

1
2
3
4
5
6
7
8
9

10 **Transcriptional control of T cell tissue adaptation and effector function in**
11 **infants and adults.**

12
13
14
15
16
17
18
19
20
21
22

23 Peter A. Szabo¹, Hanna M. Levitin², Thomas J. Connors³, David Chen², Jenny Jin⁴, Puspa Thapa¹,
24 Rebecca Guyer¹, Daniel P. Caron¹, Joshua I. Gray¹, Rei Matsumoto^{1,5}, Masaru Kubota^{1,5}, Maigan
25 Brusko⁶, Todd M. Brusko⁶, Donna L. Farber^{1,5,*}, Peter A. Sims^{2,7,*}

26
27
28
29
30
31

32 ¹ Department of Microbiology and Immunology, Columbia University Medical Center, New
33 York, NY 10032.

34 ² Department of Systems Biology, Columbia University Medical Center, New York, NY 10032.

35 ³ Department of Pediatrics, Columbia University Medical Center, New York, NY 10032.

36 ⁴ Medical Scientist Training Program, Columbia University Medical Center, New York, NY
37 10032.

38 ⁵ Department of Surgery, Columbia University Medical Center, New York, NY 10032.

39 ⁶ Department of Pathology, Immunology, and Laboratory Medicine, University of Florida,
40 Gainesville, FL 32611.

41 ⁷ Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center,
42 New York, NY 10032.

43
44

45 *Correspondence: df2396@cumc.columbia.edu, pas2182@cumc.columbia.edu

46 **ABSTRACT**

47 The first years of life are essential for the development of memory T cells, which rapidly
48 populate the body's diverse tissue sites during infancy. However, the degree to which tissue
49 memory T cell responses in early life reflect those during adulthood is unclear. Here, we use single
50 cell RNA-sequencing of resting and *ex vivo* activated T cells from lymphoid and mucosal tissues
51 of infant (aged 2-9 months) and adult (aged 40-65 years) human organ donors to dissect the
52 transcriptional programming of memory T cells over age. Infant memory T cells demonstrate a
53 unique stem-like transcriptional profile and tissue adaptation program, yet exhibit reduced
54 activation capacity and effector function relative to adults. Using CRISPR-Cas9 knockdown, we
55 define Helios (*IKZF2*) as a critical transcriptional regulator of the infant-specific tissue adaptation
56 program and restricted effector state. Our findings reveal key transcriptional mechanisms that
57 control tissue T cell fate and function in early life.

58

59 The maturation of adaptive immunity in early life is essential for the establishment of
60 protective immune memory that can last a lifetime. The first years of age represent an intense
61 period of exposure to novel antigens that generate memory T cells required for orchestrating
62 acquired immunity and vaccine-induced protection against infectious disease. However, infants
63 exhibit reduced or diminished responses to ubiquitous pathogens and vaccines relative to adults¹.
64 While recent studies uncovered distinct activation pathways and effector profiles for infant and
65 adult T cells^{2,3}, the underlying mechanisms for this discrepancy in functional capacity remain
66 unknown. Understanding the interplay between the maturation and function of T cells in infancy
67 is necessary for advancing vaccine strategies and immunotherapies targeted to early life.

68 T cell responses in infancy are distinct from those in adulthood. While initial studies
69 describe infant T cells as intrinsically impaired in effector functions relative to adults, an updated
70 paradigm holds that infant T cells exhibit distinct effector responses that are adapted to the unique
71 demands of early life⁴. Infant naïve T cells preferentially produce T helper type 2 (TH2) cytokines
72 or chemokines (e.g., CXCL8) instead of pro-inflammatory (TH1) cytokines upon activation⁵⁻⁸. We
73 and others previously showed that infant naïve T cells are more sensitive to T cell receptor (TCR)-
74 stimulation, exhibit augmented proliferative responses, and demonstrate biased differentiation
75 towards short-lived effector cells compared to adults⁹⁻¹². Features of this infant-specific response
76 may be traced to transcriptional programming or distinct progenitors within the naïve T cell pool,
77 predisposing cells towards mounting effector responses during infections at the expense of forming
78 memory^{3,10,11,13}. However, the mechanisms governing effector responses of memory T cell
79 populations that are formed during infancy are not well understood.

80 We previously showed that the generation of T cell memory in early life begins in tissues,
81 particularly in mucosal sites such as the lungs and intestines that represent the frontlines of antigen

82 exposure, while the vast majority of T cells in the blood remain naïve¹⁴⁻¹⁶. These infant tissue
83 memory T cells predominantly exhibit an effector memory (TEM) phenotype with markers of
84 tissue residency (e.g., CD69 and CD103) but show decreased expression of tissue
85 homing/adhesion molecules and reduced production of inflammatory mediators upon stimulation
86 compared to older children and adults^{16,17}. We recently defined transcriptional profiles of tissue
87 memory T cells during infancy and childhood including expression of transcription factors (TFs)
88 associated with T cell development¹⁶. How these and other transcriptional regulators control
89 maturation and effector responses of tissue memory T cells during infancy is not known.

90 Here, we use single cell RNA-sequencing (scRNA-seq) of resting and *ex-vivo* activated T
91 cells from lymphoid and mucosal tissues of infant (aged 2-9 months) and adult (aged 40-65 years)
92 human organ donors to dissect the transcriptional programming of tissue memory T cells in early
93 life. We apply a consensus-implementation of single cell Hierarchical Poisson Factorization
94 (consensus-scHPF)¹⁸ to define transcriptional states associated with T cell activation, effector
95 function, and tissue adaptation across tissues in infants and adults. We find that relative to adults,
96 infant tissue memory T cells demonstrate a stem-like transcriptional profile (*TCF1*, *LEF1*, *SOX4*),
97 yet exhibit restricted transcriptional responses to TCR-mediated stimulation. We elucidate unique
98 tissue-associated transcriptional states between infant and adult tissue memory T cells and uncover
99 drivers of these programs by gene regulatory network reconstruction. Using CRISPR-Cas9
100 knockdown in primary tissue T cells, we define Helios (*IKZF2*) as a critical regulator of an infant-
101 specific tissue adaptation program and demonstrate that Helios also restricts infant T cell effector
102 function after stimulation. Together, our results reveal key mechanisms by which age impacts T
103 cell fate and function, with important implications for targeting T cell responses during the
104 formative years of infancy.

105

106 RESULTS

107 A single cell transcriptional map of T cell activation across infant and adult tissues.

108 To define the transcriptional programming of T cells across tissues in early life, we
109 performed scRNA-seq on T cells from lymphoid and mucosal sites in infants (2-9 months old) and
110 adults (40-65 years old) (**Supplementary Table 1**). Blood and tissues were obtained from
111 deceased organ donors at the time of life-saving transplantation, including lymphoid organs (bone
112 marrow, spleen, tonsil, intestinal Peyer's patches, and lung-, jejunum-, and colon-associated lymph
113 nodes) and mucosal/barrier tissues (lungs, jejunum, ileum, colon). Purified T cell populations from
114 these sites were obtained by magnetic selection and cultured overnight in media alone ("resting")
115 or stimulated with anti-CD3 and anti-CD28 antibodies ("activated") prior to single cell sequencing
116 using the 10x Genomics Chromium platform (**Fig. 1a**). We merged this dataset with our previous
117 study of resting and activated T cells from adult human organ donors and living blood donor
118 volunteers¹⁹, for a total of ~275,000 single cell profiles of T cells across 12 tissues.

119 We first defined T cell subsets for CD4⁺ and CD8⁺ T cells in the merged dataset as
120 naïve/central memory T cells (Naive/TCM), effector memory T cells (TEM), CD4⁺ regulatory T
121 cells (CD4⁺ Tregs), and $\gamma\delta$ T cells using a Naïve Bayes classifier (**Supplementary Fig. 1, see**
122 **Methods**). Visualization of the dataset by uniform manifold approximation and projection
123 (UMAP)²⁰ revealed that age cohort and stimulation conditions were dominant sources of
124 transcriptional variability within T cell subsets (**Fig. 1b**). The expression levels of canonical
125 marker genes defining the T cell subsets were highly conserved between infants and adults: CD4⁺
126 (*CD4*) and CD8⁺ (*CD8A*) naïve/TCM were enriched in lymphoid homing molecules *CCR7* and
127 *SELL* (coding for CD62L); CD4⁺ Tregs uniquely expressed *FOXP3*; CD4⁺ and CD8⁺ TEM highly
128 expressed *CCL5* as a marker of TEM cells defined previously¹⁹; $\gamma\delta$ T cells showed increased

129 expression of *TRDC* and decreased expression of *TRAC*, which encode the eponymous δ -chain or
130 α -chain constant region of the TCR for $\gamma\delta$ or conventional $\alpha\beta$ T cells, respectively (**Fig. 1c** and
131 **Supplementary Fig. 2**).

132 For a direct comparison of T cell populations across infant and adult tissues, we focused
133 our analysis on tissue sites that were represented in both age cohorts: blood, bone marrow, lung-
134 and jejunum-associated lymph nodes, spleen, lungs, jejunum and colon. In infants, the vast
135 majority of T cells in blood and lymphoid sites were $CD4^+$ and $CD8^+$ naïve/TCM, with minor
136 populations of $CD4^+$ Tregs, and few TEM or $\gamma\delta$ T cells (**Fig. 1d**). Mucosal sites and spleen showed
137 greater proportions of TEM, particularly for $CD8^+$ T cells, and the majority of intestinal T cells
138 were either $CD4^+$ or $CD8^+$ TEM, consistent with our previous findings¹⁴⁻¹⁶. By contrast, in adults
139 $CD4^+$ and $CD8^+$ TEM predominated relative to naïve/TCM in mucosal sites and the spleen (**Fig.**
140 **1e**). Notably, we observed much lower proportions of $CD8^+$ naïve/TCM as compared to $CD4^+$
141 naïve/TCM T cells in infants across most tissues relative to adults (**Fig. 1d,e**). These findings are
142 consistent with an exponential increase in T cell memory observed across infancy and childhood
143 compared to adults¹⁶.

144

145 **Infant TEM exhibit distinct a stem-like transcriptional state relative to adults.**

146 We directly investigated changes in gene expression between infant and adult T cells in
147 the resting state using pairwise differential expression analysis across all donors and tissues with
148 adequate representation for each subset (**see Methods**). $CD4^+$ and $CD8^+$ TEM exhibited a large
149 number of differentially expressed genes (193 and 173, respectively) between the age cohorts
150 (**Extended Data Fig. 1a** and **Supplementary Table 2**). Many of these differentially expressed
151 genes were shared across $CD4^+$ and $CD8^+$ TEM in infants (**Extended Data Fig. 1b**), demonstrating

152 a conserved transcriptional state in early life. We also detected shared genes expressed by CD4⁺
153 and CD8⁺ naïve/TCM that were upregulated in adults (**Extended Data Fig. 1c**), which was likely
154 due to increased frequencies in TCM populations¹⁶, whose profiles could not be readily
155 distinguished from naïve T cells by gene expression alone²¹.

156 Infant CD4⁺ and CD8⁺ TEM showed significant upregulation of genes encoding TFs linked
157 to T cell stemness/self-renewal and quiescence, including *TCF7* (TCF1), *LEF1* and *KLF2* (**Fig.**
158 **2a,b**)²²⁻²⁴. *SOX4*, which cooperates with TCF/LEF family TFs in the Wnt signaling pathway²⁵, and
159 *IKZF2* (Helios), typically associated with Treg differentiation and function^{26,27}, were also
160 upregulated in infant TEM across sites. Infant CD4⁺ TEM were enriched for expression of the
161 TH2-driving TF *GATA3*²⁸, while infant CD8⁺ TEM expressed high levels of the resident- and
162 effector-associated TFs *ZNF683* (Hobit) and *ID3*^{29,30} relative to adults. We also observed an
163 increase in transcripts associated with innate-like T cells (*ZBTB16*, *NCR3*, *KLRB1*, *FCER1G*)³¹ in
164 infants across tissues (**Fig. 2a,b**). Lastly, we found increased expression of genes encoding T cell
165 co-stimulatory or inhibitory surface molecules (*CD27*, *CD28*, *CD38*, *KLRG1*) in both lineages of
166 infant TEM relative to those in adults.

167 Expression of genes associated with T cell effector function were variably expressed
168 between infant and adult TEM (**Fig. 2a,b**). Across tissues, genes coding for cytokines (*LTB*, *MIF*,
169 *IL16*) and chemokine receptors (*CXCR3*, *CXCR4*) were upregulated in infant TEM, while
170 chemokines (*CCL4*, *CCL5*) and mediators of cytotoxicity (*GZMB*, *SLAMF7*) were upregulated in
171 adults. Infant TEM also exhibited upregulated expression of genes associated with activation
172 (*CD38*, *CD40LG*) and effector T cell fate (*KLRG1*), consistent with ongoing activation and
173 effector differentiation in infants encountering many new antigens. By contrast, adult CD4⁺ and
174 CD8⁺ TEM showed increased expression of transcripts associated with tissue adaptation and

175 adhesion, including *LGALS1* (galectin-1), *ANXA1* (annexin-1), *VIM* (vimentin), and *ITGAI*
176 (CD49a)^{19,32}.

177 Given the critical and multi-faceted roles of TCF1 and LEF1 in T cell identity and
178 function³³, we sought to validate the augmented expression of both TFs in infant versus adult TEM
179 on the protein level by flow cytometry. The expression of TCF1 and LEF1 by CD8⁺ TEM was
180 significantly increased in infants compared to adults in the spleen, while only LEF1 was increased
181 on CD4⁺ TEM (**Extended Data Fig. 1d,e**). Taken together, our findings demonstrate that infant
182 CD4⁺ and CD8⁺ TEM exhibit a distinct transcriptional state with increased expression of
183 transcriptional regulators of stemness and memory differentiation relative to adults.

184

185 **Consensus-scHPF reveals unique signatures of tissue adaptation and effector function across**
186 **infant and adult tissue T cells.**

187 To uncover unique gene expression programs between infant and adult tissue T cells, we
188 utilized consensus-scHPF¹⁸, a probabilistic Bayesian factorization method for the *de novo*
189 discovery of latent transcriptional co-expression signatures or “factors” in scRNA-seq data
190 (diagrammed in **Fig. 3a**). We applied consensus-scHPF to all infant and adult tissue T cells and
191 identified discrete factors defined by the top genes in the consensus gene score matrix (**Fig. 3b,c**
192 and **Supplementary Table 3**). To assess whether a given factor was associated with specific
193 features of dataset, we performed multivariate linear regression using age cohort (infant or adult),
194 tissue localization (lymphoid or mucosal), T cell subset (naïve/TCM or effector), T cell lineage
195 (CD4 or CD8), and activation condition (resting or activated) as covariates and plotted regression
196 coefficients for each comparison (**Fig. 3d**). In total, we identified 18 distinct signatures

197 corresponding to T cell subsets, metabolism, tissue adaptation, activation states, and/or effector
198 functions across infant and adult tissue T cells.

