

1 **Tissue determinants of the human T cell receptor repertoire.**

2

3 **Authors:** Suhas Sureshchandra¹, James Henderson², Elizabeth Levendosky³, Sankalan Bhattacharyya²,
4 Jenna M Kastenschmidt¹, Andrew M Sorn¹, Mahina Tabassum Mitul¹, Aviv Benchorin¹, Kyle Batucal¹, Allyssa
5 Daugherty³, Samuel JH Murphy^{3,4}, Chandrani Thakur³, Douglas Trask⁵, Gurpreet Ahuja⁶, Qiu Zhong⁶, Annie
6 Moisan⁷, Andreas Tiffreau-Mayer^{*2}, Naresha Saligrama^{*3}, Lisa E Wagar^{*1}

7

8 **Affiliations**

9 1 Department of Physiology & Biophysics, Institute for Immunology, Center for Virus Research, Vaccine
10 Research & Development Center, and Cancer Research Institute, University of California Irvine, Irvine, CA,
11 USA

12 2 Division of Infection & Immunity, Institute for the Physics of Living Systems, University College London,
13 London, UK

14 3 Department of Neurology Bursky Center for Human Immunology and Immunotherapy Programs; Hope
15 Center for Neurological Disorders; Center for Brain Immunology and Glia (BIG), Siteman Cancer Center,
16 Washington University School of Medicine, St. Louis, 63110, USA

17 4 Medical Scientist Training Program, Washington University School of Medicine; St. Louis, 63110, USA

18 5 Department of Otolaryngology and Head & Neck Surgery, University of California Irvine, Irvine, CA, USA

19 6 Department of Otolaryngology, Children's Hospital of Orange County, Orange, CA, USA

20 7 Roche Pharma Research & Early Development (pRED), Basel, Switzerland

21 *equal contributions

22 Correspondence: Lisa Wagar, lwagar@hs.uci.edu

23

24 **SUMMARY**

25 98% of T cells reside in tissues, yet nearly all human T cell analyses are performed from peripheral blood.
26 We single-cell sequenced 5.7 million T cells from ten donors' autologous blood and tonsils and sought to
27 answer key questions about T cell receptor biology previously unanswerable by smaller-scale experiments.
28 We identified distinct clonal expansions and distributions in blood compared to tonsils, with surprisingly low
29 (1-7%) clonal sharing. These few shared clones exhibited divergent phenotypes across bodily sites. Analysis
30 of antigen-specific CD8 T cells revealed location as a main determinant of frequency, phenotype, and
31 immunodominance. Finally, diversity estimates from the tissue recalibrates current repertoire diversity
32 estimates, and we provide a refined estimate of whole-body repertoire. Given the tissue-restricted nature of T
33 cell phenotypes, functions, differentiation, and clonality revealed by this dataset, we conclude that tissue
34 analyses are crucial for accurate repertoire analysis and monitoring changes after perturbing therapies.

35

36 **INTRODUCTION**

37 T cells are fundamental players in the adaptive immune system that can recognize a remarkable diversity of
38 peptide antigens through their T cell receptor (TCR) repertoire^{1,2}. While blood samples are often used to
39 monitor changes in the human TCR repertoire, they represent less than 2% of the body's T cells³. Most T
40 cells reside in the lymphatics (spleen, lymph nodes, tonsils) and non-lymphoid tissues (gut, skin, lung, etc.),
41 where they surveil for infections and maintain tissue homeostasis⁴⁻⁷. Specialized T cell subsets serve unique
42 tissue-based functions⁸. For example, tissue-resident memory T cells (TRM) are crucial for local immune
43 responses^{9,10} and homeostasis¹¹, and T follicular helper cells (TFH) are essential for germinal centers¹²⁻¹⁴.
44 These specialized T cells are often absent or rarely detected in circulation, so tissue-based analyses are
45 essential to understand the full scope of T cell diversity and function¹⁵. The extent of phenotypic mixing and
46 repertoire compartmentalization between blood and tissue T cells remains unclear, largely due to low-scale
47 sampling. It is likely that T cell location influences its phenotype and TCR specificity. Current evidence
48 suggests that while T cell clones with common ancestors exhibit similar transcriptional profiles, single naive
49 clones can differentiate into various phenotypes during infection¹⁶⁻¹⁹ or tumor progression²⁰. Although much
50 progress has been made over the last decade, several critical questions remain unanswered. First, what are
51 the patterns of phenotypic concordance or divergence within clones shared between blood and tissue?
52 Second, since tissue sites maintain T cell clones that respond to localized cues, is the relationship between a
53 cell's TCR and phenotype dependent on its location? Finally, is the extent of clonal sharing dictated by the
54 nature of the antigen itself?

55

56 TCR repertoire diversity is also shaped by various host factors²¹ and environmental factors. Age is a major
57 determinant of TCR diversity²²⁻²⁵ and serves as a proxy for both changes in T cell development and
58 accumulation of immune memory, encompassing exposure to pathogens and vaccines, self-reactivity, and
59 the cumulative impact of chronic diseases²⁶. Importantly, memory T cells increase significantly in the first
60 decade of life^{27,28}, with early antigen exposures shaping TCR diversity, which is later refined by clonal
61 selection²⁹. A decline in T cell function and diversity is associated with increased susceptibility to disease³⁰.
62 An accurate estimation of TCR diversity in humans is, therefore, important for understanding effective control
63 of infections and subsequent protection^{31,32}.

64

65 Current estimates of peripheral TCR diversity have relied heavily on single (beta) chain sequencing of up to a
66 million T cells to extrapolate overall bodily diversity^{23,33,34}. While a broad range of T cells ($>10^{20}$) is
67 theoretically possible, our body has only $\sim 4.5 \times 10^{11}$ T cells³, with actual TCR diversity significantly lower due
68 to thymic selection and clonal expansion following antigen-specific response³⁵. On the other hand, a single
69 TCR can recognize a variety of peptide-MHC complexes³⁶⁻³⁸, reducing the diversity needed to recognize any
70 antigen³⁹⁻⁴¹. Estimates of T cell clonotypes in the human repertoire range from 3×10^6 (β chain diversity)⁴² to
71 $10^7 - 10^8$ ^{23,33,43}. A major challenge in reconciling these estimates is under sampling, as statistical estimators
72 struggle with poor sampling, particularly for small clones and singletons^{26,35}. However, recent improvements

73 in single-cell sequencing offer new opportunities for profiling paired TCR chains and T cell transcriptomes
74 from hundreds of thousands of single cells in an unbiased manner.

75

76 In this study, we collected paired gene expression and TCR data from a cohort of ten donors, spanning from
77 infancy to adulthood, and investigated both phenotypic and clonal segregation of T cells between blood and
78 tissue. Tissue T cells were isolated from tonsils, which is a unique and accessible site for studying the
79 influence of both lymphoid and mucosal tissue environments. By examining transcriptional and TCR
80 sequencing data from 5.7 million T cells, we show limited clonal sharing between blood and tissue and that
81 blood sampling alone significantly underestimates clonal and phenotypic T cell diversity in the human body.
82 Our analyses also revealed differences in antigen-specific T cell frequencies, phenotypes, and
83 immunodominance restricted by bodily site. These findings shed new light on our current understanding of
84 TCR repertoire biology and have significant implications for a new era of tissue-focused vaccine and
85 immunotherapy strategies.

86

87

88

89 RESULTS

90 **Comprehensive analysis of phenotypic and clonal diversity of T cells in blood and tonsils.**

91 T cells in tissues acquire unique phenotypes and functional attributes compared to their counterparts in
92 blood^{44,45}. We posited that this compartmentalization of barrier sites from circulation would result in unique
93 phenotypic and clonal characteristics reflecting tissue site-specific exposures. To date, many studies have
94 analyzed and provided approximates of the complexity of human TCR repertoire^{33,43}. However,
95 technological limitations (single chain profiling of a couple of thousands of cells) and challenges in sampling
96 (several donors, but not enough sampling depth, and lack of tissues) have, thus far, limited us from fully
97 reconciling previous assumptions and estimates of human TCR repertoire²⁶. Here, we seek to address
98 these questions by deeply characterizing paired transcriptional and T cell receptor (TCR) repertoire data
99 from $\alpha\beta$ + T cells from both blood and tonsils (**Figure S1A**) from 10 donors, averaging about 300,000 single
100 T cells per donor per tissue compartment (**Figure 1A**). Our primary cohort included demographically diverse
101 donors with different exposure histories to estimate immune variation among individuals (**Table S1**).
102 Selected donors ranged in age from infancy to adulthood, and samples from both male and female
103 participants were included. Donors were diverse in their prior exposure to herpesviruses, known to influence
104 the TCR repertoire⁴⁶, and their indication for surgery (**Figure 1A**).

105

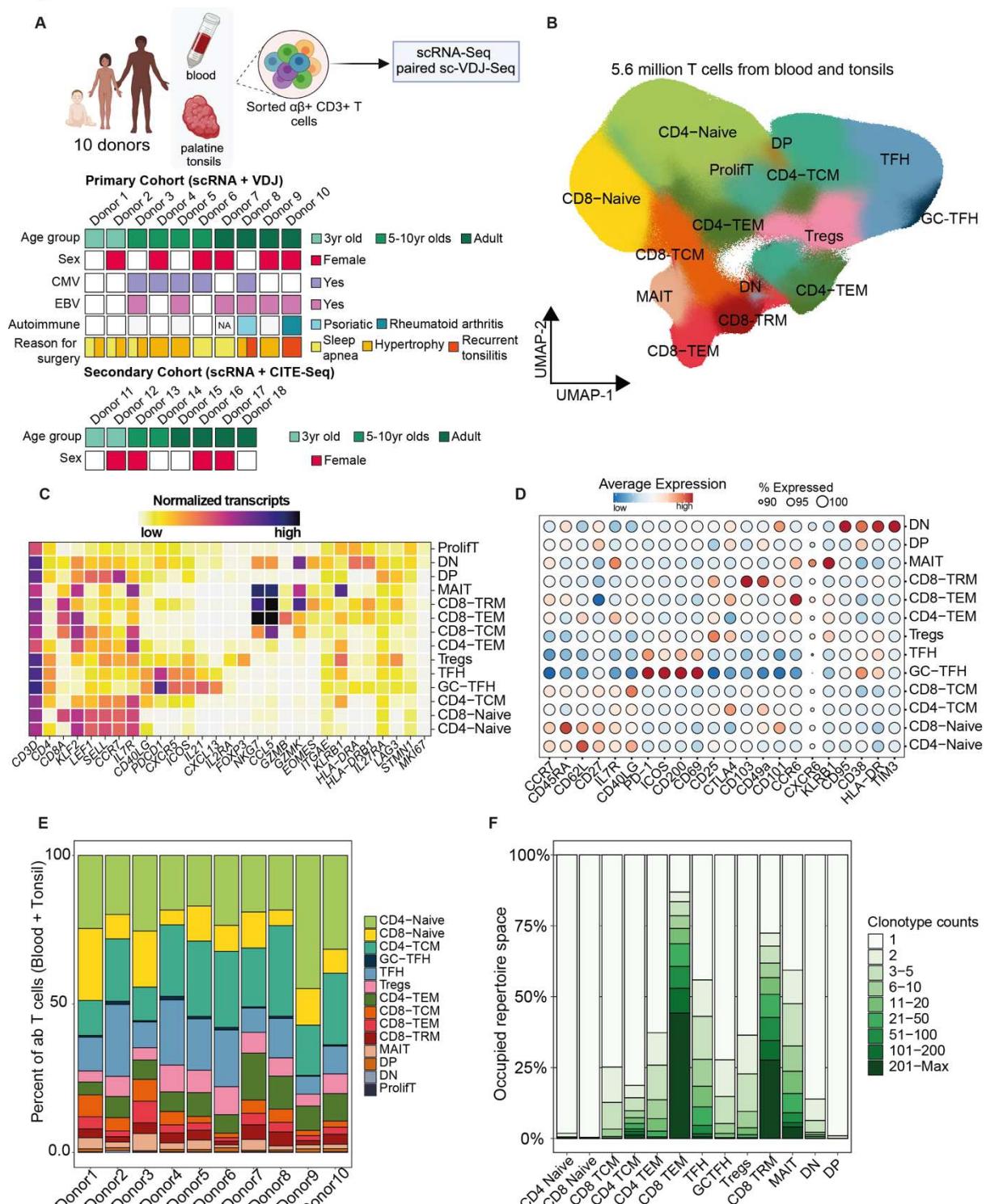
106 Data from blood and tonsils from each donor were integrated and clustered to identify and eliminate any
107 contaminating non- $\alpha\beta$ + T cells. Single cell profiles from the 10 donors were then integrated, normalized, and
108 harmonized, resulting in a combined 5,728,381 cells (**Figure 1B**; **Figure S1B**). We integrated this dataset
109 with blood and tonsillar $\alpha\beta$ + T cells (59,859 cells) from a secondary cohort of 8 donors, where both gene
110 expression and surface protein readouts of 32 T cell markers (**Table S2**) were captured using multimodal
111 immune profiling. Cells were clustered by gene expression at a high resolution to capture rare cell states,
112 resulting in 43 clusters. We carefully stratified these clusters into three levels of annotations based on both
113 gene and protein markers, starting from a broad level 1 annotation (L1: CD4, CD8, double positive (DP),
114 double negative (DN)) to level 2 annotation (L2: naive vs. memory subsets), and a refined level 3 annotation
115 (L3: central memory (TCM), effector memory (TEM), regulatory T cells (Tregs), T follicular helper cells (TFH),
116 etc. as represented by individual clusters) (**Figure 1B**). Numerous expected T cell phenotypic populations
117 were identifiable based on a manual review of gene expression (**Figure 1C**, **Table S3**) and protein-level
118 expression (**Figure 1D**) data.

119

120 All donors were well represented within each cluster (L3 annotation) (**Figure 1E**), although their proportions
121 varied substantially between individuals. TFH subsets were more abundant in pediatric donors, whereas CD4
122 effector memory (TEM) frequencies increased with age (**Figure S1C**). Although the number of cells
123 recovered from each donor was variable (ranging from 380,000 to 820,000), TCR sequence detection was
124 efficient among all donors, indicating that each donor was well represented in the TCR repertoire analysis
125 (**Figure S1D**). Overall, paired productive $\alpha\beta$ TCRs were recovered from roughly 3 million single cells.

126 Productive TCRs were well represented in all major T cell clusters, with slightly lower recovery among
 127 mucosal-associated invariant T (MAIT) cells (**Figure S1E**). As expected, naive T cell clusters were the most
 128 diverse, while memory CD8 T cells (effector memory and resident memory) and MAITs were highly clonal.
 129 TFH, which are largely restricted to lymphoid sites, were the most clonal subset within CD4 T cells (**Figure**
 130 **1F**). This high-resolution map of 5.7 million T cells is the largest (total and on a per-donor basis) dataset of
 131 immune cells profiled to date, encompassing multimodal readouts (transcriptomic, TCR repertoire, and
 132 protein-level expression) and will serve as a blueprint to compare TCR complexity in circulation and tissue.

Figure 1



134

135 **Figure 1: A high-resolution map of blood and tonsillar T cell subsets**

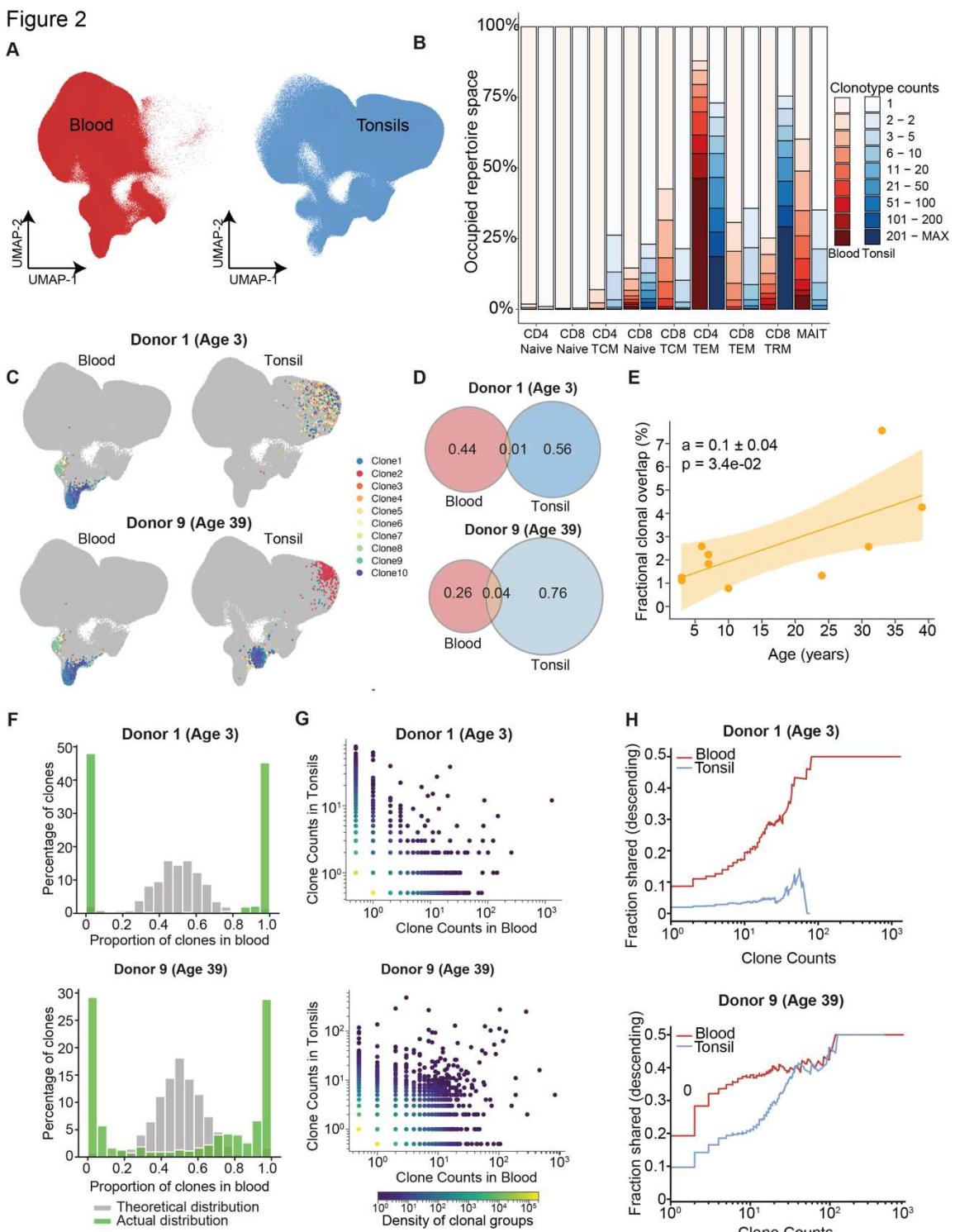
136 (A) Experimental design and brief description of the primary and secondary cohort of human donors used in
137 this study. (B) UMAP of 5.7 million cells from blood and tonsils. Level 3 cluster annotations are highlighted.
138 (C) Heatmap of top distinguishing gene markers for each annotated cluster. (D) Bubble plot highlighting the
139 relative expression of a subset of key cell surface markers from the secondary cohort of donors across L3
140 clusters. The bubble size indicates the percentage of cells in each cluster with detectable protein expression,
141 and color indicates the magnitude of expression ranging from low (blue) to high (red). (E) Stacked bar graph
142 comparing proportional breakdown of L3 clusters from each of the ten donors from the primary cohort. (F)
143 Stacked bar graph comparing clone counts across T cell clusters from all donors. Only cells with productive
144 TCRs were included in downstream analyses.
145

146 **Compartmentalization and TCR sharing between tissue sites.**

147 To assess the extent of sharing vs. segregation of T cell phenotypes and clones between the circulation and
148 tonsils, we stratified the dataset by their compartmental origin (blood vs. tonsil). As expected, naive CD4 and
149 CD8 T cells were readily detected in both sites. However, T cells with certain phenotypes, such as TFH, CD8
150 resident memory (TRM), DP, and DN, were restricted primarily to tonsils, whereas CD8 effector and central
151 memory cells were proportionally higher in circulation (**Figures 2A, S2A, and 2B**). Tregs were detected in
152 both compartments, albeit at higher frequencies in the tonsils. In contrast, most MAIT cells were captured
153 from blood rather than tonsils, as determined by both transcriptomic data and V/J gene usage.
154

155 We restricted downstream analysis to cells with productive TCRs, i.e., cells with at least one beta chain and
156 one or two alpha chains. T cells with productive TCRs were well represented in both blood (ranging from 40-
157 76%) and tonsils (68-83%) from each donor (**Figures S2C, S2D**). We examined the phenotypes of the
158 largest clones in each compartment. While central memory (TCM) subsets were more clonal in tonsils (clonal
159 proportions: 26% for CD4s, 23% for CD8s), TEMs (43% for CD4s and 88% for CD8s) and MAITs (60%) were
160 more clonal in blood (**Figure 2B**). Consistent with the literature⁴⁷, the largest clones in blood were CD8 TEM
161 and MAIT cells (**Figures 2C, S3**). In contrast, the largest expanded clones in tonsils were of TFH phenotype
162 in pediatric donors and mostly CD8 TRMs in adults (**Figures 2C, S3**), and the largest tonsil clones generally
163 occupied more diverse transcriptional profiles than blood clones (**Figure 2C**). Since millions of paired TCR
164 alpha and beta chain sequences were recovered, we assessed the proportion of T cell clones detected in
165 both tissue sites. Unique clones were identified by exact nucleotide matching for alpha and beta TCR
166 sequences, likely representing T cells from a common ancestor. Out of the total TCR repertoire, clonal
167 overlap ranged from approximately 1-7% between blood and tonsils, depending on the donor (representative
168 donors shown in **Figure 2D**). We established a clear correlation between clonal sharing across
169 compartments and donor age (**Figures 2E, S4A**), with older donors showing more shared clones between
170 tissue sites than younger donors. To determine whether the empirical clonal overlap reflected strong or weak
171 sharing among bodily sites, we compared actual repertoire sharing to a theoretically well-mixed TCR
172 distribution estimate. The estimate was generated by equalizing (via subsampling) the number of T cells in
173 each tissue site for a given clone to account for clone size distribution biases. Overall, TCR sharing among
174 tissues was heavily skewed from a theoretical perfect sharing (**Figures 2F and S4B**). Notably, many of the

175 largest expanded T cell clones identified from tonsils were not detected in blood (**Figures 2G and S5A**),
 176 indicating that even highly expanded tonsil T cell clones are not always detectable by peripheral blood
 177 sampling. In contrast, most of the largest blood-derived T cell clones were identified in tonsils, though at
 178 much lower frequencies than in blood. This disparity was consistent across the clonal spectrum, with
 179 medium-sized clones from blood more likely to appear in tonsils (**Figures 2H and S5B**) than vice versa.
 180 Collectively, these findings suggest that T cell clonal expansions are often restricted to discrete bodily
 181 compartments and that blood represents a subsample of T cells from various sites.



