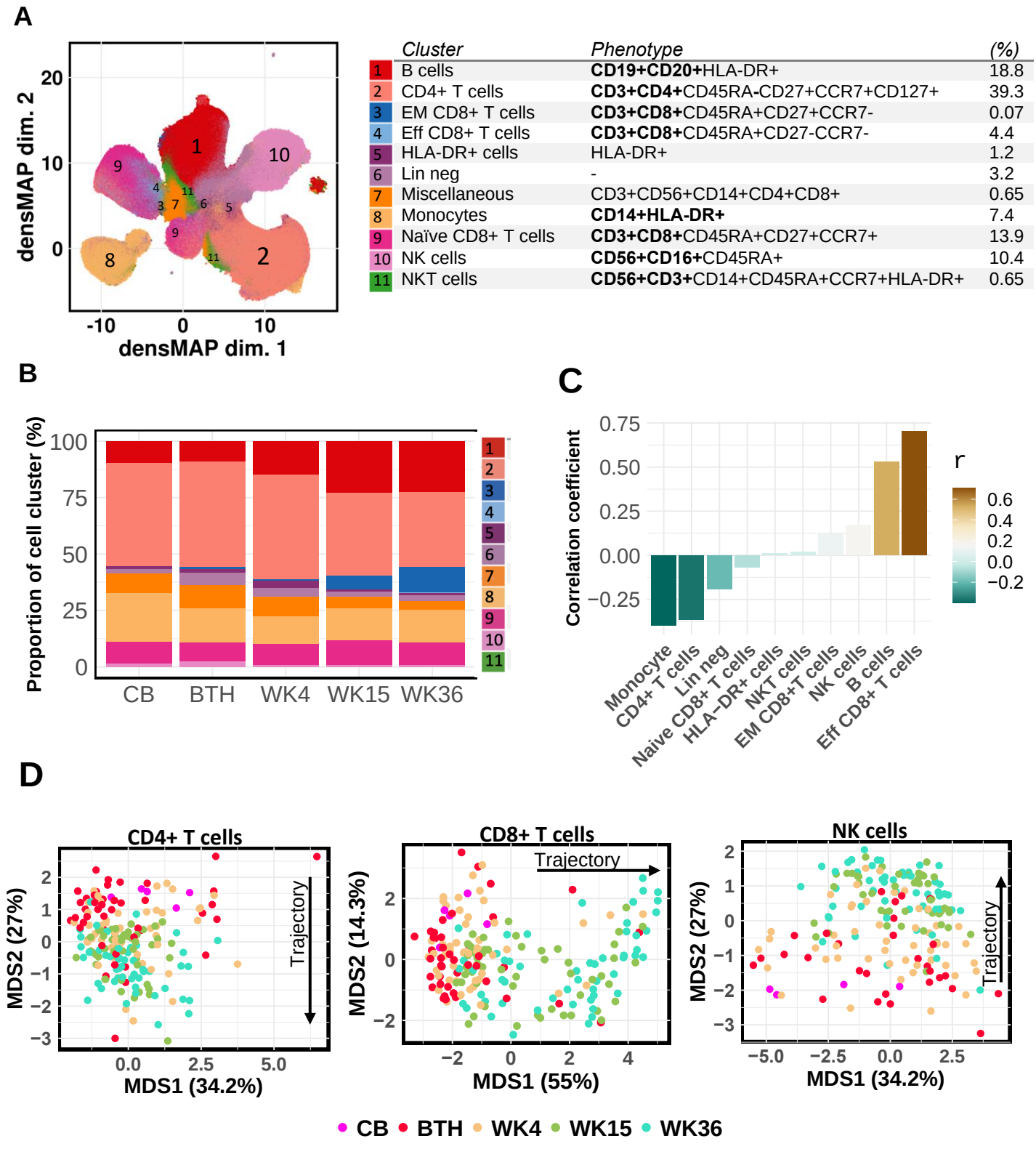
Table 2: Phenotypic description of $CD8^+$ T cell clusters identified by unsupervised cell clustering.