199 Consensus-scHPF revealed three factors associated with tissue localization and adaptation
200 that differed between infants and adults. The first tissue factor (“Infant Tissue”) was defined by
201 transcripts that were highly differentially expressed in infant TEM relative to adults from our
202 previous analysis (from **Fig. 2**), including *SMC4*, *NCR3*, *CXXC5*, *LAYN*, and *IKZF2*, and was
203 strongly enriched in infant CD8⁺ TEM from mucosal tissues (**Fig. 3c,d**). The second tissue factor
204 (“Tissue Signature”) was characterized by genes that we previously identified as a signature of T
205 cells residing in tissues compared to the blood^{19,32}, including *S100A4/6*, *CRIP1*, *LGALS1*, *KLRB1*,
206 and *ANXA2*, and was associated with adult TEM (**Fig. 3c,d**). The third tissue factor (“Gut
207 Residency”) was strongly biased towards adult TEM and was distinguished by markers of
208 intestinal homing and adhesion (*CCR9*, *ITGAI*, *CTNNA1*)^{32,34} and tissue-resident memory T cell
209 (TRM) development (*AHR*, *JUN*, *FOSB*)^{35,36} (**Fig. 3c,d**).

210 Three scHPF factors were associated with distinct T cell effector states. The “Cytotoxicity”
211 factor was defined by cytolytic molecules *GZMK*, *GNLY*, *GZMA*, *NKG7* and *PRF1*, and was highly
212 enriched in mucosal CD8⁺ TEM associated with the resting condition (**Fig. 3c,d**), reflecting a
213 poised cytotoxic state. Relatedly, the “Chemokine/Cytotoxic” factor included highly ranked
214 transcripts for cytotoxic mediators (*GZMB*, *GZMH*, *FASLG*, *PRF1*) and potent chemoattractants
215 (*CCL3*, *CCL4*, *CCL3L1*, *CCL3L3*), and was strongly associated with mucosal CD8⁺ TEM in the
216 activated condition (**Fig. 3c,d**). By contrast, the “Inflammatory Cytokine” factor was biased
217 towards CD4⁺ TEM and characterized by genes for *IFNG*, *IL2*, *TNF*, *CSF2*, and *LTA* (**Fig. 3c,d**).
218 Importantly, all three factors relating to effector functions and indeed most factors associated with

219 activated conditions in general (e.g., Activation/Survival, Immunoregulatory, Treg), were more
220 strongly associated with adults compared to infants (**Fig. 3d**).

221

222 **Infant T cells demonstrate a reduced activation capacity relative to adults.**

223 We uncovered several scHPF factors related to T cell activation and effector function that
224 were more strongly associated with adult T cells compared to those in infancy. To interrogate these
225 apparent differences in functional capacity, we investigated the expression of the top genes in both
226 Inflammatory Cytokine and Chemokine/Cytotoxic factors across TCR-simulated CD4⁺ and CD8⁺
227 TEM from all paired tissues in infants and adults. We found moderate differences in the expression
228 of genes in the Inflammatory Cytokine factor in CD4⁺ TEM between age groups, but strikingly
229 increased expression of these genes among CD8⁺ TEM in adults (**Fig. 4a**). For the
230 Chemokine/Cytotoxic factor, we also observed a prominent increase in expression for its top genes
231 in adult T cells compared to infants, especially for CD8⁺ TEM (**Fig. 4b**). As orthogonal
232 confirmation, we directly assessed the functional capacity of infant and adult TEM via intracellular
233 cytokine staining by flow cytometry after a short term *ex-vivo* stimulation. Both splenic CD4⁺ and
234 CD8⁺ TEM in adults exhibited much greater frequencies of IFN γ , IL-2 and TNF α -producing cells
235 relative to infants (**Fig. 4c,d**), consistent with our findings from scRNA-seq. We also found
236 increased intracellular production of the cytotoxic mediator granzyme B from unstimulated
237 conditions in adult TEM relative to infants, reflecting their augmented poised cytotoxic state
238 suggested by our scHPF analysis (**Fig. 4c,d**).

239 For unbiased comparison of T cell activation across age groups, we modeled activation
240 trajectories of T cells from both resting and activated conditions using diffusion maps, for CD4⁺
241 or CD8⁺ TEM separately. These diffusion maps separated resting T cells (in blue) on the *left* and

242 activated T cells (in red) projecting out to the *right* (**Fig. 4e**). The trajectories for CD4⁺ TEM
243 showed moderate differences in activation between age cohorts, with adults exhibiting an
244 increased number of cells along the activation axis (i.e., x-axis) relative to infants; however, the
245 CD8⁺ TEM trajectories showed strikingly more adult cells along the activation axis compared to
246 infants. Visualizing cell scores of the Inflammatory Cytokine (**Fig. 4f**) and Chemokine/Cytotoxic
247 (**Fig. 4g**) factors from scHPF further highlighted the increased expression of effector signatures on
248 adult CD4⁺ and CD8⁺ T cells compared to infants. These findings collectively demonstrate reduced
249 transcriptional responses to stimulation and a restricted capacity for effector function in infant
250 memory T cells relative to adults.

251

252 **Gene regulatory network inference uncovers distinct transcriptional regulators of tissue** 253 **adaptation in infant and adult tissue T cells.**

254 Consensus-scHPF identified three factors related to tissue residency and adaptation that
255 differed across infants and adult mucosal T cells (**Fig. 3c**). The top genes from the Infant Tissue
256 factor were highly enriched among infant CD4⁺ and CD8⁺ TEM (**Extended Data Fig. 2a**), while
257 genes from the Tissue Signature factor were expressed across both infant and adult tissues
258 (**Extended Data Fig. 2b**). By contrast, top genes from the Gut Residency factor were exclusively
259 enriched in the intestinal sites (jejunum, colon, jejunum-associated lymph nodes) of adult CD4⁺
260 and CD8⁺ TEM (**Extended Data Fig. 2c**), demonstrating a unique transcriptional state.

261 To investigate the differences in tissue adaptation between age cohorts, we used diffusion
262 maps to model maturation trajectories of resting infant and adult TEM from the intestines
263 (jejunum), where virtually all TEM are tissue-resident in both age groups¹⁶. For both CD4⁺ and
264 CD8⁺ TEM, the trajectories reflected a continuous transition from an infant state (*left*, purple) to

265 an adult state (*right*, green) and visualizing cell scores for the three tissue-associated scHPF factors
266 on the trajectories highlighted the features of this transition (**Fig. 5a**). The Infant Tissue factor was
267 largely specific to the *left*-most population of infant TEM, whereas high cell scores for the Tissue
268 Signature factor appeared at an intermediate position expressed by subpopulations of both infant
269 and adult TEM. High cell scores for the Gut Residency factor were enriched in adult TEM in the
270 *right*-most population. This analysis demonstrated a continuous tissue adaptation process in infant
271 and adult mucosal TEM, including a shared intermediate state.

272 We next sought to identify the putative TFs responsible for driving the functional and
273 tissue-associated transcriptional states in infant and adult T cells. We applied the Algorithm for
274 the Reconstruction of Accurate Cellular Networks (ARACNe), which reverse engineers a gene
275 regulatory network from gene expression data by inferring direct relationships between TFs and
276 their target genes^{37,38}. ARACNe generated a set of target genes for each TF, known as a “regulon”
277 (**Supplementary Table 4**). We performed gene set enrichment analysis³⁹ between the positively
278 regulated genes in each TF’s regulon and the ranked list of genes for each scHPF factor to associate
279 individual TFs with the cell states defined by each factor. The top two TF regulons with the highest
280 normalized enrichment scores for each scHPF factor are shown in a heatmap in **Fig. 5b**. This
281 analysis identified many TFs previously known to be linked to their respective T cell states
282 including, IRF1 and STAT1 for responses to IFN signaling^{40,41}, NFKB1 for T cell activation⁴²,
283 IRF7 regulating IFN responses and Tregs^{43,44}, ZEB2 for cytotoxic T cell function⁴⁵, and KLF2,
284 TCF1 (*TCF7*) and LEF1 regulating naïve T cell stemness and quiescence²⁴.

285 Importantly, this analysis identified Helios (*IKZF2*) as the top regulator of the Infant Tissue
286 factor and KLF6 as the top regulator of both the Tissue Signature and Gut Residency factors (**Fig.**
287 **5b**). We visualized the relationship between the top genes in the tissue adaptation factors with each

288 TF's regulon using a network, with the mutual information between each TF-gene pair as a
289 measure of their interaction strength. Among the top 50 genes in the Infant Tissue factor, 24 genes
290 were inferred to be regulated by Helios (**Fig. 5c**). Of the top 50 genes in the Tissue Signature and
291 Gut Residency factors, KLF6 was inferred to regulate 21 and 23 genes, respectively (**Extended**
292 **Data Fig. 2d**).

293 To assess differences in the activities of TFs across infants and adult T cells, we next
294 performed Virtual Inference of Protein-activity by Enriched Regulon (VIPER) analysis, which
295 utilizes the relative expression of a TF's up- and down-regulated targets to infer its activity in a
296 given cell⁴⁶. We plotted the TF activities of both Helios and KLF6 on the tissue adaptation
297 trajectories we generated previously and found that Helios activity was restricted to infants, while
298 KLF6 activity was aligned with the transition from infant to adults and increased on all adult CD4⁺
299 and CD8⁺ TEM (**Fig. 5d**). We next compared the differences in gene expression and activities for
300 each TF in infants versus adults to identify those that were both highly differentially expressed and
301 highly active in each age cohort. Helios was among the top differentially expressed and
302 differentially active TFs in infants CD4⁺ and CD8⁺ TEM (**Extended Data Fig. 2e,f**), along with
303 the other stem-like TFs (TCF1, LEF1, SOX4) that we identified by differential gene expression
304 alone. KLF6 was only mildly enriched in activity and expression in adults, suggesting that
305 differences in the function of KLF6 in infants and adults is not controlled on the level of
306 transcription. Together, these analyses facilitated the discovery of individual TFs associated with
307 T cell states defined by scHPF and identified Helios as putative regulator of the infant-specific T
308 cell tissue adaptation program.

309

310 **Helios (*IKZF2*) drives an infant-specific transcriptional program and restrains T cell effector**
311 **function in early life.**

312 To interrogate the functional role of Helios (*IKZF2*) in infant T cells, we first confirmed
313 its expression on the protein level by intracellular staining via flow cytometry. Helios was highly
314 expressed by all subsets of CD8⁺ T cells in infants relative to adults, while only naïve T cells were
315 significantly enriched for Helios in infants among CD4⁺ T cells (**Fig. 6a**).

316 We directly investigated the role of Helios in regulating its inferred targets by disrupting
317 Helios expression in primary infant T cells using CRISPR-Cas9 gene editing. We isolated CD3⁺
318 T cells from the spleen of infant donors (ages 2 and 3 months old), transfected T cells with Cas-9
319 ribonucleoproteins targeting Helios without prior T cell stimulation, confirmed protein knockout
320 (KO) by flow cytometry, and assessed the transcriptional profiles of Helios-KO infant T cells
321 relative to controls at rest or following TCR-stimulation using scRNA-seq (**Fig. 6b**). CRISPR-
322 Cas9 KO of Helios resulted in ~60% reduction of Helios-expressing cells by flow cytometry for
323 both infants (**Fig. 6c**). To validate the targets of Helios in our regulatory network, we assessed
324 changes in gene expression of Helios-activated and Helios-repressed targets in Helios-KO CD8⁺
325 TEM relative to negative controls from the resting condition in the scRNA-seq data. We observed
326 a marked inversion in expression of Helios' gene targets in KO cells compared to controls, where
327 Helios-activated targets were decreased and Helios-repressed targets were increased in both infants
328 (**Fig. 6d**). These results experimentally confirm Helios' ARACNe-inferred regulon, which is
329 highly enriched in genes from the Infant Tissue factor that defines an infant-specific transcriptional
330 state.

331 Following TCR-stimulation, we found only minor differences in gene expression for
332 Helios-KO cells relative to controls in both CD4⁺ and CD8⁺ TEM (**Extended Data Fig. 3a,b**).

333 However, Helios-KO in CD8⁺ naïve/TCM resulted in reduced expression of genes associated with
334 IFN signaling (*MX1*, *GBP1*, *GBP4*, *GBP5*) and TFs regulating CD8⁺ memory T cell differentiation
335 and function, including *ID2*⁴⁷, *PLAC8*⁴⁸, *HIC*⁴⁹, *TOX2*⁵⁰ (**Fig. 6e**). Conversely, we observed
336 increased expression of an array of chemokines (*XCL1*, *XCL2*, *CCL3*, *CCL4*, *CCL4L2*, *CCL20*),
337 pro-inflammatory cytokines (*IL2*, *CSF2*), cytotoxic mediators (*GZMB*), and co-stimulatory
338 molecules (*TNFRSF9*, *TNFSF9*, *TNFRSF14*) (**Fig. 6f**). TCR-stimulated CD4⁺ naïve/TCM
339 populations showed a similar pattern of increased expression for cytokines (*IL2*, *CSF2*, *EBI3*),
340 chemokines (*XCL1*, *XCL2*) and co-stimulatory molecules and receptors (*TIGIT*, *KLRB1*,
341 *TNFRSF4*, *TNFRSF9*, *TNFRSF18*, *TNFSF14*, *FCER1G*, *IL18R1*, *IL1R1*) in the Helios-KO
342 condition relative to negative controls (**Extended Data Fig. 6c,d**). Together, our data demonstrate
343 that Helios restricts effector functions of infant naïve/TCM cells after TCR-mediated activation,
344 in addition to regulating a transcriptional program associated with T cell adaptation to tissues in
345 infants.

346 **DISCUSSION**

347 Our study provides in-depth transcriptional and functional analyses of tissue T cell
348 responses and their underlying programming during a critical period of immune development in
349 infancy relative to adulthood. We reveal unique transcriptional programs and multiple TFs
350 expressed by infant memory T cells across tissues relative to adults. Infant memory T cells retain
351 high levels of expression LEF1, TCF1, and KLF2, typically associated with stemness and
352 quiescence among naïve T cells^{51,52}. Also upregulated in infant memory T cells across tissues is
353 the SRY-related HMG-box family TF SOX4, a critical transcriptional regulator that cooperates
354 with TCF1 and LEF1 to facilitate T cell differentiation in the thymus⁵³. Previous studies
355 demonstrate that SOX4 regulates CD8⁺ memory T cell development⁵⁴, antagonizes TH2
356 development in tandem with LEF1 in CD4⁺ T cells^{55,56}, and facilitates stemness in cancer cells⁵⁷,
357 suggesting this TF network may play an important role in maintaining quiescence and
358 differentiation potential among infant memory T cells. Furthermore, a recent study identified a
359 small population of blood naïve T cells in healthy young adults expressing SOX4 and Helios
360 (*IKZF2*) as recent thymic emigrants (RTEs)⁵⁸. Several genes in RTE signature overlap with the
361 top genes in the Infant Tissue signature (e.g., *TOX*, *SMC4*, *PDE7B*, *IKZF2*), raising the intriguing
362 possibility that the infant-specific transcriptional state may arise from the generation of memory T
363 cells from RTEs⁵⁹.