183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200

Figure 2: Compartmentalization of TCRs between tonsils and circulation

(A) UMAP of T cells from all donors split by tissue origin, i.e., blood or tonsil. (B) Stacked bar graph comparing the distribution of clone counts within each T cell cluster (L3 annotation) between blood and tonsils. (C) UMAP of T cells from blood (left) or tonsils (right) from two donors, Donor 1 and 9, with their top 10 clones from each compartment highlighted in color. (D) Venn diagrams comparing clonal overlap of T cells between blood and tonsils in the two representative donors. Values indicate the fraction of unique nucleotide sequences per donor. Overlap values indicate the number of unique nucleotide sequences present in both compartments as a fraction of the total number of unique sequences per donor. (E) Line graph comparing the fraction of clones (intersect values from D) shared between blood and tonsils in relation to donor age (represented on the x-axis) (F) Histograms of well-mixed (theoretical, grey) and true clonal sharing (green) of memory T cell compartments between blood and tonsils of the two representative donors. The x-axis represents the proportion of clonal frequencies in blood. (G) Scatter plots directly comparing clone counts in blood and tonsils from the two representative donors. Clones with members only in one compartment were given a clone count 0.5 in the other tissue. The color intensity of scatter points represents the density of clones within each compartment. (H) Clone count distribution for the two representative donors. y-axis represents the average fraction of clones of a particular size with a member in the second compartment. Clones are ranked by size, smallest to largest, from left to right.

201

TCR diversity estimates from tonsils and blood.

202

Current estimates of human TCR diversity were derived primarily from peripheral blood sampling and estimated to span several orders of magnitude, from 3×10^6 (β chain diversity)⁴² to $10^7 - 10^8$ ^{23,33,43}, based on modeling of both alpha and beta chains. However, given the limited clonal sharing we observed between blood and tonsils, we aimed to determine a more accurate estimate of TCR diversity that covered the distribution of clones in both circulation and in tissue sites. A major bottleneck in diversity analysis of human TCR repertoire is undersampling²⁶, coupled with the wide variability of clonal abundances. The distribution of clone counts within the repertoire is heavy-tailed⁴⁸, with the proportion of clones exceeding clonal count thresholds decreasing with approximate power-law scaling (**Figure 3A**). While the largest clones comprised up to 10,000 cells (>1% of total cells), more than 90% of TCRs were represented by only a single cell in the sampled repertoire. Therefore, even at the scale of this dataset (about 20-fold higher per sample than a typical single-cell dataset for a single donor), rarefaction analysis showed no saturation of discovered clonotypes with increasing cell numbers (**Figure 3B**). Moreover, due to the broad distribution of clonal expansion sizes, we expect many expanded clones would not have been detected as such at the typically smaller dataset sizes of prior scTCRseq studies. Indeed, when subsampling cells for each donor, we observed that between 30-60% of memory clonotypes identified as singletons at the average sampling size of prior studies were part of expanded clonotypes when considered within the entire dataset (**Figure 3C**). These results underscore the challenges of insufficient sampling for accurate T cell diversity assessments.

219

220

Next, we assessed the clone count distributions within CD4 and CD8 T cells, each separated by naive and memory phenotypes. We detected clonal expansions in all subsets (defined here by a minimal clone count of two cells), including within the naive CD4 and CD8 compartments (**Figures 3D, 3E**). Within naive CD4 T cells, as previously described^{25,22,23}, these expansions were more frequent with increased donor age (**Figure 3D**) and were found independent of tissue site, albeit at a higher magnitude in blood. As expected, memory T cell populations contained significantly more clonally expanded TCRs (**Figures 3A, 3E**). Memory CD8 T cells

226 were more clonal than memory CD4s, underscoring their divergent proliferative responses to antigens^{49,50}.
227 Differences in clonality were also observed based on tissue compartment, where memory CD4 (all donors)
228 and memory CD8 (in adults) T cells were more clonal in tonsils compared to their blood counterparts (**Figure**
229 **3E**). Collectively, these findings highlight the role of homeostatic mechanisms, antigen exposure, and tissue-
230 specific factors against the backdrop of aging in shaping the T cell repertoire.

231

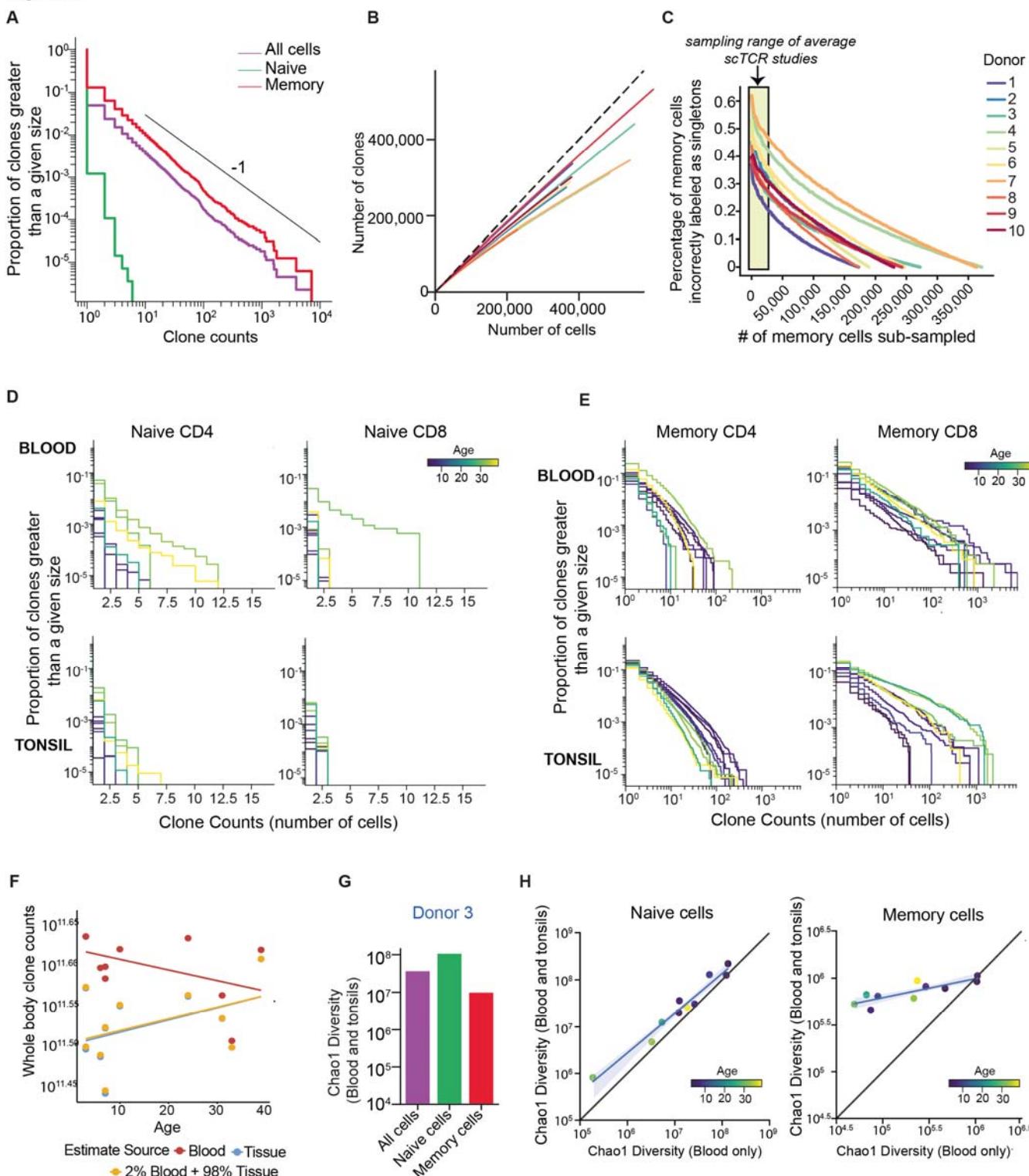
232 Roughly 98% of T cells (of the 4.5×10^{11}) in the human body are in lymphatics or non-lymphoid tissues³. We,
233 therefore, asked if simple extrapolation of tonsils vs. blood repertoire rarefaction curves to whole body clone
234 counts would predict different overall TCR richness estimates. Our analyses revealed a consistent trend of
235 higher estimates from blood repertoire (average 3.95×10^{11} clones) vs. tonsils (average 3.35×10^{11} clones)
236 (**Figure 3F**). However, differences in estimates were highly variable, ranging from 2-43% higher when using
237 blood compared to tissue repertoire. To obtain statistically more robust lower bounds on TCR richness, we
238 applied the nonparametric Chao151 estimator of species (i.e., clones) diversity to sampled T cell clone
239 counts. Considering the differences in clonal distributions between naive and memory T cells, we analyzed
240 Chao1 estimates not only for the entire repertoire but also separately for these subsets. Unexpectedly, we
241 found that naive diversity estimates surpassed the diversity estimates from the total repertoire
242 (representative donor in **Figure 3G**). These results indicate that applying diversity estimators based on T cell
243 phenotype provides a more accurate estimate of total T cell diversity than considering the entire T cell pool.
244 While memory T cells in the tonsils maintained similar diversity levels across donors, memory T cell diversity
245 in the blood decreased with age (**Figure S6**). These findings suggest a preferential accumulation of
246 expanded clones over time in the blood compared to tonsils, while naive T cell diversity declines with age in
247 both compartments (**Figure S6**).

248

249 Finally, we investigated how relying solely on blood T cell data could influence diversity estimates compared
250 to incorporating repertoire data from both blood and tissue. To this end, we compared Chao1 estimates for
251 the naive and memory repertoire calculated using either all sampled cells from each donor or cells from
252 blood alone (**Figure 3H**). While naive TCR diversity estimates were similar, memory diversity was
253 underestimated by up to an order of magnitude in some donors when only blood T cells were sampled
254 (**Figure 3H**).

255

Figure 3



256

257

Figure 3: Diversity estimates of human blood and tissue TCR repertoire.

(A) Clone count distribution of naive, memory, or total T cells for a representative donor, Donor 3. The y-axis represents the proportion of clones greater than a given size of the clone. (B) Rarefaction curves compare the number of clones relative to the number of cells sampled. Each line represents a donor, with blood and tonsillar T cells combined and singletons highlighted as a dashed line. (C) Comparison of fraction of memory T cell clones incorrectly assigned as singletons as a function of the numbers of cells sampled. The sampling size of a typical single-cell TCR dataset is highlighted in yellow. (D-E) Clone count distribution for (D) naive and (E) memory T cells from blood (top) or tonsils (bottom) for all ten donors (colored by age). The y-axis represents the proportion of clones greater than a given size of the clone. (F) Comparisons of whole-body TCR estimates from rarefaction analyses of each compartment extrapolated to total T cells in the human

267 body. The x-axis represents the donor's age. (G) Nonparametric (Chao1) estimates of TCR diversity for T
268 cells (blood and tonsils combined) from Donor 3, split by their naive/memory phenotype. (H) Comparisons of
269 Chao1 diversity of naive (left) or memory (right) T cells from blood and tonsils combined (y-axis) vs. blood
270 cells alone (x-axis). Each dot represents a donor, and the color represents their age.
271

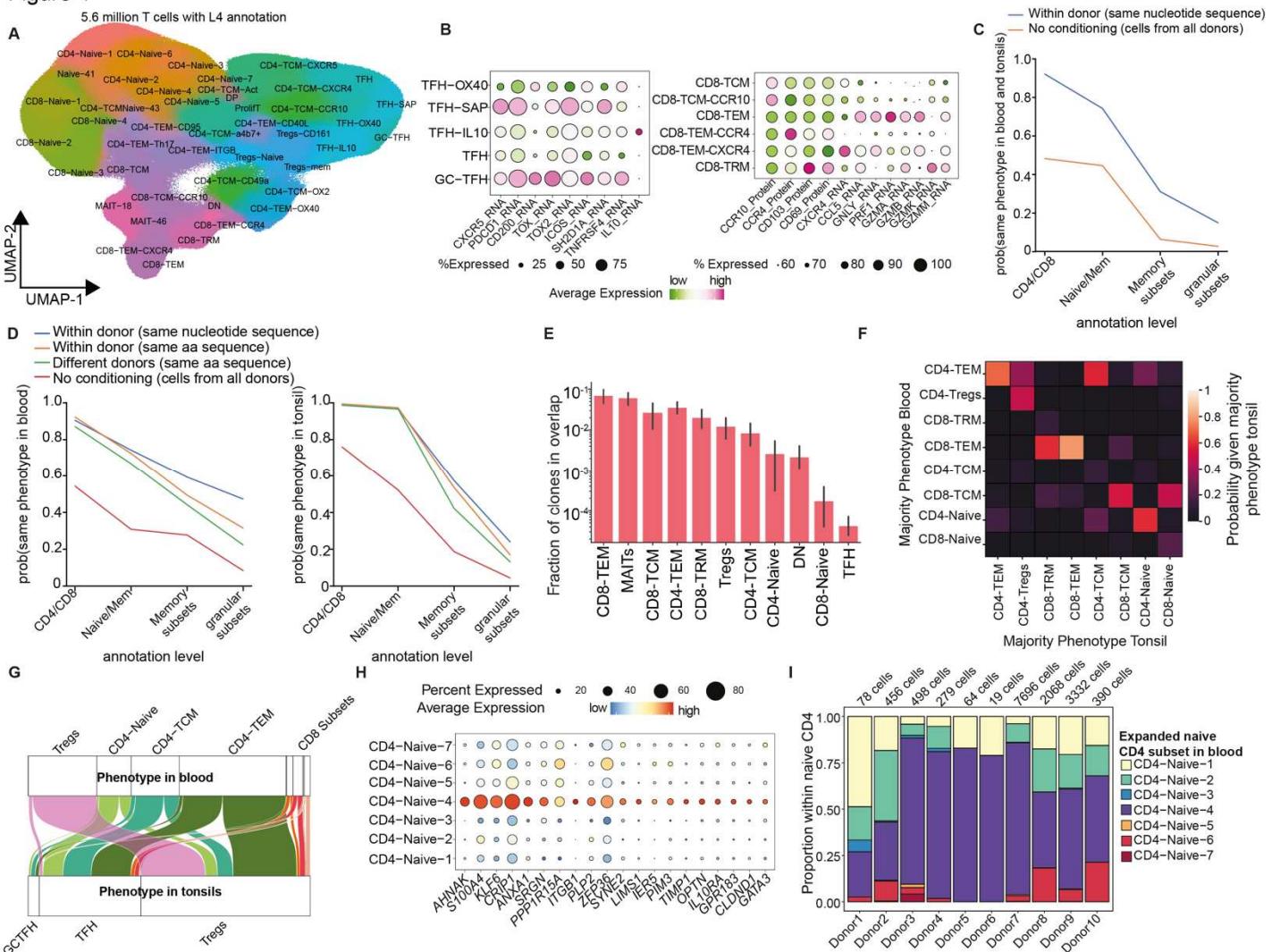
272 **TCR-driven biases in T cell transcriptional state**

273 Prior work has established that T cells differentiating from a common naive T cell are likely but not
274 guaranteed to acquire similar differentiation profiles^{16,17,19,52}. However, only a handful of studies have
275 systematically examined the link between TCR sequence and T cell phenotype potential^{53–55}. Consequently,
276 we asked if there exists a predictable relationship between a cell's TCR and its phenotype and whether that
277 relationship is influenced by the cell's presence in either blood or tissue. To address this question, we used
278 the most detailed phenotypic annotation of each T cell cluster in the dataset, extending the annotations to
279 Level 4 (L4) phenotypes (Figure 4A) based on the expression of specific costimulatory markers, activation
280 states, and cytokines (Figures 4B, S7A, and S7B). Given the limited clone sharing between blood and
281 tonsils, we first asked whether the few shared clones between tonsils and blood shared phenotypic features.
282 Overall, we observed high concordance associated with CD4 vs. CD8 T cell lineage fate and naive vs.
283 memory T cell differentiation status (Figure 4C). As our phenotypic comparisons became more fine-grained
284 (e.g., specific central/effector memory phenotypes and transcriptional profiles), concordance dropped for
285 both blood and tonsil T cells (Figure 4D). CD8 memory subsets and MAITs exhibited high clonal sharing
286 between blood and tonsils (Figure 4E). However, it was more probable that a clone found in both
287 compartments acquired distinct transcriptional profiles based on its tissue site. The most common pairs with
288 converging phenotypes in blood and tonsils were CD4 and CD8 TEMs, whereas the most common pairs with
289 diverging phenotypes were clones with CD8 TRM phenotype in tonsil and CD8 TEM phenotype in blood, or
290 CD4 TCM phenotype in tonsil and CD4 TEM phenotype in blood (Figure 4F).
291

292 To assess the extent of clonal sharing in different T cell phenotypes, we looked at subsets represented in
293 both tonsils and blood. Tregs, for example, exhibited relatively high rates of clonal sharing between the
294 sampled sites (Figures 4E, 4F)⁵⁶. However, our analysis revealed non-Treg phenotypes in blood that were
295 clonally related to tonsil Tregs, including CD4 TEM cells expressing tissue homing markers and a subset of
296 naive CD4s (naive CD4-4) (Figure 4G). Conversely, T cell phenotypes restricted to tissue, such as TFH and
297 GC-TFH, had matching TCR sequences to non-TFH cells in circulation, where they mainly manifested as
298 TEM and TCM, expressed homing markers and, to some degree, were related to a naive cluster (CD4-4)
299 (Figure 4G). Differential gene expression (DEG) analysis of this naive CD4-4 cluster relative to other naive
300 CD4 clusters revealed up-regulation of homing (such as *ITGB1*) and activation (*KLF6* and *ANHAK*) markers
301 (Figure 4H). Interestingly, this subset of naive CD4s accounted for most of the expanded naive CD4 T cells
302 and was particularly enriched in adults (Figure 4I). Taken together, these results show not only a significant
303 degree of differentiation plasticity for T cells, but also highlight the influence of location on phenotypic fate.
304 We conclude that T cell phenotype in the blood does not reliably reflect the differentiation state and functional
305 role of identical clones in tissue.