Cluster	Phenotype	(%)
(1)	Per+CD45RA-CD27-CCR7-CD57+ (TD)	7.5
(2)	CD56+CD16+NkpK6+NKG2A+ (NKT-like)	0.4
(3)	HLA-DR+	2.7
(4)	CD45RA+CD27+CCR7+ (Naïve)	55.3
(5)	Per+CD57+	15.6
(6)	Per+CD45RA-CD27-CCR7-	18.4
(7)	Ki67+CD45RA+CD27+	0.1

Table 3: Phenotypic description of NK cell clusters identified by unsupervised cell clustering.

Cluster	Phenotype	(%)
(1)	($CD56^{dim}CD16^{dim}$) Per+CD45RA+FcERIy	30.5

(2)	($CD56^{-}CD16^{-}$) DNAM1+	20.7
(3)	($CD56^{-}CD16^{+}$) FcERIy	1.8
(4)	($CD56^{dim}CD16^{+}$) Per+CD45RA+CD38+2B4+	11.7
(5)	($CD56^{dim}CD16^{+}$) Per+CD45RA+CD38+CD57+	12.4
(6)	($CD56^{dim}CD16^{dim}$) FcERIy+DNAM1+HLA-DR+CXCR3+CCR4+	1.4
(7)	($CD56^{-}CD16^{+}$) CD38+CD39+CD27+Ki67+	0.9
(8)	($CD56^{+}CD16^{bri}$) Per+CD38+CD45RA+DNAM1+ 2B4+Nkp30+NKp46+NKG2A+Siglec-7+	15.0
(9)	($CD56^{bri}CD16^{-}$) DNAM1+NKG2D+Nkp46+NKG2A+	4.9
(10)	($CD56^{+}CD16^{+}$) Per+CD38+CD45RA+LILRB1+NTBA+2B4+Nk30+ NKG2D+HLA-DR+CCR4+CCR7+PD-1+TIGIT+NKg2C+	0.7



	Cluster	Phenotype	(%)
1	B cells	**CD19+CD20+**HLA-DR+	18.8
2	CD4+ T cells	**CD3+CD4+**CD45RA-CD27+CCR7+CD127+	39.3
3	EM CD8+ T cells	**CD3+CD8+**CD45RA+CD27+CCR7-	0.07
4	Eff CD8+ T cells	**CD3+CD8+**CD45RA+CD27-CCR7-	4.4
5	HLA-DR+ cells	HLA-DR+	1.2
6	Lin neg	-	3.2
7	Miscellaneous	CD3+CD56+CD14+CD4+CD8+	0.65
8	Monocytes	**CD14+HLA-DR+**	7.4
9	Naïve CD8+ T cells	**CD3+CD8+**CD45RA+CD27+CCR7+	13.9
10	NK cells	**CD56+CD16+**CD45RA+	10.4
11	NKT cells	**CD56+CD3+**CD14+CD45RA+CCR7+HLA-DR+	0.65

Figure 1: Immunophenotypic trajectory of infant peripheral blood mononuclear cells (PBMC) in the first 9 months of life. **A)** Dimensional reduction of infant PBMC Immune lineage clusters embedded on uniform manifold approximation and proximation (UMAP) with density preservation and table summarising immune cell cluster phenotype. **B)** Relative abundance of immune lineage clusters measured in cord blood (CB) and infant peripheral blood at birth (BTH), weeks (WK) 4, 15 and 36. **C)** Spearman's rank correlation between proportion of immune lineage cell clusters and infants age from birth until week 36. **D)** Age related maturation trajectory of CD4+ and CD8+ T cells and NK cells depicted using multidimensional scaling coordinates derived from median marker expressions for each infant samples. **E)** Boxplots summarizing MDS coordinate associated with infant age for CD4+ and CD8+ T cells and NK cells.