364 Our findings demonstrate a globally reduced capacity for TCR-mediated activation among
365 infant memory T cells relative to those in adults. This work expands on previous observations by
366 our group and others showing that mucosal memory T cells from infants exhibit decreased
367 production of inflammatory cytokines relative to older individuals^{16,17,60}. These findings provide
368 an intriguing contrast to our earlier work showing that naïve T cells from infants are more sensitive

369 to TCR-stimulation and are biased towards differentiation into short-lived effector cells relative to
370 adults^{9,11}. Together, these data suggest that the infant immune system preferentially utilizes short-
371 lived effector responses from naïve T cells to respond to infections, rather than effector memory
372 responses in tissues.

373 We identify transcriptional programs strongly enriched in mucosal CD4⁺ and CD8⁺ TEM
374 associated with tissue adaptation and residency. These findings extend and unify our previous
375 work describing a shared tissue-associated signature in TEM from the bone marrow, lungs, and
376 lymph nodes¹⁹ and tissue-specific adaptation signatures across barrier tissues³². Here, we elucidate
377 a common tissue signature in TEM across multiple tissues in both infants and adults as well as a
378 highly intestine-specific signature associated with gut homing, adhesion, and residency unique to
379 adults. We identify KLF6 as a putative transcriptional regulator of both common and intestine-
380 specific resident signatures using gene regulatory network inference. Accordingly, a recent study
381 in mice also defines *Klf6* as a TF specific to the tissue-resident cell state using populational level
382 RNA-seq and the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)
383 data from a total of 10 murine studies⁶¹. Our trajectory analysis further suggests a role for KLF6
384 in the acquisition of a mature adult-like TRM phenotype over age, providing critical context to our
385 previous work demonstrating staged maturation of TRM during infancy and childhood¹⁶.

386 Our study uncovers a distinct infant-specific transcriptional state in CD4⁺ and CD8⁺ TEM
387 defined by expression of Helios (*IKZF2*), epigenetic modifiers *SMC4* and *CXXC5*, and regulators
388 of tissue adhesion (*LAYN*, *CD9*). We provide direct experimental evidence of Helios' role in
389 driving the expression of genes in this transcriptional state using CRISPR-Cas9 knockdown, which
390 also reveals Helios as a repressor of T cell effector function after TCR-stimulation. This finding is
391 consistent with a previous report of human patients with Helios loss-of-function mutations that

392 show enhanced IL-2 production by T cells from the peripheral blood⁶² or small populations of
393 Helios-expressing T cells in the blood with reduced effector capacity⁶³. Helios also promotes
394 human fetal Treg differentiation and phenotype, where experimental knockdown of Helios
395 enhances expression of pro-inflammatory genes in *ex-vivo* induced fetal Tregs²⁷. Taken together,
396 these data provide compelling evidence for tolerogenic state maintained by Helios in infant T cells.

397 This investigation reveals critical differences in transcriptional programming and
398 activation capacity between infant and adult tissue T cells, with important implications for the
399 generation and maintenance of protective immunity throughout the body. The expression of stem-
400 like TFs in infant T cells may represent a developmental adaptation that facilitates rapid
401 establishment of the tissue memory T cell niche. Conversely, the reduced activation capacity and
402 effector function of infant memory T cells may serve to limit excessive inflammatory responses in
403 tissues during this vulnerable period of development, but may also impair the ability to mount
404 protective immune responses upon re-infection. Our work provides direct evidence for a cell-
405 intrinsic mechanism regulating the infant-specific transcriptional state. This study lays the
406 foundation for understanding the mechanisms governing infant T cells responses, which may aid
407 identification of novel therapeutic targets to modulate T cell function in early life and promote
408 long-lasting, protective adaptive immunity.

409

410 **METHODS**

411 **Human organ donors and tissue acquisition**

412 We obtained human organ tissue from deceased (brain-dead) organ donors directly at the
413 time of acquisition for life-saving clinical transplantation through approved research protocols and
414 materials transfer agreements with organ procurement organizations (OPOs) in the United States.
415 Tissues from adult organ donors were obtained from an approved protocol with the OPO for the
416 New York metropolitan area, LiveOnNY and included donors from our previous study¹⁹. Tissues
417 from infant organ donors were obtained from LiveOnNY as well as through the Human Atlas for
418 Neonatal Development (HANDEL) program based on the network for Pancreatic Organ Donors
419 (nPOD). Consent for use of tissues for research was obtained by next of kin. All organ donors used
420 in the study (**Supplementary Table 1**) were free of cancer, chronic disease, seronegative for
421 hepatitis B, C and HIV and did not show evidence for active infection based on blood, urine,
422 respiratory and radiological surveillance testing. The use of tissues from deceased organ donors
423 does not qualify as human subjects research as confirmed by the Columbia University institutional
424 review board.

425

426 **Tissue processing, T cell isolation and stimulation**

427 Organ donor tissue samples were maintained in saline or University of Wisconsin solution
428 on ice for transport to the laboratory and processing, typically within 2-24 hours of acquisition.
429 Processing of infant and adult tissues to single cell suspensions was performed as previously
430 described^{16,19}. Briefly, blood was obtained by venipuncture, bone marrow was aspirated from the
431 superior iliac crest, and mononuclear cells from both sites was obtained by density gradient
432 centrifugation using Lymphocyte Separation Medium (Corning) or Ficoll-Paque Plus (Cytiva).

433 Lymph nodes were isolated by dissection from the intestinal mesentery or the tracheobronchial
434 tree of the lungs. Isolated lymph nodes, spleen and tonsil samples were placed in complete media
435 composed of IMDM or RPMI 1640 (Gibco), 10% fetal bovine serum (GeminiBio), and 1% L-
436 Glutamine:Pen:Strep Solution (GeminiBio), and mechanically dissociated with surgical scissors.
437 Lung parenchymal tissue was dissected from the large airways, mechanically dissociated, and
438 placed in digestion media composed of complete media, 1mg/ml Collagenase D (Millipore Sigma),
439 0.1 mg/ml DNase (Millipore Sigma) in a shaker at 37°C for 30 minutes. Intestinal tissues were
440 dissected by their anatomical locations (jejunum, ileum, colon, Peyer's patches), washed with
441 sterile PBS (Corning) to remove luminal content, mechanically dissociated with scissors, and
442 placed into digestion media in a shaker at 37°C for 30 minutes. Dissociated and/or digested cell
443 suspensions of the lymph nodes, spleen, tonsils, lungs, and intestines were filtered using a 100 um
444 filter (VWR) and centrifuged on a density gradient as above to remove debris and isolate
445 mononuclear cells, followed by resuspension in complete media.

446 T cells from individual tissue single cell suspensions were enriched via magnetic negative
447 selection (EasySep Human T cell Enrichment Kit; STEMCELL Technologies) followed by dead
448 removal (Milteyni Biotec) resulting in +80-95% purity. We cultured 0.5-1 million T cell-enriched
449 cells from each tissue for 16 hours at 37°C in complete medium, with or without TCR stimulation
450 using ImmunoCult Human CD3/CD28 T Cell Activator (STEMCELL Technologies), after which
451 dead cells were removed (as above) before single cell encapsulation.

452

453 **Single cell RNA-sequencing and data processing**

454 T cell-enriched samples were loaded onto the Next GEM Chromium Controller using the
455 Chromium Next GEM Single Cell 3' Reagent kit v3.1 from 10x Genomics for single cell

456 encapsulation and library construction as per manufacturer's suggested protocols. Libraries were
457 sequenced on an Illumina NovaSeq 6000, targeting ~300M raw reads per sample (~60,000 raw
458 reads per cell).

459 scRNA-seq data were aligned and demultiplexed as described in Szabo et al¹⁹ using a
460 publicly available pipeline (<https://github.com/simslab/DropSeqPipeline8>). Briefly, for each
461 sample we trimmed read 2 to remove 3'-poly(A) tails (>7 A's in length), discarded reads with
462 fewer than 24 nucleotides remaining after trimming, and aligned the rest to GRCh38 (GENCODE
463 v.24 annotation) using STAR v.2.5.0⁶⁴. We assigned an address comprised of a cell-identifying
464 barcode, unique molecular identifier (UMI) barcode, and gene identifier to each read with a unique,
465 strand-specific exonic alignment. We followed the method in Griffiths et al.⁶⁵ to filter the reads
466 for index swapping and collapsed PCR duplicates using the UMIs after correcting sequencing
467 errors in both the cell-identifying and UMI barcodes to generate an initial, unfiltered count matrix
468 for each sample.

469 To identify cell-identifying barcodes that correspond to actual cells and to filter low-quality
470 single-cell profiles, we used the methodology described in Zhao et al.⁶⁶. Briefly, we used the
471 EmptyDrops algorithm⁶⁷ to remove cell-identifying barcodes that primarily contain ambient RNA.
472 We then filtered the resulting count matrix to remove cell barcodes with high mitochondrial
473 alignment rates (>1.96 standard deviations above the mean for a sample), high ratio of whole gene
474 body to exonic alignment (>1.96 standard deviations above the mean for a sample), high average
475 number of reads per transcript or transcripts per gene (>2.5 standard deviations above the mean
476 for a sample), or cells where >40% of UMI bases are T or where the average number of T-bases
477 per UMI is at least four.

478

479 **In silico T cell purification**

480 While the EmptyDrops algorithm used above is adept at removing cell-identifying
481 barcodes that correspond to ambient RNA, debris or molecular aggregates are more problematic⁶⁷.
482 We used a Gaussian mixture model (GMM)-based filter to further identify low-quality cells that
483 could result from unfiltered index-swapping artifacts. We noticed that histograms of the number
484 of transcripts per cell for some samples contained a lower mode, particularly for samples that were
485 co-sequenced with samples external to this study (**Supplementary Fig. 3a**). To identify and isolate
486 these modes, we modeled each sample's log-scale transcript per cell distribution as a two-
487 component GMM using *sklearn.mixture.GaussianMixture* (scikit-learn, version 0.21.3) and
488 considered samples with a high-mode to low-mode ratio of at least 1.2 (on a log₂-scale) as
489 candidates for further filtering at the single-cell level. For these samples, we set a cutoff at three
490 standard deviations above the mean for the high mode, using the square-root of the GMM's
491 estimate of the high-mode's variance (*sklearn.mixture.GaussianMixture.covariance*), and
492 removed cells below this cutoff. Differential expression analysis between the filtered and
493 unfiltered cells showed that these low-coverage, filtered cells were likely contaminants and
494 enriched in neural markers like *NNAT*, *BSN*, *NCAM2*, *NRXNI*, *GRIA1*, *GABRA1*, and *BDNF*.

495 With count matrices for high-quality cells in hand for each sample, we removed all non-T
496 cells from the data, including contaminating cells, multiplets, and cells in which apparent T cell
497 marker expression was likely an artifact of cross-talk¹⁹. We first performed unsupervised Louvain
498 clustering using Phenograph⁶⁸ with highly variable marker selection and k-nearest-neighbor graph
499 construction as described in Levitin et al⁶⁹. Highly variable marker selection was performed at the
500 donor level and applied to each sample within a given donor. Next, we labeled each cluster as a
501 putative T cell cluster (t_1), contaminant cluster (t_{-1}), or unknown cluster (t_0) based on the average

502 normalized expression of *CD3D*, *CD3E*, and *CD3G*. For each sample, we then examined the
503 distribution of the number of reads per transcript (RPT), which estimates the number of PCR
504 amplicons per transcript. PCR recombination can result in a multi-modal RPT distribution as
505 described in Szabo et al¹⁹. To determine an RPT threshold for non-artifactual transcripts, we
506 modeled this distribution as a two-component Geometric-Negative Binomial using the Scipy
507 functions *scipy.stats.genom.pmf*, *scipy.stats.nbinom.pmf*, and *scipy.optimize.curve_fit* and took the
508 intercept of the two components. We defined high-confidence T cells within each t_1 cluster as cells
509 that expressed any of *CD3D*, *CD3E*, *CD3G*, *TRAC*, *TRBC1*, *TRBC2*, *TRDC*, *TRGC1*, and *TRGC2*
510 with RPT greater than the sample-specific threshold. Next, we performed pairwise differential
511 expression analysis (as described below) between the high-confidence cells in the t_1 cluster and
512 the cells in the t_{-1} cluster for each sample. We used this analysis to construct a contaminant gene
513 list for each donor. We considered a gene to be contaminant-specific if it was at least 10-fold
514 enriched in at least two t_{-1} clusters across the donor with $FDR < 10^{-5}$ and if it was not enriched in
515 more than one t_1 cluster with greater than 10% enrichment and $FDR < 10^{-2}$. For each sample, we
516 computed the proportion of the contaminant gene list that each cell expresses above the sample-
517 specific RPT, which we call p_c . We then fit a truncated Gaussian distribution to the distribution of
518 p_c specifically for the high-confidence T cells from t_1 clusters to establish a threshold proportion
519 p_t at three standard deviations above the mean of the fit. Finally, to call T cells, we identified cells
520 in t_1 clusters with any of *CD3D*, *CD3E*, *CD3G*, *TRAC*, *TRBC1*, *TRBC2*, *TRDC*, *TRGC1*, and
521 *TRGC2* detected with RPT greater than the sample-specific threshold and $p_c < p_t$ as T cells. We
522 also identified cells in t_0 clusters with any of *CD3D*, *CD3E*, *CD3G*, *TRAC*, *TRBC1*, *TRBC2*, *TRDC*,
523 *TRGC1*, and *TRGC2* detected with RPT greater than the sample-specific threshold and $p_c < p_t$ as

524 T cells as long as the cluster's first quartile for p_c was also less than p_t . **Supplementary Fig. 3b-g**
525 contains a graphical depiction of each step in this procedure for a sample.

526 To validate the above procedure, we clustered all T cells that resulted with Phenograph and
527 found no clusters with systematically lower expression of *CD3D*, *CD3E*, or *CD3G* (i.e., less than
528 half of their mean expression). Furthermore, we clustered and performed differential expression
529 analysis on all cells for each sample that were not called T cells to verify that most populations
530 represented non-T cells and that all *CD3D*, *CD3E*, or *CD3G*-expression populations also expressed
531 contaminants in common with non-T cells (i.e., possible multiplets). We did not find any clusters
532 without clear contaminants.

533

534 **Consensus single cell Hierarchical Poisson Factorization**

535 scHPF is a Bayesian algorithm for probabilistic factorization of scRNA-seq count matrices
536 that produces highly interpretable factors or gene co-expression signatures⁶⁹. Here, we applied the
537 consensus implementation of scHPF, which generates and integrates many independent models of
538 large scRNA-seq data sets, identifies recurrent factors from these models, and learns a final
539 consensus model¹⁸. We used consensus scHPF to generate a single factor model for the entire T
540 cell data set presented here including infant, adult, resting, and activated T cells from all tissue
541 sites.

542 To construct the class-balanced dataset for scHPF, we randomly sampled 1,000 cells from
543 each sample from tissues where we had at least one adult and one infant donor (blood, bone
544 marrow, jejunum, jejunum lymph node, colon, lung lymph node, and spleen). To correct for
545 coverage differences between samples, we downsampled the count matrices in this training set to
546 the same mean number of transcripts per cell (2,086 transcripts/cell). We then filtered the training

547 data to contain only protein coding genes detected in at least 1% of cells after downsampling. We
548 additionally created an equivalently downsampled test set of cells that were not in the training set,
549 with up to 200 cells for each experimental sample in the training data.

