The functional and phenotypic diversity of single T-cell infiltrates in human colorectal cancer as correlated with clinical outcome

Kazuya Masuda[1], Adam Kornberg[1,2,3], Sijie Lin[1], Patricia Ho[1,3], Kerim Secener[1], Nathan Suek[1], Alyssa M. Bacarella[4], Matthew Ingham[5,6], Vilma Rosario[6,7], Ahmed M. Al-Masrou[6,7], Steven A. Lee-Kong[6,7], P. Ravi Kiran[6,7], Kelley S. Yan[8,9], Marlon Stoeckius[10], Peter Smibert[10], Paul E. Oberstein[5,6], Peter A. Sims[11], Arnold Han[,1,2,6,8*]

[1]Columbia Center for Translational Immunology, Columbia University, New York, NY

[2]Department of Microbiology and Immunology, Columbia University, New York, NY

[3]Integrated Program in Cellular, Molecular and Biomedical Studies

[4]Department of Pediatrics, Columbia University, New York, NY

[5]Department of Medicine, Division of Hematology and Oncology, Columbia University, New York, NY

[6]Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY

[7]Department of Surgery, Division of Colorectal Surgery, Columbia University, NY

[8]Department of Medicine, Digestive and Liver Diseases, Columbia University, New York, NY

[9]Department of Genetics and Development, Columbia University, New York, New York

[10]New York Genome Center, New York, NY

[11]Departments of Systems Biology and Biochemistry and Molecular Biophysics, Columbia University, New York, NY

*Correspondence: Arnold Han, ash3@cumc.columbia.edu

Abstract

Although degree of T-cell infiltration in CRC was shown to correlate with a positive prognosis, the contribution of phenotypically and functionally distinct T cell subtypes within tumors remains unclear. We analyzed 37,931 single T cells with respect to transcriptome, TCR sequence and 23 cell surface proteins, from tumors and adjacent normal colon of 16 patients. Our comprehensive analysis revealed two phenotypically distinct cytotoxic T cell populations within tumors. Its main population transitioned from one positively prognostic effector memory to the other non-prognostic resident memory population in a stage-dependent manner. We further defined several Treg subpopulations within tumors. While Tregs overall were associated with positive clinical outcomes, the $CD38^+$ peripherally-derived Treg subpopulation, phenotypically related to Th17 cells, correlated with poor outcomes independent of cancer stage. Thus, we demonstrated the prognostic significance of single $CD8^+$ and $CD4^+$ T-cell infiltrates in CRC, which could inform therapeutic strategies.

Introduction

Landmark studies performed on histological samples correlated T cell-infiltration with cancer survival in colorectal cancer (CRC), a highly prevalent cancer and leading cause of cancer deaths worldwide[1,2]. However, the success of immunotherapy in CRC has been largely limited to a small subset of CRC with high microsatellite instability (MSI-H), characterized by higher degree of mutation due to defective DNA mismatch repair and accounting for ~15% of CRC[3-7]. Several traits of MSI-H tumors have been proposed to account for its efficacy, including increased infiltration of effector $CD45RO^{+}CD8^{+}$ T cells and T-bet^{+} lymphocytes (like T-helper 1 cells), and higher degree of PD-1-PDL1 expression[5,8,9]. However, none of these traits are unique to MSI-H CRC, which is also not universally responsive to anti-PD-1 therapy[5,10]. These observations imply the feasibility of novel T cell-directed immunotherapies to treat CRC.

In contrast to $CD8^{+}$ T cells, the contribution of $CD4^{+}$ T-cell infiltrates in CRC to clinical outcome, in particular regulatory T cells (Tregs) and T helper 17 (Th17) cells, is controversial[11-18]. In several cancers including breast, lung, gastric and hepatocellular carcinomas, Treg infiltration negatively correlates with patient survival[19-23]. In contrast, the presence of Tregs in CRC represents a conundrum with some groups concluding that Treg density predicts positive outcomes[24,25], although a recent report suggested different clinical outcomes were associated with heterogenous Tregs in CRC[26].

To probe the functional and phenotypic heterogeneity of T-cell infiltrates and its contribution to tumor growth in CRC, we performed unbiased droplet-based single-cell RNA sequencing (scRNA-seq) of T cells isolated from tumors or adjacent normal colon of 16 CRC patients[18]. We used a modified workflow for scRNA-seq incorporating oligonucleotide-tagged antibodies to assess cell surface expression levels in T cells, termed CITE-seq (cellular indexing of

73 transcriptomes and epitopes by sequencing) and also distinguish cells from different samples for
74 multiplexing, termed cell hashing[27,28].

75 Although recent scRNA-seq technologies have enabled cell-type specific RNA profiling, they still
76 lack the scalability of bulk RNA-seq[29]. We applied semi-quantitative gene set enrichment analysis
77 (GSEA) to quantify specific T cell types characterized by scRNA-seq from bulk RNA-seq
78 databases in the TCGA cohort of several types of tumors including CRC, breast cancer (BRC) and
79 melanoma[30-33]. Collectively, our study shows the dynamic relationship between phenotypically
80 and functionally distinct T cell subtypes in CRC, and demonstrates the prognostic significance of
81 certain T cell subtypes, providing additional insights regarding their role in tumor progression and
82 potentially enabling novel avenues for immunotherapies.
83
84 .
85
86
87
88
89
90
91
92
93
94
95

96 **Results**

97 **Single-cell profiling of T cells in CRC and adjacent normal colon**

98 We analyzed 37,931 single T cells in CRC and adjacent normal colon from total 16 patients
99 including its single-cell transcriptome, 23 cell surface marker expressions, and TCR sequences
100 (Fig. 1a-c, and Extended Data Fig.1a). Viable TCRαβ$^+$ CD3$^+$ T cells were sorted and single-cell
101 RNA sequencing (scRNA-seq) and TCR sequencing (TCR-seq) were performed[27] (Fig. 1a and
102 Extended Data Fig.1a). Cells from different normal colon/tumor samples or stimulated/non-
103 stimulated fractions were resolved through labeling with distinct antibody hashtags[28] (Fig. 1c,h,
104 Extended Data Fig. 1c, Methods). After stringent quality filtering, we obtained 35,145 single T
105 cell transcriptomes. TCR-seq analysis assigned annotated TCR sequences for 32,550 single T cells,
106 of which 27,999 cells had paired TCR sequences (Methods).

107 Unsupervised clustering identified 9 distinct T cell clusters consisting of non-stimulated and
108 stimulated clusters (Fig. 1b,e, Extended Data Fig. 1c). Stimulated T cell clusters (T-Stim1 and T-
109 Stim2) that produced effector molecules and cytokines such as *GZMB, IL2, TNFA, IFNG*, *IL21*
110 and *IL17A*, were composed primarily of CD8$^+$ and CD4$^+$ T cells, respectively (Fig. 1b,f, Extended
111 Data Fig.1c,d).

112 The T_Tn/Tcm (naïve/central memory) cluster was enriched for markers associated with naïve T
113 cell phenotypes (*TCF7*, *CCR7,* CD62L$^+$, CD45RA$^+$, CD27$^+$). There was little clonal expansion
114 present within this population (Fig. 1c,d). The T_Treg (regulatory T) cluster consisted primarily
115 of CD4$^+$ T cells, and was enriched for markers associated with Tregs (*FOXP3*, *BATF*, CD25$^+$,
116 CD62L$^+$, CTLA4$^+$, CD27$^+$, ICOS$^+$, TIGIT$^+$, OX40$^+$, TIM3$^+$) (Fig. 1e,g, Extended Data Fig.
117 1b,d,e). Both T_Trm (resident memory T) and T_Tact (activated T) clusters were enriched for
118 markers of Trm cells (CD69$^+$, CD103$^+$) (Fig. 1g, Extended Data 1e). The T_Trm cluster, the most

119 diverse in terms of gene expression, consisted of both $CD4^+$ and $CD8^+$ T cells with distinct
120 signature genes (*IL7R*, *JAML, ITGA1*), and was widely present in both normal colon and tumors
121 (Fig. 1c,h, Extended Data Fig. 1b,d). The T_Tact cluster exhibited signatures of T cell activation
122 (*RELB*, *NFKB2*, $CD25^+$, $CD69^+$, $CD27^+$) (Fig. 1e,g, Extended Data Fig. 1e). The T_Tex
123 (exhausted T) cluster was enriched for expression of genes (*CXCL13*, *HAVCR2*) and surface
124 proteins (PD-1, ICOS) associated with T cell exhaustion or dysfunction[34,35] (Fig. 1e,g, Extended
125 Data Fig. 1b). Cells within this cluster were highly clonally expanded and co-expressed CD39 and
126 CD103, consistent with tumor-infiltrating T cells with tumor-reactivity[36,37] (Fig. 1g, Extended
127 Data Fig. 1e).