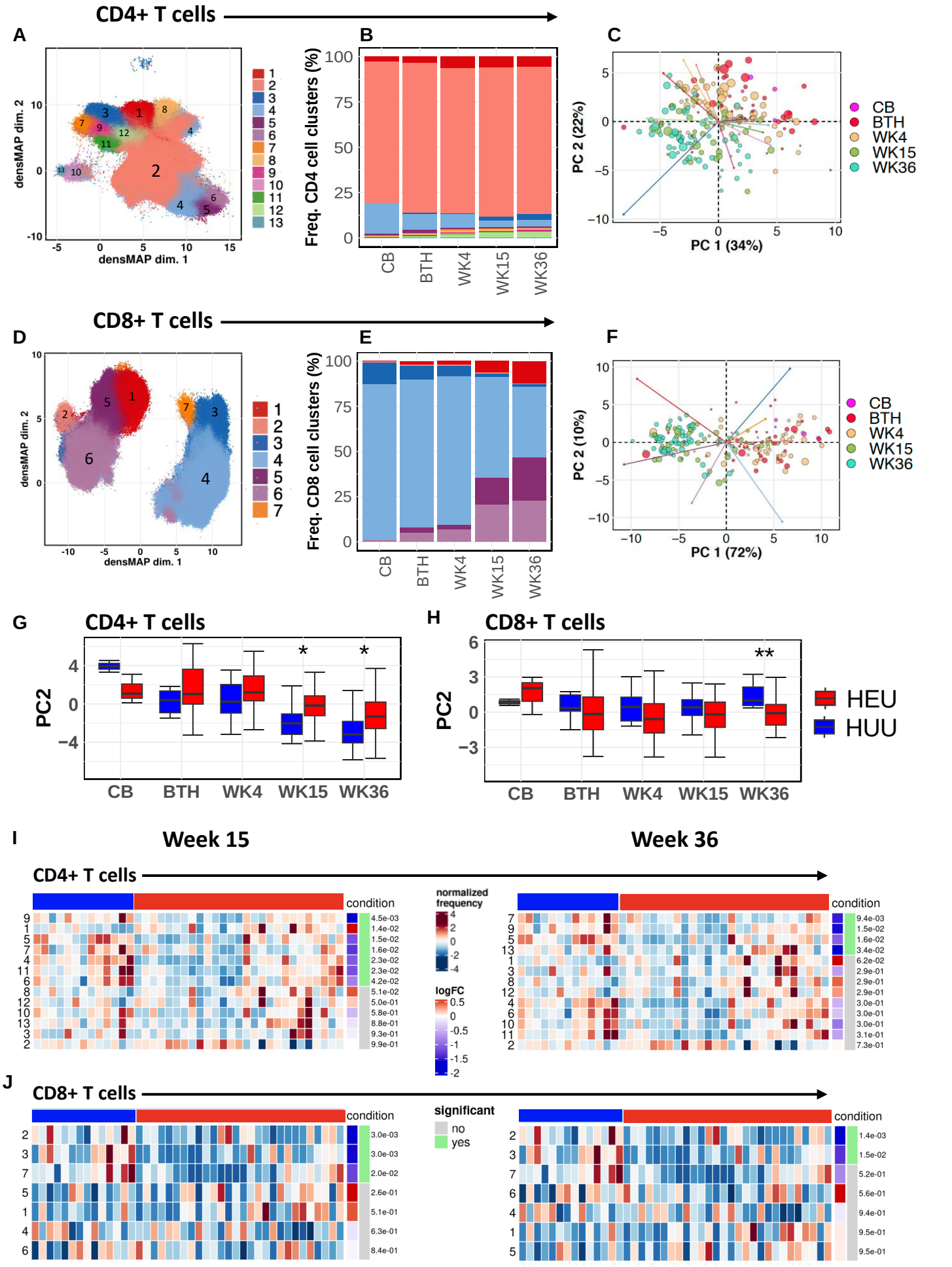