550 Due to scHPF's highly multi-modal posterior on this complex dataset, we used the
551 consensus approach described previously¹⁸, which allowed us to capture highly robust patterns of
552 expression that consistently appear across scHPF training models with different random
553 initializations, while still giving the model the freedom to approximate the parameter values the
554 best explain the data. First, we ran scHPF with 10 random initializations for $k=10$ through 20 and
555 selected three out of each set of 10 models with the lowest mean negative log likelihood on the
556 training data for each value of k . Using Walktrap clustering, we identified 29 modules of similar
557 factors that were observed in multiple models, and used their median gene weights to reinitialize
558 scHPF and learn a refined, consensus model with 29 factors.

559 We evaluated the consensus-initialized model as compared to randomly initialized models
560 with the same number of factors using the mean negative log-likelihood of the held-out test set.
561 The consensus-initialized model had significantly better loss than any of the randomly initialized
562 models (2.3953 with consensus-initialized vs 2.4006 \pm 0.0002 SEM for the 10 randomly
563 initialized models). Thus, the consensus model achieved better log-likelihood for the held-out data,
564 ensured that the factors were robust against random initializations, and effectively automated the
565 selection of the number of factors k .

566 To project the full dataset onto the reference model obtained above, we downsampled all
567 cells that were not included in training to have the same mean number of transcripts per cell as the
568 training data. We then used the scHPF command *prep-like* to generate an appropriately filtered

569 and formatted count matrix for these remaining cells and projected them into the consensus model
570 using the scHPF command *project* with default parameters.

571

572 **T cell subset classification and analysis**

573 We used scHPF's embeddings in combination with gene expression values to annotate T
574 cell subsets. CD4⁺ and CD8⁺ T cells can be difficult to distinguish based on transcriptional profiles
575 alone due to transcript drop-out, particularly for CD4⁺ T cells, and because CD4 vs. CD8 status is
576 highly correlated with effector status for the cells profiled here. This problem is further exacerbated
577 in stimulated T cells where subset-specific markers are downregulated. Because scHPF breaks
578 transcriptional profiles down into component expression programs, we can leverage its
579 representations to distinguish between subsets, even for activated T cells.

580 We used a Naïve Bayes classifier on scHPF's cell scores concatenated with expression
581 values for several key markers: *CD4*, *CD8A*, *CD8B*, *CCL5*, *SELL*, *TRDC*, *TYROBP*, *CCR7*,
582 *CTLA4*, and *FOXP3*. We first defined separate training sets for infant and adult resting T cells
583 from within the scRNA-seq dataset based on co-expression of these markers for CD4⁺ naïve/TCM,
584 CD8⁺ naïve/TCM, CD4⁺ TEM, CD8⁺ TEM, CD4⁺ Tregs, and $\gamma\delta$ T cells, according to the scheme
585 in **Supplementary Table 5**. Next, we generated a concatenated matrix comprised of the scHPF
586 cell scores for the consensus model described above and the log-normalized expression of the
587 above markers. Size factors for normalized counts were computed using the *computeSumFactors*
588 function in *scrn* as described by Lun et al⁷⁰. We then standardized the resulting concatenated
589 feature matrix using the *sklearn.preprocessing.StandardScaler* function (scikit-learn, version
590 0.23.2), trained a Naïve Bayes classifier separately for resting infant and adult T cells with
591 *sklearn.naive_bayes.GaussianNB.fit*, and predicted the remaining resting infant and adult T cells

592 separately using *sklearn.StandardScaler.transform*. We repeated this exact procedure for the
593 infant and adult T cells separately, but this time including both resting and activated T cells and
594 omitting the CD4⁺ Treg class. We could not classify CD4⁺ Tregs from the activated T cells in our
595 dataset, because conventional T cells upregulate many canonical CD4⁺ Treg markers upon
596 stimulation. Thus, this second round of classification allowed us to share information between the
597 resting and activated T cells while obtaining an annotation for the activated T cells that excluded
598 the CD4⁺ Treg class. Finally, for activated T cells, we used the annotation obtained from this
599 second round of classification for downstream analysis. For resting T cells, we used this same
600 annotation, but substituted the CD4⁺ Treg class for any cell classified as a CD4⁺ T cell in the
601 second round of classification that was also classified as a CD4⁺ Treg in the first round.

602 We used several approaches to validating this classifier. First, we obtained excellent
603 agreement between the expression patterns of canonical T cell subset markers and the classifier
604 results as shown in **Fig. 1c**. Second, because the resting and activated T cells originate from
605 matched samples, the number of resting and activated cells in each class should be roughly equal
606 to each other for cells from the same sample. As shown in **Supplementary Fig. 4** the median
607 absolute deviation between resting and activated cell frequencies for CD4⁺ and CD8⁺ T cells is 2.3
608 and 1.6%, respectively. Similarly, for naïve/TCM, TEM, and $\gamma\delta$ T cells we obtain 3.4%, 3.1%, and
609 0.2%, respectively.

610 Our third approach was to estimate the accuracy of the classifier and therefore we applied
611 it to a multi-tissue immune cell dataset from an organ donor from which we obtained CITE-seq
612 data from blood, bone marrow, lung, lung lymph node, jejunum, and spleen^{21,71}. We used the
613 surface protein data from CITE-seq as an orthogonal ground truth to the corresponding RNA
614 expression data from each cell to which we applied the Naïve Bayes classifier. First, we defined

615 cells into high-confidence subsets based on the surface protein data. We defined these subsets as:
616 $CD4^+$ naïve ($CD4^+$ $CD8^-$ $TCR\gamma\delta^-$ $CD45RA^+$ $CD45RO^-$ $CCR7^+$ $CD62L^+$ $CD27^+$ $CD25^-$), $CD8^+$
617 naïve ($CD4^-$ $CD8^+$ $TCR\gamma\delta^-$ $CD45RA^+$ $CD45RO^-$ $CCR7^+$ $CD62L^+$ $CD27^+$), $CD4^+$ TEM ($CD4^+$
618 $CD8^-$ $TCR\gamma\delta^-$ $CD45RA^-$ $CD45RO^+$ $CCR7^-$ $CD62L^-$ $CD27^-$ $CD25^-$), $CD8^+$ TEM ($CD4^-$ $CD8^+$
619 $TCR\gamma\delta^-$ $CD45RA^-$ $CD45RO^+$ $CCR7^-$ $CD62L^-$ $CD27^-$), $CD4^+$ Treg ($CD4^+$ $CD8^-$ $TCR\gamma\delta^-$ $CD45RA^+$
620 $CD45RO^-$ $CCR7^+$ $CD62L^+$ $CD27^+$ $CD127^-$ $CD25^+$), and $\gamma\delta$ T cells ($CD4^-$ $TCR\gamma\delta^+$) and ensured
621 that these classes were mutually exclusive. To define positive and negative cell populations for
622 each surface protein marker, we log-transformed the marker's expression level ($\log_2(\text{counts per}$
623 $\text{thousand} + 1)$) and fit a two-component Gaussian mixture model to the transformed expression
624 distribution (**Supplementary Fig. 1a**). For each fit, we computed defined three expression
625 thresholds: 1.96 standard deviations below the mean of the higher mode Gaussian (L_1), 1.96
626 standard deviations above the mean of lower mode (L_2), and the local minimum of the Gaussian
627 mixture fit between the means of the two components (L_3). We then set our thresholds for marker-
628 negative and positive subpopulations as $\min(L_1, L_2, L_3)$ and $\max(L_1, L_2, L_3)$, respectively, to establish
629 our ground truth T cell subsets. Finally, we trained a consensus scHPF model on the scRNA-seq
630 component of the CITE-seq dataset, trained the Naïve Bayes classifier using the same procedure
631 as described above based only on the scRNA-seq, and classified the high-confidence cells
632 established using CITE-seq. Importantly, the Naïve Bayes classifier was blinded to the surface
633 protein data used to define the high-confidence subsets and to the high-confidence subset
634 annotations themselves. By comparing the Naïve Bayes classifier results to the high-confidence
635 subsets established from surface protein expression, we found that the Naïve Bayes classifier was
636 highly performant with favorable results for sensitivity, specificity, precision, and accuracy across
637 T cell subsets (**Supplementary Fig. 1b**).

638 For downstream analysis, we interrogated T cells that were classified as CD4⁺ naïve/TCM,
639 CD8⁺ naïve/TCM, CD4⁺ TEM, CD8⁺ TEM, CD4⁺ Tregs, and $\gamma\delta$ T cells. We visualized the data
640 using supervised dimensionality reduction and embedded the data into two dimensions by
641 UMAP²⁰, with T cell subsets as classes. Plotting UMAP embeddings and markers for each T cell
642 subset was performed with *scanpy*.

643

644 **Differential gene expression between infant and adult T cells**

645 We performed differential expression analysis across age groups for each conventional T
646 cell subset (CD4⁺ naïve/TCM, CD8⁺ naïve/TCM, CD4⁺ TEM, CD8⁺ TEM, CD4⁺ Treg), using all
647 tissue samples that met criteria of at least 100 cells per subset-donor-tissue combination. For CD4⁺
648 and CD8⁺ naïve/TCM, this included the blood, jejunum- and lung-associated lymph nodes, and
649 spleen; for CD4⁺ and CD8⁺ TEM, this included the jejunum, lung and spleen; and for CD4⁺ Tregs,
650 this included jejunum- and lung-associated lymph nodes as well as the spleen. For each tissue
651 group within a subset, we performed pairwise differential expression using *scanpy v1.9.3*
652 (*rank_genes_groups*; Wilcoxon with tie correction) between each infant donor versus every adult
653 donor, using equalized cell counts (subsampling) and total counts (downsampling) for each group.
654 We used the intersection of differentially expressed genes (FDR adjusted p-value < 0.05, log-fold
655 change > 1) for every infant-adult comparison (i.e., must be differentially expressed in each infant
656 donor compared to every adult donor) within each tissue to generate a list of differentially
657 expressed genes across age in a given tissue. We next used the union of differentially expressed
658 genes across all tissue comparisons within a subset to generate a final list of genes by T cell subset
659 (**Supplementary Table 2**). UpSet plots were generated using the python package UpSetPlot
660 *v0.8.0*.

661

662 **Activation trajectories by diffusion maps**

663 We constructed activation trajectories for CD4⁺ and CD8⁺ TEM for mucosal tissues, where
664 these subsets are highly enriched in TRM using diffusion component analysis as described
665 previously¹⁹. Briefly, for CD4⁺ TEM and CD8⁺ TEM separately, we randomly sampled the same
666 cell numbers from resting infant, resting adult, activated infant, and activated adult conditions from
667 the mucosal sites lung, jejunum, ileum, and colon. Next, we computed a pairwise Euclidean
668 distance matrix for the scHPF cell scores of the sampled mucosal CD4⁺ and CD8⁺ TEM, and used
669 the DMAPS package (<https://github.com/hsidky/dmaps>) to embed the scHPF model into its first
670 two diffusion components, which consistently separated the cells based on activation status.

671

672 **ARACNe and VIPER analysis**

673 We used ARACNe-AP (<https://github.com/califano-lab/ARACNe-AP>) to infer gene
674 regulatory networks from the scRNA-seq dataset^{37,38} using the metacell workflow described by
675 Vlahos et al⁷². While ARACNe has been widely used for regulatory network inference from bulk
676 RNA-seq data, the sparsity of scRNA-seq data requires the construction of pseudo-bulk profiles
677 that average the expression profiles of multiple individual cells called metacells. To quantify cell-
678 cell similarity for generating metacells, we used the cell score matrix from the consensus scHPF
679 model described above into a Pearson correlation matrix from which we generated a k-nearest
680 neighbors graph with k=50 to aggregate scRNA-seq profiles into metacell profiles by averaging
681 over 50 similar cells as described previously⁷². Finally, we used ARACNe-AP to compute a
682 transcription factor-target gene regulatory network consolidated from 200 rounds of bootstrapping

683 using the metacell matrix and a list of 688 transcription factors that met the criteria for inclusion
684 in our scHPF model (see above).

685 We used the gene regulatory network to associate transcription factors with scHPF factors.
686 For this, we identified the subset of ARACNe-inferred targets that were activated by a given
687 transcription factor using the *aracne2regulon* function in the VIPER package⁴⁶. Then, we
688 performed GSEA for each scHPF factor-transcription factor pair where the ranked gene list was
689 obtained by ranking all genes by their scHPF gene score for a given scHPF factor and the gene
690 sets were the set of activated targets for each transcription factor. This calculation yielded a
691 normalized enrichment score reflecting the enrichment of a given transcription factor's activated
692 targets among the top-ranked genes in a given scHPF factor. We also used the gene regulatory
693 network to infer transcription factor activities at the single-cell level using the VIPER algorithm
694 (v1.26.0), a companion tool for calculating protein activity from ARACNe-inferred networks⁴⁶.
695 Specifically, we used the *viper* function with the z-scored, log-normalized scRNA-seq expression
696 matrix (using *scrna* as described above) to compute transcription factor activities with default
697 parameters. Similarly, to assess differential VIPER activity between two conditions (e.g., for the
698 Helios KO experiments described below), we generated a gene signature for the two conditions
699 using the *rowTtest* function in VIPER, a null model using the *ttestNull* function in VIPER with
700 1,000 permutations, and the *msviper* function in VIPER with default parameters.

701

702 **Flow cytometry and intracellular staining**

703 Single cell suspensions of tissue mononuclear cells washed with staining buffer comprised
704 of PBS (Corning), 2% FBS (GeminiBio) and 2 mM EDTA (Gibco), incubated with Human
705 TruStain FcX (BioLegend) for 10 minutes on ice. Cells were then stained with fluorescently

706 labeled antibodies (**Supplementary Table 6**) for 30 minutes on ice and washed with staining
707 buffer to remove unbound antibodies. We used the True-Nuclear Transcription Factor Buffer Set
708 (BioLegend) for fixation, permeabilization and intracellular transcription factor antibody staining
709 (**Supplementary Table 6**) according to manufacturer's recommended protocols.

710 For stimulation assays, single cell suspensions of tissue mononuclear cells were cultured
711 in complete media in 96-well U-bottom plates (Corning) at ~2-5 million cells per well and
712 stimulated with 50 ng/ml PMA (Sigma) and 1 ug/ml Ionomycin (Sigma) in the presence of
713 GolgiStop and GolgiPlug (BD Biosciences) for 4 hours at 37°C. Cells were then washed with
714 staining buffer and stained with surface antibodies as above. For fixation, permeabilization and
715 intracellular cytokine antibody staining (**Supplementary Table 6**), we used the BD
716 Cytotfix/Cytoperm Fixation/Permeabilization Kit (BD Biosciences) as per manufacturer's
717 recommended protocols.

718 For all flow cytometry assays, we acquired cell fluorescence data using the Cytex Aurora
719 spectral flow cytometer and analyzed data using FlowJo v10.10 (BD Life Sciences). For a gating
720 strategy to identify T cell subsets refer to **Supplementary Fig. 5**.