306

Figure 4



307

308 **Figure 4: Relationship between TCR and T cell phenotype**

309 (A) UMAP of 5.7 million T cells colored by annotation level 4 (L4). Cluster labels highlight L3 annotation in
310 addition to highly expressed markers (tissue homing, cytokine, chemokine, costimulatory molecules, etc.).
311 (B) Bubble plots highlight distinct subsets of TFH (left) and memory CD8 T cells (right) subsets by differential
312 RNA or surface protein expression. (C) Probability of a clone having identical phenotype between tissue
313 compartments. Probabilities are compared across annotation levels (y-axis). A clone here is defined by
314 identical nucleotide sequence matches. The probability of phenotypic sharing independent of the donor origin
315 is highlighted in orange. (D) Compartment-specific probability (blood-left; tonsil-right) of two identical clones
316 having identical phenotypes across annotation levels (x-axis). The blue line represents cells from the same
317 donor with identical nucleotide sequences; orange represents cells from the same donor with identical amino
318 acid sequences; green represents cells from all donors with identical amino acid sequences; red - cells from
319 all donors with no conditioning. (E) Bar graph highlighting the fraction of clones that have members in both
320 blood and tonsils. Data are aggregated across donors. (F) Confusion matrix highlighting T cell phenotypes of
321 shared clones in blood (y-axis) and tonsils (x-axis). Only clones shared between both compartments are
322 highlighted here. Colors indicate the number of shared clones as a proportion of clones in tonsils. (G) Alluvial
323 plot tracking blood phenotype of cells clonally related to tonsillar TFH, GC-TFH, and Tregs. (H) Bubble plot of
324 top differentially expressed genes within naive CD4-4 subset relative to other naive CD4 subsets within L4
325 annotation. The size of the bubble represents the percent of cells within each cluster expressing the marker,
326 whereas color represents normalized, scale transcript counts. (I) Phenotype of expanded naive CD4 T cells
327 (annotation level L4) across donors. Number of cells is highlighted.

328

329 **Divergent differentiation fates of influenza-specific CD8 T cells derived from naive and**
330 **memory pools following vaccination and infection.**

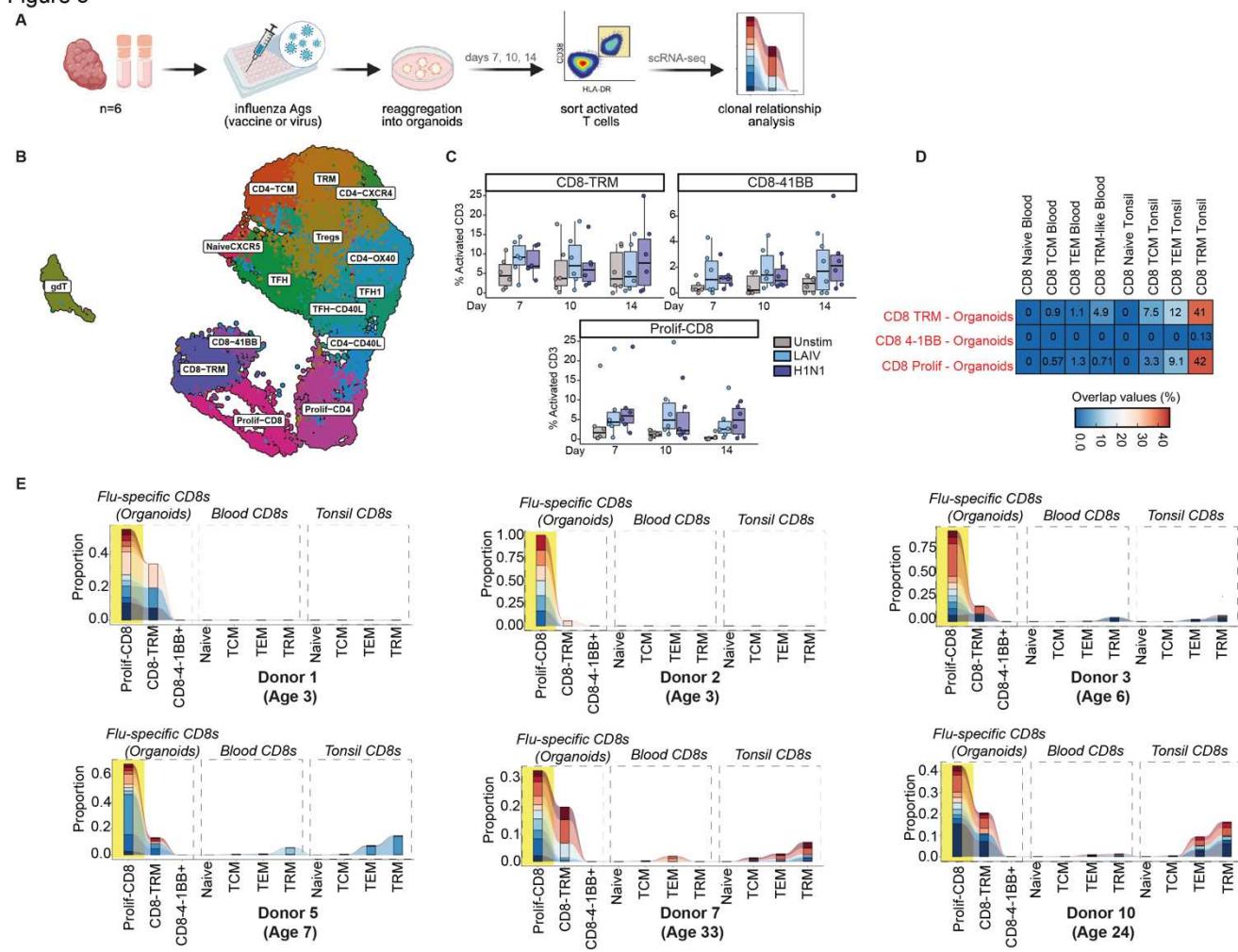
331 We next asked if memory CD8 T cell clones activated by pathogen or vaccine antigens maintain their
332 phenotype or acquire distinct transcriptional states. Tonsil organoids (from 6 of the 10 analyzed donors) were
333 stimulated with live-attenuated influenza vaccine (LAIV) or wild-type A/California/2009 H1N1 virus in replicate
334 cultures for temporal sampling of influenza-specific CD8 T cells. On days 7, 10, and 14 post-stimulation, all
335 activated T cells (CD3+ HLA-DR+ CD38++) were sorted and analyzed using scRNA-seq and TCR-seq
336 (**Figures 5A, S8A**) to capture antigen-responsive T cells. Transcriptional clustering of the cells revealed 3
337 dominant profiles of activated CD8 T cells from tonsil organoids: CD8-TRM, CD8-4-1BB+, and proliferating
338 CD8 (**Figures 5B, S8B**). We analyzed the temporal dynamics of these populations and found that both 4-
339 1BB+ and proliferating CD8 populations expanded relative to unstimulated controls (**Figure 5C**).
340

341 We then compared the clonal relationships between T cells from organoids with CD8 T cell subsets in direct
342 ex vivo blood and tonsils from the same individual. Notably, 4-1BB expressing cells were clonally unrelated
343 to other activated CD8 T cells from organoids or any CD8 T cell subsets assessed ex vivo, suggesting they
344 differentiated from the naive compartment (**Figures 5D, S8C**). In contrast, proliferating CD8 T cells from
345 antigen-stimulated organoids exhibited significant clonal sharing with CD8 TRM and, to a lesser extent, with
346 CD8 TEM and TCM phenotypes from tonsils (**Figure 5D**). Influenza-specific expanded clones from organoids
347 were under-represented within blood CD8 compartments. Since influenza-specific CD8 T cells are readily
348 detected in circulation following vaccination or infection⁵⁷, our findings suggest that under steady-state
349 conditions, the pool of memory CD8 T cells from the tissue that can respond to influenza antigens mostly
350 consists of resident memory cells.
351

352 Granular clonal tracking revealed donor-specific heterogeneity in the relationship between starting and
353 acquired phenotypes after stimulation, presumably influenced by their previous antigen exposure. Organoids
354 from pediatric donors (3 years old) showed no clonal sharing among tonsil memory CD8 subsets, while
355 organoids from adult tissues demonstrated the most sharing (**Figure 5E**). Taken together, these findings
356 indicate that antigen-specific CD8 T cells acquire different phenotypes post-vaccination/infection and that
357 their fate is influenced by their naive or memory origin.

358

Figure 5



359

360

361 Figure 5: Clonal origins of antigen-specific CD8 T cells following vaccination/infection

362 (A) Experimental design - tonsil immune organoids were generated from 6 of the 10 donors from the primary
 363 cohort and stimulated with live-attenuated influenza vaccine (LAIV) or wild-type virus H1N1. Organoids were
 364 harvested on days 7, 10, and 14, and activated CD3+ T cells were analyzed using scRNA-seq. (B) UMAP of
 365 activated T cells and their associated annotations based on gene and protein expression. (C) Box plots
 366 compare the proportions of CD8 T cell subsets within the pool of total activated T cells over time across
 367 stimulation conditions. (D) Heatmap of clonal overlap of flu-specific activated CD8 T cells sampled (x-axis) on
 368 days 7, 10, and 14 compared to CD8 T cell subsets from blood and tonsils sampled directly ex vivo (y-axis).
 369 The data shown here is aggregated across donors. (E) Clonal tracking analysis of top 10 flu-specific
 370 proliferating CD8 T cell clones within each of the 6 donors and their occupied repertoire within CD8 subsets
 371 sampled ex vivo from blood and tonsils.

372

373 **Tissue-specific phenotypic and clonal segregation of antigen-specific T cells**

374 Recent work demonstrated that antigen specificity is strongly linked to CD8 T cell phenotype^{54,58-60}. While
 375 this association has been well established in peripheral blood T cells, it remains unclear whether a similar
 376 relationship spans different bodily sites. If true, one would expect that CD8 T cells with a tissue-restricted
 377 transcriptional profile would have unique TCR specificities compared to blood. To address this, we searched
 378 for TCR sequence matches between our data and VDJdb, a public repository of TCRs with known
 379 specificities⁶¹ (Figure S9). We identified 22 exact matching TCRs (3,195 CD8 T cells) with VDJdb, some of

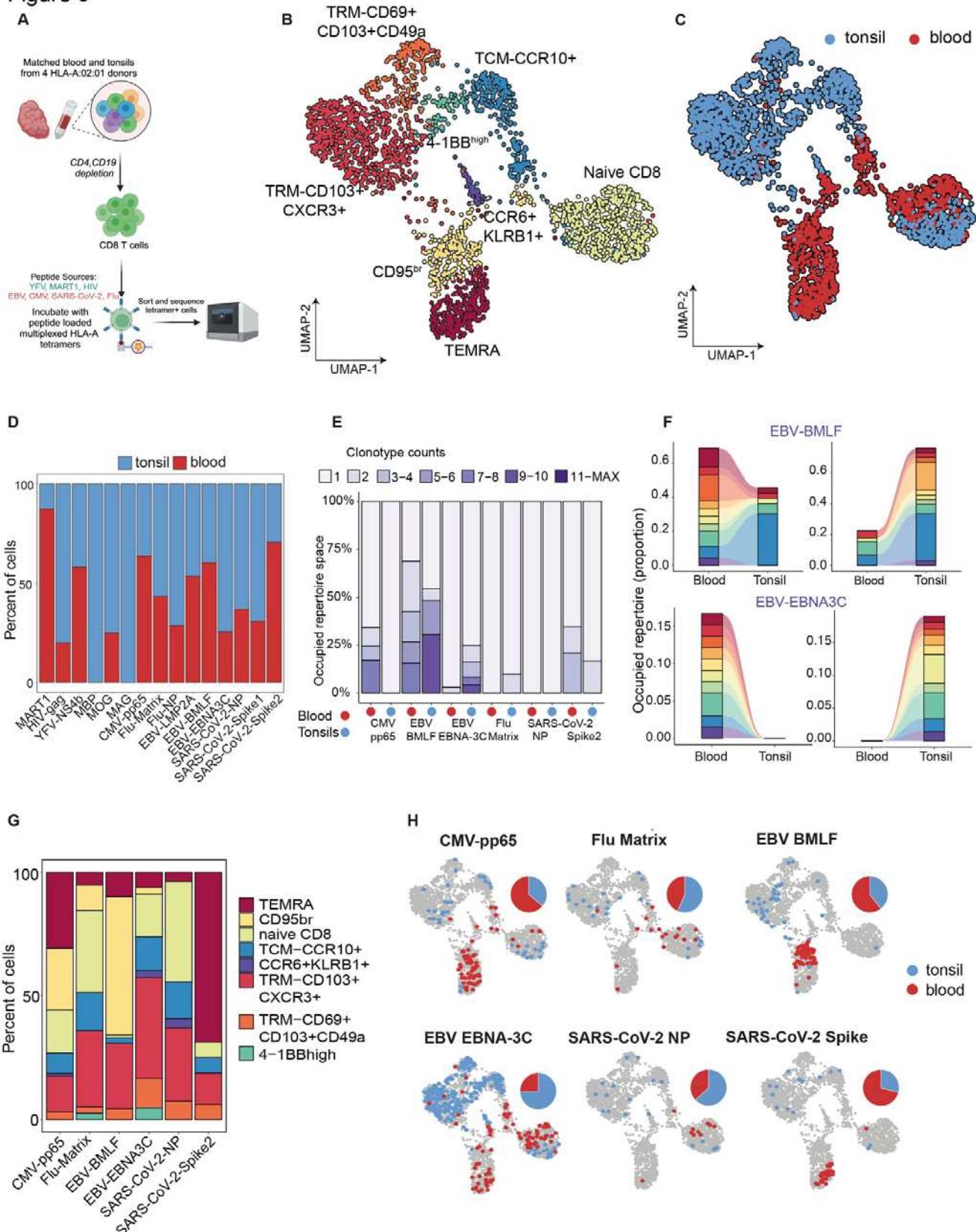
380 which were expanded clones in our dataset. We further analyzed the expanded clones by tissue site. For one
381 particular specificity (AVFDRSDAK, derived from EBV), one donor had three different expanded clones,
382 which were present at varying frequencies in blood and tonsils but conserved in phenotype (**Figure S9A**). In
383 another case, two clones specific for a different EBV peptide (FLRGRAYGL) were only found in the tonsils
384 and had TRM and TEM phenotypes (**Figure S9B**). Finally, a sequence identified as a dual binder for a CMV
385 and an influenza peptide (CMV-NLVPVMVATAV; influenza-GILGFVFTL) in VDJdb was also found in our
386 dataset and represented a single clone found in circulation (**Figure S9C**), with limited representation in
387 tonsils.

388
389 Given these observations from database matching, we next asked if compartmentalization of antigen-specific
390 T cell clones or even specificities extend to other antigens. To test this, we sorted and single-cell sequenced
391 CD8 T cells of known specificities using oligo-barcoded tetramers (**Figure 6A**) with 15 different peptides
392 covering naive antigens (melanoma, yellow fever virus vaccine, human immunodeficiency virus) and a
393 variety of common memory exposures such as Epstein-Barr virus (EBV), cytomegalovirus (CMV), influenza,
394 and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2). Matched blood and tonsils
395 from an independent cohort of donors (2 pediatric and 2 adults) were used for this analysis. Dimensionality
396 reduction and clustering of cells with single antigen specificities revealed clear separation of antigen-specific
397 CD8 T cells by tissue source (**Figures 6B, S10A**). Additionally, these clusters were consistently represented
398 among all donors (**Figure S10B**). While naive and TCM cells were observed in both compartments, TRMs
399 expressing CD69 and CD103 were only found in tonsils, and CD95+ and effector memory RA (EMRA)
400 phenotypes were restricted to blood (**Figures 6B, 6C, and S10A**). We analyzed the tissue
401 compartmentalization and clonality within each antigen specificity and found that clone sizes varied
402 substantially between tissue compartments. EBV and influenza-specific CD8 T cells were more clonal in
403 tonsils than in blood, while CMV-specific cells were clonally expanded in circulation (**Figures 6D, 6E**). Clonal
404 tracking analysis of the top ten clones for each specificity revealed the extent of clonal sharing between
405 compartments (**Figure 6F**). While the top EBV-BMLF clones in either blood (left) or tonsil (right) had
406 representation in the opposing compartment, the top clones specific for EBV EBNA-3C (**Figure 6F**), CMV-
407 pp65, influenza-matrix, and SARS-CoV-2 NP/Spike (**Figure S10B**) were unique within each compartment.
408 These data, combined with our VDJdb matching analysis, demonstrate that most memory TCR clones are
409 compartmentalized to different bodily sites and that this is due, at least in part, to their specificity and the
410 influence of the tissue microenvironment.

411
412 In humans, significant clonal expansions within the CD8 T cell repertoire are dominated by exposure to
413 chronic infections such as EBV and CMV^{62,63}. Our analysis of antigen-specific T cells against these viruses
414 demonstrates that their frequency, phenotype, and clonality vary considerably between blood and tonsils.
415 While CMV pp65-specific cells were largely restricted to blood and acquired an EMRA-like phenotype
416 (**Figures 6G, 6H**), EBV-specific CD8 T cells showed a more complex behavior that was dependent on both
417 the tissue site and peptide specificity. EBNA-3C-specific CD8 T cells (which bind a peptide from a viral

418 latency factor) were found in both the naive and memory CD8 T cell compartments of the tonsil. In contrast,
419 BMLF-specific CD8 T cells (which bind a peptide from an EBV lytic factor) were more likely to be memory
420 cells in circulation (**Figures 6G, 6H**). A similar phenomenon was also true for SARS-CoV-2 proteins
421 (nucleocapsid vs. spike) (**Figures 6G, 6H**). Taken together, these observations suggest that factors such as
422 the site of infection, protein accessibility, and immune evasion mechanisms likely influence the prevalence
423 and differentiation profiles of antigen-specific T cells.

Figure 6



424

425

Figure 6: Clonal and phenotypic compartmentalization of Ag+ CD8 T cells

426 (A) Experimental design for multiplexed assessment of tetramer-specific CD8 T cells from an independent
 427 cohort of HLA-A*02:01 donors with tonsils and autologous blood samples. CD4 T cells and B cells were
 428 depleted from tonsils and PBMC, stained with multiplexed tetramer assemblies, sorted, pooled, and analyzed
 429 using scRNA-seq. (B) UMAP of Ag+ CD8 T cells with associated cluster annotations based on gene and
 430 protein markers. (C) UMAP of Ag+ CD8 T cells colored by their tissue origin (D) Distribution of CD8 T cells

431 with different specificities (x-axis) in blood and tonsils. (E) Tissue-specific distribution of clone counts of CD8
432 T cells with specificities against viral peptides. Data shown here are aggregated across all donors. (F)
433 Tracking of top 10 clones specific to EBV-BMLF (top) and EBV-EBNA3C (bottom) in blood (left) and tonsils
434 (right) and their occupied repertoire in the other tissue compartment. (G) Stacked bar graph comparing the
435 phenotypic distribution of CD8 T cells with specificities against viral peptides. (H) UMAP of CD8 T cells
436 (aggregated across all donors) against memory antigens in blood and tonsils. Individual pie charts highlight
437 the proportion of cells in blood (red) or tonsils (blue).
438

439 **DISCUSSION**

440 A major impetus behind this study was to define the key determinants of TCR biology by deeply sampling T
441 cells from both blood and tonsils. By comparing these two sources, we aimed to ground our findings with
442 prior research on the peripheral blood TCR repertoire and to contrast them with T cells residing in a mixed
443 lymphoid and mucosal tissue environment. This unprecedented sampling depth analyzed hundreds of
444 thousands to nearly a million single cells from each donor. We juxtaposed this atlas of 5.7 million T cells with
445 cell surface protein readouts from a subset of cells to confidently assign phenotypes widely accepted by the
446 immunology community⁶⁴. Our study lays the groundwork for a definitive understanding of TCR diversity in
447 healthy individuals across various ages and bodily compartments. This foundational knowledge will be a
448 benchmark for identifying disturbances in T cell phenotypes and TCR repertoires related to various human
449 diseases. More importantly, it will aid in the development of advanced analysis tools aimed to quantify and
450 interpret TCR repertoires and reveal complex relationships between TCR sequence, specificity, cell fate, and
451 their respective locations and functions.