128 We identified two clusters of cytotoxic T cells, T_Tcyto1 (cytotoxic T1) and T_Tcyto2 (cytotoxic
129 T2), expressing signature genes of effector function and cytotoxicity (Fig. 1b,e, Extended Data
130 Fig.1b). The T_Tcyto1 cluster was enriched for markers of effector memory T cells (Tem; *EOMES*,
131 *GZMK*, $KLRG1^{hi}$, $CD57^{hi}$, $CD27^{hi}$, $CD44^+$), while the T_Tcyto2 cluster (*GNLY, GZMH, PRF1*)
132 was enriched for expression of the resident memory marker CD103 (Fig. 1e,g, Extended Data Fig.
133 1b,d). Both clusters consisted primarily of $CD8^+$ T cells with high clonal expansion, although
134 $CD4^+$ T cells were present in both populations (Extended Data Figs. 1d and 5a). Cells in the
135 T_Tcyto1 cluster were relatively enriched within tumors compared to normal colon with high
136 clonal expansion (Fig. 1c,d,h).

137 Normal T cells exhibited relatively less phenotypic diversity, and were essentially absent in
138 T_Treg and T_Tex clusters (Fig. 1h, Extended Data Fig. 2d). In contrast, intratumoral T cells were
139 widely distributed over all the clusters and each T cell type was observed in all the patients (Fig.
140 1c,h and Extended Data Fig. 2), suggesting that functional insights and gene signatures gleaned
141 through this analysis are broadly applicable in human CRC (Fig. 1c,h).

142 **Prognostic significance of distinct T cell populations with effector functions in CRC**

143 To investigate the prognostic significance of distinct T cell subtypes, we applied semi-quantitative
144 gene set enrichment analysis (GSEA) to The Cancer Genome Atlas (TCGA) database[31] (Methods).
145 We queried the top 30-40 differentially expressed genes in non-stimulated clusters with effector
146 function (T_Tcyto1, T_Tcyto2, T_Tex, and T_Treg) within bulk RNAseq data from the CRC
147 TCGA cohort (Supplementary Table 5). We subsequently compared survival between the
148 statistically gene-enriched and -depleted patients through Kaplan-Meier analysis (Methods). CRC
149 patients who were enriched with the T_Tcyto1 gene set (n= 33) showed a favorable prognosis
150 compared to the depleted patients (n=14, $P < 0.05$) (Fig. 2a), whereas there was no statistically
151 significant difference in survival between T_Tcyto2 gene-enriched (n=36) and depleted patients
152 (n=22, $P = 0.2$) (Fig. 2b). These results suggest that T_Tcyto1 and T_Tcyto2 clusters contain
153 functionally distinct T cells (mainly $CD8^+$ T cells). In support of this, the proportion of the
154 T_Tcyto1 gene-enriched patients decreased with late-stage cancer, while the proportion of
155 T_Tcyto2 gene-enriched patients increased ($P < 0.004$; two-sided fisher exact test) (Fig. 2f). As
156 expected, stage influenced survival (Fig. 2e). Thus, it is likely that the positive prognostic
157 significance of the T_Tcyto1 cluster is at least partially dependent on its relative enrichment in
158 early-stage cancer (Fig. 2f).

159 We also queried gene sets from the T_Tcyto1 and T_Tcyto2 clusters in gene expression data from
160 the TCGA cohort of melanoma (Supplementary Table 5, Methods). In the melanoma cohort,
161 positive clinical outcomes were associated with both gene sets from the T_Tcyto1 ($P < 0.05$) and
162 the T_Tcyto2 cluster ($P < 0.05$) (Extended Data Fig. 3a,b). As expected, survival correlated with
163 stage (Extended Data Fig. 3d). In contrast to CRC, the proportion of the gene-enriched patients
164 did not correlate with tumor-stage in melanoma (Extended Data Fig. 3e). Moreover, there was

165 higher overlap between gene-enriched patients from these clusters in melanoma (86 %; T_Tcyto1,
166 78 %; T_Tcyto2) compared to CRC, likely reflecting differences in cytotoxic T cell populations
167 between these two cancers (Extended Data Fig. 3f,g).

168 The gene set from the T_Treg cluster also correlated with a positive prognosis ($P < 0.05$) in CRC,
169 but had no significant correlation ($P = 0.5$) with prognosis in melanoma (Fig. 2c, Extended Data
170 Fig. 3c). These findings are consistent with the preponderance of histologic studies that concluded
171 that Tregs are associated with favorable outcomes in CRC[38,39]. Notably, relative enrichment of
172 T_Treg genes in the patients was highest in early-stage cancer, indicating that the improved
173 prognosis may be related to stage (Fig. 2f). Analysis of the prognostic value of the T_Tex gene
174 set did not reach statistical significance ($P = 0.1$) in CRC (Fig. 2d).

175

176 **Distinct differentiation pathways of CD8$^+$ T-cell infiltrates including prognostic effector**
177 **memory and non-prognostic resident memory cells**

178 To define T cells with improved granularity, we segregated CD8$^+$ and CD4$^+$ T cells in total 35,145
179 single cells based on ADT (antibody-derived tag) signal and transcriptome, and analyzed them
180 separately (Extended Data Fig. 4a). Within 12,642 single CD8$^+$ T cells, unsupervised clustering
181 based on transcriptome identified 13 distinct clusters consisting of non-simulated and stimulated
182 CD8$^+$ T cells (Fig. 3a). Based on 21 ADT signals (except CD4 and CD8), we were able to link
183 non-stimulated clusters to its cognate stimulated (Fig. 3c, Methods). Overall, all CD8 subtypes
184 were observed within tumors, whereas normal cells were largely confined to the clusters
185 CD8_Trm, CD8_IEL, and CD8_Tact (Fig. 3a).

186 Within non-stimulated cells, the CD8_Tn/Tcm cluster contained cells with relatively higher *TCF7*
187 expression and the least clonal expansion (Fig. 3a,b,d,e, Extended Data Fig. 4b,e). The CD8_Trm,

188 CD8_IEL (intraepithelial lymphocyte), and CD8_Tact clusters all consisted predominantly of
189 CD8$^+$ Trm cells (CD69$^+$, CD103$^+$) (Fig. 3a,e). The CD8_Trm cluster (*GPR15*, *IL7R*) mainly
190 contained cells with relatively higher expression of CD161, described to define a particular
191 memory cell-type in CD8$^+$ T cells including MAIT cells (CD161^{hi}, *RORC*, *IKZF2*) expressing the
192 semi-invariant TCR Vα7.2-Jα33/12/20[40], which were present at low frequency (~ 4 %) within this
193 cluster (Fig. 3d,j, Extended Data Fig.4b, Supplementary Table 7). Cells within its cognate
194 stimulated clusters (CD8_Stim1.1.Trm, CD8_Stim1.2.Trm) expressed cytokines including *TNF*,
195 *IL2*, and *IFNG* (Extended Data Fig. 4d).
196 Cells within the CD8_IEL cluster expressed effectors (*GNLY, PRF1*) with expression of IEL
197 (intraepithelial lymphocyte) gene signatures such as KIRs (Killer cell immunoglobulin-like
198 receptors; *KIR3DL1/2*)[41], and its cognate stimulated cluster (CD8_Stim3.IEL) contained cells
199 expressing effector molecules, *GNLY, GZMB,* and *PRF1*. (Fig. 3d, Extended Data Fig. 4b,d). The
200 CD8_IEL cluster overlapped with the T_Tcyto2 cluster, which accounted for 38% of cells within
201 the cluster (Fig. 3f,g). The CD8_Tact cluster consisted of activated T cells (*NFKB2*, *RELB*) and
202 contained both stimulated and non-stimulated cells (Fig. 3a,d, Extended Data Fig. 4b). The
203 CD8_IEL and CD8_Tact clusters were associated with relatively high co-expression of CD39 and
204 CD103, proposed to denote tumor-reactivity[36,37] (Extended Data Fig. 4c).
205 The CD8_Tem cluster contained Tem cells (*EOMES*, *GZMK,* CXCR3, KLRG1$^+$, CD57$^+$, CD44$^+$,
206 CD27$^+$) with relatively higher expression of naïve T cell markers (*CCR7*, *TCF7*), similar to
207 "effector" T cells previously described in several types of tumors[9,42,43] (Fig. 3d,e,h, Extended Data
208 Fig. 4b, Supplementary Table 8). Cells within its cognate stimulated cluster, CD8_Stim.Tem,
209 expressed *IL2*, *TNFA*, and *IFNG* (Extended Data Fig. 4d). Importantly, the CD8_Tem cluster

210 largely overlapped with the positively prognostic T-Tcyto1 cluster, which constituted 62 % of the
211 cells within this cluster (Fig. 3f,g).