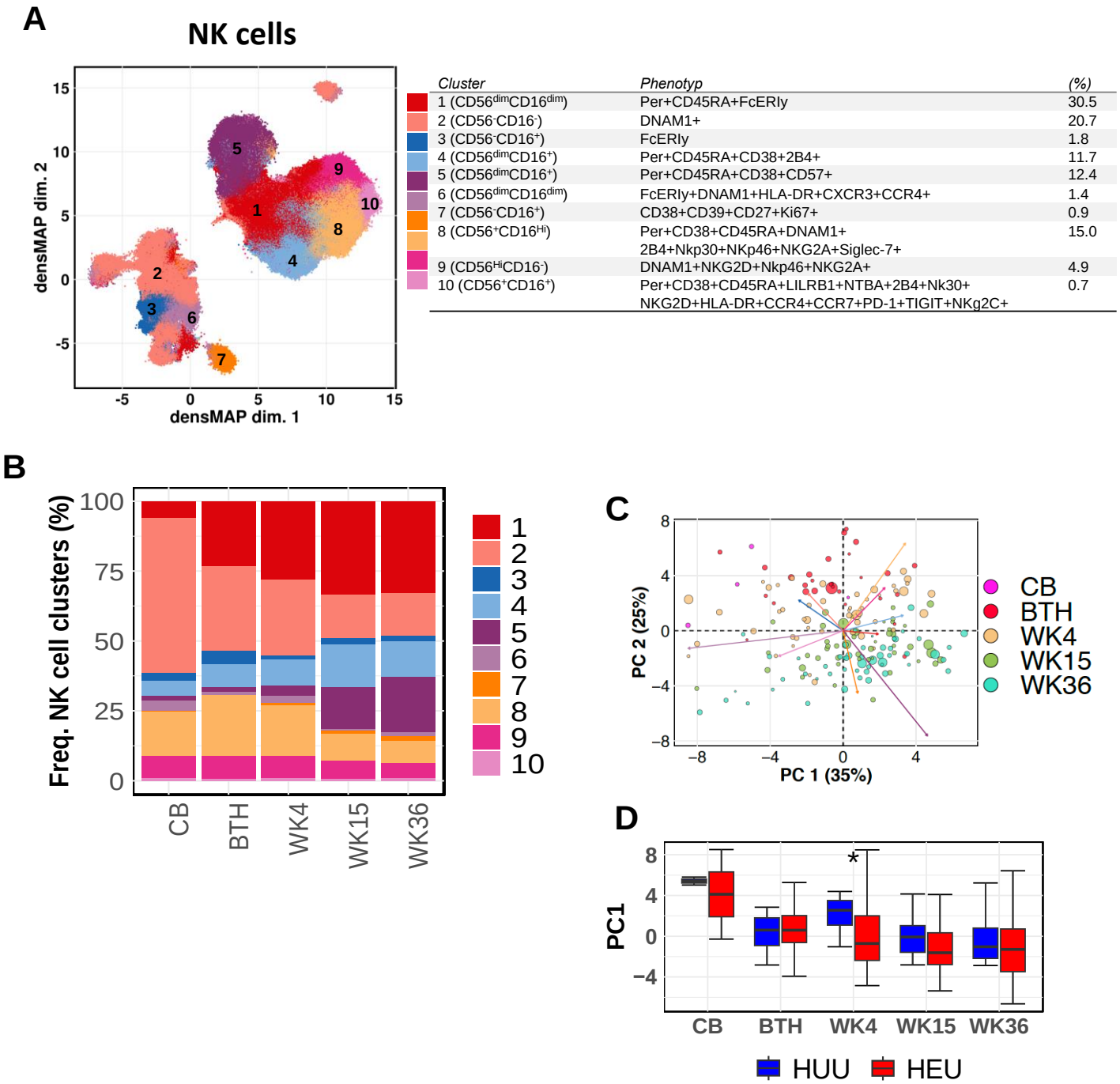