721

722 **CRISPR-Cas9 deletion of Helios (*IKZF2*) in primary human tissue T cells**

723 Mononuclear cell suspensions from infant spleens were obtained and T cell magnetic
724 negative selection was performed as described above. For CRISPR-Cas9 deletion of Helios
725 (*IKZF2*) we used a Cas9 RNP transfection approach⁷³. Briefly, 3 Alt-R CRISPR-Cas9 crRNAs
726 targeting Helios or negative controls (**Supplementary Table 6**) were individually complexed to
727 Alt-R CRISPR-Cas9 tracrRNAs (IDT) in equimolar concentrations. Cas9 RNPs were prepared by
728 combining crRNA:tracrRNA duplexes with TrueCut Cas9 Protein v2 (Thermo Fisher Scientific)

729 at a molar ratio of 3:2. RNP nucleofection of T cells was performed using a Lonza 4D-Nucleofector
730 X unit using the Lonza P3 Primary Cell 4D-Nucleofector X Kit with 3 ul of each RNP complex in
731 20 ul of nucleofection buffer and a pulse code of EH100. Nucleofected T cells were immediately
732 placed in complete media with 5% human AB serum (Millipore Sigma) and cultured for 3 days at
733 37°C in an incubator for 3 days. Helios protein expression in Helios-KO T cells or negative controls
734 was assessed after 3 days in culture by flow cytometry as described above. To assess the effects of
735 Helios-KO on the T cell transcriptome, we first removed dead cells (Miltenyi Biotech) and either
736 rested Helios-KO cells or negative controls overnight in complete media or stimulated cells with
737 anti-CD3 and anti-CD28 and performed scRNA-seq as above.

738 We processed scRNA-seq data, identified T cells and categorized cells into T cell subsets
739 using the Naïve Bayes classifier described above. We performed differential expression analysis
740 between Helios-KO and negative control T cells in the activated condition, for each T cell subset
741 with in each donor individually. We used the same method as above for identifying differentially
742 expressed genes between KO and negative control cells (equalized cell numbers and counts;
743 differential expression by Wilcoxon with tie correction). For visualizing differentially expressed
744 genes (averaged FDR adjusted p-value < 0.05, averaged log₂-fold change > 1) in volcano plots in
745 **Fig. 6e**, we plotted averaged FDR-adjusted p-values and log-fold changes from both donors. Only
746 genes that were differentially expressed between KO and negative control samples in both donors
747 were plotted in **Fig. 6f**.

748

749 **Statistical Analysis**

750 Descriptive analyses and statistical testing of flow cytometry data were performed using
751 GraphPad Prism (v9.5.2) and comparisons between groups were made using statistical tests

752 indicated in the figure legends. We considered comparisons as statistically significant for $p < 0.05$.
753 For multivariate linear regression analysis between scHPF factor gene scores and relevant binary
754 covariates (age, tissue type, subset, lineage, activation), we performed an ordinary least squares
755 regression (statsmodels.OLS). Factor gene scores were z-scored for input and resulting regression
756 coefficients were plotted for each factor and covariate.

757

758 **Data and Code Availability**

759 Raw scRNA-seq data and metadata have been deposited in GEO (accession number GSE195844).
760 Original source code with tutorials for scHPF, which is used to build individual models in
761 consensus scHPF can be found at <https://github.com/simslab/scHPF>. Code for running consensus-
762 scHPF along with helper scripts and instructions can be found
763 at https://github.com/simslab/consensus_scHPF_wrapper.

764

765 **Acknowledgements**

766 We extend our gratitude to the families of organ donors whose gift of donation made this work
767 possible. We acknowledge the dedication and exceptional efforts of transplant coordinators and
768 staff of LiveOnNY and the network for Pancreatic Organ Donors with Diabetes (nPOD) for tissue
769 acquisition. This work was supported by US National Institutes of Health (NIH) grants to D.L.F.
770 (AI106697, AI168634), D.L.F. and P.A.Si. (AI128949), and Helmsley Charitable Trust to D.L.F.
771 and T.M.B. P.A.Sz. was supported by a Canadian Institutes of Health Research Fellowship and an
772 AAI Intersect Fellowship. T.J.C. was supported by a K23 (AI141686) Award. J.J. was supported
773 by the NIH Ruth L. Kirschstein Institutional National Research Service Award (5T32GM145440).

774 D.P.C. was supported by the Columbia University Graduate Training Program in Microbiology
775 and Immunology (T32AI106711).

776

777 Research reported in this publication was performed in the Columbia Center for Translational
778 Immunology Flow Cytometry Core (supported by NIH S10OD030282) and the Sulzberger
779 Columbia Genome Center and Single Cell Analysis core (supported by NIH P30CA013696).
780 Experimental schematics in **Fig. 1a** and **Fig. 6b** were created with BioRender.com. Content is
781 solely the responsibility of the authors and does not necessarily represent official views of the NIH.

782

783 **Author Contributions**

784 Conceptualization: P.A.Sz, D.L.F., P.A.Si; Data Curation: P.A.Sz, H.M.L., T.J.C., P.A.Si;
785 Formal Analysis: P.A.Sz, H.M.L., P.A.Si; Funding Acquisition: D.L.F., P.A.Si; Investigation:
786 P.A.Sz, H.M.L, T.J.C., D.C., J.J., P.T., R.G., D.P.C., J.G., R.M., M.K; Methodology: P.A.Sz,
787 H.M.L., P.A.Si; Resources: T.J.C., M.B., T.M.B., D.L.F., P.A.Si; Project Administration: D.L.F.,
788 P.A.Si; Supervision: D.L.F., P.A.Si; Visualization: P.A.Sz., D.P.C., P.A.Si; Writing – original
789 draft: P.A.Sz, D.L.F., P.A.Si; Writing – review & editing: all authors.

790

791

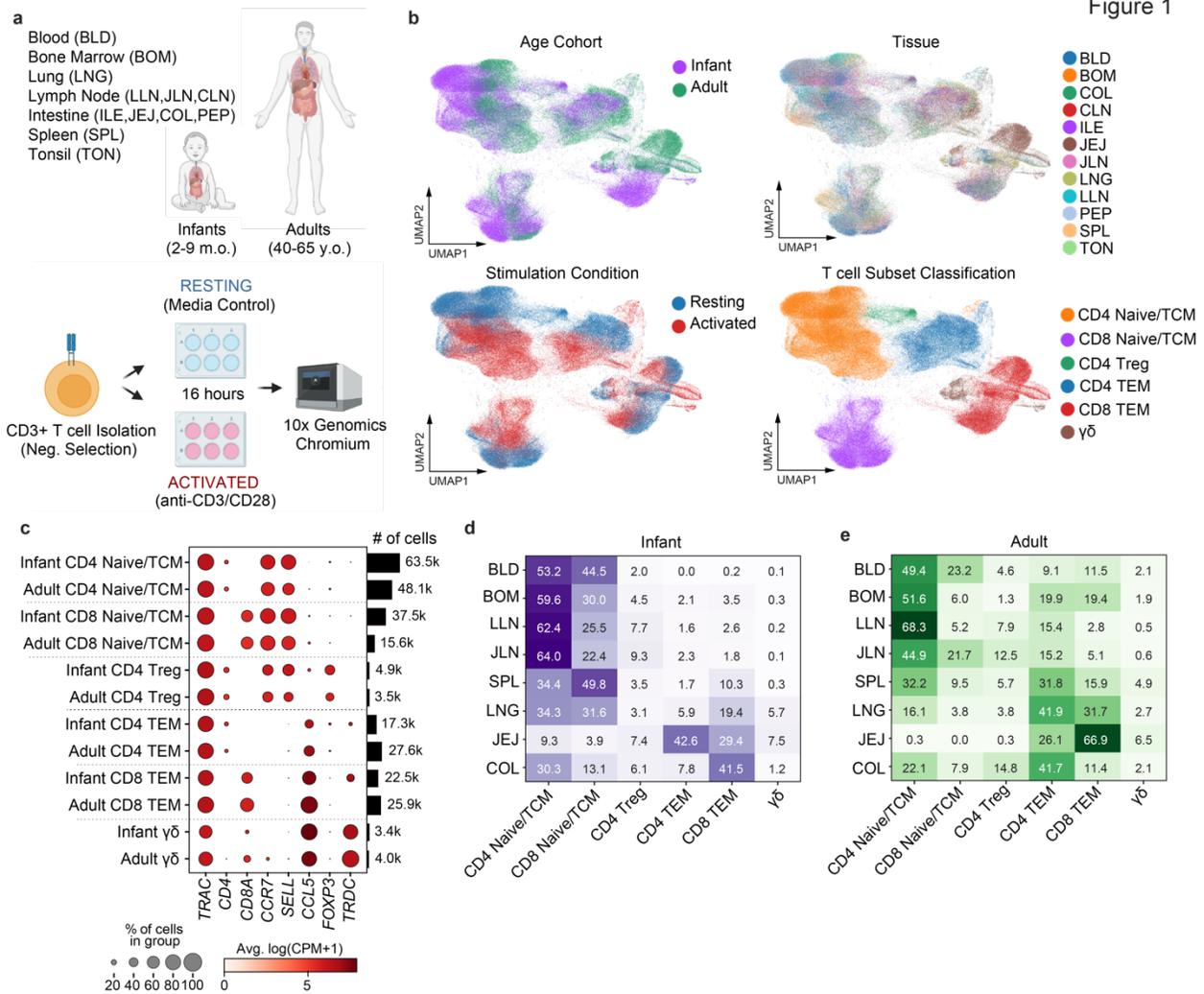
792

793

794

795

796 FIGURES



797

798 **Figure 1: A single cell transcriptomic map of resting and activated T cells from human**

799 **lymphoid and mucosal tissues. a) Schematic of scRNA-seq experimental design and workflow**

800 including T cell isolation from infant and adult tissues using negative selection, overnight rest or

801 TCR-mediated stimulation with anti-CD3 and anti-CD28 antibodies, and single cell encapsulation

802 with the 10x Genomics Chromium system. **b) UMAP embeddings of merged scRNA-seq profiles**

803 from resting and activated T cells from all samples, colored by age cohort, tissue of origin,

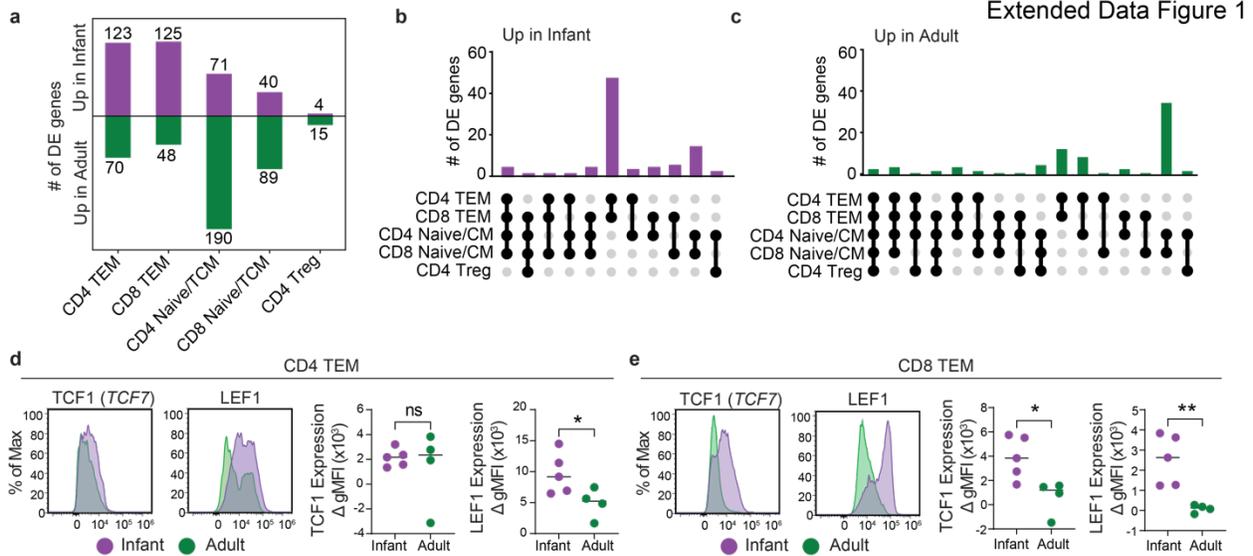
804 stimulation condition and T cell subset classification. **c) Dot plot displaying expression of T cell**

805 lineage defining markers across T cell subsets for infants and adults. Color intensity reflects mean

806 gene expression by group and dot size reflects percentage of cells in each group expressing

807 indicated marker genes. Number of cells in each T cell subset across all donors and tissues for each

808 age cohort is indicated in the bar plot on the right. **d)** Heatmap indicating percentage of cells for
809 each T cell subset within a tissue (i.e., row-normalized) for infants where color intensity indicates
810 higher frequencies. **e)** Heatmap same as **(d)** but for adults. Abbreviations: blood (BLD), bone
811 marrow (BOM), colon (COL), colon-associated lymph node (CLN), ileum (ILE), jejunum (JEJ),
812 jejunum-associated lymph node (JLN), lung (LNG), lung-associated lymph node (LLN), Peyer's
813 patch (PEP), spleen (SPL), tonsil (TON).
814



815

816

817

818

819

820

821

822

823

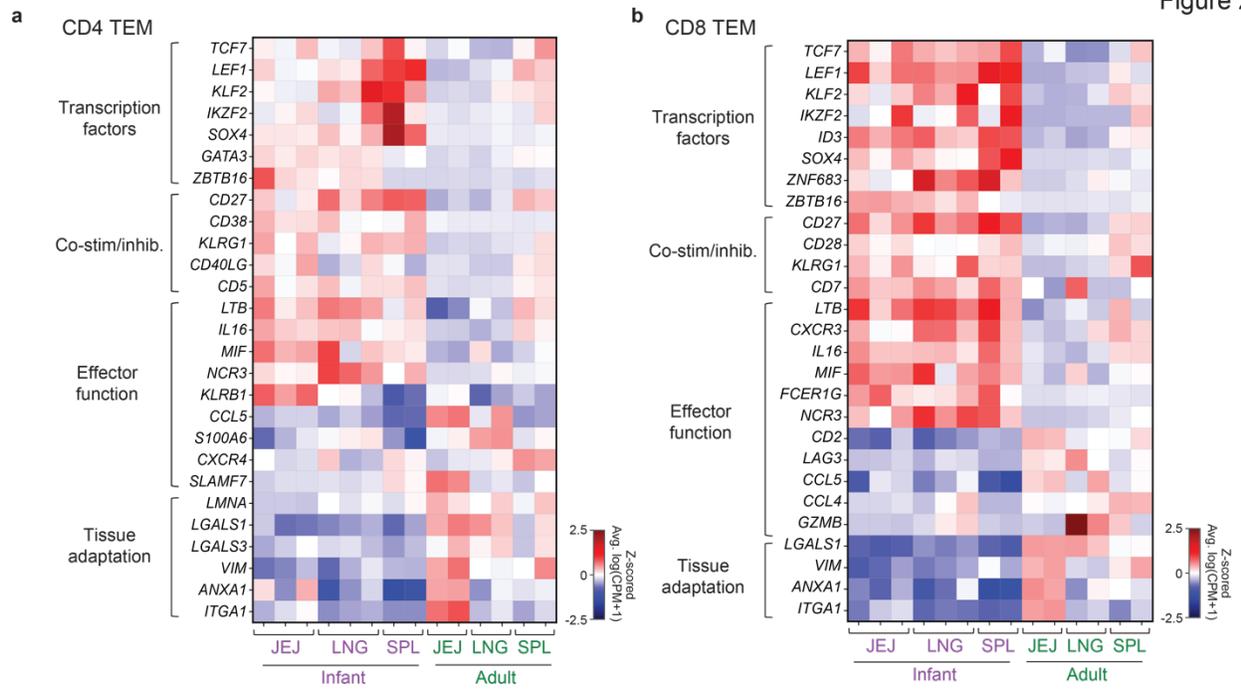
824

825

826

827

Extended Data Figure 1: Differential gene expression analysis and TF staining by flow cytometry across infant and adult T cells. **a**) Bar plot showing number of differentially expressed genes upregulated in infants or adults by T cell subset across all tissues with adequate cell representation (see **Methods**). **b**) UpSet plot showing number of differentially expressed genes in infants relative to adults shared across at least two T cell subsets. **c**) UpSet plot similar to **(b)** but for differentially expressed genes upregulated in adults. **d**) Representative histograms showing TCF1 or LEF1 expression and quantification of geometric mean fluorescence intensity relative to isotype controls (Δ gMFI) in infant ($n = 5$) and adult ($n = 4$) CD4⁺ TEM from the spleen by flow cytometry. **e**) Representative histograms and quantification for TCF1 and LEF1 same as **(d)** but for CD8⁺ TEM. For panels **(d)** and **(e)**, statistical comparisons between indicated groups made by Student's *t*-test; "ns" denotes not significant, * $p < 0.05$, and ** $p < 0.01$.