212 The CD8_Temra cluster contained terminally differentiated Temra (effector memory CD45RA^{+})
213 cells (*GZMH*, *GNLY*, CD45RA^{hi}, KLRG1^{hi}, CD57^{hi}, TIGIT^{hi}, CD27^{-}) (Fig. 3d,e and Extended Data
214 Fig. 4b). This cluster was enriched for "cytotoxic" signature genes (*FGFBP2, GZMH*) and
215 contained highly clonally expanded cells[42] (Fig. 3b,h, Extended Data Fig. 4b, Supplementary Table
216 8). The CD8_Tex cluster was enriched for Tex-cell markers (*HAVCR2*, CXCL13, PD-1^{+}, TIM3^{+})
217 and contained the most clonally expanded cells with co-expression of CD103 and CD39 (Fig.
218 3d,e,h, Extended Data Fig. 4b,c, Supplementary Table 8). Cells within its cognate stimulated
219 CD8_Stim4.Tex cluster were capable of producing certain effectors (*PRF1*, *GZMB*, *IFNG)*, while
220 relatively impaired for expression of *IL2* and *TNFA* (Extended Data Fig. 4d). Notably, the non-
221 prognostic T_Tcyto2 cluster contained a large number of Tex cells (36%) as well as IEL(-like)
222 cells (38%) (Fig. 3f,g).

223 To visualize the relationship between CD8^{+} T cell clusters, we used diffusion maps and ordered
224 cells in pseudotime within non-stimulated clusters (Extended Data Fig. 4f, Methods). In
225 combination with TCR sharing data, we found 4 differentiation pathways of CD8^{+} T cells within
226 tumors, which culminated in terminally differentiated cell-types present within CD8_Temra or
227 CD8_Tex clusters (Fig. 3i-k, Extended Data Fig. 4f,g). In one pathway (pathway 1), CD8_Tem
228 cells differentiated into CD8_Temra cells. The other pathways culminated into three distinct Tex
229 subpopulations: (pathway 2) CD8_Tem cells towards CD8_Tex1 subpopulation (*NKG7*, *GZMK*,
230 *GZMA*), (pathway 3) CD8_Trm cells towards CD8_Tex2 subpopulation (CD161^{hi}, *KLRB1*,
231 *IGFLR1*), or (pathway 4) IELs towards CD8_Tex3 subpopulation (CD161^{+}, *CXCR6*, *GNLY*,
232 *CCL5*). Interestingly, cells within the positively prognostic T_Tcyto1 cluster were largely present

233 in the CD8_Tem and Tex1 clusters with more effector function, while cells within the non-
234 prognostic T_Tcyto2 cluster were present in the CD8_IEL and Tex3 clusters containing more
235 dysfunctional cells (Fig. 3f,g,i). Thus, diverse types of $CD8^+$ T cells (mainly prognostic Tem and
236 non-prognostic Trm cells) transitioned to terminally differentiated cells through different pathways
237 within tumors.
238
239 **Single $CD4^+$ T-cell infiltrates with prognostic significance are phenotypically and**
240 **functionally diverse.**
241 Unsupervised clustering of 19,183 single $CD4^+$ T cells based on its transcriptome identified 10
242 distinct clusters consisting of non-stimulated and stimulated T cells (Fig. 4a,b, Extended Data Fig.
243 5a).
244 Normal $CD4^+$ T cells with a resident memory phenotype ($CD69^+$, $CD161^+$, *IL7R*) were present in
245 non-stimulated clusters (CD4_Trm, CD4_Trm.CCL5, CD4_Tact1) and stimulated clusters
246 (CD4_Stim1, CD4_Stim3), and were clonally expanded mostly within the CD4_Trm.CCL5
247 cluster with expression of CCL5, known to regulate proliferation and homeostatic maintenance of
248 tissue Trm cells[44] (Fig. 4a,c,d,j, Extended Data Fig. 5b). Stimulated normal cells expressed *IL2,*
249 *IFNG, TNFA* and *IL10* and were associated with expression of Th1 marker (*Tbx21*; CD4_Stim1)
250 or Tfh marker (*BCL6*, *IL21*; CD4_Stim3), whereas gene markers for Th17 and Treg cells, were
251 mostly absent in normal cells (Extended Data Fig. 5c,d).
252 Intratumoral cells primarily contained effector $CD4^+$ T cells expressing the activation marker
253 CD39 (Fig. 4c,e). Many canonical genes that define effector $CD4^+$ Th subsets were elicited through
254 stimulation, including Th1, Th17, and Tfh signature genes (Extended Data Fig. 5c). The
255 CD4_Stim2 cluster was further subdivided into 5 distinct subclusters by unsupervised clustering

256 (Extended Data Fig. 5f). Through GSEA, we were able to link these subclusters with its cognate
257 non-stimulated clusters and named them accordingly (Fig. 4k, Extended Data Fig. 5f,g).
258 Interestingly, CD4^{+} T cells within T-Tcyto1 and T-Tcyto2 clusters were observed in the
259 CD4_Trm.CCL5 cluster, where further computational analysis identified 3 distinct subclusters
260 including the CD4_Tcyto1 and CD4_Tcyto2 clusters predominantly containing T-Tcyto1 cells and
261 T-Tcyto2 cells, respectively (Fig. 4f-I, Extended Data Fig. 5a). Cells within the CD4_Tcyto1
262 cluster (*NKG7*, *EOMES*, *GZMK*, *GZMA,* KLRG1^{+}, CD57^{+}, CD27^{+}, PD-1^{+}) were enriched within
263 tumors and exhibited a similar phenotype to the CD8_Tem cluster (Fig. 4f-i, Extended Data
264 Fig5a,e). Its cognate stimulated cluster (CD4_Stim2.Tcyto1) contained cells expressing effector
265 cytokines, *IFN* and *IL2,* and chemokines (*CCL3, CCL4*) that can mediate T cell recruitment[45,46]
266 (Extended Data Fig. 5f,g). Likewise, cells within CD4_Tcyto2 exhibited a similar phenotype to
267 the T_Tcyto2 cluster, and were not confined to tumors (Fig. 4f-i, Extended Data Fig. 5e).
268 The CD4_Th17-like cluster contained cells with Th17 cell signature genes (*KLRB1, RORA,*
269 *CXCR6*) with high clonal expansion, and its cognate stimulated subclusters, (CD4_Stim2.Th17.1
270 and CD4_Stim2.Th17.2) contained cells expressing canonical Th17 genes (*RORC*, *IL17A*) (Fig.
271 4b, Extended Data Fig. 5b,c,f,g). Cells within the CD4_Stim2.Th17.1 cluster expressed genes
272 associated with regulatory function (*IL10, ITGA2*)[47-49], whereas cells within the
273 CD4_Stim2.Th17.2 cluster expressed proinflammatory effectors (*IFNG, IL23R, PRF1*), described
274 to be produced by pathogenic Th17 cells in autoimmunity[50] (Extended Data Fig. 5g).
275 The positively prognostic FOXP3^{+} T cell population (T-Treg) was largely divided into 2 CD4
276 subpopulations (CD4_Treg, CD4_pTreg,) (Fig. 4a,b). The CD4_Treg cluster contained cells
277 expressing distinct signature genes (*BATF*, *GATA3*, *IKZF2*) and high levels of ICOS (Fig. 5b,c,
278 Extended Data Fig. 7a). The transcription factor Helios (encoded by the gene *IKZF2*) has been

279 proposed to mark thymically-derived Tregs although currently controversial[51-53]. In contrast,
280 Helios^{-} Tregs within the CD4_pTreg cluster exhibited a distinct phenotype (CD161^{+}, CD69^{+},
281 LAG3^{+}, CD38^{hi}) and shared genes (*IL26*, *KLRB1*, *IKZF3, LAG3*) with peripherally-derived CD4^{+}
282 FOXP3^{-} T conventional (Tconv) cells, known as peripherally-derived Tregs (pTregs)[54-57] (Figs. 4j
283 and 5b,c,e, Extended Data Figs. 5h and 7a).