452

453 **Tissue compartment differences**

454 We hypothesized that T cell phenotypes would differ between tissue compartments. Consistent with our
455 expectations, certain phenotypes, such as TRMs and TFH, were restricted to tissue and rarely detected in
456 circulation during homeostasis. Other memory subtypes, including TCM, TEM, MAITs, and Tregs⁵⁶, were
457 observed in both compartments in varying proportions, sometimes influenced by the donor's age.
458 Surprisingly, MAITs were more abundant in blood than in tonsils, which likely reflects their rarity in human
459 palatine tonsils rather than discrepancies in cell annotation. Within phenotypes shared between
460 compartments, we observed differences in clonal distributions based on compartmental origin. For example,
461 TCM subsets were more clonal in tissues, while TEMs and MAITs were more clonal in blood, highlighting
462 location-dependent expansion and distribution patterns within phenotypically related cells⁶⁵. While tonsils are
463 just one of the many possible tissue sites, clonal expansions will probably differ based on the tissue and
464 extent of antigen exposure.

465

466 Consistent with prior studies, CD8 TEMs were the most clonal among circulating T cell subsets⁵⁴. In tonsils,
467 the largest TCR clones were TFH in the youngest donors and CD8 TEM/TRM in older children and adults.
468 These expanded TFH clones in pediatric donors were transcriptionally diverse and spanned various TFH
469 phenotypes, whereas CD8 TEM clones in adults were transcriptionally homogeneous. This aligns with recent
470 studies showing that clonally related CD8 T cells are more likely to be transcriptionally similar⁵⁴. In contrast,
471 TFH are highly plastic and can co-opt various flavors of Th programs (TFH1, TFH2, TFH17) depending on
472 the immunological context and extracellular cues they sense⁶⁶. We posit this as a possible explanation for
473 the phenotypic diversity of TFH clones.

474

475 **Clonal sharing and antigen-specific clones.**

476 A major strength of the study was the in-depth sampling of hundreds of thousands of T cells across
477 autologous blood and tissue. While previous studies have investigated the clonality of specific human T cell
478 subsets (TRM, TEM, iNKT) across tissue compartments in humans on a much smaller scale^{47,65,67}, the
479 unparalleled sampling in our study (100X deeper on a per donor basis) allowed us to truly understand the
480 extent of clonal sharing and compartmentalization of T cells between blood and tissue. Even at this sampling
481 depth, a key finding from our analyses was low clonal sharing between the compartments, underscoring a
482 consistent theme in patterns of clonal and phenotypic segregation between blood and tissue. The fraction of
483 shared clones was much lower than theoretical estimates of a well-mixed TCR distribution, ranging from 1-
484 7%. Higher levels of clonal sharing were observed in older donors and correlated with increased clonality of
485 memory CD4 and CD8 subsets from tonsils. Another intriguing finding from our analyses was the presence
486 of a population of naive CD4 T cells with small but detectable clonal expansions in both blood and tonsils
487 from adult donors. A similar population has been described for CD8 T cells in older adults^{23,68}. While the
488 mechanisms regulating this clonal expansion are unclear^{34,69}, we demonstrate that a subset of these cells are
489 clonally related to memory CD4 T cells in tissue. Together, these findings reflect the overall impact of
490 accumulated memory and continuous immunological interactions between blood and tissue with age²⁵. Given
491 the differential distribution of T cell phenotypes and clones by location, it is crucial we incorporate additional
492 tissue analyses in human T cell aging studies.

493
494 An important discovery from our experiments was the distinct clonality and phenotypic patterns of antigen-
495 specific T cells between blood and tissue. CMV-specific CD8 T cells were often restricted to blood, whereas
496 EBV-specific clones were distributed in blood, tonsils, or both, depending on their specificity. Interestingly,
497 large EBNA3C-specific (latency factor) CD8 T cell clones were confined to tonsils and were undetectable in
498 blood, whereas BMLF-specific (lytic factor) T cells were more likely to be present in circulation. Furthermore,
499 using an *in vitro* organoid model, we demonstrated that influenza-specific CD8 T cells that expanded in
500 stimulated cultures were under-represented in autologous blood repertoire^{47,65}. Our analysis shows that
501 localized expansions contribute to disparities in clonal representation, likely dictated by viral
502 presence/function, and suggests that tissue location may play a critical role in establishing
503 immunodominance. It will be crucial to understand the relationship between immunodominant T cell
504 responses and disease or therapeutic outcomes, particularly as they might be influenced by the tissue
505 microenvironment.

506
507 One of the key insights from our analysis is that while migration and tissue patrolling are vital T cell functions,
508 this movement is insufficient to establish a full mixing of T cell clones across different bodily sites. Some
509 evidence for this comes from parabiosis experiments in mice^{4,70,71}, where tissue compartmentalization was a
510 major contributing factor for immune protection in skin, gut, and lungs during infection. Furthermore, while
511 early activated T cells can migrate to all non-lymphoid organs, memory cell trafficking becomes more
512 restricted over time⁷². The extent of this “division of labor” between circulating and bona fide tissue resident
513 subsets in humans is only beginning to emerge^{47,73}. The dynamic nature of T cell trafficking, influenced by

514 factors such as infection site, stage of infection, inflammation, and homeostasis, poses challenges in
515 capturing a consistent picture of T cell distribution and clonal diversity based solely on blood samples^{47,65}.
516

517 **A renewed estimate of TCR repertoire**

518 The size and complexity of the human immune repertoire have significant implications for areas such as
519 vaccine immunotherapy design, autoimmunity, and immune aging. With increasing age, memory CD8 T cells
520 were more clonal in tonsils than blood, suggesting higher rates of clonal expansion in tissue relative to
521 circulation. In contrast, the CD4 T cell memory subset was consistently more clonal in tissue in all donors.
522 These observations suggest that, in addition to qualitative differences between tissue and blood repertoire,
523 there exists significant differences in clonal distribution between blood and tissue. Current estimates of
524 diversity are derivatives of blood sampling, and range from 10^6 - 10^8 , depending on beta vs. paired chain
525 sequencing^{23,33,43}. Our analyses suggest that, at least for the memory compartment, the addition of tissue
526 repertoire provided a better estimate of diversity than blood alone. Furthermore, since blood represents only
527 2% of total T cells in the human body³, we argue that using the tissue repertoire to model the remaining 98%
528 of T cells in the human body would provide a more accurate estimate of TCR diversity. Towards this, our
529 calculations indicate an average total repertoire size of 3.3×10^{11} clones based on real T cell distribution in
530 the bodily sites vs. 3.95×10^{11} clones based on extrapolation of blood repertoire. This discrepancy is more
531 prominent in young children, with overestimations reaching as high as 42% when only blood was used to
532 predict the overall diversity. Collectively, these findings reiterate that in addition to estimating diversity on a
533 per phenotype basis (naive vs. memory), the source of T cells (blood vs. various tissues and the extent of
534 sharing) are essential factors that should be considered while evaluating the total TCR repertoire in humans
535 at steady state. We argue that current methodologies, which rely on sampling a few thousand cells from
536 circulation, might effectively capture ongoing immune responses to vaccination or immunotherapy, but
537 accurately identifying the peak of T cell response remains challenging.
538

539 **Relationship between TCR sequence and phenotype**

540 It has long been believed that T cells with a given specificity differentiating from a common naive cell tend to
541 acquire similar transcriptional or functional profiles. With the advent of single-cell technologies, which permit
542 the integration of protein, gene, and TCR readouts from the same cell, there has been a renewed interest in
543 testing a potential deterministic relationship between a cell's TCR and its phenotype⁵⁴. It has been reported
544 that CD8 T cells for a given specificity have identical transcriptional profiles in circulation⁵⁴. On the contrary,
545 several studies in mouse models have demonstrated that a single naive CD4⁵² or CD8 T cell¹⁶ can give rise
546 to multiple fates, suggesting significant interclonal and intraclonal functional heterogeneity in T cell
547 responses. The ultimate fate of the cell depends on the quality of early priming⁷⁴, interactions between T cells
548 and antigen-presenting cells (APC)⁷⁵, or subsequent signals during infection⁷⁶. We hypothesized that the
549 tissue microenvironment could potentially shape the phenotype of clonally related cells in humans. Our
550 analysis revealed high concordance between CD4 vs. CD8 lineages and naive vs. memory status across
551 compartments, suggesting some degree of phenotypic consistency. More specifically, CD8 memory subsets

552 and MAIT cells were more likely to retain similar phenotypic features across blood and tonsils. However, we
553 also observed clonally related T cells with phenotypic disparities across compartments, such as CD8 TRMs
554 in tonsils and CD8 TEMs in blood. This highlights the possibility that identical TCR-bearing clones can adopt
555 divergent functionalities based on location.

556
557 We observed a higher degree of clonal sharing between compartments for T cell subtypes present in both
558 blood and tonsils, such as Tregs. This is consistent with a recent report indicating that most Treg clones are
559 specific for ubiquitous self-antigens and widely circulate and recirculate among various organs⁵⁶.
560 Surprisingly, our study points to a significant number of non-Treg cells (TEM, TCM) from blood, that are
561 clonally related to tonsillar Tregs⁷⁷. Similar trends were observed in tissue-restricted TFH phenotypes, which
562 showed significant clonal sharing with non-TFH phenotypes (TEM and TCM) in blood. This is a surprising
563 finding, given that influenza-specific TFH cells found transiently in circulation are clonally related to TFH cells
564 but not to non-TFH cells in the tonsils⁷⁸. Although this may be true transiently during acute antigen
565 stimulation, our analysis highlights the possibility that at a steady state, a TCR clone can acquire distinct
566 phenotypes depending on their localization. More importantly, this finding further underscores the adaptability
567 and plasticity of T cells in response to the microenvironment. Consistent with this idea, a recent study reports
568 the phenotypic plasticity of Spike S₁₆₇₋₁₈₀ specific CD4 T cells in circulation following COVID-19 mRNA
569 vaccination. Mudd et al. showed that spike-specific TFH persisted in draining lymph nodes for 200 days post-
570 vaccination but circulating TFH-like cells were detected only transiently in peripheral blood and acquired a
571 Th1 phenotype 3-4 weeks post-vaccination.

572
573 The tissue compartment also had a strong influence on the phenotype of antigen-specific CD8 T cells against
574 common viruses (EBV, CMV, SARS-CoV-2) where cells with identical specificities had TEM or TEMRA
575 phenotypes in circulation but a TRM phenotype in tonsils. The phenotypic fate of antigen-specific CD8 T cells
576 following vaccination or infection was also dictated by the phenotype of their precursors. Analysis of tonsil
577 organoids stimulated with influenza vaccines or viruses allowed us to identify a stimulation-specific pool of
578 resting and proliferating TRMs on day 7 post-stimulation that were clonally related to the pre-existing pool of
579 CD8 TRM or TEM in older children and adults but not in younger children. On the other hand, organoid CD8
580 T cells expressing the activation marker 4-1BB were clonally unrelated to CD8 T cells sampled on day 0,
581 suggesting their differentiation from the naive pool of TCRs. Taken together, these results highlight the
582 complex relationship between TCR, naive or memory status, and phenotypic fate in the context of antigen
583 stimulation in tissues.

584
585 **Limitations of the study**
586 Tonsils are tissue sites that combine lymphoid and mucosal characteristics, making them an ideal location to
587 capture immune responses occurring in these distinct but interconnected environments. However, we
588 recognize that our findings may not be completely generalizable to other tissue sites such as the gut, lung,
589 and skin. Despite the large number of cells studied from a few donors, the observed patterns may not apply

590 to older adult donors, where major age-associated changes in TCR repertoire occur^{68,79}. T cell responses are
591 dynamic, and the extent of clonal sharing between tissues might likely be underestimated, particularly during
592 an ongoing infection. Additionally, phenotypic convergence of antigen-specific T cells between blood and
593 tissue during infection is a likely possibility, which was not tested in this study. Finally, although the organoid
594 system aims to capture this dynamic response, microenvironmental factors such as local cytokines and
595 tissue-specific interactions might not be fully understood or easily replicable *in vitro*. Despite these limitations,
596 *in toto*, this study underscores the limited sharing and selective compartmentalization of T cells and antigen-
597 specific T cell responses, with implications for immune surveillance and tissue-specific immunity. Our study
598 highlights the importance of considering tissue localization when assessing T cell responses, whether for
599 chronic infections, immunotherapy, or vaccination and reinforces the need for paired analysis of blood and
600 affected tissues, when possible, especially for organ-restricted infections.

601

602

603 **RESOURCE AVAILABILITY**

604 **Lead contact**

605 Requests for further information or access to data should be directed to Lisa E. Wagar (lwagar@hs.uci.edu).

606

607 **Materials availability**

608 This study did not generate new unique reagents.

609

610 **ACKNOWLEDGEMENTS**

611 The authors thank the tonsillectomy patients and their families for participating in this study, Drs. Jennifer
612 Atwood and Michael Hou at the University of California, Irvine Institute for Immunology flow cytometry core
613 for technical assistance with flow sorting, Drs. Robert Edwards and Delia Tifrea at the University of
614 California, Irvine Medical Center pathology for sample coordination, UCI Genomics Research and
615 Technology Hub (GRT-Hub) for assistance with library preparation, and Drs. Eric Pearlman, Shivashankar
616 Othy, Robert Schreiber, Wayne Yokoyama, Michael Paley, Chyi Hsieh, Ulrike Lorenz, and Maxim Artyomov
617 for critical feedback on the work.

618

619 This work is supported by funding from the Wellcome Leap HOPE Program (to L.E.W.) and the National
620 Institutes of Health (NIH) grant R01AI173023 (to L.E.W.). This work was also made possible, in part, by the
621 Genome Technology Access Center at the McDonnell Genome Institute at Washington University School of
622 Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support
623 Grant P30 CA91842 to the Siteman Cancer Center from the National Center for Research Resources
624 (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research.
625 This work is supported by funding from the Children's Discovery Institute of Washington University and St.
626 Louis Children's Hospital (MI-LI-2020-914 to N.S.). This work was also made possible, in part, through
627 access to the following: the Genomics Research and Technology Hub (formerly Genomics High-Throughput
628 Facility) Shared Resource of the Cancer Center Support Grant (P30CA-062203), the Single Cell Analysis
629 Core shared resource of Complexity, Cooperation, and Community in Cancer (U54CA217378), the
630 Genomics-Bioinformatics Core of the Skin Biology Resource Based Center (P30AR075047) at the University
631 of California, Irvine and NIH shared instrumentation grants 1S10RR025496-01, 1S10OD010794-01, and
632 1S10OD021718-01. This publication is solely the responsibility of the authors and does not necessarily
633 represent the official view of the funding agencies.

634

635 **AUTHOR CONTRIBUTIONS**

636 Designed, optimized, and performed experiments: SS, EL, JMK, AD, MTM, SJHM, CT

637 Analyzed and interpreted data: SS, JH, SB, AB, KB, LEW, AT-M, NS

638 Sample procurement: AMS, DT, GA, QZ

639 Conceived of study, study design, secured funding: SS, AM, AT-M, NS, LEW

640 Writing the original manuscript: SS, LEW

641 Edited, reviewed, and approved manuscript: all authors.

642

643 **DECLARATION OF INTERESTS**

644 JMK is currently an employee of F. Hoffman La Roche, Basel, Switzerland. AB is currently an employee of
645 Amazon Inc., Seattle, USA. LEW declares inventor status on a US patent (US-20230235284-A1) describing
646 the immune organoid technology. The other authors declare no competing interests.

647 **FIGURE TITLES AND LEGENDS**

648 **Figure 1: A high-resolution map of blood and tonsillar T cell subsets**

649 (A) Experimental design and brief description of the primary and secondary cohort of human donors used in
650 this study. (B) UMAP of 5.7 million cells from blood and tonsils. Level 3 cluster annotations are highlighted.
651 (C) Heatmap of top distinguishing gene markers for each annotated cluster. (D) Bubble plot highlighting the
652 relative expression of a subset of key cell surface markers from the secondary cohort of donors across L3
653 clusters. The bubble size indicates the percentage of cells in each cluster with detectable protein expression,
654 and color indicates the magnitude of expression ranging from low (blue) to high (red). (E) Stacked bar graph
655 comparing proportional breakdown of L3 clusters from each of the ten donors from the primary cohort. (F)
656 Stacked bar graph comparing clone counts across T cell clusters from all donors. Only cells with productive
657 TCRs were included in downstream analyses.

658

659 **Figure 2: Compartmentalization of TCRs between tonsils and circulation**

660 (A) UMAP of T cells from all donors split by tissue origin, i.e., blood or tonsil. (B) Stacked bar graph
661 comparing the distribution of clone counts within each T cell cluster (L3 annotation) between blood and
662 tonsils. (C) UMAP of T cells from blood (left) or tonsils (right) from two donors, Donor 1 and 9, with their top
663 10 clones from each compartment highlighted in color. (D) Venn diagrams comparing clonal overlap of T
664 cells between blood and tonsils in the two representative donors. Values indicate the fraction of unique
665 nucleotide sequences per donor. Overlap values indicate the number of unique nucleotide sequences
666 present in both compartments as a fraction of the total number of unique sequences per donor. (E) Line
667 graph comparing the fraction of clones (intersect values from D) shared between blood and tonsils in relation
668 to donor age (represented on the x-axis) (F) Histograms of well-mixed (theoretical, grey) and true clonal
669 sharing (green) of memory T cell compartments between blood and tonsils of the two representative donors.
670 The x-axis represents the proportion of clonal frequencies in blood. (G) Scatter plots directly comparing clone
671 counts in blood and tonsils from the two representative donors. Clones with members only in one
672 compartment were given a clone count 0.5 in the other tissue. The color intensity of scatter points represents
673 the density of clones within each compartment. (H) Clone count distribution for the two representative
674 donors. y-axis represents the average fraction of clones of a particular size with a member in the second
675 compartment. Clones are ranked by size, smallest to largest, from left to right.