Figure 2: Divergent T cell memory differentiation in HIV-exposed uninfected infants (iHEU) compared to HIV-unexposed uninfected infants (iHUU). **A & D)** Uniform manifold approximation and proximation (UMAP) with density preservation showing dimensional reduction of FlowSOM clusters of CD4+ and CD8 T cells **B &E)** Relative abundances of CD4+ and CD8 + T cell clusters measured in cord blood (CB) and infant peripheral blood at birth (BTH), weeks (WK) 4, 15 and 36. **C &F)** Age related phenotypic composition of CD4+ and CD8+ T cells depicted as principal component (PC) coordinates of centred log-odd ratios of FlowSOM clusters for each infant samples and arrows indicating contribution of each cell cluster in scatter of PC components. **G &H)** Boxplots comparing PC coordinates of CD4+ and CD8+ T cells between iHEU and iHUU longitudinally. **I & J)** Generalized linear mixed model (GLMM) comparing the abundances of CD4+ and CD8+ T cell clusters between iHEU and iHUU at weeks 15 and 36, Positive log Fold change (FC) signified cell cluster frequencies were higher in iHEU compared to iHUU.



Cluster	Phenotyp	(%)
1 (CD56^{dim}CD16^{dim})	Per+CD45RA+FcERIy	30.5
2 (CD56^{-}CD16^{-})	DNAM1+	20.7
3 (CD56^{-}CD16^{+})	FcERIy	1.8
4 (CD56^{dim}CD16^{+})	Per+CD45RA+CD38+2B4+	11.7
5 (CD56^{dim}CD16^{+})	Per+CD45RA+CD38+CD57+	12.4
6 (CD56^{dim}CD16^{dim})	FcERIy+DNAM1+HLA-DR+CXCR3+CCR4+	1.4
7 (CD56^{-}CD16^{+})	CD38+CD39+CD27+Ki67+	0.9
8 (CD56^{+}CD16^{Hi})	Per+CD38+CD45RA+DNAM1+ 2B4+Nkp30+NKp46+NKG2A+Siglec-7+	15.0
9 (CD56^{Hi}CD16^{-})	DNAM1+NKG2D+Nkp46+NKG2A+	4.9
10 (CD56^{+}CD16^{+})	Per+CD38+CD45RA+LILRB1+NTBA+2B4+Nk30+ NKG2D+HLA-DR+CCR4+CCR7+PD-1+TIGIT+NKg2C+	0.7

Figure 3: Early life immunophenotypic alteration in NK cells in HIV-exposed uninfected infants (iHEU) compared to HIV-unexposed uninfected infants (iHUU). **A)** Uniform manifold approximation and proximation (UMAP) with density preservation showing dimensional reduction of FlowSOM clusters on NK cells and table summarising immune cell cluster phenotype. **B)** Relative abundances of NK cell clusters measured in cord blood (CB) and infant peripheral blood at birth (BTH), weeks (WK) 4, 15 and 36. **C)** Age related phenotypic composition of NK cells depicted as principal component (PC) coordinates of centred log-odd ratios of FlowSOM clusters for each infant samples and arrows indicating contribution of each cell cluster in scatter of PC components. **D)** Boxplots comparing PC coordinates of NK cells between iHEU and iHUU.