284 The CD4_Tfh cluster, containing PD-1^{hi}ICOS^{hi}CD27^{hi} *CXCL13*$^{+}$ T cells with expression of naïve
285 T cell markers (*CCR7*, *TCF7*), correlated with its stimulated clusters (CD4_Stim2.Tfh,
286 CD4_Stim3) with expression of Tfh signature genes (*BCL6*, *CXCR5, IL21*) (Fig. 4b,h, Extended
287 Data Figs. 5b,c,d, and 7). The CD4_Tex cluster (*CXCL13*, *GNLY, GZMA*, TIGIT^{hi}, ICOS^{hi}, CD27^{-}
288) with expression of co-inhibitory molecules (PD-1^{hi}, TIM3^{hi}) and exhausted marker genes
289 (*HAVCR2*, *TOX)*[40,58] contained cells with high clonal expansion, and correlated with the stimulated
290 CD4_Stim2.Tex cluster with expression of *IFNG* and *IL21* (Fig.4b-f,h, Extended Data Fig. 5b).
291 The CD4_Tact2 cluster contained CD4^{+} T cells with high relative expression of type-1 interferon
292 (IFN)-inducible genes (*IFIT3*, *MX1*, and *HERC5*)[59,60] (Extended Data Fig. 5b).

293 We found that cells within the CD4_Tfh cluster with a more naïve phenotype (CD39^{lo}, *CCR7*,
294 *TCF7*) shared TCRs with effector cells within the several clusters (CD4_Tex, CD4_Th17-like,
295 CD4_pTreg, CD4_Treg) (Fig. 4a,c,d,j, Extended Data Fig. 6). These results suggest that the
296 CD4_Tfh cluster could contain a transitional population that gives rise to more differentiated
297 effector CD4^{+} T cell populations. Through GSEA, we did not observe any correlation of CD4_Tfh,
298 CD4_Th17-like, or CD4_Tex gene sets with clinical outcomes (Extended Data Fig. 6).

299

300 **CD38^{+} pTreg-infiltrates correlate with poor clinical outcomes in CRC**

301 To investigate the relationship between non-stimulated and stimulated clusters, Treg populations
302 were further subdivided by unsupervised clustering, resulting in 4 distinct non-stimulated and 5
303 distinct stimulated subpopulations. We designated non-stimulated subpopulations cTreg (central
304 memory), pTreg (peripherally-derived), eTreg (effector), and Tfr (follicular) (Fig. 4a, Fig. 5a,b,c,
305 Extended Data Fig. 7a). Based on 21 ADT signals, we were able to link these Treg subpopulations
306 with its cognate subpopulations within stimulated subclusters (Stim4.1, Stim4.2, Stim4.3 within
307 Stim4 cluster; Stim5.1, Stim5.2 within Stim5 cluster) (Fig. 5d).
308 Helios^{+} Tregs within the CD4_Treg cluster were subdivided into the CD4_eTreg and CD4_Tfr
309 clusters (Fig. 5a, b). The CD4_eTreg cluster, containing cells highly expressing CD62L, ICOS,
310 and CD25 with high clonal expansion (>30%), correlated with its cognate stimulated subclusters
311 (CD4_Stim4.2 and CD4_Stim5.1) containing cells with expression of genes related to trafficking
312 (*CCR6, CCL22*) (Fig. 5c,f,l). The CD4_Tfr cluster was phenotypically similar to the CD4_eTreg,
313 but distinguished by a Tfr marker gene *CXCR5* (Extended Data Fig. 7a). Tfr cells also expressed
314 the gene *PDLM1,* encoding the transcription factor Blimp-1, a marker of eTregs[61], and had high
315 relative expression of OX40 and HLA-DR (Fig. 5b,c). The CD4_Tfr cluster correlated with
316 cognate stimulated CD4_Stim4.3 cluster with genes (*IL18R1*, *FOXP1*, *BCL2*, *TRAF4*) related to
317 the activation and maintenance of Tregs[62-65] (Fig. 5d,l).
318 The CD4_pTreg cluster was linked to cognate stimulated clusters CD4_Stim4.1 and CD4_Stim5.2,
319 which expressed Th17 signature genes, *IL17A* and/or *RORC* and a higher level of *IL10* compared
320 to other Treg clusters (Fig. 5b,k,l, Extended Data Fig. 7a). Through GSEA, the gene set from the
321 CD4_pTreg was associated with poor clinical outcomes in not only the CRC but also the BRC
322 cohort (Fig. 5g,i, Methods). Moreover, the gene set enrichment was observed in patients with
323 early-stage cancer (except Stage IV), indicating that pTreg enrichment was independent of stage

324 (Fig. 5g). Interestingly, there was substantial overlap between CD4_pTreg-enriched and
325 CD4_Th17-like-enriched patients within the CRC cohort (76%; 25 of 33 enriched patients), but
326 no overlap between CD4_pTreg-enriched and CD4_Th17-like-depleted patients (Extended Data
327 Fig. 7c). In addition, pTregs and Th17-like clusters shared T cell clones, although TCR sharing
328 was not confined to these clusters (Extended Data Fig. 7f). These results indicate that negatively
329 prognostic pTregs and Th17 cells in CRC are functionally and phenotypically linked, and enriched
330 in the same patients.
331 Although neither CD4_eTreg nor CD4_Tfr gene sets were associated with a favorable prognosis
332 CRC, leading-edge genes within the positively prognostic T_Treg cluster were relatively enriched
333 in eTreg and Tfr cells, but not pTregs (Fig. 5j, Extended Data Fig. 7d,e,g). Also, both eTreg and
334 Tfr cells had significantly higher frequency of TCR sharing, compared to pTregs ($P < 0.0001$)
335 (Extended Data Fig. 7b). These results suggest that Helios^{+} Tregs, as opposed to Helios^{-} Tregs,
336 account for Tregs associated with favorable outcomes in CRC. Taken together, these data
337 demonstrate substantial phenotypic and functional heterogeneity of Tregs within CRC, and
338 provides clues on how these cells influence outcomes.
339
340
341
342
343
344
345
346

347 **DISCUSSION**

348 Prior studies, performed through histology or analysis of bulk RNA, have demonstrated a
349 relationship between T-cell infiltrates and clinical outcomes in CRC[1,2,8,11,66]. However, as a
350 relatively limited number of genes were assessed in these studies, the resolution to assess the
351 significance of distinct T cell subtypes was limited.

352 Our single-cell analysis delineated the functional and phenotypic diversity of TILs and its
353 prognostic significance in CRC. For $CD8^+$ T-infiltrates in CRC, we found that $KLRG1^+$ *GZMK*$^+$
354 Tem cell population could differentiate into 2 different cell-types. One contained a Tex population
355 at a less dysfunctional transcriptional state ($CD39^+$, $CD103^+$, *GZMA*, *GZMK*, *PRF1*), enriched in
356 early-stage cancer. The other contained $CD45RA^+$ Temra cells with senescent cell markers
357 ($KLRG1^{hi}$, $CD57^{hi}$)[67]. Moreover, 2 different cell-types of non-prognostic Trm cells (IELs and
358 $CD161^+$ T cells) were relatively enriched in late-stage cancer. These cells also differentiated into
359 2 different Tex populations primarily within tumors. Of these four pathways, the differentiation
360 pathway including *GZMK*$^+$ Tem and the more functional Tex subcluster accounted for a majority
361 of positively prognostic cytotoxic T cells. These findings imply that factors driving distinct
362 pathways of differentiation within CRC may be a key for anti-tumor immunity and a therapeutic
363 avenue. Further, our data indicate that $CD8^+$ T cell populations within CRC evolve with stage,
364 transitioning from effector to IEL phenotypes with tumor progression.