676

677 **Figure 3: Diversity estimates of human blood and tissue TCR repertoire.**

678 (A) Clone count distribution of naive, memory, or total T cells for a representative donor, Donor 3. The y-axis
679 represents the proportion of clones greater than a given size of the clone. (B) Rarefaction curves compare
680 the number of clones relative to the number of cells sampled. Each line represents a donor, with blood and
681 tonsillar T cells combined and singletons highlighted as a dashed line. (C) Comparison of fraction of memory
682 T cell clones incorrectly assigned as singletons as a function of the numbers of cells sampled. The sampling
683 size of a typical single-cell TCR dataset is highlighted in yellow. (D-E) Clone count distribution for (D) naive
684 and (E) memory T cells from blood (top) or tonsils (bottom) for all ten donors (colored by age). The y-axis

685 represents the proportion of clones greater than a given size of the clone. (F) Comparisons of whole-body
686 TCR estimates from rarefaction analyses of each compartment extrapolated to total T cells in the human
687 body. The x-axis represents the donor's age. (G) Nonparametric (Chao1) estimates of TCR diversity for T
688 cells (blood and tonsils combined) from Donor 3, split by their naive/memory phenotype. (H) Comparisons of
689 Chao1 diversity of naive (left) or memory (right) T cells from blood and tonsils combined (y-axis) vs. blood
690 cells alone (x-axis). Each dot represents a donor, and the color represents their age.
691

692 **Figure 4: Relationship between TCR and T cell phenotype**

693 (A) UMAP of 5.7 million T cells colored by annotation level 4 (L4). Cluster labels highlight L3 annotation in
694 addition to highly expressed markers (tissue homing, cytokine, chemokine, costimulatory molecules, etc.).
695 (B) Bubble plots highlight distinct subsets of TFH (left) and memory CD8 T cells (right) subsets by differential
696 RNA or surface protein expression. (C) Probability of a clone having identical phenotype between tissue
697 compartments. Probabilities are compared across annotation levels (y-axis). A clone here is defined by
698 identical nucleotide sequence matches. The probability of phenotypic sharing independent of the donor origin
699 is highlighted in orange. (D) Compartment-specific probability (blood-left; tonsil-right) of two identical clones
700 having identical phenotypes across annotation levels (x-axis). The blue line represents cells from the same
701 donor with identical nucleotide sequences; orange represents cells from the same donor with identical amino
702 acid sequences; green represents cells from all donors with identical amino acid sequences; red - cells from
703 all donors with no conditioning (E) Bar graph highlighting the fraction of clones that have members in both
704 blood and tonsils. Data are aggregated across donors. (F) Confusion matrix highlighting T cell phenotypes of
705 shared clones in blood (y-axis) and tonsils (x-axis). Only clones shared between both compartments are
706 highlighted here. Colors indicate the number of shared clones as a proportion of clones in tonsils. (G) Alluvial
707 plot tracking blood phenotype of cells clonally related to tonsillar TFH, GC-TFH, and Tregs. (H) Bubble plot of
708 top differentially expressed genes within naive CD4-4 subset relative to other naive CD4 subsets within L4
709 annotation. The size of the bubble represents the percent of cells within each cluster expressing the marker,
710 whereas color represents normalized, scale transcript counts (I) Phenotype of expanded naive CD4 T cells
711 (annotation level L4) across donors. Number of cells is highlighted.
712

713 **Figure 5: Clonal origins of antigen-specific CD8 T cells following vaccination/infection**

714 (A) Experimental design - tonsil immune organoids were generated from 6 of the 10 donors from the primary
715 cohort and stimulated with live-attenuated influenza vaccine (LAIV) or wild-type virus H1N1. Organoids were
716 harvested on days 7, 10, and 14, and activated CD3+ T cells were analyzed using scRNA-seq. (B) UMAP of
717 activated T cells and their associated annotations based on gene and protein expression. (C) Box plots
718 compare the proportions of CD8 T cell subsets within the pool of total activated T cells over time across
719 stimulation conditions. (D) Heatmap of clonal overlap of flu-specific activated CD8 T cells sampled (x-axis) on
720 days 7, 10, and 14 compared to CD8 T cell subsets from blood and tonsils sampled directly ex vivo (y-axis).
721 The data shown here is aggregated across donors. (E) Clonal tracking analysis of top 10 flu-specific

722 proliferating CD8 T cell clones within each of the 6 donors and their occupied repertoire within CD8 subsets
723 sampled ex vivo from blood and tonsils.

724

725 **Figure 6: Clonal and phenotypic compartmentalization of Ag+ CD8 T cells**

726 (A) Experimental design for multiplexed assessment of tetramer-specific CD8 T cells from an independent
727 cohort of HLA-A*02:01 donors with tonsils and autologous blood samples. CD4 T cells and B cells were
728 depleted from tonsils and PBMC, stained with multiplexed tetramer assemblies, sorted, pooled, and analyzed
729 using scRNA-seq. (B) UMAP of Ag+ CD8 T cells with associated cluster annotations based on gene and
730 protein markers. (C) UMAP of Ag+ CD8 T cells colored by their tissue origin (D) Distribution of CD8 T cells
731 with different specificities (x-axis) in blood and tonsils. (E) Tissue-specific distribution of clone counts of CD8
732 T cells with specificities against viral peptides. Data shown here are aggregated across all donors. (F)
733 Tracking of top 10 clones specific to EBV-BMLF (top) and EBV-EBNA3C (bottom) in blood (left) and tonsils
734 (right) and their occupied repertoire in the other tissue compartment. (G) Stacked bar graph comparing the
735 phenotypic distribution of CD8 T cells with specificities against viral peptides. (H) UMAP of CD8 T cells
736 (aggregated across all donors) against memory antigens in blood and tonsils. Individual pie charts highlight
737 the proportion of cells in blood (red) or tonsils (blue).

738

739

740

741

742 **METHODS**

743 **Informed consent and sample collection.** Tonsils from healthy consented individuals undergoing surgery
744 for obstructive sleep apnea, hypertrophy, or recurrent tonsillitis were collected in accordance with the
745 University of California, Irvine Institutional Review Board (IRB). Ethics approval was granted by the University
746 of California, Irvine IRB (protocol #2020-6075), and all participants provided written informed consent.
747 Participants in the primary cohort used for the study were aged 3-39 (**Table S1**). Overall, tonsil tissue was
748 healthy in appearance. Although most donors were otherwise healthy, two of the participants self-reported
749 prior diagnosis with an autoimmune disease (psoriatic arthritis in Donor 8 and rheumatoid arthritis in Donor
750 10).

751

752 **Sample processing.** Samples were processed as previously described. Briefly, whole tonsils were collected
753 in saline after surgery and immersed in an antimicrobial bath of Ham's F12 medium (Gibco) containing
754 Normocin (InvivoGen), penicillin, and streptomycin for 30-60 min at 4°C for decontamination of the tissue.
755 Tonsils were then briefly rinsed with PBS and mechanically dissociated; debris was removed using gradient
756 centrifugation (Lymphoprep, Stemcell). PBMCs were isolated from blood (diluted 1:1 in PBS) by standard
757 density gradient centrifugation over Ficoll. Cells were counted, and samples were cryopreserved in fetal
758 bovine serum (FBS) with 10% DMSO and stored in nitrogen until use.

759

760 **Pathogen exposure history.** Prior exposure to select viruses was determined using semi-quantitative
761 ELISAs measuring IgG antibodies against CMV (Abcam, cat #ab108724), EBV-VCA (Abcam, cat
762 #ab108730), and EBV-EBNA1 (Abcam, cat #108731) per manufacturer's instructions. Briefly, diluted serum
763 samples were added to pre-coated wells and incubated at 37°C for 1 hour. Samples were then washed,
764 treated with HRP-conjugated secondary antibodies, washed, and TMB substrate was used for signal
765 development. Experimental readouts were measured colorimetrically and seropositivity was established
766 using the thresholds defined by the manufacturer for each kit.

767

768 **T cell isolation for scRNA-seq.** Cryopreserved tonsil and PBMC samples (donors 1-10) were revived using
769 pre-warmed media: RPMI1640 supplemented with glutamax, 10% heat-inactivated FBS, 1x nonessential
770 amino acids, 1x sodium pyruvate, and 1x penicillin-streptomycin. Cells were washed then washed with FACS
771 buffer (PBS + 2% FBS + 0.05% Sodium Azide). Roughly 10^7 PBMC or tonsil cells were stained with
772 Live/Dead NIR (1/1500) in PBS for 30 minutes in the dark. Cells were washed and prestained with Fc Block
773 (1/50) for 20 minutes on ice and stained with a cocktail of antibodies - CD3-BB700 (1/50), TCR-gd-PE-Cy7
774 (1/100), CD14-BV605 (1/100), CD19-BV605 (1/100), and CD16 (1/100) for an additional 30 minutes on ice.
775 After washing, cell pellets were resuspended in 200 uL of 1X PBS + 0.04% BSA and placed on ice prior to
776 sorting. ab+ T cells were sorted based on viable single cells with markers CD19/CD14/CD16- CD3+ TCRgd-
777 (**Figure S1A**).

778

779

780

781 scRNAseq with scTCR

782 Sorted cells were washed and resuspended in 1X PBS with 0.04% BSA, counted, and resuspended in a final
783 concentration at 1200 cells/uL. Single-cell suspensions were loaded on a Chromium X controller (10X
784 Genomics) with a loading target of 30,000 - 50,000 cells in several replicates (4-24 GEM samples depending
785 on the sort yield) with a goal of sequencing ~500,000 cells per donor per compartment (blood or tonsils).
786 Gene expression and TCR libraries were generated using the Chromium Next Gem Single Cell 5' Reagent
787 Kit v2 (Dual Index) per the manufacturer's instructions. Quality and quantity of libraries were measured on
788 tapestation, qubit, and bioanalyzer and sequenced on Illumina NovaSeq 6000 with a sequencing target of
789 30000 reads per cell for gene expression libraries and 5000 reads per cell for TCR.

790

791 scRNAseq with cell surface protein analysis

792 Cryopreserved tonsil and PBMC samples from 8 independent donors (donors 11-18, ages 3-58) (**Table S1**)
793 were revived in pre-warmed RPMI1640 (supplemented with 10% FBS), washed with FACS buffer, and
794 stained using a cocktail containing Fc block, CD19-APC (1/50), CD3-AF488 (1/50), and TCRgd-PE (1/50).
795 Additionally, during the antibody staining step, samples were stained with unique hashing antibodies
796 (TotalSeqC, BioLegend CA) for sample barcoding. Cells were stained for 30 minutes on ice, washed twice
797 with FACS buffer, and sorted on a FACSAria Fusion (BD Biosciences). Roughly 50,000 ab+ T cells were
798 collected from viable single cells with CD19-CD3+TCRgd- phenotype, pooled by compartment, washed with
799 FACS buffer, and stained with a cocktail of oligo-tagged cell surface antibodies (**Table S2**). Cell pellets were
800 washed thoroughly in 4 mL FACS buffer for a total of 4 washes. After the final wash, pellets were
801 resuspended in 100 uL 10X buffer, counted, and resuspended to a final concentration of 2,400 cells/uL.
802 Single-cell suspensions were loaded on a 10X Genomics Chromium controller in duplicates with a loading
803 target of 60,000 cells per sample. Libraries were generated using the Chromium Next Gem Single Cell 5'
804 Reagent Kit v2 (Dual Index) per the manufacturer's instructions with the addition of 5' Feature Barcode
805 libraries. Quality and quantity of libraries were measured on tapestation, qubit, and bioanalyzer, and
806 sequenced on Illumina NovaSeq 6000 with a sequencing target of 30,000 reads per cell for gene expression
807 libraries and 10,000 reads per cell for TCR and feature barcode libraries.

808

809 Single-cell data QC

810 Raw reads from gene expression, TCR, and for a subset of donors, feature barcode libraries were aligned
811 using Cell Ranger Single Cell Software Suite with Feature Barcode addition (version 6.1.1; 10X Genomics)
812 against the GRCh38 human reference genome (GRCh38-2020-A) using the STAR aligner (version 2.7.2a).
813 Alignment was performed using feature and vdj options (vdj_GRCh38_alts_ensembl-5.0.0) in Cell Ranger.
814 Samples were aggregated, resulting in one composite alignment file for each donor and compartment. For
815 the 8 donors with paired cell surface data, feature files from Cell Ranger were manually updated to separate
816 hashing antibodies (HTO) vs. other cell surface antibodies (Protein) to facilitate independent normalization
817 procedures. Next, data frames from each donor/compartment were split by libraries for QC. For each library,

818 the gene expression matrix was loaded in Seurat (version 5.0.2) and TCR genes were summarized into
819 custom *TCRA*, *TCRB*, *TCRG*, and *TCRD* genes to avoid downstream clustering based on V/D/J gene usage.
820 Cellular features such as mitochondrial and ribosomal gene expression and the number of housekeeping
821 genes were added as additional meta-features. Next, within each library, doublets were identified
822 computationally, with an expected doublet rate of 0.4% per 1,000 cells. DoubletFinder (version 3)⁸⁰ was run
823 on default settings (pN = 0.25, pK = 0.09, PCs=1:20) and removed from subsequent QC steps. Poor quality
824 cells and droplets with ambient RNA (number of expressed genes less than 200 and housekeeping genes
825 less than 25) and additional doublets (number of expressed genes more than 4000) were excluded from
826 subsequent analysis. Finally, libraries from each donor and compartment were re-merged to generate a QC-
827 filtered dataset.

828

829 **Donor specific analysis**

830 Before integrating data from all 10 donors, QC-filtered data from blood and tonsils from each donor were
831 analyzed independently. Donor specific objects were generated by merging blood and tonsil datasets from
832 each donor using the *merge* function in Seurat. Data normalization and variance stabilization were performed
833 on the merged donor-specific object using *NormalizeData* and *ScaleData* functions in Seurat. Dimension
834 reduction was performed using the *RunPCA* function to obtain the first 30 principal components, of which the
835 first 20 were used for clustering with Seurat's *RunUMAP* function with a resolution of 0.8. At this stage, cell
836 types were broadly assigned as *CD3D* expressing T cells and *CD19* or *MS4A1* expressing B cells.
837 Contaminating B cell clusters from each UMAP were filtered before downstream integration. For samples
838 with paired protein information, data were normalized using centered log-ratio transformation (CLR), and
839 donor assignment and doublet removal were performed using the *HTODemux* function in Seurat⁸¹. For cell
840 surface protein analysis, read counts were normalized using CLR, and any contaminating *CD19* protein-
841 expressing clusters were excluded from downstream integration.

842

843 **Data Integration and harmonization**

844 To enable efficient and large-scale integrative analysis of millions of cells, we used atomic sketch integration
845 in Seurat⁸². This involved a) iterating through each donor, sampling 50,000 cells (atoms) each from blood
846 and tonsils with equal representation, b) generating a dictionary representation to reconstruct each cell
847 based on the atoms, c) integrating the atoms from each dataset using *FastRPCAIntegration*, and d) for each
848 donor, reconstructing each cell from the integrated atoms using *IntegrateSketchEmbeddings*. Since this
849 approach does not require loading or processing the entire dataset simultaneously, it easily enables analysis
850 of all 5.7 million cells. For the integrated object, we ran *FindVariableFeatures* to extract the top 2,000 variable
851 genes, performed z-score transformation using *ScaleData*, and then performed principal component analysis
852 using *RunPCA*. To correct batch effects, we used Harmony v1.0.107 to scale millions of cells across donors
853 but not compartments. We used *RunHarmony*, with the top 30 PCs as input, and corrected batch effects
854 across donors⁸³. The success of data integration and harmonization was tested qualitatively with the

855 *FindClusters* and *RunUMAP* function using the top 15 dimensions and resolution 2.0, resulting in 47 clusters
856 encompassing 5.7 million cells from blood and tonsils combined.

857

858 **Cluster annotation**

859 Slight over-clustering of the dataset allowed us to annotate clusters to a finer level (annotation level 4), which
860 were progressively consolidated to broad T cell annotations (levels 3, 2, and 1). Gene and protein markers
861 for each of the 47 clusters were identified using the *FindMarkers* function in Seurat. For gene expression, we
862 used Wilcoxon Rank Sum tests for differential marker detection using a log2 fold change cutoff of at least
863 0.4, and clusters were identified using a known catalog of markers defined for T cells from human tonsil⁶⁴. A
864 similar approach was followed for protein expression. Both positive and negative protein markers from each
865 cluster were used to aid manual annotation. In certain cases, clusters were merged when no distinguishing
866 markers were identified between cell states (E.g., TFH subsets). A list of the final cluster-specific markers is
867 provided in **Table S3**.

868

869 **TCR preprocessing**

870 Only QC-filtered cells with confirmed annotation were included in downstream TCR analysis. In addition,
871 cells were included in the TCR analysis only if they contained a single TRB chain and a maximum of 2 TRA
872 chains as productive TCRs. In a further processing step, only the TRA with the highest UMI was retained for
873 cells with multiple TRA to simplify downstream analyses. A secondary dataset was created to analyze clonal
874 expansion, aggregating cells with the same clonotype. Clonotypes were defined by identical full TRA+TRB
875 nucleotide sequences. Clonotypes were annotated by the majority origin (blood or tonsil), and phenotypic
876 annotations of corresponding cells. Basic features of the repertoire (clonality by donor, cluster, compartment)
877 was described using standard workflows in Immunarch⁸⁴.

878

879 **Clonal overlap and mixing**

880 Data were filtered for clonal overlap/mixing analysis to retain only cells annotated as memory phenotype. The
881 following filtering steps were performed for clonal mixing/sharing analysis to equalize blood and tonsil clone
882 size distributions. For each donor, only clones with sizes from 10-100 cells were retained. The number of
883 unique clones in blood and tonsils was equalized by subsampling, followed by equalizing the total number of
884 cells in blood and tonsils. Cells from blood or tonsils were randomly permuted among all cells for a given
885 donor to create a well-mixed dataset. The number of counts for a given clone in blood and tonsils was
886 calculated to compute mixing fractions for the true and well-mixed sets. The probability that a member of a
887 given clone group would be found in a blood sample was then computed as:

$$p(\text{blood}) = (\text{counts_blood}) / (\text{counts_blood} + \text{counts_tonsil})$$

888 and vice versa. This subsampling was performed 10 times for each donor. The data from each run were
889 aggregated to produce the distribution of mixing.

890

892 **Diversity Estimates and Phenotypic matching**

893 T-cell richness was computed using the Chao1 diversity measure⁵¹. Phenotypic coincidence probabilities
894 were calculated using an unbiased Simpson's diversity estimator⁸⁵, with conditional coincidence probabilities
895 calculated by an equally weighted average across conditioning groups. Sampling variance was calculated
896 using the unbiased variance estimator described in⁸⁶. Rarefaction curves were generated by repeated
897 random subsampling of cells and calculating the number of clones within each sample. For each donor,
898 these curves were extrapolated to total T cells in the body, based on estimates published in³ solely from
899 tonsils or blood repertoire or a combination of both (98% tonsil and 2% blood).

900

901 **VDJ-db queries**

902 A matching function was used to query our dataset and associated metadata (donor-specific TCRs, HLA,
903 and annotations) against a set of reference TCRs with known specificities from VDJ-db (cite). Intersecting
904 matches were based on common features such as TRB amino acid sequences and paired TRA-TRB amino
905 acid sequences. Once matched, the epitope and epitope species from the reference and a list of matching
906 cell barcodes were retrieved. Matching sequences were further filtered to ensure either class-I or class-II
907 allele matching for a particular donor.

908

909 **Immune organoids for analysis of activated CD8 T cells**

910 Cryopreserved tonsil cells from 6 of the original 10 donors (donors 1, 2, 4, 5, 7, and 10) were thawed,
911 enumerated, and roughly 1.5 e6 cells were plated in 200 μ l final volume in ultra-low attachment plates
912 (Corning). Organoid media was composed of RPMI1640 with glutamax, 10% FBS, 1x nonessential amino
913 acids, 1x sodium pyruvate, 1x penicillin-streptomycin, 1x Normocin (InvivoGen), 1x insulin, selenium,
914 transferrin supplement (Gibco), and 0.5 μ g/ml recombinant human BAFF. The following amounts of influenza
915 antigens were added to replicate cultures on day 0 based on our prior titration⁸⁷: A/California/07/2009 H1N1
916 virus - 2.5 hemagglutination units (HAU) per culture; 2019/20 live attenuated influenza vaccine (LAIV
917 FluMist® Quadrivalent) - 1/2,000 final dilution. Cultures were incubated at 37°C, 5% CO₂ with humidity, and
918 media was replenished every other day by exchanging 30% of the volume with fresh organoid media.
919 Organoids were cultured for two weeks and harvested on days 7, 10, and 14. On harvest days, organoids
920 were washed in FACS buffer and surface stained in a cocktail containing CD19-BV605 (1/100), CD3-APC-
921 Cy7 (1/100), HLA-DR-PerCP-Cy5.5 (1/100), CD38-FITC (1/100), and Zombie Aqua (1/200), in addition oligo-
922 tagged surface antibodies (TotalSeq-C, **Table S2**). Additionally, each organoid was labeled with a unique
923 hashing antibody (TotalSeq-C), allowing for sample multiplexing. Samples were stained for 30 minutes in the
924 dark on ice, washed twice with FACS buffer, and left on ice until sorting. Activated T cells were sorted from
925 viable singlets from the CD3+HLA-DR++CD38++ gate. Cells were washed with 10X buffer, counted,
926 resuspended in a final concentration of 2,400 cells/uL, and immediately loaded into a 10X Genomics
927 Chromium Controller. Libraries were generated using the Chromium Next Gem Single Cell 5' Reagent Kit v2
928 (Dual Index) per the manufacturer's instructions with the addition of TCR and 5' Feature Barcode libraries.
929 Quality and quantity of libraries were measured on tapestation, qubit, and Bioanalyzer and sequenced on

930 Illumina NovaSeq 6000 with a sequencing target of 30,000 reads per cell for gene expression libraries and
931 10,000 reads per cell for TCR and feature barcode libraries.