828

829 **Figure 2: Differentially expressed genes across infant and adult TEM across tissues. a)**

830 Heatmap of selected differentially expressed genes in infant versus adult CD4⁺ TEM from

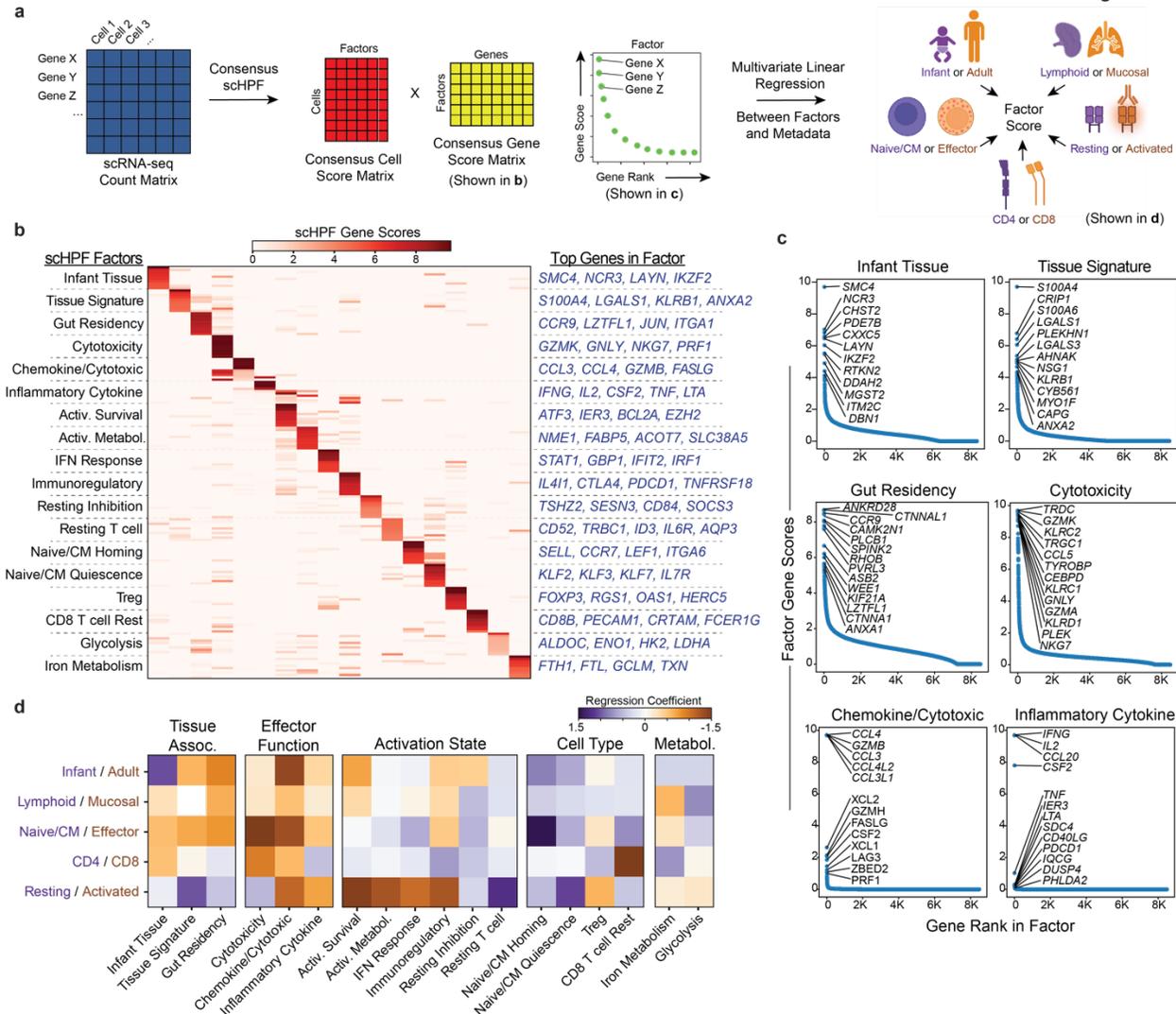
831 indicated tissues in the resting condition. Color intensity reflects Z-scored average log(counts per

832 million+1) expression by row. **b)** Heatmap of differentially expressed genes in infants and adults

833 as in (a) but for CD8⁺ TEM.

834

Figure 3



835

836 **Figure 3: Consensus single cell Hierarchical Poisson Factorization (scHPF) reveals**

837 **transcriptional co-expression patterns in infant and adult T cells. a)** Schematic of consensus-

838 scHPF analysis to identify co-expression patterns or “factors” across the dataset and multivariate

839 linear regression to associate factors with metadata. **b)** Heatmap showing scHPF gene scores of

840 the top 10 genes in each factor. Selected genes for each factor are indicated to the right and a

841 ranked list of the top 100 genes in all factors is provided in **Supplementary Table 3**. **c)** Dot plots

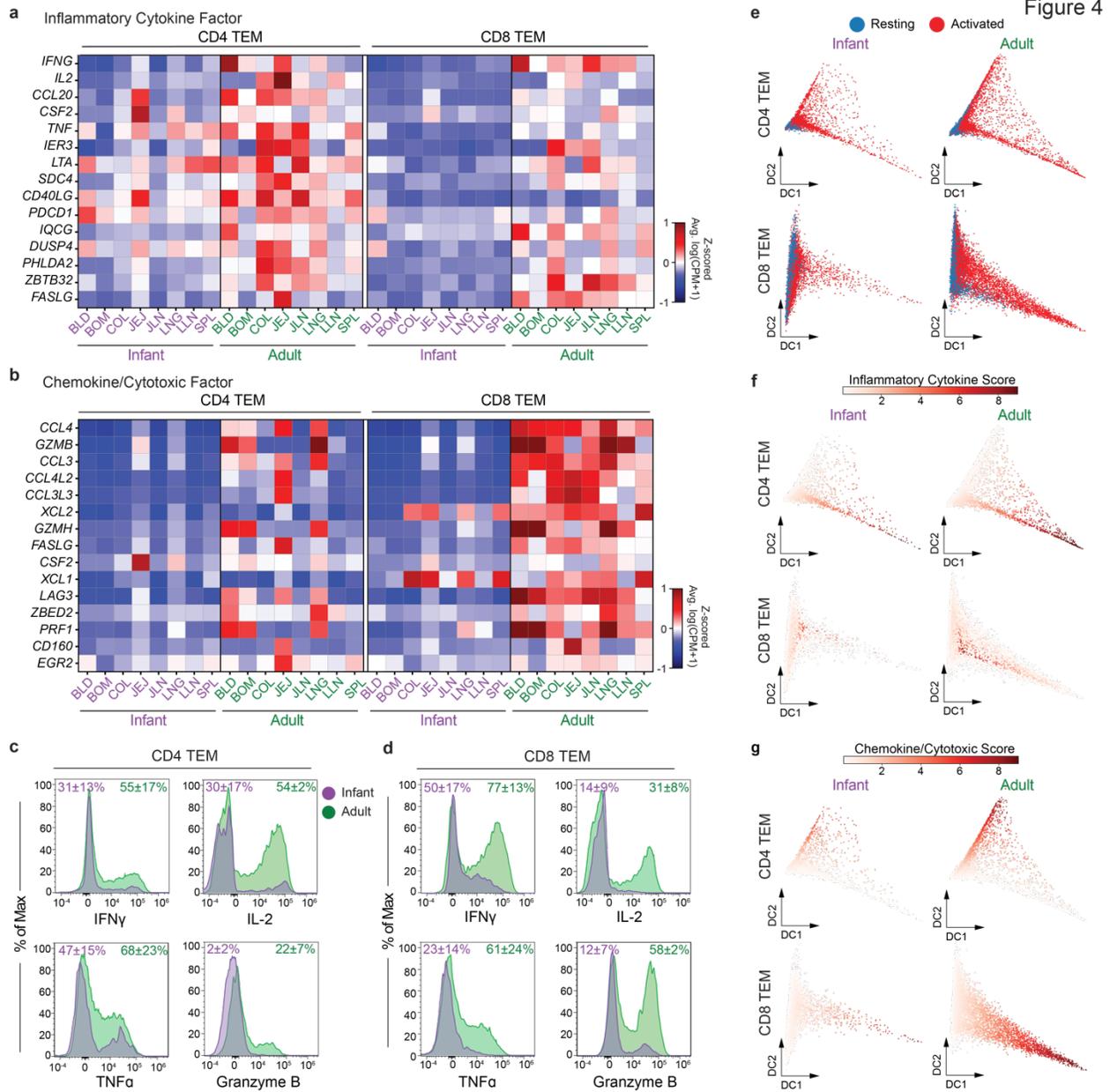
842 showing the rank and gene score for genes in selected scHPF factors with labels for the top genes.

843 **d)** Heatmap of regression coefficients for multivariate linear regression between scHPF factor cell

844 scores and covariates: age cohort (infant/adult), tissue localization (lymphoid/mucosal), T cell

845 subset (naïveCM/effector), T cell lineage (CD4/CD8), or activation conditions (resting/activated).

846



847

848 **Figure 4: T cell activation capacity and effector function of CD4⁺ and CD8⁺ TEM in**

849 **infants and adults. a)** Heatmap showing z-scored average gene expression as log(counts per

850 million+1) of the top 15 genes in the Inflammatory Cytokine factor in CD4⁺ and CD8⁺ TEM across

851 tissues in infants and adults from activated conditions. **b)** Heatmap as in (a) but showing the top

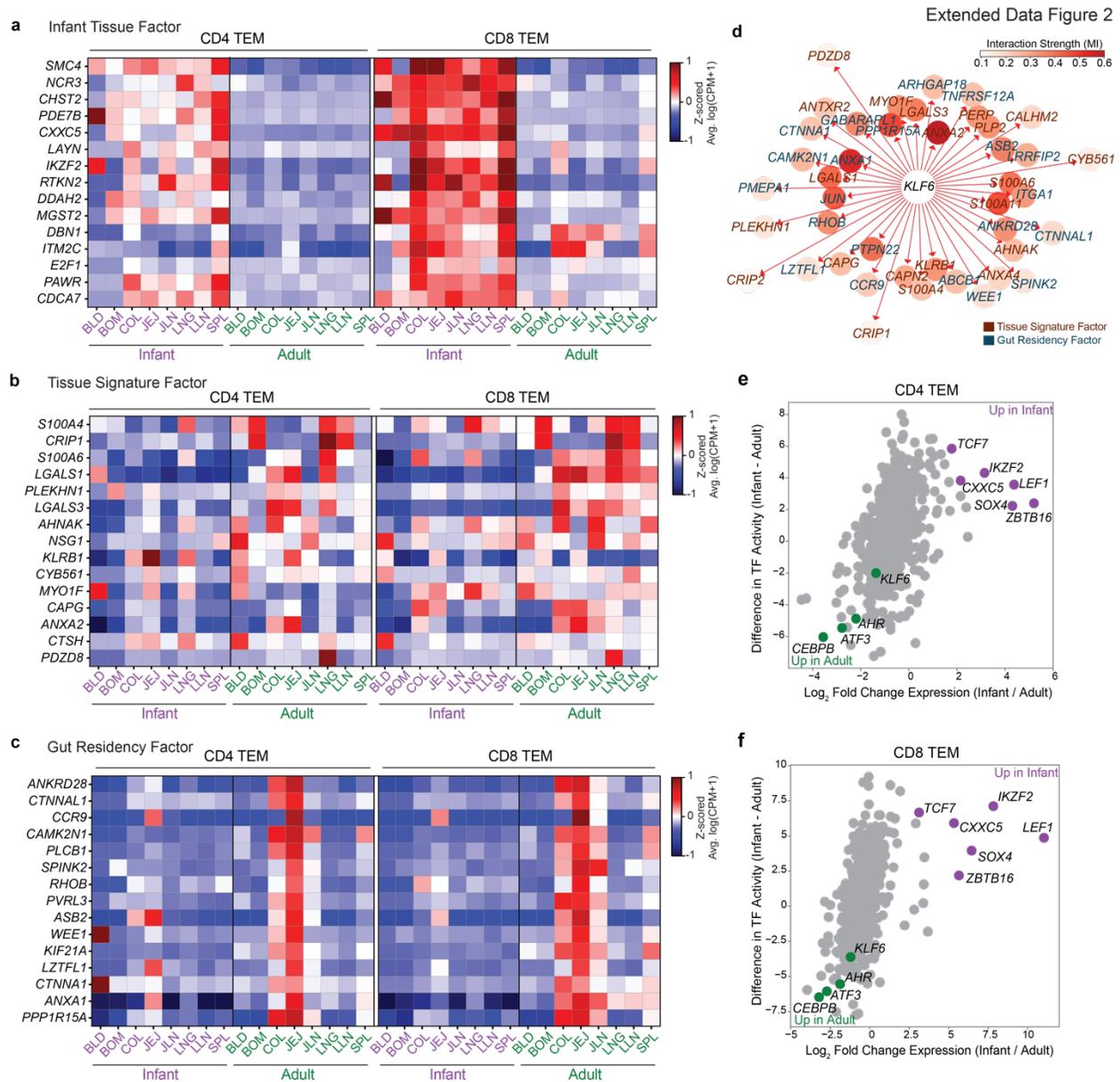
852 15 genes in the Chemokine/Cytotoxic factor. **c)** Representative histograms of effector molecule

853 expression in infant and adult CD4⁺ TEM from the spleen assessed by intracellular flow cytometry

854 staining. Cytokine (IFN γ , IL-2 and TNF α) expression was evaluated after 4-hour stimulation with