365 Compared to $CD8^+$ T cells, $CD4^+$ T cell subtypes within tumors exhibited more complexity and
366 diversity. Notably, we found that the prognostic Tem cells contain cytotoxic $CD4^+$ T cells (but not
367 Th1 cells) with a similar phenotype of $CD8^+$ T cells. As a recent report indicated anti-tumor
368 reactivity of cytotoxic $CD4^+$ T cells in bladder cancer, further investigation could detail their role
369 in CRC.

370 Although T-bet^{+} T cell infiltration was reported to be associated with a favorable prognosis in
371 CRC[1,2], our single-cell analysis showed substantial diversity in CRC, with expression of T-bet
372 observed in not only Th1 cells but also Th17 cells and Tregs[68,69]. T-bet^{+}Th17 cells were primarily
373 enriched within tumor rather than canonical Th1 cells. We were not able to stratify such gene sets-
374 enriched or -depleted patients from bulk-seq data of TCGA cohort (possibly due to loss of effector
375 gene expression without any stimulation); it is worthwhile validating the contribution of these
376 specific effector T cells to clinical outcome for future.
377 *CXCL13*$^{+}$CD4^{+} T cells including Tfh and CD4^{+} Tex cells were mainly observed within tumors
378 and both populations highly expressed PD-1. In addition, cells within the CD4_Tex cluster were
379 enriched in MSI-H patients (53 %; 16 of 30 enriched patients) consistent with the prior study
380 suggesting these cells could respond to PD-1 blockade[9]. Based on GSEA analysis, CD4_Tex-cell
381 infiltrates were not predictive of prognosis.
382 For tumor-infiltrating Tregs in CRC, our data showed that almost all Tregs were specifically
383 enriched within tumors (but not normal colon) and highly expressed the activated markers of CD39
384 and ICOS. Phenotypically, these Tregs were further subdivided into Helios^{+} Tregs and Helios^{-}
385 Tregs (pTregs). pTreg signature genes were associated with a poor prognosis, while the gene set
386 from the T-Treg cluster (before subclustering) was associated with a favorable prognosis. These
387 results could be explained by our data that gene signatures of Helios^{+} Tregs accounted for a large
388 number of prognostic leading-edge genes of the T-Treg cluster, suggesting that Helios^{+} Tregs
389 rather than pTregs could correlate with positive clinical outcomes.
390 pTregs exhibited a Th17-cell phenotype with expression of *IL17A*, *RORC*, and *IKZF3* and shared
391 TCRs with cells within the CD4_Th17-like cluster, although Th17(-like) cell-infiltrates were not
392 associated with clinical outcomes. These results suggest that IL-17A^{+} pTregs rather than Th17

393 cells itself, positively involve tumor progression, and therefore may represent a target of
394 immunotherapy. In support of this, no studies to date have considered the cellular source of IL-17,
395 although IL-17 has been correlated with poor outcomes in CRC[9,11]. pTregs also showed a unique
396 phenotype and function with higher expression of CD38 and LAG3 compared to Helios^{+} Tregs,
397 and potentially highly produce IL-10. It is intriguing that Daratumumab, a monoclonal antibody
398 used to treat myeloma, has been shown to target a CD38^{+} Treg subpopulation[70].
399
400
401 ,
402
403
404
405
406
407
408
409
410
411

Acknowledgement

We thank patient volunteers and providers in the Departments of Surgery, Medicine, Pathology and the Herbert Irving Comprehensive Cancer Center at Columbia University Irving Medical Center for assistance in obtaining patient samples. We thank Hanina Hibshoosh and tissue bank for processing of patient samples. Erin Bush and Izabella Krupska within the JP Sulzberger Columbia Genome Center helped with single-cell analysis. We thank Michael Kissner and the flow cytometry cores within the Columbia Stem Cell Initiative and the Columbia Center for Translational Medicine (S10OD020056). This work was supported by the Damon Runyon Cancer Research Foundation, The Louis V. Gerstner Jr. Scholars Program, and K08 DK100738 (Han). We thank members of the Han lab and Yan lab for comments and suggestions.

Author contributions

Study Concept and Design, K.M., A.H.; Patient and Sample Acquisition, K.M., S.L., M.I., V.R., A.M.A., S.A.L., P.R.K., K.S.Y., P.E.O, and A.H.; K.M., A.H. analyzed Data; Performed and designed experiments, K.M., S.L., K.S., M.S., P.S; Computational Analysis, K.M., A.K., P.H., A.M.B. and P.A.S; Wrote manuscript, K.M., N.S. A.H.; Study oversight, A.H.

Competing interests

The authors declare no competing interests.

Supplementary Data

Extended Data Figures and Supplementary Tables are available at Supplementary Materials.

434 **Figure legends**

435 **Fig. 1 Characterization of T cells in CRC and normal adjacent colon by multiplexed**
436 **scRNAseq.**

437 **a**, Workflow for preparation of multiplexed scRNA-seq and TCR-seq libraries using 10x
438 Genomics platform.

439 **b**, Uniform Manifold Approximation and Projection (UMAP) of 35,145 single T cells in CRC and
440 adjacent normal colon from total 16 patients. Distinct clusters based on unsupervised clustering of
441 single-cell transcriptomes are indicated. Unstimulated T cells populate within 7 clusters (left),
442 stimulated cells T cells populate two clusters (right). The number of tumor-infiltrating T cells or
443 normal cells within each cluster is indicated in Supplementary Table 4.

444 **c**, Single T cells within tumor (left) or adjacent normal colon (middle) from each patient (as
445 indicated) are plotted on the total UMAP in Fig. 1b. The number of tumor-infiltrating T cells or
446 normal cells from each patient is indicated in Supplementary Table 4. Cells are colored based on
447 TCR clone frequency (right), normalized to total number of T cells per patient sample.

448 **d**, Bar graph depicting clonal composition of T cells within each cluster.

449 **e**, Dot plot depicting relative expression of identifying marker genes for each non-stimulated
450 cluster in Fig. 1b

451 **f**, Heatmap of cytokine and effectors based on relative gene expression (as indicated) on the total
452 UMAP in Fig. 1b

453 **g**, Heatmap depicting average ADT signal in the CITE-seq antibody panel within each cluster.

454 **h**, Bar graphs showing relative distribution of cells from each cluster (as indicated) within tumor
455 (left) or normal adjacent colon (middle) of each patient (as indicated), and proportion of cells
456 within each cluster from tumor or normal tissue (right).

457 **Fig.2 Clinical outcomes associated with T cell subtypes identified by scRNA-seq in**
458 **CRC.**
459 (**a**-**d**) Kaplan–Meier curves of overall survival in the CRC TCGA cohort for patients enriched or
460 depleted for the following gene sets using GSEA (Methods): T-Tcyto1 (**a**) T-Tcyto2 (**b**) T-Treg
461 (**c**) or T-Tex clusters (**d**). #denotes $P < 0.05$ (log-rank test) compared to non-enriched, non-depleted
462 patients (curve at black). *denotes $P < 0.05$ (log-rank test) when enriched vs. depleted patients are
463 compared. See detailed p-values (Supplementary Table 6).
464 **e**, Kaplan-Meier curves of overall survival by clinical tumor-stage for the CRC TCGA cohort.
465 **f**, Bar graph depicting relative proportion of patients by stage enriched or depleted for T-Tcyto1,
466 T-Tcyto2 or T-Treg gene sets. Average stage of patients is indicated in table on bottom.
467
468 **Fig. 3 Dynamic phenotypic changes of CD8$^+$ T-cell infiltrates in CRC.**
469 **a**, UMAP of 19,183 single CD8$^+$ T cells in CRC and adjacent normal colon from 16 patients (left).
470 Unstimulated T cells populate within clusters on left side, stimulated cells T cells populate clusters
471 on the right side. Cells from tumor (red) or adjacent normal colon (blue) are indicated (middle).
472 Cells are colored based on TCR clone frequency (right), normalized to total number of T cells per
473 patient sample.
474 **b**, Bar graph depicting clonal composition of T cells within each cluster.
475 **c**, Pearson correlation map between each cluster in non-stimulated and stimulated cells, correlated
476 based on 21 ADT signals.
477 **d**, Dot plot depicting relative expression of identifying marker genes for each non-stimulated
478 cluster in Fig. 3a.