932

933

934 **scRNA analysis of activated CD8 T cells**

935 Sequencing reads were aligned using Cell Ranger, with feature barcode and hashing antibody readouts.
936 Data QC was performed in Seurat using cutoffs described above. Additional doublets were removed based
937 on co-expression of multiple hashing antibodies. QC filtered objects from day 7, 10, and 14 libraries were
938 merged in Seurat. For the integrated object, we ran *FindVariableFeatures* to extract the top 10,000 variable
939 genes, performed z-score transformation using *ScaleData*, and then performed principal component analysis
940 using *RunPCA*. Finally, we ran *FindClusters* and *RunUMAP* function using the top 20 dimensions and
941 resolution 0.8, resulting in 15 clusters of activated T cells encompassing 26,723 cells from tonsil organoids.
942 Clusters were annotated based on protein readouts and top gene markers calculated using *FindAllMarkers* in
943 Seurat. Three CD8 T cell subsets were further analyzed for clonal sharing and tracking with *ex vivo* CD8 T
944 cell subsets from donor-matched blood and tonsils using Immunarch⁸⁴.

945

946 **Antigen-specific CD8 T cell analysis**

947 Antigen-specific CD8 T cells were profiled using Chromium Single Cell 5's Barcode Enabled Antigen
948 Mapping (BEAM) technology (10X Genomics). Peptides (Genscript, 100 uM each) from commonly circulating
949 viruses (EBV, CMV, influenza, SARS-CoV-2) or sources naive to most individuals (myelin, melanoma, YFV,
950 HIV) were prescreened on an HLA-A*02:01 donor per the manufacturer's instructions and analyzed by flow
951 cytometry. On experiment day, assemblies were generated for 15 peptides (**Table S4**), and a negative
952 control and incubated overnight at 4C. The following day, each of the assemblies (negative controls and
953 target peptides) was quenched for 15 minutes on ice per the manufacturer's instructions, pooled, and left on
954 ice in the dark. CD8 T cells from autologous blood and tonsils from 4 HLA-A*02:01 donors were enriched
955 using negative selection magnetic beads (CD4, CD19 depletion, Miltenyi Biotec). Roughly 1.5 e6 enriched
956 CD8s were stained with Fc block for 10 minutes on ice, mixed with pooled BEAM assemblies, and incubated
957 for an additional 15 minutes on ice in the dark. Samples were stained with a cocktail containing Zombie Aqua
958 (1/200), CD3-BV605 (1/50), CD19-APC (1/50), CD4-Pacific Blue (1/50), CD8-APC-Cy7 (1/100), CD45RA
959 (1/50) prepared in FACS buffer. The staining cocktail also contained a universal cocktail of oligo-tagged
960 antibodies (**Table S2**) and unique hashing antibodies (BioLegend) for sample barcoding. Samples were
961 incubated for 30 minutes on ice in the dark, washed twice with FACS buffer, and sorted on FACSaria Fusion
962 (BD Biosciences). CD19-CD3+CD4-CD8+ PE+ cells were sorted into 10X buffer, washed twice, resuspended
963 in 25 uL, and loaded on the 10X Genomics Chromium Controller. Libraries were generated using the
964 Chromium Next Gem Single Cell 5' Reagent Kit v2 (Dual Index) per the manufacturer's instructions with the
965 addition of TCR, BEAM, and 5' Feature Barcode libraries. Quality and quantity of libraries were measured on
966 tapestation, qubit, and Bioanalyzer and sequenced on Illumina NovaSeq 6000 with a sequencing target of
967 30,000 reads per cell for gene expression libraries and 10,000 reads per cell for TCR, BEAM, and feature

968 barcode libraries. Data were aligned and analyzed as described for activated T cells, but with the
969 incorporation of BEAM libraries in Cell Ranger multi alignment, which provided antigen-specificity scores for
970 each cell after correcting for signals from negative control peptide. Only cells with a single dominant peptide
971 specificity score (cutoff 0.5) were included in downstream analyses.

972 **SUPPLEMENTAL INFORMATION**

973 Figure S1: T cell profiles from blood and tonsils.

974 Figure S2: Donor heterogeneity in T cell profiles.

975 Figure S3: Phenotypes of largest T cell clones in individual donors.

976 Figure S4: Clonal mixing between blood and tonsils.

977 Figure S5: Distribution of shared clones between blood and tonsils.

978 Figure S6: Diversity estimates of human blood and tissue TCR repertoire.

979 Figure S7: Granular annotations of T cell clusters.

980 Figure S8: Clonal origins of CD8 T cells responding to influenza vaccination/infection.

981 Figure S9: Predicted specificities of expanded CD8 T cell clones.

982 Figure S10: Phenotype and clonal distribution of Ag+ CD8 T cells in blood and tonsils.

983

984 Table S1: Donor characteristics and exposure history.

985 Table S2: List of TotalSeq-C antibodies used in this study.

986 Table S3: Top gene and protein markers across clusters at annotation level 4.

987 Table S4: List of peptides used in multiplexed class I tetramer experiment.

988

REFERENCES

989

1. Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S.P., Aleksander, I., and Mak, T.W. (1984). A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* 308, 145–149. <https://doi.org/10.1038/308145a0>.
2. Davis, M.M., and Bjorkman, P.J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature* 334, 395–402. <https://doi.org/10.1038/334395a0>.
3. Sender, R., Weiss, Y., Navon, Y., Milo, I., Azulay, N., Keren, L., Fuchs, S., Ben-Zvi, D., Noor, E., and Milo, R. (2023). The total mass, number, and distribution of immune cells in the human body. *Proc. Natl. Acad. Sci.* 120, e2308511120. <https://doi.org/10.1073/pnas.2308511120>.
4. Masopust, D., Choo, D., Vezys, V., Wherry, E.J., Duraiswamy, J., Akondy, R., Wang, J., Casey, K.A., Barber, D.L., Kawamura, K.S., et al. (2010). Dynamic T cell migration program provides resident memory within intestinal epithelium. *J. Exp. Med.* 207, 553–564. <https://doi.org/10.1084/jem.20090858>.
5. Okhrimenko, A., Grün, J.R., Westendorf, K., Fang, Z., Reinke, S., Von Roth, P., Wassilew, G., Kühl, A.A., Kudernatsch, R., Demski, S., et al. (2014). Human memory T cells from the bone marrow are resting and maintain long-lasting systemic memory. *Proc. Natl. Acad. Sci.* 111, 9229–9234. <https://doi.org/10.1073/pnas.1318731111>.
6. Turner, D.L., Bickham, K.L., Thome, J.J., Kim, C.Y., D'Ovidio, F., Wherry, E.J., and Farber, D.L. (2014). Lung niches for the generation and maintenance of tissue-resident memory T cells. *Mucosal Immunol.* 7, 501–510. <https://doi.org/10.1038/mi.2013.67>.
7. Cheuk, S., Schlums, H., Gallais Sérézal, I., Martini, E., Chiang, S.C., Marquardt, N., Gibbs, A., Detlofsson, E., Introini, A., Forkel, M., et al. (2017). CD49a Expression Defines Tissue-Resident CD8 + T Cells Poised for Cytotoxic Function in Human Skin. *Immunity* 46, 287–300. <https://doi.org/10.1016/j.immuni.2017.01.009>.
8. Farber, D.L. (2021). Tissues, not blood, are where immune cells function. *Nature* 593, 506–509. <https://doi.org/10.1038/d41586-021-01396-y>.
9. Kumar, B.V., Ma, W., Miron, M., Granot, T., Guyer, R.S., Carpenter, D.J., Senda, T., Sun, X., Ho, S.-H., Lerner, H., et al. (2017). Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Rep.* 20, 2921–2934. <https://doi.org/10.1016/j.celrep.2017.08.078>.
10. Szabo, P.A., Miron, M., and Farber, D.L. (2019). Location, location, location: Tissue resident memory T cells in mice and humans. *Sci. Immunol.* 4, eaas9673. <https://doi.org/10.1126/sciimmunol.aas9673>.
11. Weisberg, S.P., Carpenter, D.J., Chait, M., Dogra, P., Gartrell-Corrado, R.D., Chen, A.X., Campbell, S., Liu, W., Saraf, P., Snyder, M.E., et al. (2019). Tissue-Resident Memory T Cells Mediate Immune Homeostasis in the Human Pancreas through the PD-1/PD-L1 Pathway. *Cell Rep.* 29, 3916–3932.e5. <https://doi.org/10.1016/j.celrep.2019.11.056>.
12. Breitfeld, D., Ohl, L., Kremmer, E., Ellwart, J., Sallusto, F., Lipp, M., and Förster, R. (2000). Follicular B Helper T Cells Express Cxc Chemokine Receptor 5, Localize to B Cell Follicles, and Support Immunoglobulin Production. *J. Exp. Med.* 192, 1545–1552. <https://doi.org/10.1084/jem.192.11.1545>.
13. Schaerli, P., Willimann, K., Lang, A.B., Lipp, M., Loetscher, P., and Moser, B. (2000). Cxc Chemokine Receptor 5 Expression Defines Follicular Homing T Cells with B Cell Helper Function. *J. Exp. Med.* 192, 1553–1562. <https://doi.org/10.1084/jem.192.11.1553>.
14. Yu, D., Rao, S., Tsai, L.M., Lee, S.K., He, Y., Sutcliffe, E.L., Srivastava, M., Linterman, M., Zheng, L., Simpson, N., et al. (2009). The Transcriptional Repressor Bcl-6 Directs T Follicular Helper Cell Lineage Commitment. *Immunity* 31, 457–468. <https://doi.org/10.1016/j.immuni.2009.07.002>.
15. Connors, T.J., Matsumoto, R., Verma, S., Szabo, P.A., Guyer, R., Gray, J., Wang, Z., Thapa, P., Dogra, P., Poon, M.M.L., et al. (2023). Site-specific development and progressive maturation of human tissue-resident memory T cells over infancy and childhood. *Immunity* 56, 1894–1909.e5. <https://doi.org/10.1016/j.immuni.2023.06.008>.
16. Stemberger, C., Huster, K.M., Koffler, M., Anderl, F., Schiemann, M., Wagner, H., and Busch, D.H. (2007). A Single Naive CD8+ T Cell Precursor Can Develop into Diverse Effector and Memory Subsets. *Immunity* 27, 985–997. <https://doi.org/10.1016/j.immuni.2007.10.012>.
17. Gerlach, C., Van Heijst, J.W.J., Swart, E., Sie, D., Armstrong, N., Kerkhoven, R.M., Zehn, D., Bevan, M.J., Schepers, K., and Schumacher, T.N.M. (2010). One naive T cell, multiple fates in CD8+ T cell differentiation. *J. Exp. Med.* 207, 1235–1246. <https://doi.org/10.1084/jem.20091175>.
18. Tubo, N.J., Pagán, A.J., Taylor, J.J., Nelson, R.W., Linehan, J.L., Ertelt, J.M., Huseby, E.S., Way, S.S.,

044 and Jenkins, M.K. (2013). Single Naive CD4+ T Cells from a Diverse Repertoire Produce Different
045 Effector Cell Types during Infection. *Cell* 153, 785–796. <https://doi.org/10.1016/j.cell.2013.04.007>.

046 19. Khatun, A., Kasmani, M.Y., Zander, R., Schauder, D.M., Snook, J.P., Shen, J., Wu, X., Burns, R., Chen,
047 Y.-G., Lin, C.-W., et al. (2021). Single-cell lineage mapping of a diverse virus-specific naive CD4 T cell
048 repertoire. *J. Exp. Med.* 218, e20200650. <https://doi.org/10.1084/jem.20200650>.

049 20. Prokhnevskaya, N., Cardenas, M.A., Valanparambil, R.M., Sobierajska, E., Barwick, B.G., Jansen, C.,
050 Reyes Moon, A., Gregorova, P., delBalzo, L., Greenwald, R., et al. (2023). CD8+ T cell activation in
051 cancer comprises an initial activation phase in lymph nodes followed by effector differentiation within the
052 tumor. *Immunity* 56, 107–124.e5. <https://doi.org/10.1016/j.immuni.2022.12.002>.

053 21. Krishna, C., Chowell, D., Gönen, M., Elhanati, Y., and Chan, T.A. (2020). Genetic and environmental
054 determinants of human TCR repertoire diversity. *Immun. Ageing* 17, 26. <https://doi.org/10.1186/s12979-020-00195-9>.

055 22. Naylor, K., Li, G., Vallejo, A.N., Lee, W.-W., Koetz, K., Bryl, E., Witkowski, J., Fulbright, J., Weyand,
056 C.M., and Goronzy, J.J. (2005). The Influence of Age on T Cell Generation and TCR Diversity. *J. Immunol.* 174, 7446–7452. <https://doi.org/10.4049/jimmunol.174.11.7446>.

057 23. Qi, Q., Liu, Y., Cheng, Y., Glanville, J., Zhang, D., Lee, J.-Y., Olshen, R.A., Weyand, C.M., Boyd, S.D.,
058 and Goronzy, J.J. (2014). Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl. Acad. Sci.* 111, 13139–13144. <https://doi.org/10.1073/pnas.1409155111>.

059 24. Nikolich-Žugich, J. (2018). The twilight of immunity: emerging concepts in aging of the immune system.
060 *Nat. Immunol.* 19, 10–19. <https://doi.org/10.1038/s41590-017-0006-x>.

061 25. Sun, X., Nguyen, T., Achour, A., Ko, A., Cifello, J., Ling, C., Sharma, J., Hiroi, T., Zhang, Y., Chia, C.W.,
062 et al. (2022). Longitudinal analysis reveals age-related changes in the T cell receptor repertoire of human
063 T cell subsets. *J. Clin. Invest.* 132, e158122. <https://doi.org/10.1172/JCI158122>.

064 26. Laydon, D.J., Bangham, C.R.M., and Asquith, B. (2015). Estimating T-cell repertoire diversity: limitations
065 of classical estimators and a new approach. *Philos. Trans. R. Soc. B Biol. Sci.* 370, 20140291.
066 <https://doi.org/10.1098/rstb.2014.0291>.

067 27. Cossarizza, A., Ortolani, C., Paganelli, R., Barbieri, D., Monti, D., Sansoni, P., Fagiolo, U., Castellani, G.,
068 Bersani, F., Londei, M., et al. (1996). CD45 isoforms expression on CD4+ and CD8+ T cells throughout
069 life, from newborns to centenarians: implications for T cell memory. *Mech. Ageing Dev.* 86, 173–195.
070 [https://doi.org/10.1016/0047-6374\(95\)01691-0](https://doi.org/10.1016/0047-6374(95)01691-0).

071 28. Farber, D.L., Yudanin, N.A., and Restifo, N.P. (2014). Human memory T cells: generation,
072 compartmentalization and homeostasis. *Nat. Rev. Immunol.* 14, 24–35. <https://doi.org/10.1038/nri3567>.

073 29. Gaimann, M.U., Nguyen, M., Desponts, J., and Mayer, A. (2020). Early life imprints the hierarchy of T
074 cell clone sizes. *eLife* 9, e61639. <https://doi.org/10.7554/eLife.61639>.

075 30. Messaoudi, I., Patiño, J.A.G., Dyall, R., LeMaoult, J., and Nikolich-Žugich, J. (2002). Direct Link Between
076 *mhc* Polymorphism, T Cell Avidity, and Diversity in Immune Defense. *Science* 298, 1797–1800.
077 <https://doi.org/10.1126/science.1076064>.

078 31. Kim, S.-K., Cornberg, M., Wang, X.Z., Chen, H.D., Selin, L.K., and Welsh, R.M. (2005). Private
079 specificities of CD8 T cell responses control patterns of heterologous immunity. *J. Exp. Med.* 201, 523–
080 533. <https://doi.org/10.1084/jem.20041337>.

081 32. Turner, S.J., La Gruta, N.L., Kedzierska, K., Thomas, P.G., and Doherty, P.C. (2009). Functional
082 implications of T cell receptor diversity. *Curr. Opin. Immunol.* 21, 286–290.
083 <https://doi.org/10.1016/j.coib.2009.05.004>.

084 33. Arstila, T.P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., and Kourilsky, P. (1999). A Direct
085 Estimate of the Human $\alpha\beta$ T Cell Receptor Diversity. *Science* 286, 958–961.
086 <https://doi.org/10.1126/science.286.5441.958>.

087 34. De Greef, P.C., Oakes, T., Gerritsen, B., Ismail, M., Heather, J.M., Hermsen, R., Chain, B., and De Boer,
088 R.J. (2020). The naive T-cell receptor repertoire has an extremely broad distribution of clone sizes. *eLife*
089 9, e49900. <https://doi.org/10.7554/eLife.49900>.

090 35. Mora, T., and Walczak, A.M. (2019). How many different clonotypes do immune repertoires contain?
091 *Curr. Opin. Syst. Biol.* 18, 104–110. <https://doi.org/10.1016/j.coisb.2019.10.001>.

092 36. Mason, D. (1998). A very high level of crossreactivity is an essential feature of the T-cell receptor.
093 *Immunol. Today* 19, 395–404. [https://doi.org/10.1016/S0167-5699\(98\)01299-7](https://doi.org/10.1016/S0167-5699(98)01299-7).

094 37. Wooldridge, L., Ekeruche-Makinde, J., Van Den Berg, H.A., Skowron, A., Miles, J.J., Tan, M.P., Dolton,
095 G., Clement, M., Llewellyn-Lacey, S., Price, D.A., et al. (2012). A Single Autoimmune T Cell Receptor
096 Recognizes More Than a Million Different Peptides. *J. Biol. Chem.* 287, 1168–1177.

097 098 099

100 https://doi.org/10.1074/jbc.M111.289488.

101 38. Birnbaum, M.E., Mendoza, J.L., Sethi, D.K., Dong, S., Glanville, J., Dobbins, J., Özkan, E., Davis, M.M.,
102 Wucherpfennig, K.W., and Garcia, K.C. (2014). Deconstructing the Peptide-MHC Specificity of T Cell
103 Recognition. *Cell* 157, 1073–1087. <https://doi.org/10.1016/j.cell.2014.03.047>.

104 39. Sewell, A.K. (2012). Why must T cells be cross-reactive? *Nat. Rev. Immunol.* 12, 669–677.
105 <https://doi.org/10.1038/nri3279>.

106 40. Su, L.F., Kidd, B.A., Han, A., Kotzin, J.J., and Davis, M.M. (2013). Virus-Specific CD4+ Memory-
107 Phenotype T Cells Are Abundant in Unexposed Adults. *Immunity* 38, 373–383.
108 <https://doi.org/10.1016/j.jimmuni.2012.10.021>.

109 41. Yu, W., Jiang, N., Ebert, P.J.R., Kidd, B.A., Müller, S., Lund, P.J., Juang, J., Adachi, K., Tse, T.,
110 Birnbaum, M.E., et al. (2015). Clonal Deletion Prunes but Does Not Eliminate Self-Specific $\alpha\beta$ CD8+ T
111 Lymphocytes. *Immunity* 42, 929–941. <https://doi.org/10.1016/j.jimmuni.2015.05.001>.

112 42. Robins, H.S., Campregher, P.V., Srivastava, S.K., Wacher, A., Turtle, C.J., Kahsai, O., Riddell, S.R.,
113 Warren, E.H., and Carlson, C.S. (2009). Comprehensive assessment of T-cell receptor β -chain diversity
114 in $\alpha\beta$ T cells. *Blood* 114, 4099–4107. <https://doi.org/10.1182/blood-2009-04-217604>.

115 43. Keşmir, C., Borghans, J.A.M., and De Boer, R.J. (2000). Diversity of Human $\alpha\beta$ T Cell Receptors.
116 *Science* 288, 1135–1135. <https://doi.org/10.1126/science.288.5469.1135a>.

117 44. Mackay, L.K., and Kallies, A. (2017). Transcriptional Regulation of Tissue-Resident Lymphocytes. *Trends
118 Immunol.* 38, 94–103. <https://doi.org/10.1016/j.it.2016.11.004>.