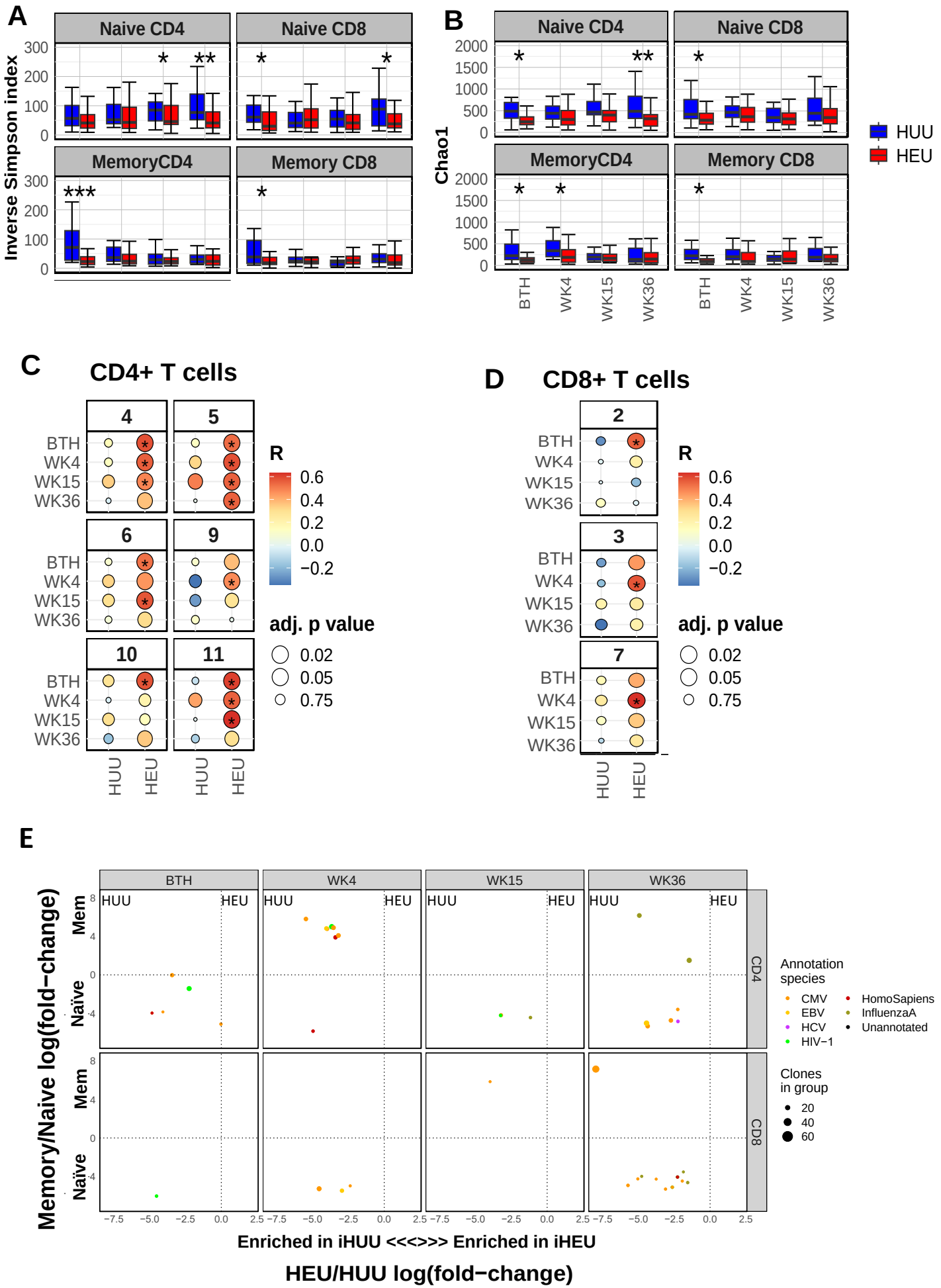


Figure 4: Premature CD4+ and CD8+ T cell receptor (TCR) repertoire skewing in HIV-exposed uninfected infants (iHEU) relative to HIV-unexposed uninfected infants (iHUU). **A)** Boxplots comparing Inverse Simpson TCR diversity scores between iHUU and iHEU at birth and weeks 4, 15 and 36. **B)** Boxplots comparing Chao1 TCR clonotype richness between iHUU and iHEU at birth and weeks 4, 15 and 36. **C &D**) Spearman's rank correlation between naive CD4+ and CD8+ T cell Inverse Simpson scores and frequencies of FlowSOM clusters for CD4+ and CD8+ T cell clusters respectively. **E**) . GLIPH analysis showing antigen specificity groups of the top 100 TCR clones that are significantly different between iHEU and iHUU and between naïve and memory CD4+ and CD8+ T cell subsets.