479 **e**, Heatmap depicting average ADT signal in the CITE-seq antibody panel within each non-
480 stimulated cluster.
481 **f**, Distribution of cells within the clusters T-Tcyto1, T-Tcyto2, T-Tex (in Fig. 1b) plotted on the
482 CD8 UMAP.
483 **g**, Pie graphs depicting composition of cells within the clusters T-Tcyto1 or T-Tcyto2 as a
484 proportion of CD8 clusters CD8_Tem, CD8_Temra, CD8_IEL, CD8_Tex.
485 **h**, Dot plots of cells from each CD8 non-stimulated cluster by naïve vs. exhaustion score or
486 cytotoxic vs. exhaustion score (Methods).
487 **i**, Distinct differentiation pathways of $CD8^+$ T cells based on cell trajectory (relate to Extended
488 Data Fig. 4f,g).
489 **j**, Heatmap of ADT signals (as indicated) on the CD8 UMAP.
490 **k**, Gene heatmap of differentially expressed genes within Tex subpopulations (Tex1-Tex3) in Fig.
491 3i.
492
493 **Fig. 4 The functional and phenotypic diversity of single $CD4^+$ T cells with**
494 **prognostic significance in CRC.**
495 **a**, UMAP of 12,642 single $CD4^+$ T cells in CRC and adjacent normal colon from total 16 patients.
496 Unstimulated T cells populate within clusters on right side, stimulated cells T cells populate
497 clusters on the left side.
498 **b**, Dot plot depicting relative expression of identifying marker genes for each non-stimulated
499 cluster in Fig. 4a.
500 **c**, Cells from tumor or adjacent normal colon are indicated on the CD4 UMAP. Heatmap of CD39
501 ADT signal and heatmap of TCR clone frequency in normal cells. on the CD4 UMAP.

502 **d**, Heatmap of TCR clone frequency in intratumoral cells on the CD4 UMAP.

503 **e**, Bar graph depicting clonal composition of T cells within each cluster.

504 **f**, Distribution of cells within the clusters of T-Tcyto1, T-Tcyto2, T-Tex, and Treg (identified in

505 Fig. 1b) on the CD4 UMAP.

506 **g**, Subclustering of CD4_Trm.CCL5 cluster. 3 subpopulations are labeled.

507 **h**, Heatmap depicting average ADT signal in the CITE-seq antibody panel within each

508 CD4_Trm.CCL5 subcluster shown in Fig. 4g.

509 **i**, Heatmap showing relative gene expression of the genes indicated within each CD4_Trm.CCL5

510 subcluster shown in Fig. 4g.

511 **j**, Heatmap of ADT signals (as indicated) on the CD4 UMAP.

512 **k**, GSEA based on gene expression between non-stimulated clusters and stimulated subclusters (as

513 indicated).

514

515 **Fig. 5 Prognostic significance of Treg subpopulations in CRC.**

516 **a**, Treg subpopulations on the CD4 UMAP from Fig. 4a.

517 **b**, Heatmap of differentially expressed genes (as indicated) within Treg subpopulations.

518 **c**, Heatmap depicting average ADT signal in the CITE-seq antibody panel within each non-

519 stimulated Treg subpopulation.

520 **d**, Heatmap of pearson correlation between non-stimulated and stimulated cells within each Treg

521 subcluster, correlated based on 21 ADT signals.

522 **e**, Heatmap showing pearson correlations between 21 ADT signals from each non-stimulated Treg

523 subcluster.

524 **f**, Bar graph depicting relative proportion of cells within each Treg subpopulation.

525 (**g**,**i**) Kaplan–Meier curves of overall survival in the CRC (**g**) or BRC (**i**) TCGA cohort for patients
526 enriched or depleted for the CD4_pTreg gene set by GSEA (Methods). #denotes $P < 0.05$ (log-
527 rank test) compared to non-enriched, non-depleted patients (curve at black). *denotes $P < 0.05$
528 (log-rank test) when enriched vs. depleted patients are compared. See detailed p-values
529 (Supplementary Table 6).
530 **h**, Bar graphs depicting relative proportion of patients by stage enriched or depleted for
531 CD4_pTreg gene sets in CRC. Average stage is indicated on right.
532 **j**, NES and FDR results from GSEA between the leading-edge genes (in Extended Data Fig. 7g)
533 of the positively prognostic T-Treg cluster (in Fig. 2c) and signature genes in each Treg
534 subpopulation. FDR threshold value indicated by the orange line is 0.05.
535 **k**, Bargraphs depicting proportion of intratumoral *IL10*$^{+}$ and *IL10*$^{-}$ Tregs within each stimulated
536 Treg subcluster. The number of cells within each cluster were indicated.
537 **l**, Heatmap of differentially expressed genes in stimulated Treg subclusters.
538
539
540

541 ## Methods

542 **Human colon biopsy and single-cell dissociation**

543 Tumors and adjacent normal colon tissues from the patients, diagnosed with colorectal
544 adenocarcinoma at Columbia University Irving Medical Center, were collected with informed
545 consent, as approved by the Columbia University Institutional Review Board (IRB). All protocols
546 were performed in accordance with the guidelines provided by the IRB. All patients were
547 treatment naïve at time of sample acquisition (Supplementary Table 1).

548 Fresh tissue was collected the same day of surgery. Tissue was dissociated into single-cell
549 suspensions with enzymic digestions using a gentleMACS dissociate per protocol (Miltenyi
550 Biotec). Single-cell suspensions were cryopreserved (10% DMSO) and stored in a liquid nitrogen.

551

552 **Antibody-oligo conjugates**

553 Antibodies for CITE-seq and Cell Hashing were purchased from BioLegend, and were covalently
554 conjugated to barcoded oligos by click-chemistry tools in accordance with the protocol of CITE-
555 seq Hyper Antibody-Oligo Conjugation (https://cite-seq.com/protocol/). The list of antibodies,
556 clones and barcodes are available in Supplementary Table 2.

557

558 **T cell isolation, T cell stimulation and cell sorting from CRC and adjacent normal tissues**

559 Cryopreserved single-cell suspensions were thawed at 37 degrees, and single T cells were isolated
560 by percoll (Millipore-Sigma) density gradient (20/40/80%) centrifugation (2200 rpm for 30 min).
561 T cells were rested in complete RPMI in 96 well plates overnight. Next day, T cells from each
562 patient were split into two wells, incubated with or without PMA/Ionomycin for 3h. After 3
563 washes, cells were incubated for 10 min with Fc receptor block (BioLegend) to block nonspecific

564 antibody binding, each sample was stained with Cell hashing antibody cocktail for 30 min. After
565 3 washes, cells were stained with CITE-seq antibodies and FACS antibody cocktails for 30 min
566 (Supplementary Table 2). Viable T cells were sorted by FACS Aria II by gating $DAPI^-$
567 $ZombiAqua^-$ $CD3^+$ $TCR\alpha\beta^+$ T cells (Extended Data Fig. 1a).
568
569 **Single cell 5' transcript, TCRαβ enriched, and Cell hashing/CITE-seq library preparation**
570 **on the 10x chromium, and sequencing.**
571 The scRNA-seq, scTCR-seq and HTO/ADT-seq libraries as mentioned below were prepared using
572 the 10xChromium single cell 5' library and gel bead Kit, per manufacturer's instructions. In brief,
573 oligo-conjugated antibody-tagged single cell suspensions were loaded into the Chromium™
574 controller to make nanoliter-scale droplets with uniquely barcoded 5' gel beads called GEMs (gel
575 bead-in emulsions). After GEM-RT, GEMs were cleaned up by Dyna beads MyOne™ Silane
576 beads. At the cDNA amplification step, HTO/ADT additive oligos were added to amplify oligos
577 derived from Cell hashing and CITE-seq antibodies. The products were size separated with SPRI
578 Beads (Beckman Coulter): < 300 nt fragments containing the oligos derived from CITE-seq
579 antibodies and Cell Hashing, and > 300 nt fragments containing cDNAs derived from cellular
580 mRNA. HTO/ADTs were amplified using specific primers that append P5 and P7 sequences for
581 illumina sequencing.
582 The cDNA was then pooled for downstream processing and library preparation according to the
583 manufacturer's instructions. The 5' transcript library was sequenced with Illumina Novaseq. The
584 HTO/ADT libraries were sequenced with Illumina Miseq or Nextseq. The single cell TCR
585 enriched library was sequenced with Illumina Miseq using 150 paired-end reads.
586