119 45. Buquicchio, F.A., Fonseca, R., Yan, P.K., Wang, F., Evrard, M., Obers, A., Gutierrez, J.C., Raposo, C.J.,
120 Belk, J.A., Daniel, B., et al. (2024). Distinct epigenomic landscapes underlie tissue-specific memory T cell
121 differentiation. *Immunity*, S1074761324003200. <https://doi.org/10.1016/j.jimmuni.2024.06.014>.

122 46. Poon, M.M.L., Byington, E., Meng, W., Kubota, M., Matsumoto, R., Grifoni, A., Weiskopf, D., Dogra, P.,
123 Lam, N., Szabo, P.A., et al. (2021). Heterogeneity of human anti-viral immunity shaped by virus, tissue,
124 age, and sex. *Cell Rep.* 37, 110071. <https://doi.org/10.1016/j.celrep.2021.110071>.

125 47. Poon, M.M.L., Caron, D.P., Wang, Z., Wells, S.B., Chen, D., Meng, W., Szabo, P.A., Lam, N., Kubota,
126 M., Matsumoto, R., et al. (2023). Tissue adaptation and clonal segregation of human memory T cells in
127 barrier sites. *Nat. Immunol.* 24, 309–319. <https://doi.org/10.1038/s41590-022-01395-9>.

128 48. Cvijović, I., Swift, M., and Quake, S.R. (2023). Long-term B cell memory emerges at uniform relative
129 rates in the human immune response. Preprint, <https://doi.org/10.1101/2023.11.27.568934>
130 <https://doi.org/10.1101/2023.11.27.568934>.

131 49. Kaech, S.M., Wherry, E.J., and Ahmed, R. (2002). Effector and memory T-cell differentiation: implications
132 for vaccine development. *Nat. Rev. Immunol.* 2, 251–262. <https://doi.org/10.1038/nri778>.

133 50. Foulds, K.E., Zenewicz, L.A., Shedlock, D.J., Jiang, J., Troy, A.E., and Shen, H. (2002). Cutting Edge:
134 CD4 and CD8 T Cells Are Intrinsically Different in Their Proliferative Responses. *J. Immunol.* 168, 1528–
135 1532. <https://doi.org/10.4049/jimmunol.168.4.1528>.

136 51. Chiu, C., Wang, Y., Walther, B.A., and Chao, A. (2014). An improved nonparametric lower bound of
137 species richness via a modified good-turing frequency formula. *Biometrics* 70, 671–682.
138 <https://doi.org/10.1111/biom.12200>.

139 52. Becattini, S., Latorre, D., Mele, F., Foglierini, M., De Gregorio, C., Cassotta, A., Fernandez, B.,
140 Kelderman, S., Schumacher, T.N., Corti, D., et al. (2015). Functional heterogeneity of human memory
141 CD4⁺ T cell clones primed by pathogens or vaccines. *Science* 347, 400–406.
142 <https://doi.org/10.1126/science.1260668>.

143 53. Zemmour, D., Zilionis, R., Kiner, E., Klein, A.M., Mathis, D., and Benoist, C. (2018). Single-cell gene
144 expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat. Immunol.* 19,
145 291–301. <https://doi.org/10.1038/s41590-018-0051-0>.

146 54. Schattgen, S.A., Guion, K., Crawford, J.C., Souquette, A., Barrio, A.M., Stubbington, M.J.T., Thomas,
147 P.G., and Bradley, P. (2022). Integrating T cell receptor sequences and transcriptional profiles by
148 clonotype neighbor graph analysis (CoNGA). *Nat. Biotechnol.* 40, 54–63. [https://doi.org/10.1038/s41587-021-00989-2](https://doi.org/10.1038/s41587-
149 021-00989-2).

150 55. Lagattuta, K.A., Kang, J.B., Nathan, A., Pauken, K.E., Jonsson, A.H., Rao, D.A., Sharpe, A.H., Ishigaki,
151 K., and Raychaudhuri, S. (2022). Repertoire analyses reveal T cell antigen receptor sequence features
152 that influence T cell fate. *Nat. Immunol.* 23, 446–457. <https://doi.org/10.1038/s41590-022-01129-x>.

153 56. Burton, O., Bricard, O., Tareen, S., Gergelits, V., Andrews, S., Roca, C.P., Whyte, C., Junius, S., Brajic,
154 A., Pasciuto, E., et al. (2023). The tissue-resident regulatory T cell pool is shaped by transient multi-
155 tissue migration and a conserved residency program. Preprint,

156 https://doi.org/10.1101/2023.08.14.553196 https://doi.org/10.1101/2023.08.14.553196.

157 57. He, X., Holmes, T.H., Mahmood, K., Kemble, G.W., Dekker, C.L., Arvin, A.M., and Greenberg, H.B.
158 (2008). Phenotypic Changes in Influenza-Specific CD8⁺ T Cells after Immunization of Children and
159 Adults with Influenza Vaccines. *J. Infect. Dis.* 197, 803–811. <https://doi.org/10.1086/528804>.

160 58. Kuhn, R., Sandu, I., Agrafiotis, A., Hong, K.-L., Shlesinger, D., Neimeier, D., Merkler, D., Oxenius, A.,
161 Reddy, S.T., and Yermanos, A. (2022). Clonally Expanded Virus-Specific CD8 T Cells Acquire Diverse
162 Transcriptional Phenotypes During Acute, Chronic, and Latent Infections. *Front. Immunol.* 13, 782441.
163 <https://doi.org/10.3389/fimmu.2022.782441>.

164 59. Schmidt, F., Fields, H.F., Purwanti, Y., Milojkovic, A., Salim, S., Wu, K.X., Simoni, Y., Vitiello, A.,
165 MacLeod, D.T., Nardin, A., et al. (2023). In-depth analysis of human virus-specific CD8+ T cells
166 delineates unique phenotypic signatures for T cell specificity prediction. *Cell Rep.* 42, 113250.
167 <https://doi.org/10.1016/j.celrep.2023.113250>.

168 60. Chen, D.G., Xie, J., Su, Y., and Heath, J.R. (2023). T cell receptor sequences are the dominant factor
169 contributing to the phenotype of CD8+ T cells with specificities against immunogenic viral antigens. *Cell
170 Rep.* 42, 113279. <https://doi.org/10.1016/j.celrep.2023.113279>.

171 61. Shugay, M., Bagaev, D.V., Zvyagin, I.V., Vroomans, R.M., Crawford, J.C., Dolton, G., Komech, E.A.,
172 Sycheva, A.L., Koneva, A.E., Egorov, E.S., et al. (2018). VDJdb: a curated database of T-cell receptor
173 sequences with known antigen specificity. *Nucleic Acids Res.* 46, D419–D427.
174 <https://doi.org/10.1093/nar/gkx760>.

175 62. Khan, N., Shariff, N., Cobbold, M., Bruton, R., Ainsworth, J.A., Sinclair, A.J., Nayak, L., and Moss, P.A.H.
176 (2002). Cytomegalovirus Seropositivity Drives the CD8 T Cell Repertoire Toward Greater Clonality in
177 Healthy Elderly Individuals. *J. Immunol.* 169, 1984–1992. <https://doi.org/10.4049/jimmunol.169.4.1984>.

178 63. Klarenbeek, P.L., Remmerswaal, E.B.M., Ten Berge, I.J.M., Doorenspleet, M.E., Van Schaik, B.D.C.,
179 Esveldt, R.E.E., Koch, S.D., Ten Brinke, A., Van Kampen, A.H.C., Bemelman, F.J., et al. (2012). Deep
180 Sequencing of Antiviral T-Cell Responses to HCMV and EBV in Humans Reveals a Stable Repertoire
181 That Is Maintained for Many Years. *PLoS Pathog.* 8, e1002889.
182 <https://doi.org/10.1371/journal.ppat.1002889>.

183 64. Massoni-Badosa, R., Aguilar-Fernández, S., Nieto, J.C., Soler-Vila, P., Elosua-Bayes, M., Marchese, D.,
184 Kulis, M., Vilas-Zornoza, A., Bühler, M.M., Rashmi, S., et al. (2024). An atlas of cells in the human tonsil.
185 *Immunity* 57, 379-399.e18. <https://doi.org/10.1016/j.immuni.2024.01.006>.

186 65. Miron, M., Meng, W., Rosenfeld, A.M., Dvorkin, S., Poon, M.M.L., Lam, N., Kumar, B.V., Louzoun, Y.,
187 Luning Prak, E.T., and Farber, D.L. (2021). Maintenance of the human memory T cell repertoire by
188 subset and tissue site. *Genome Med.* 13, 100. <https://doi.org/10.1186/s13073-021-00918-7>.

189 66. Kumar, S., Basto, A.P., Ribeiro, F., Almeida, S.C.P., Campos, P., Peres, C., Pulvirenti, N., Al-Khalidi, S.,
190 Kilbey, A., Tosello, J., et al. (2024). Specialized Tfh cell subsets driving type-1 and type-2 humoral
191 responses in lymphoid tissue. *Cell Discov.* 10, 64. <https://doi.org/10.1038/s41421-024-00681-0>.

192 67. Jimeno, R., Lebrusant-Fernandez, M., Margreitter, C., Lucas, B., Veerapen, N., Kelly, G., Besra, G.S.,
193 Fraternali, F., Spencer, J., Anderson, G., et al. (2019). Tissue-specific shaping of the TCR repertoire and
194 antigen specificity of iNKT cells. *eLife* 8, e51663. <https://doi.org/10.7554/eLife.51663>.

195 68. Goronzy, J.J., Fang, F., Cavanagh, M.M., Qi, Q., and Weyand, C.M. (2015). Naive T Cell Maintenance
196 and Function in Human Aging. *J. Immunol.* 194, 4073–4080. <https://doi.org/10.4049/jimmunol.1500046>.

197 69. Milighetti, M., Peng, Y., Tan, C., Mark, M., Nageswaran, G., Byrne, S., Ronel, T., Peacock, T., Mayer, A.,
198 Chandran, A., et al. (2023). Large clones of pre-existing T cells drive early immunity against SARS-CoV-
199 2 and LCMV infection. *iScience* 26, 106937. <https://doi.org/10.1016/j.isci.2023.106937>.

200 70. Teijaro, J.R., Turner, D., Pham, Q., Wherry, E.J., Lefrançois, L., and Farber, D.L. (2011). Cutting Edge:
201 Tissue-Selective Lung Memory CD4 T Cells Mediate Optimal Protection to Respiratory Virus Infection. *J.
202 Immunol.* 187, 5510–5514. <https://doi.org/10.4049/jimmunol.1102243>.

203 71. Jiang, X., Clark, R.A., Liu, L., Wagers, A.J., Fuhlbrigge, R.C., and Kupper, T.S. (2012). Skin infection
204 generates non-migratory memory CD8+ TRM cells providing global skin immunity. *Nature* 483, 227–231.
205 <https://doi.org/10.1038/nature10851>.

206 72. Masopust, D., Vezys, V., Usherwood, E.J., Cauley, L.S., Olson, S., Marzo, A.L., Ward, R.L., Woodland,
207 D.L., and Lefrançois, L. (2004). Activated Primary and Memory CD8 T Cells Migrate to Nonlymphoid
208 Tissues Regardless of Site of Activation or Tissue of Origin. *J. Immunol.* 172, 4875–4882.
209 <https://doi.org/10.4049/jimmunol.172.8.4875>.

210 73. Buggert, M., Vella, L.A., Nguyen, S., Wu, V.H., Chen, Z., Sekine, T., Perez-Potti, A., Maldini, C.R.,
211 Manne, S., Darko, S., et al. (2020). The Identity of Human Tissue-Emigrant CD8+ T Cells. *Cell* 183,

212 1946-1961.e15. <https://doi.org/10.1016/j.cell.2020.11.019>.

213 74. Kaech, S.M., and Ahmed, R. (2001). Memory CD8+ T cell differentiation: initial antigen encounter
214 triggers a developmental program in naïve cells. *Nat. Immunol.* 2, 415–422.
215 <https://doi.org/10.1038/87720>.

216 75. Chang, J.T., Palanivel, V.R., Kinjyo, I., Schambach, F., Intlekofer, A.M., Banerjee, A., Longworth, S.A.,
217 Vinup, K.E., Mrass, P., Oliaro, J., et al. (2007). Asymmetric T Lymphocyte Division in the Initiation of
218 Adaptive Immune Responses. *Science* 315, 1687–1691. <https://doi.org/10.1126/science.1139393>.

219 76. Buchholz, V.R., Flossdorf, M., Hensel, I., Kretschmer, L., Weissbrich, B., Gräf, P., Verschoor, A.,
220 Schiemann, M., Höfer, T., and Busch, D.H. (2013). Disparate Individual Fates Compose Robust CD8⁺ T
221 Cell Immunity. *Science* 340, 630–635. <https://doi.org/10.1126/science.1235454>.

222 77. Mensink, M., Verleng, L.J., Schrama, E., Janssen, G.M.C., Tjokrodirijo, R.T.N., Van Veelen, P.A., Jiang,
223 Q., Pascutti, M.F., Van Der Hoorn, M.-L., Eikmans, M., et al. (2024). Treg cells from human blood
224 differentiate into non-lymphoid tissue-resident effector cells upon TNFR2 costimulation. *JCI Insight*.
225 <https://doi.org/10.1172/jci.insight.172942>.

226 78. Brenna, E., Davydov, A.N., Ladell, K., McLaren, J.E., Bonaiuti, P., Metsger, M., Ramsden, J.D., Gilbert,
227 S.C., Lambe, T., Price, D.A., et al. (2020). CD4+ T Follicular Helper Cells in Human Tonsils and Blood
228 Are Clonally Convergent but Divergent from Non-Tfh CD4+ Cells. *Cell Rep.* 30, 137-152.e5.
229 <https://doi.org/10.1016/j.celrep.2019.12.016>.

230 79. Goronzy, J.J., Qi, Q., Olshen, R.A., and Weyand, C.M. (2015). High-throughput sequencing insights into
231 T-cell receptor repertoire diversity in aging. *Genome Med.* 7, 117. <https://doi.org/10.1186/s13073-015-0242-3>.

233 80. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell
234 RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* 8, 329-337.e4.
235 <https://doi.org/10.1016/j.cels.2019.03.003>.

236 81. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby,
237 C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573-3587.e29.
238 <https://doi.org/10.1016/j.cell.2021.04.048>.

239 82. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G.,
240 Madad, S., Fernandez-Granda, C., et al. (2024). Dictionary learning for integrative, multimodal and
241 scalable single-cell analysis. *Nat. Biotechnol.* 42, 293–304. <https://doi.org/10.1038/s41587-023-01767-y>.

242 83. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh,
243 P., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with
244 Harmony. *Nat. Methods* 16, 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>.

245 84. Aleksandr Popov, Maria Samokhina, Ivan Balashov, immunarch. bot, Eugene Rumynskiy, Vadim I.
246 Nazarov, gracecodeadventures, tsvvas, and Maksym Zarodniuk (2024). immunomind/immunarch:
247 Immunarch 0.9.1. Version 0.9.1 (Zenodo). <https://doi.org/10.5281/ZENODO.3367200>
248 <https://doi.org/10.5281/ZENODO.3367200>.

249 85. Mayer, A., and Callan, C.G. (2023). Measures of epitope binding degeneracy from T cell receptor
250 repertoires. *Proc. Natl. Acad. Sci.* 120, e2213264120. <https://doi.org/10.1073/pnas.2213264120>.

251 86. Tiffreau-Mayer, A. (2024). Unbiased estimation of sampling variance for Simpson's diversity index. *Phys.*
252 *Rev. E* 109, 064411. <https://doi.org/10.1103/PhysRevE.109.064411>.

253 87. Kastenschmidt, J.M., Sureshchandra, S., Jain, A., Hernandez-Davies, J.E., de Assis, R., Wagoner, Z.W.,
254 Sorn, A.M., Mitul, M.T., Benchorin, A.I., Levendosky, E., et al. (2023). Influenza vaccine format mediates
255 distinct cellular and antibody responses in human immune organoids. *Immunity* 56, 1910-1926.e7.
256 <https://doi.org/10.1016/j.immuni.2023.06.019>.

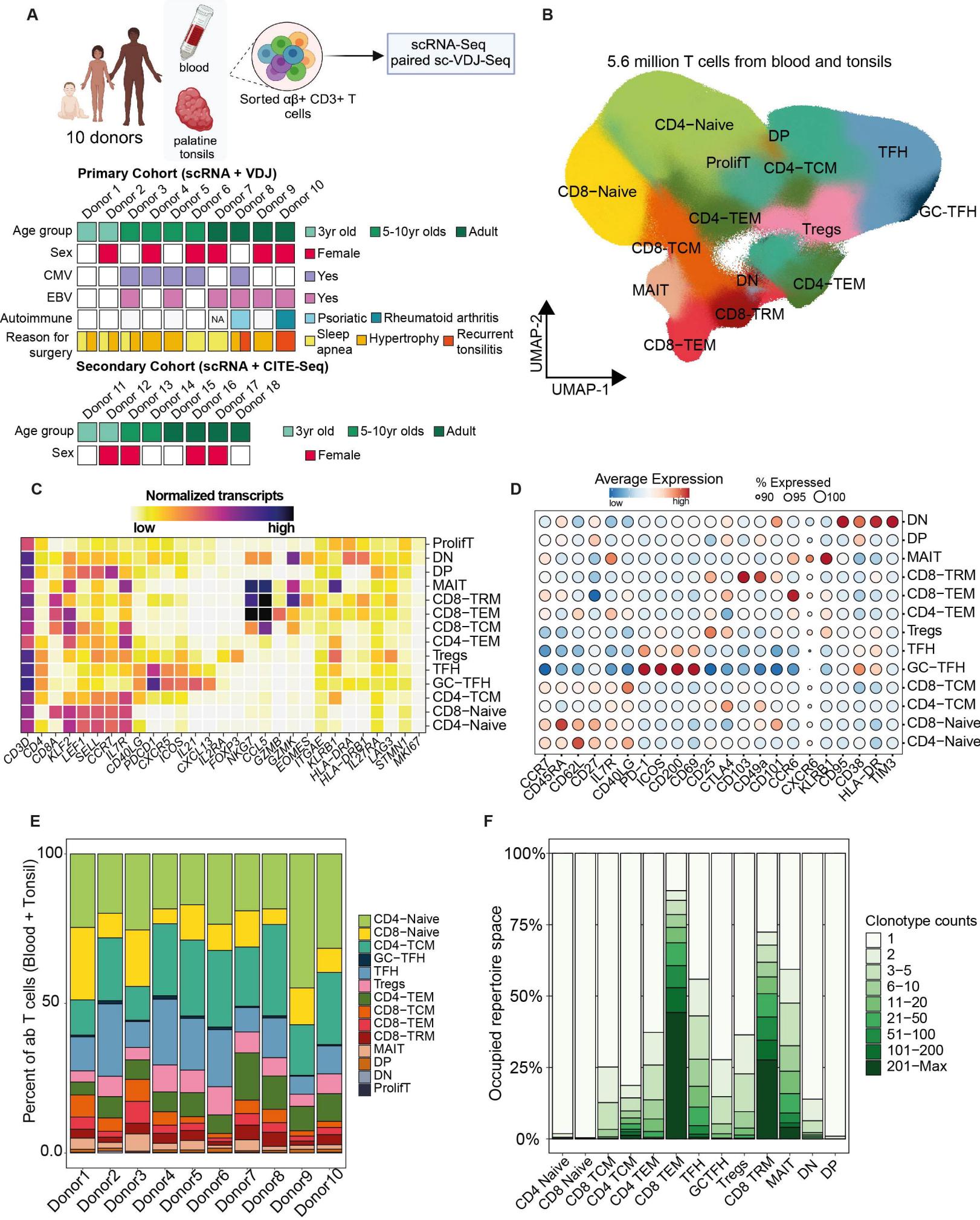
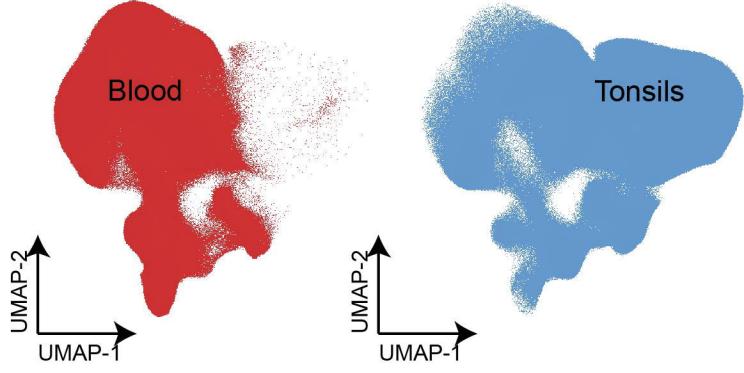
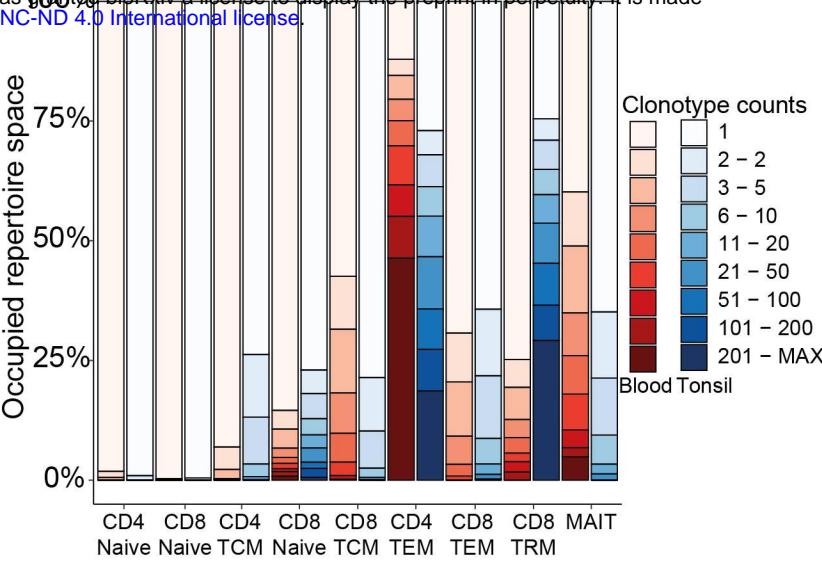