855 PMA and Ionomycin, while granzyme B expression was evaluated in unstimulated conditions.

856 Values represent mean +/- standard deviation of percent positive cells of each target across 3
857 individual donors in each age cohort. **d)** Representative histograms and percent expression values
858 as in **(c)** but for CD8⁺ TEM. **e)** Diffusion maps of CD4⁺ and CD8⁺ TEM from mucosal tissues
859 (jejunum and lung), with cells colored by activation condition as resting (blue) or activated (red).
860 **f, g)** Diffusion maps as in **(e)** but colored by cell scores for the Inflammatory Cytokine factor **(f)**
861 or Chemokine Cytotoxic factor **(g)** for CD4⁺ TEM (*top*) or CD8⁺ TEM (*bottom*).
862



863

864 **Extended Data Figure 2: Expression of tissue-associated transcriptional programs in infants**

865 **and adults. a)** Heatmap showing Z-scored average gene expression as log(counts per million+1)

866 of the top 15 genes in the Infant Tissue factor in CD4⁺ and CD8⁺ EM across tissues in infants and

867 adults. **b, c)** Heatmap same as **(a)** but for the top genes in the Tissue Signature factor **(b)** and Gut

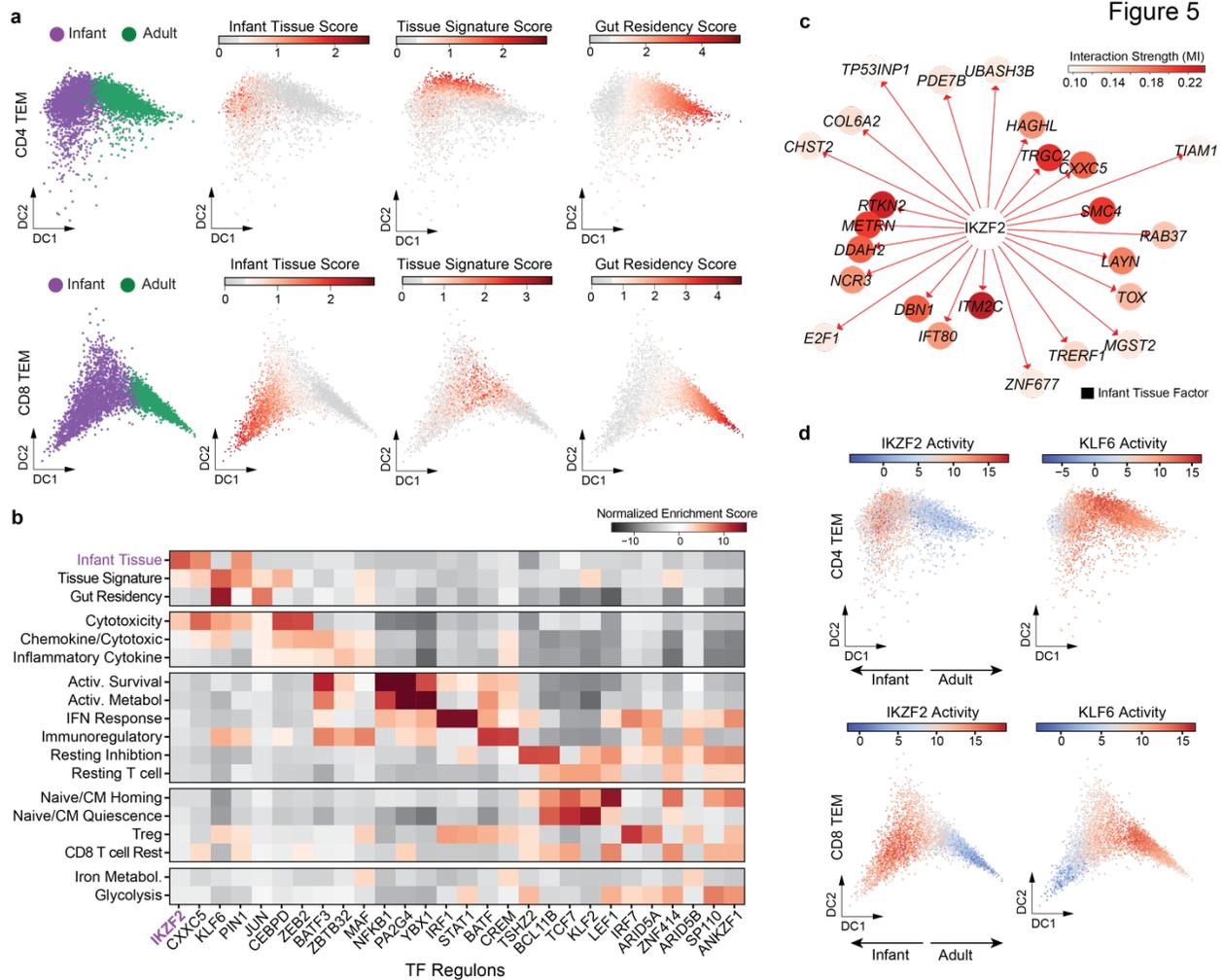
868 Residency factor **(c)**. **d)** TF network for KLF6 showing overlap of inferred targets within the top

869 50 genes in the Tissue Signature (*deep red*) and Gut Residency (*deep blue*) factors. Node color

870 intensity and inverse distance in the network represents interaction strength (mutual information)

871 between TF and target gene. **e, f)** Dot plots showing differences in TF activity (infant - adult) and

872 \log_2 fold change in TF gene expression (infant / adult) in CD4⁺ (e) and CD8⁺ (f) TEM from the
873 jejunum.
874



875

876 **Figure 5: Gene regulatory network reconstruction to identify transcription factors driving**

877 **tissue adaptation in infant and adult tissue T cells. a)** Diffusion maps of CD4⁺ and CD8⁺ TEM

878 from the jejunum with cells colored by age cohort as infant (*purple*) and adult (*green*) or by cell

879 scores for the indicated scHPF factors. **b)** Heatmap showing normalized enrichment score from

880 GSEA between the genes in each TF's regulon and the ranked list of genes for each scHPF factor.

881 The top two TFs with the highest scores for each factor (excluding duplicates) are shown. **c)** TF

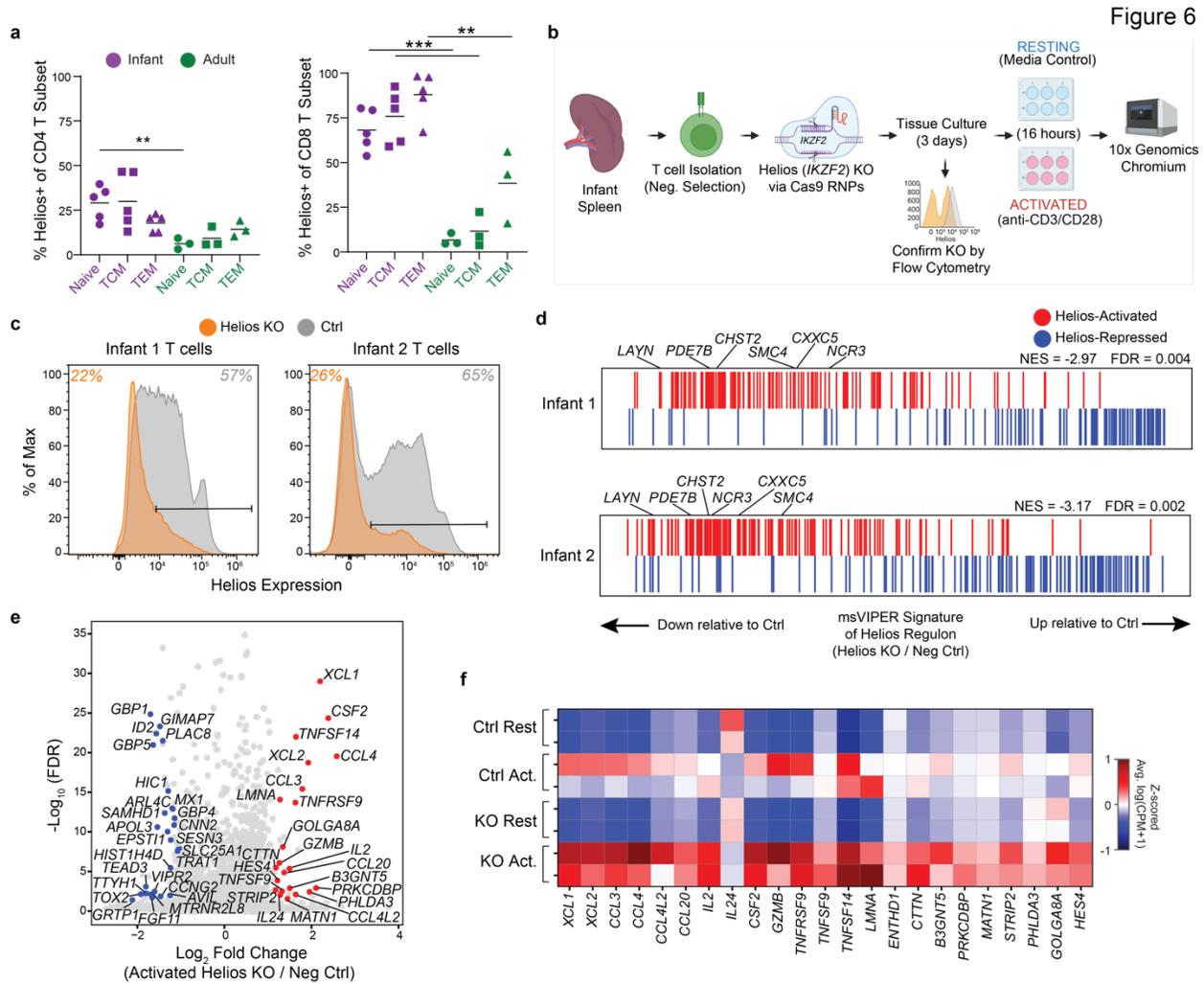
882 network of Helios (*IKZF2*) showing inferred TF targets overlapping with the top 50 genes in the

883 Infant Tissue factor. Node color intensity and inverse distance in the network represents interaction

884 strength (mutual information) between TF and target gene. **d)** Diffusion maps from (a) but colored

885 by Helios (*IKZF2*) or KLF6 activity determined by VIPER (see **Methods**).

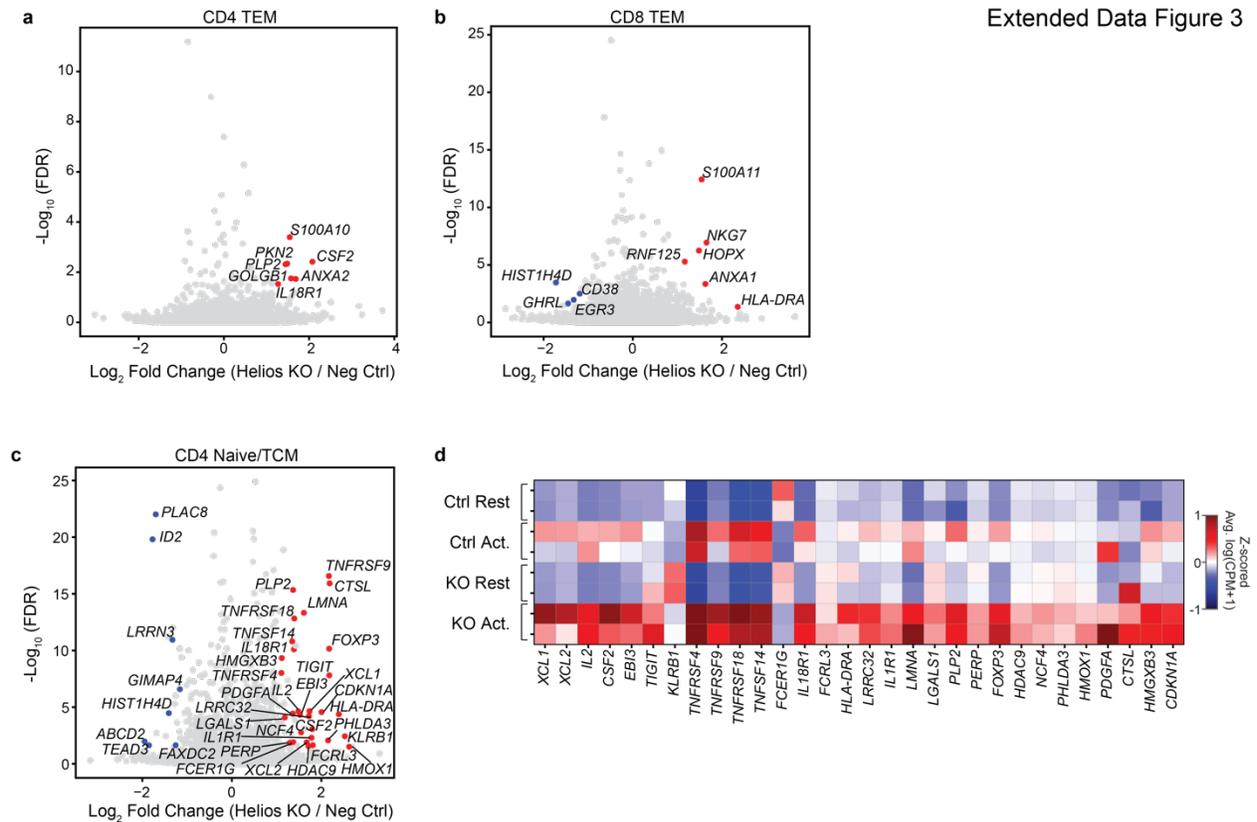
886



887

888 **Figure 6: Helios expression and function in infant T cells.** **a)** Percentage of Helios-expressing
 889 cells among conventional ($CD3^+ \gamma\delta TCR^- FOXP3^-$) $CD4^+$ and $CD8^+$ T cell subsets from the spleen
 890 in infant ($n = 5$) and adults ($n = 3$) by flow cytometry. Statistical comparisons between infant and
 891 adult subsets made by Students' t -test, where ** denotes $p < 0.01$, *** denotes $p < 0.001$. **b)**
 892 Schematic of CRISPR-Cas9 knockout (KO) of Helios (*IKZF2*) in infant splenic T cells after 3 days
 893 in culture, overnight rest or activation with anti-CD3 and anti-CD28 antibodies, and single cell
 894 sequencing. **c)** Histograms showing protein expression of Helios among KO and negative control
 895 infant splenic T cells ($\gamma\delta TCR^- FOXP3^- CD3^+$ cells) as determined by intracellular flow cytometry.
 896 Percentage of Helios-expressing cells in each group indicated on the top for both infant donors. **d)**
 897 msVIPER plots showing relative expression of genes from the Helios regulon (from ARACNe) in
 898 Helios KO relative to negative control cells by scRNA-seq. Genes positively regulated by Helios
 899 ("activated") are in red and genes negatively regulated by Helios ("repressed") are in blue. Top 6

900 genes from the Infant Tissue factor are labeled. **e)** Volcano plot showing FDR-adjusted p -value
901 and \log_2 fold change in gene expression between Helios-KO and negative control CD8⁺
902 naïve/TCM from infant spleen from the TCR-stimulated condition. Data is averaged over both
903 donor experiments for plotting and colored for genes with an FDR adjusted p -value < 0.05 and
904 \log_2 fold change >1 (red) or <1 (blue) in both donors. **f)** Heatmap showing Z-scored average gene
905 expression as $\log(\text{counts per million}+1)$ of up-regulated genes in Helios-KO versus negative
906 control CD8⁺ naïve/TCM from the spleen in both infant donors.
907



908

909 **Extended Data Figure 3: Differential expression in Helios-KO versus negative control infant**

910 **T cells. a,b)** Volcano plots showing FDR-adjusted *p*-value and log₂ fold change in gene expression

911 between Helios KO and negative control CD4⁺ TEM **(a)** and CD8⁺ TEM **(b)** from infant spleen in

912 the activated condition. Data are averaged over both donor experiments for plotting and colored

913 for genes with an FDR adjusted *p*-value < 0.05 and log₂ fold change >1 (red) or <1 (blue) in both

914 donors. **c)** Volcano plot as in **(a,b)** but for CD4⁺ naïve/TCM. **d)** Heatmap showing Z-scored

915 average gene expression as log(counts per million+1) of up-regulated genes in Helios-KO versus

916 negative control CD4⁺ naïve/TCM from the spleen in both infant donors.