587 **Data processing of HTO/ADT libraries**

588 We counted antibody-derived-tags (ADTs) or Cell Hashing tags (HTOs) in raw sequencing reads
589 and built a count matrix using CITE-seq-Count as previously described, available at
590 (https://github.com/Hoohm/CITE-seq-Count). For HTO quantification, each HTO count was
591 normalized by total counts of HTOs per cell barcode (total UMI counts were set at 100). Then, cell
592 barcode identity was assigned to the most highly expressed HTO, otherwise assigned to "negative"
593 or "doublet", available at (https://github.com/akornberg/data_analysis_ak). Each CITE-seq UMI
594 count was normalized by the average of total UMI counts from all cell barcodes, and then a count
595 matrix was set into the Seurat as the ADT data, which was normalized based on the Seurat package,
596 CLR normalization, and scaled.

597

598 **Data processing of scRNA-seq libraries.**

599 The total reads from scRNA-seq were aligned to GRCh38 reference genome and quantified using
600 cellranger count (10x Genomics, version 3.1.0). In brief, only cells which passed the threshold set
601 by the pipeline (more than 500 UMIs), were considered for further analysis. On average, we
602 obtained reads from 1,862 genes per cell (median: 1,572) and 4,908 unique transcripts per cell
603 (median: 4,178). The analysis of the created filtered feature-barcode matrices was performed using
604 Seurat (version 3.1.3), available at (https://satijalab.org/seurat/).

605

606 **Data processing of single-cell TCR-seq libraries.**

607 The total reads from scRNA-seq were aligned to GRCh38 reference genome and consensus TCR
608 annotation was conducted by cellranger vdj (10x Genomics, version 2.1.0). In total, 86 % (27,999

609 out of 32,550 cells) of annotated T cells were assigned a TCR (TRB and/or TRA), and 20,875
610 clonotypes were detected.
611
612 **Dimension reduction and clustering**
613 For the downstream analysis, four multiple scRNAseq datasets (seurat objects) with different
614 individuals and different conditions (Supplementary Table 3) were created by the Seurat, in which
615 the metadata for HTO assignment (patients, stimulated or non-stimulated cells), normalized ADT
616 counts (23 CITE-seq antibodies), and TCR clonotypes and sequences were combined into the
617 datasets using the function AddMetadata and CreateSeuratobject, and were integrated using an
618 anchor-based single-cell data integration method in Seurat[71]. In the method, the batch effects
619 among the samples in four different datasets were normalized using the function the SCtransform
620 in Seurat. Greater than 10% mitochondrial RNA content were excluded. Variable feature genes
621 were set at 5,000 genes. Once the datasets were integrated, the data was used input into a principal
622 component analysis (PCA) on the basis of variable genes. The same principal components were
623 used to generate the UMAP projections. Clusters were identified using shared nearest neighbor
624 (SNN)-based clustering. For the clustering analysis, the function RunUMAP, FindClusters, and
625 FindNeighbors in Seurat were used, in which "dims" or "resolution" were set at between 10 and
626 30 or between 0.1 and 0.5, respectively.
627
628 **Analysis of TCR sharing and clone frequency**
629 Identical TCR clonotypes (regarded as clonal identifier) among single-cells were counted through
630 the Seurat metadata. Cells missing an identifier were considered as not observed (shown as NA).
631 TCR sharing between two distinct clusters or different cell states on the UMAP, was assessed by

632 the counts of identical TCRβ clonotypes. After $CD4^+/CD8^+$ single-cell separation, clone frequency
633 in each single-cell was calculated as a relative value of TCRβ clonotype counts to total single-cells
634 from each patient (TILs were separated from normal cells). Code is available at
635 (https://github.com/akornberg/data_analysis_ak).
636
637
638 **Gene-gene Correlation Analysis and gene signature scoring**
639 The stem-like, exhausted and cytotoxic signature genes were selected from variable genes across
640 all CD8 T cells, which were computed using the Seurat: Those signature genes consisted of the
641 top 20-30 genes with the highest correlation with the reference genes *TCF7*, *HAVCR2*, or
642 *FGFBP2*, respectively as previously described[42] (Supplementary Table 8). Individual cells on the
643 CD8 UMAP (non-stimulated) were scored by the average of the selected gene expression, and
644 shown as dot plots.
645
646 **Gene set enrichment analysis (GSEA) using TCGA cohort.**
647 Gene expression data of the TCGA cohort were obtained from cBioPortal
648 (https://www.cbioportal.org/). The TCGA cohort data of CRC, BRC, and melanoma were
649 downloaded from Colorectal Adenocarcinoma (TCGA, PanCancer Atlas), Breast Cancer
650 (METABRIC), and Skin Cutaneous Melanoma[72,73] (TCGA, PanCancer Atlas). Gene expression
651 was normalized by log2-transformation and an average z-score across patient samples was
652 calculated per gene. For each patient, genes were ranked by the average z-score in each gene.
653 Regarding original gene sets in distinct T cell subsets delineated by unsupervised clustering, genes
654 were ranked by differential expression test which was carried out using Seurat and the top genes

655 (~ 100 genes) were selected (min.diff.pct was set at > 0.25) after removing specific non-coding
656 RNA such as lincRNA and miRNA, and genes linked with poorly supported transcriptional models
657 (annotated with the prefix "AP-") (Supplementary Table 5). The correlation between the gene sets
658 and bulk databases of the TCGA cohort was calculated using pre-ranked gene set enrichment
659 analysis method fgsea package (https://github.com/ctlab/fgsea).
660
661 **Prognosis prediction.**
662 The CRC, melanoma, or BRC TCGA cohort were grouped on the basis of GSEA, respectively.
663 One group (NES, normalized enrichment score, >0 & a BH-adjusted p-value < 0.05) was
664 categorized into "Enriched". The other (NES <0 & a BH-adjusted p-value < 0.05) was assigned
665 into "Depleted". Kaplan-Meier survival curves were generated and log-rank test for two groups
666 was performed using the R package survival (3.1.8).
667
668 **Diffusion map trajectory analysis.**
669 The Seurat object including the cluster CD8_Tn/Tcm, CD8_Trm, CD8_Tem, CD8_Temra,
670 CD8_IEL, and CD8_Tex, which were characterized by unsupervised clustering in Seurat (Fig. 3a),
671 was subjected into single-cell trajectory analysis using the Monocle (2.14.0), available at
672 (http://cole-trapnell-lab.github.io/monocle-release/). RNA expression data from the Seurat object was
673 converted into Monocle as a sparse matrix. The meta data from the Seurat object (total single cells
674 in each patient, and cluster annotation) were incorporated as "phenodata" in Monocle.
675 Then, newCellDataSet was created with a function "expressionFamily = negbinomial.size()". In
676 order to infer a single-cell trajectory, the total 650 genes (Supplementary Table 9) which were
677 assembled from each cluster in the Seurat object using differential expression test, were chosen

678 and then the cells were ordered after dimension reduction with "DDR-tree" function, and
679 visualized.
680
681 **Gene or ADT expression heatmap and correlation analysis**
682 For single-cell gene heatmaps, the representative marker genes were selected from differentially
683 expressed genes within each cluster that was identified by unsupervised clustering. Gene heatmaps
684 between cells within each cluster was created using the function DoHeatmap in Seurat (Extended
685 Data Figs. 4b and 5b).
686 For ADT heatmaps and gene heatmaps among subclusters, each average ADT signal of cells
687 within each cluster or average expression of each differentially expressed gene in each subcluster
688 was calculated using the function AverageExpression in Seurat. Then, heatmap of each average
689 ADT signal or gene expression was shown in selected clusters using the R package pheatmap
690 (1.0.12).
691 The relationship between stimulated and non-stimulated clusters, was assessed by a statistical
692 correlation between gene expression or average ADT signals in single-cells. In brief, for ADT
693 correlation analysis, a matrix including each average ADT signal in stimulated and non-stimulated
694 clusters was created and its correlation was examined by Pearson correlations. For gene expression
695 correlation analysis, the representative marker genes were selected from differentially expressed
696 genes within each non-stimulated cluster, respectively. And a matrix including each average
697 variable gene expression in each stimulated cluster was created, respectively. Then, a correlation
698 between marker genes in non-stimulated clusters and average gene expression in stimulated
699 clusters was examined by GSEA using the R package fgsea.
700

701 ## Code and Data Availability

702 The accession numbers for the sequencing results reported in this paper are **ArrayExpress**: E-
703 MTAB-9455 (under revision). Fastq files from each sequencing lane (L1, L2) and corresponding
704 to each fragment end (R1, R2) were concatenated into a single file to generate 2 fastq files (R1,
705 R2) per biological replicate. Code is available at the links as indicated (Methods).