Figure 2 bioRxiv preprint doi: <https://doi.org/10.1101/2024.08.17.608295>; this version posted August 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

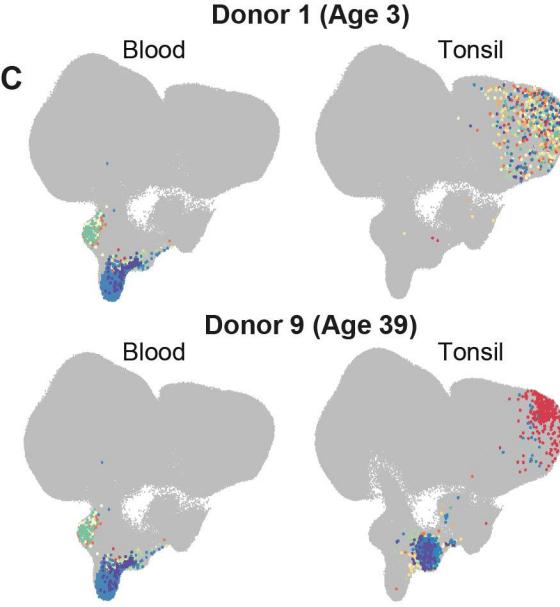
A



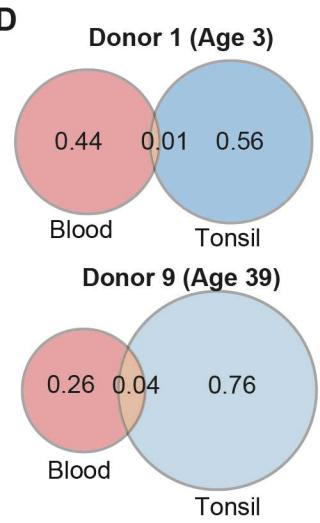
B



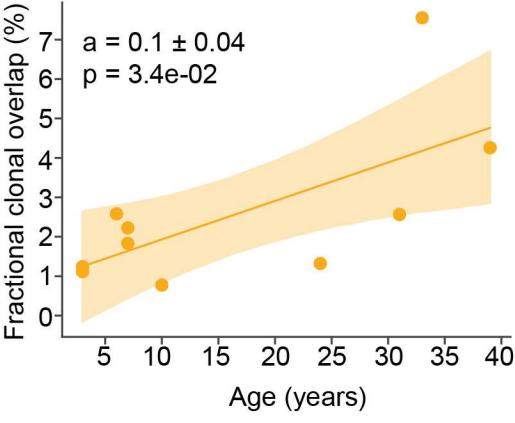
C



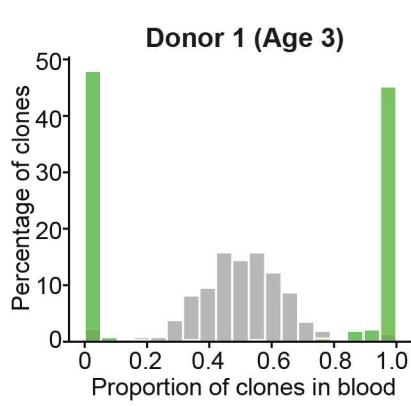
D



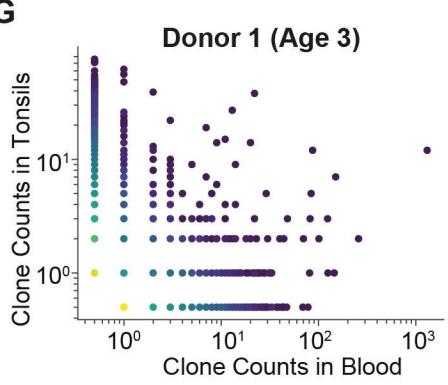
E



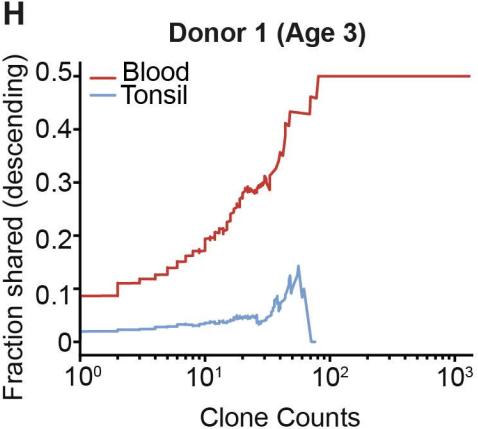
F



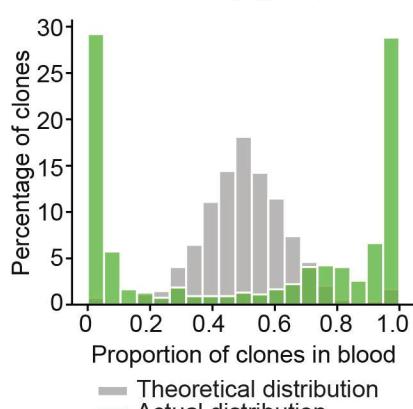
G



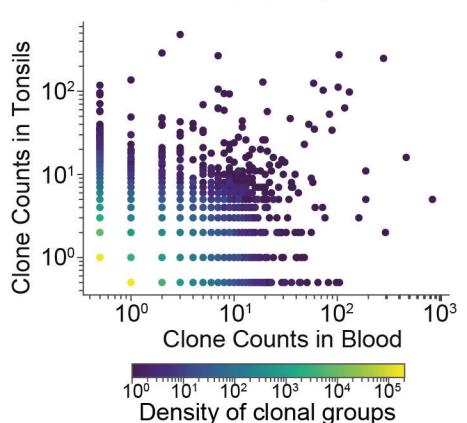
H



Donor 9 (Age 39)



Donor 9 (Age 39)



Donor 9 (Age 39)

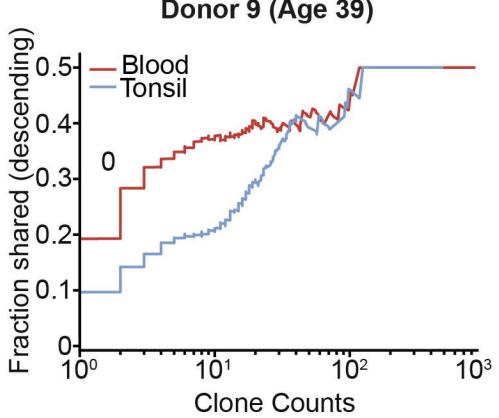


Figure 3

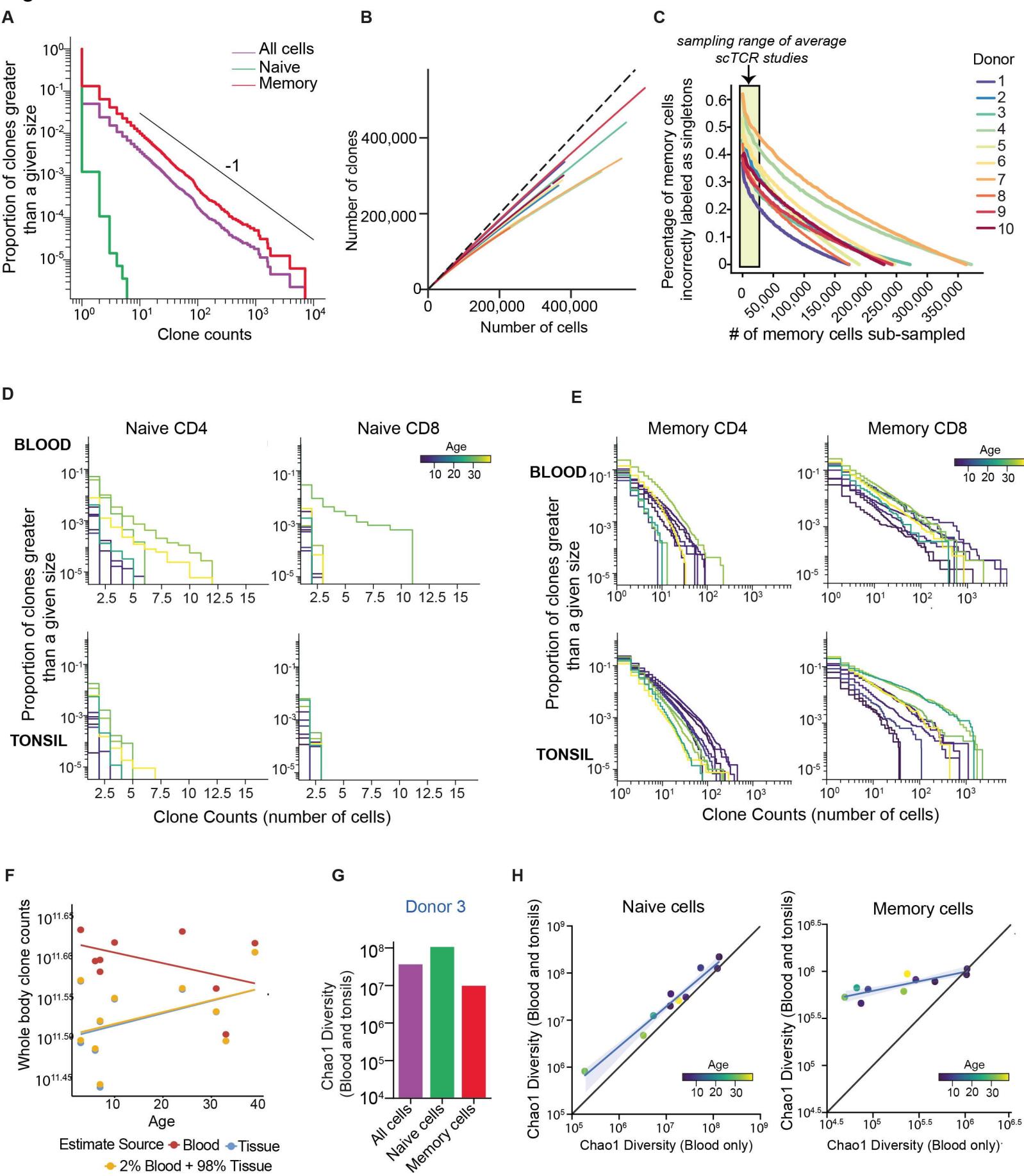


Figure 4

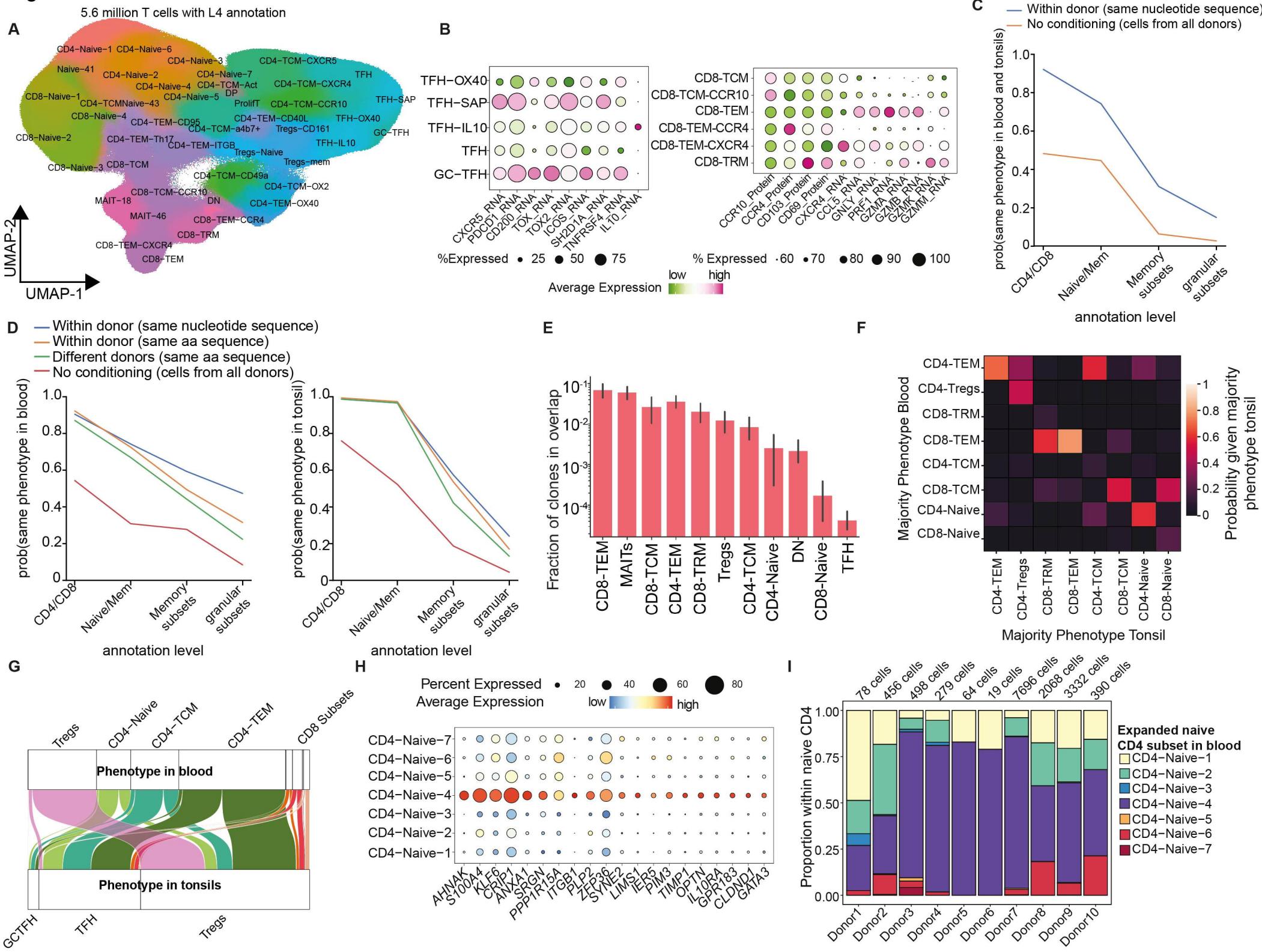
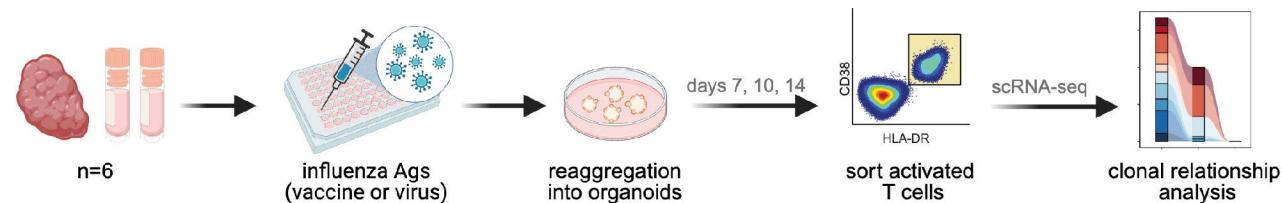
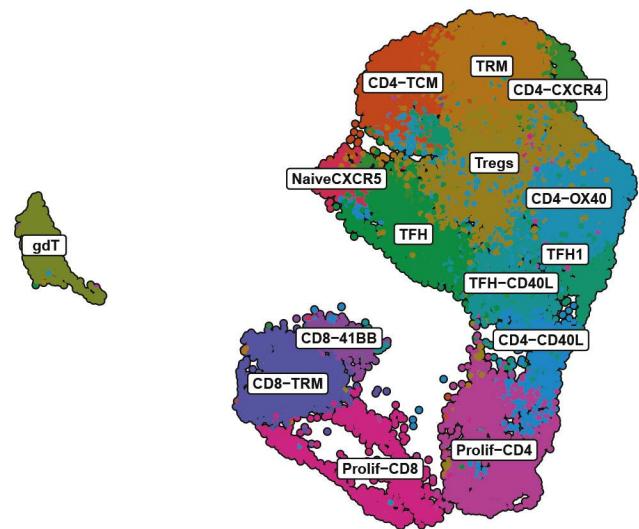


Figure 5

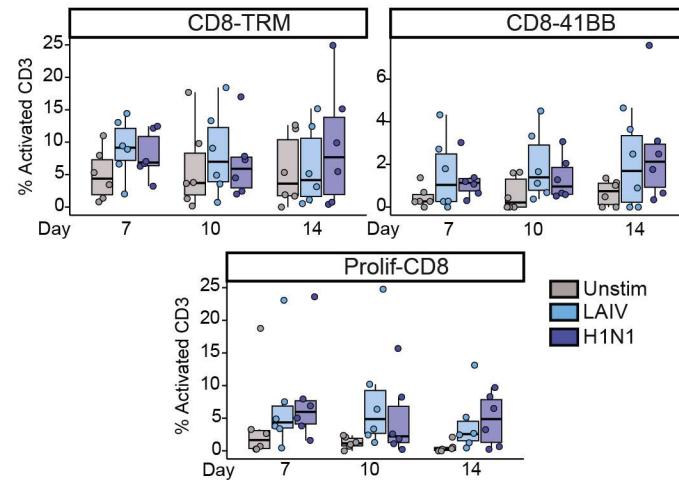
A



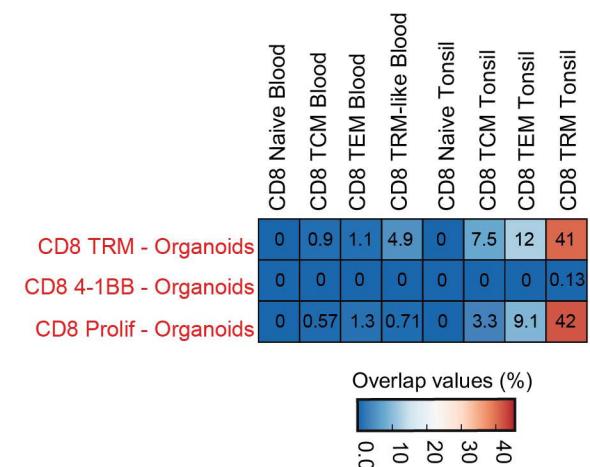
B



C



D



E