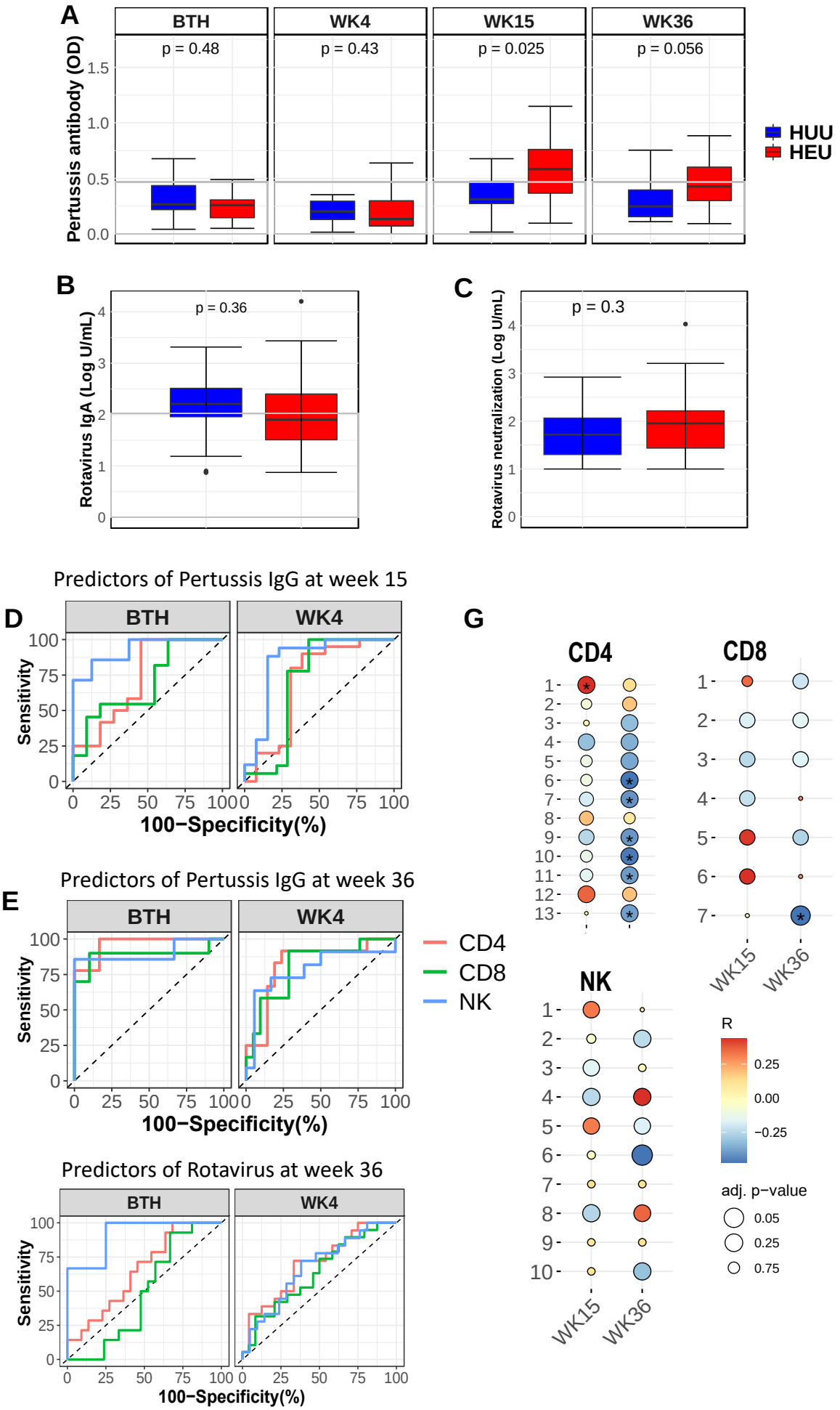


Figure 5 Association of vaccine antibody responses to immune cell phenotypes. **A)** Comparing IgG levels against pertussis between HIV-exposed uninfected infants (iHEU) and HIV-unexposed uninfected infants (iHUU), grey shaded area indicate threshold IgG levels for protective pertussis vaccine response. **B & C**) Boxplots comparing rotavirus specific IgA and neutralization titres between iHEU and iHUU at week 36. **D, E & F)** Summary of ROC analysis using the latent variable axis-1 derived from partial least square discriminate analysis (PLS-DA) of NK, CD4 and CD8 T cell clusters determined to be best predictors at birth and week 4 of pertussis antibody responses at weeks 15 and 36 and rotavirus antibody response at week 37. **G)** Spearman's correlation between abundances FlowSOM clusters of CD4+ and CD8+ T cells and NK cells and anti-pertussis IgG titres measured at week 15 and 36.