917

918

919

920

921

922 **REFERENCES**

- 923 1 Semmes, E. C. *et al.* Understanding Early-Life Adaptive Immunity to Guide Interventions
924 for Pediatric Health. *Front Immunol* **11**, 595297.
- 925 2 Watson, N. B. *et al.* The gene regulatory basis of bystander activation in CD8(+) T cells.
926 *Sci Immunol* **9**, eadf8776.
- 927 3 Smith, N. L. *et al.* Developmental Origin Governs CD8(+) T Cell Fate Decisions during
928 Infection. *Cell* **174**, 117-130 e114.
- 929 4 Rudd, B. D. Neonatal T Cells: A Reinterpretation. *Annu Rev Immunol* **38**, 229-247.
- 930 5 Forsthuber, T., Yip, H. C. & Lehmann, P. V. Induction of TH1 and TH2 immunity in
931 neonatal mice. *Science* **271**, 1728-1730.
- 932 6 Chen, N. & Field, E. H. Enhanced type 2 and diminished type 1 cytokines in neonatal
933 tolerance. *Transplantation* **59**, 933-941.
- 934 7 Adkins, B., Bu, Y. & Guevara, P. The generation of Th memory in neonates versus adults:
935 prolonged primary Th2 effector function and impaired development of Th1 memory
936 effector function in murine neonates. *J Immunol* **166**, 918-925.
- 937 8 Gibbons, D. *et al.* Interleukin-8 (CXCL8) production is a signatory T cell effector function
938 of human newborn infants. *Nat Med* **20**, 1206-1210.
- 939 9 Thapa, P. *et al.* Infant T cells are developmentally adapted for robust lung immune
940 responses through enhanced T cell receptor signaling. *Sci Immunol* **6**, eabj0789.
- 941 10 Smith, N. L. *et al.* Rapid proliferation and differentiation impairs the development of
942 memory CD8+ T cells in early life. *J Immunol* **193**, 177-184.
- 943 11 Zens, K. D. *et al.* Reduced generation of lung tissue-resident memory T cells during
944 infancy. *J Exp Med* **214**, 2915-2932.
- 945 12 Reynaldi, A. *et al.* Modeling the dynamics of neonatal CD8(+) T-cell responses. *Immunol*
946 *Cell Biol* **94**, 838-848.
- 947 13 Wissink, E. M., Smith, N. L., Spektor, R., Rudd, B. D. & Grimson, A. MicroRNAs and
948 Their Targets Are Differentially Regulated in Adult and Neonatal Mouse CD8+ T Cells.
949 *Genetics* **201**, 1017-1030.
- 950 14 Thome, J. J. *et al.* Early-life compartmentalization of human T cell differentiation and
951 regulatory function in mucosal and lymphoid tissues. *Nat Med* **22**, 72-77.
- 952 15 Senda, T. *et al.* Microanatomical dissection of human intestinal T-cell immunity reveals
953 site-specific changes in gut-associated lymphoid tissues over life. *Mucosal Immunol* **12**,
954 378-389.

- 955 16 Connors, T. J. *et al.* Site-specific development and progressive maturation of human tissue-
956 resident memory T cells over infancy and childhood. *Immunity* **56**, 1894-1909 e1895.
- 957 17 Schreurs, R. *et al.* Intestinal CD8(+) T cell responses are abundantly induced early in
958 human development but show impaired cytotoxic effector capacities. *Mucosal Immunol*
959 **14**, 605-614.
- 960 18 Levitin, H. M., Zhao, W., Bruce, J. N., Canoll, P. & Sims, P. A. Consensus scHPF Identifies
961 Cell Type-Specific Drug Responses in Glioma by Integrating Large-Scale scRNA-seq.
962 *bioRxiv*, 2023.2012.2005.570193.
- 963 19 Szabo, P. A. *et al.* Single-cell transcriptomics of human T cells reveals tissue and activation
964 signatures in health and disease. *Nat Commun* **10**, 4706.
- 965 20 Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat*
966 *Biotechnol.*
- 967 21 Caron, D. P. *et al.* Multimodal hierarchical classification of CITE-seq data delineates
968 immune cell states across lineages and tissues. *bioRxiv*, 2023.2007.2006.547944.
- 969 22 Kratchmarov, R., Magun, A. M. & Reiner, S. L. TCF1 expression marks self-renewing
970 human CD8(+) T cells. *Blood Adv* **2**, 1685-1690.
- 971 23 Weinreich, M. A. *et al.* KLF2 transcription-factor deficiency in T cells results in
972 unrestrained cytokine production and upregulation of bystander chemokine receptors.
973 *Immunity* **31**, 122-130.
- 974 24 Gattinoni, L., Klebanoff, C. A. & Restifo, N. P. Paths to stemness: building the ultimate
975 antitumour T cell. *Nat Rev Cancer* **12**, 671-684.
- 976 25 Sinner, D. *et al.* Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity
977 and proliferation of colon carcinoma cells. *Mol Cell Biol* **27**, 7802-7815.
- 978 26 Lam, A. J., Uday, P., Gillies, J. K. & Levings, M. K. Helios is a marker, not a driver, of
979 human Treg stability. *Eur J Immunol* **52**, 75-84.
- 980 27 Ng, M. S. F., Roth, T. L., Mendoza, V. F., Marson, A. & Burt, T. D. Helios enhances the
981 preferential differentiation of human fetal CD4(+) naive T cells into regulatory T cells. *Sci*
982 *Immunol* **4**.
- 983 28 Zheng, W. & Flavell, R. A. The transcription factor GATA-3 is necessary and sufficient
984 for Th2 cytokine gene expression in CD4 T cells. *Cell* **89**, 587-596.
- 985 29 Szabo, P. A. Axes of heterogeneity in human tissue-resident memory T cells. *Immunol Rev*
986 **316**, 23-37.
- 987 30 Yang, C. Y. *et al.* The transcriptional regulators Id2 and Id3 control the formation of
988 distinct memory CD8+ T cell subsets. *Nat Immunol* **12**, 1221-1229.

- 989 31 Correia, M. P. *et al.* Distinct human circulating NKp30(+)FcepsilonRIgamma(+)CD8(+) T cell population exhibiting high natural killer-like antitumor potential. *Proc Natl Acad Sci U S A* **115**, E5980-E5989.
- 990
991
- 992 32 Poon, M. M. L. *et al.* Tissue adaptation and clonal segregation of human memory T cells in barrier sites. *Nat Immunol*.
- 993
- 994 33 Zhao, X., Shan, Q. & Xue, H. H. TCF1 in T cell immunity: a broadened frontier. *Nat Rev Immunol* **22**, 147-157.
- 995
- 996 34 Kobiela, A. & Fuchs, E. Alpha-catenin: at the junction of intercellular adhesion and actin dynamics. *Nat Rev Mol Cell Biol* **5**, 614-625.
- 997
- 998 35 Dean, J. W. *et al.* The aryl hydrocarbon receptor cell intrinsically promotes resident memory CD8(+) T cell differentiation and function. *Cell Rep* **42**, 111963.
- 999
- 1000 36 Chen, Y., Shen, J., Kasmani, M. Y., Topchyan, P. & Cui, W. Single-Cell Transcriptomics Reveals Core Regulatory Programs That Determine the Heterogeneity of Circulating and Tissue-Resident Memory CD8(+) T Cells. *Cells* **10**.
- 1001
1002
- 1003 37 Lachmann, A., Giorgi, F. M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **32**, 2233-2235.
- 1004
1005
- 1006 38 Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7 Suppl 1**, S7.
- 1007
- 1008 39 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550.
- 1009
1010
- 1011 40 Crouse, J., Kalinke, U. & Oxenius, A. Regulation of antiviral T cell responses by type I interferons. *Nat Rev Immunol* **15**, 231-242.
- 1012
- 1013 41 Sumida, T. S. *et al.* Type I interferon transcriptional network regulates expression of coinhibitory receptors in human T cells. *Nat Immunol* **23**, 632-642.
- 1014
- 1015 42 Hayden, M. S. & Ghosh, S. NF-kappaB in immunobiology. *Cell Res* **21**, 223-244.
- 1016 43 Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **434**, 772-777.
- 1017
- 1018 44 Wang, Z. *et al.* DNA methylation impairs TLR9 induced Foxp3 expression by attenuating IRF-7 binding activity in fulminant type 1 diabetes. *J Autoimmun* **41**, 50-59.
- 1019
- 1020 45 Omilusik, K. D. *et al.* Transcriptional repressor ZEB2 promotes terminal differentiation of CD8+ effector and memory T cell populations during infection. *J Exp Med* **212**, 2027-2039.
- 1021
1022

- 1023 46 Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using
1024 network-based inference of protein activity. *Nat Genet* **48**, 838-847.
- 1025 47 Cannarile, M. A. *et al.* Transcriptional regulator Id2 mediates CD8+ T cell immunity. *Nat*
1026 *Immunol* **7**, 1317-1325.
- 1027 48 Slade, C. D., Reagin, K. L., Lakshmanan, H. G., Klonowski, K. D. & Watford, W. T.
1028 Placenta-specific 8 limits IFN γ production by CD4 T cells in vitro and promotes
1029 establishment of influenza-specific CD8 T cells in vivo. *PLoS One* **15**, e0235706.
- 1030 49 Crowl, J. T. *et al.* Tissue-resident memory CD8(+) T cells possess unique transcriptional,
1031 epigenetic and functional adaptations to different tissue environments. *Nat Immunol* **23**,
1032 1121-1131.
- 1033 50 Seo, H. *et al.* TOX and TOX2 transcription factors cooperate with NR4A transcription
1034 factors to impose CD8(+) T cell exhaustion. *Proc Natl Acad Sci U S A* **116**, 12410-12415.
- 1035 51 Escobar, G., Mangani, D. & Anderson, A. C. T cell factor 1: A master regulator of the T
1036 cell response in disease. *Sci Immunol* **5**.
- 1037 52 ElTanbouly, M. A. & Noelle, R. J. Rethinking peripheral T cell tolerance: checkpoints
1038 across a T cell's journey. *Nat Rev Immunol* **21**, 257-267.
- 1039 53 Zens, K. D., Chen, J. K. & Farber, D. L. Vaccine-generated lung tissue-resident memory
1040 T cells provide heterosubtypic protection to influenza infection. *JCI Insight* **1**.
- 1041 54 Hu, G. & Chen, J. A genome-wide regulatory network identifies key transcription factors
1042 for memory CD8(+) T-cell development. *Nat Commun* **4**, 2830.
- 1043 55 Kuwahara, M. *et al.* The transcription factor Sox4 is a downstream target of signaling by
1044 the cytokine TGF-beta and suppresses T(H)2 differentiation. *Nat Immunol* **13**, 778-786.
- 1045 56 Hossain, M. B. *et al.* Lymphoid enhancer factor interacts with GATA-3 and controls its
1046 function in T helper type 2 cells. *Immunology* **125**, 377-386.
- 1047 57 Ikushima, H. *et al.* Autocrine TGF-beta signaling maintains tumorigenicity of glioma-
1048 initiating cells through Sry-related HMG-box factors. *Cell Stem Cell* **5**, 504-514.
- 1049 58 Bohacova, P. *et al.* Multidimensional profiling of human T cells reveals high CD38
1050 expression, marking recent thymic emigrants and age-related naive T cell remodeling.
1051 *Immunity* **57**, 2362-2379 e2310.
- 1052 59 Staton, T. L. *et al.* CD8+ recent thymic emigrants home to and efficiently repopulate the
1053 small intestine epithelium. *Nat Immunol* **7**, 482-488.
- 1054 60 Schreurs, R. *et al.* Human Fetal TNF-alpha-Cytokine-Producing CD4(+) Effector Memory
1055 T Cells Promote Intestinal Development and Mediate Inflammation Early in Life.
1056 *Immunity* **50**, 462-476 e468.

- 1057 61 Chung, H. K. *et al.* Multi-Omics Atlas-Assisted Discovery of Transcription Factors for
1058 Selective T Cell State Programming. *bioRxiv*.
- 1059 62 Hetemaki, I. *et al.* Loss-of-function mutation in IKZF2 leads to immunodeficiency with
1060 dysregulated germinal center reactions and reduction of MAIT cells. *Sci Immunol* **6**,
1061 eabe3454.
- 1062 63 Neyens, D. *et al.* HELIOS-expressing human CD8 T cells exhibit limited effector
1063 functions. *Front Immunol* **14**, 1308539.
- 1064 64 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21.
- 1065 65 Griffiths, J. A., Richard, A. C., Bach, K., Lun, A. T. L. & Marioni, J. C. Detection and
1066 removal of barcode swapping in single-cell RNA-seq data. *Nat Commun* **9**, 2667.
- 1067 66 Zhao, W. *et al.* Deconvolution of cell type-specific drug responses in human tumor tissue
1068 with single-cell RNA-seq. *Genome Med* **13**, 82.
- 1069 67 Lun, A. T. L. *et al.* EmptyDrops: distinguishing cells from empty droplets in droplet-based
1070 single-cell RNA sequencing data. *Genome Biol* **20**, 63.
- 1071 68 Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like
1072 Cells that Correlate with Prognosis. *Cell* **162**, 184-197.
- 1073 69 Levitin, H. M. *et al.* De novo gene signature identification from single-cell RNA-seq with
1074 hierarchical Poisson factorization. *Mol Syst Biol* **15**, e8557.
- 1075 70 Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA
1076 sequencing data with many zero counts. *Genome Biol* **17**, 75.
- 1077 71 Dominguez Conde, C. *et al.* Cross-tissue immune cell analysis reveals tissue-specific
1078 features in humans. *Science* **376**, eabl5197.
- 1079 72 Vlahos, L. *et al.* Systematic, Protein Activity-based Characterization of Single Cell State.
1080 *bioRxiv*, 2021.2005.2020.445002.
- 1081 73 Oh, S. A., Seki, A. & Rutz, S. Ribonucleoprotein Transfection for CRISPR/Cas9-Mediated
1082 Gene Knockout in Primary T Cells. *Curr Protoc Immunol* **124**, e69.
1083