706

707 ## References

708 1. Galon, J.*, et al.* Type, density, and location of immune cells within human colorectal
709 tumors predict clinical outcome. *Science* **313**, 1960-1964 (2006).

710 2. Pages, F.*, et al.* Effector memory T cells, early metastasis, and survival in colorectal
711 cancer. *N Engl J Med* **353**, 2654-2666 (2005).

712 3. Boland, C.R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology*
713 **138**, 2073-2087 e2073 (2010).

714 4. Gryfe, R.*, et al.* Tumor microsatellite instability and clinical outcome in young patients
715 with colorectal cancer. *N Engl J Med* **342**, 69-77 (2000).

716 5. Le, D.T.*, et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med*
717 **372**, 2509-2520 (2015).

718 6. Le, D.T.*, et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1
719 blockade. *Science* **357**, 409-413 (2017).

720 7. Gentles, A.J.*, et al.* The prognostic landscape of genes and infiltrating immune cells
721 across human cancers. *Nat Med* **21**, 938-945 (2015).

722 8. Mlecnik, B.*, et al.* Integrative Analyses of Colorectal Cancer Show Immunoscore Is a
723 Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity* **44**, 698-
724 711 (2016).

725 9. Zhang, L.*, et al.* Lineage tracking reveals dynamic relationships of T cells in colorectal
726 cancer. *Nature* **564**, 268-272 (2018).

727 10. Ribas, A. & Wolchok, J.D. Cancer immunotherapy using checkpoint blockade. *Science*
728 **359**, 1350-1355 (2018).

729 11. Tosolini, M.*, et al.* Clinical impact of different classes of infiltrating T cytotoxic and
730 helper cells (Th1, th2, treg, th17) in patients with colorectal cancer. *Cancer Res* **71**, 1263-
731 1271 (2011).

732 12. Blatner, N.R.*, et al.* Expression of RORgammat marks a pathogenic regulatory T cell
733 subset in human colon cancer. *Sci Transl Med* **4**, 164ra159 (2012).

734 13. Grivennikov, S.I.*, et al.* Adenoma-linked barrier defects and microbial products drive IL-
735 23/IL-17-mediated tumour growth. *Nature* **491**, 254-258 (2012).

736 14. Chae, W.J.*, et al.* Ablation of IL-17A abrogates progression of spontaneous intestinal
737 tumorigenesis. *Proc Natl Acad Sci U S A* **107**, 5540-5544 (2010).

738 15. Numasaki, M.*, et al.* Interleukin-17 promotes angiogenesis and tumor growth. *Blood*
739 **101**, 2620-2627 (2003).

740 16. Wang, K.*, et al.* Interleukin-17 receptor a signaling in transformed enterocytes promotes
741 early colorectal tumorigenesis. *Immunity* **41**, 1052-1063 (2014).

742 17. Friedrich, M., Pohin, M. & Powrie, F. Cytokine Networks in the Pathophysiology of
743 Inflammatory Bowel Disease. *Immunity* **50**, 992-1006 (2019).

744 18. Zheng, G.X.*, et al.* Massively parallel digital transcriptional profiling of single cells. *Nat*
745 *Commun* **8**, 14049 (2017).

746 19. Bates, G.J.*, et al.* Quantification of regulatory T cells enables the identification of high-
747 risk breast cancer patients and those at risk of late relapse. *J Clin Oncol* **24**, 5373-5380
748 (2006).

749 20. Gao, Q.*, et al.* Intratumoral balance of regulatory and cytotoxic T cells is associated with
750 prognosis of hepatocellular carcinoma after resection. *J Clin Oncol* **25**, 2586-2593 (2007).

751 21. Perrone, G.*, et al.* Intratumoural FOXP3-positive regulatory T cells are associated with
752 adverse prognosis in radically resected gastric cancer. *Eur J Cancer* **44**, 1875-1882
753 (2008).

754 22. Petersen, R.P.*, et al.* Tumor infiltrating Foxp3+ regulatory T-cells are associated with
755 recurrence in pathologic stage I NSCLC patients. *Cancer* **107**, 2866-2872 (2006).

756 23. Shang, B., Liu, Y., Jiang, S.J. & Liu, Y. Prognostic value of tumor-infiltrating FoxP3+
757 regulatory T cells in cancers: a systematic review and meta-analysis. *Sci Rep* **5**, 15179
758 (2015).

759 24. Hu, G., Li, Z. & Wang, S. Tumor-infiltrating FoxP3(+) Tregs predict favorable outcome in
760 colorectal cancer patients: A meta-analysis. *Oncotarget* **8**, 75361-75371 (2017).

761 25. Kuwahara, T.*, et al.* Intratumoural-infiltrating CD4 + and FOXP3 + T cells as strong
762 positive predictive markers for the prognosis of resectable colorectal cancer. *Br J Cancer*
763 **121**, 659-665 (2019).

764 26. Saito, T.*, et al.* Two FOXP3(+)CD4(+) T cell subpopulations distinctly control the
765 prognosis of colorectal cancers. *Nat Med* **22**, 679-684 (2016).

766 27. Stoeckius, M.*, et al.* Simultaneous epitope and transcriptome measurement in single
767 cells. *Nat Methods* **14**, 865-868 (2017).

768 28. Stoeckius, M.*, et al.* Cell Hashing with barcoded antibodies enables multiplexing and
769 doublet detection for single cell genomics. *Genome Biol* **19**, 224 (2018).

770 29. Lahnemann, D.*, et al.* Eleven grand challenges in single-cell data science. *Genome Biol*
771 **21**, 31 (2020).

772 30. Ma, Y.*, et al.* Integrative differential expression and gene set enrichment analysis using
773 summary statistics for scRNA-seq studies. *Nat Commun* **11**, 1585 (2020).

774 31. Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon
775 and rectal cancer. *Nature* **487**, 330-337 (2012).

776 32. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours.
777 *Nature* **490**, 61-70 (2012).

778 33. Cancer Genome Atlas, N. Genomic Classification of Cutaneous Melanoma. *Cell* **161**,
779 1681-1696 (2015).

780 34. Thommen, D.S.*, et al.* A transcriptionally and functionally distinct PD-1(+) CD8(+) T cell
781 pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade.
782 *Nat Med* **24**, 994-1004 (2018).

783 35. Zheng, C.*, et al.* Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell
784 Sequencing. *Cell* **169**, 1342-1356 e1316 (2017).

785 36. Simoni, Y.*, et al.* Bystander CD8(+) T cells are abundant and phenotypically distinct in
786 human tumour infiltrates. *Nature* **557**, 575-579 (2018).

787 37. Duhen, T.*, et al.* Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells
788 in human solid tumors. *Nat Commun* **9**, 2724 (2018).

789 38. Salama, P.*, et al.* Tumor-infiltrating FOXP3+ T regulatory cells show strong prognostic
790 significance in colorectal cancer. *J Clin Oncol* **27**, 186-192 (2009).

791 39. Frey, D.M.*, et al.* High frequency of tumor-infiltrating FOXP3(+) regulatory T cells
792 predicts improved survival in mismatch repair-proficient colorectal cancer patients. *Int J*
793 *Cancer* **126**, 2635-2643 (2010).

794 40. Fergusson, J.R.*, et al.* CD161 defines a transcriptional and functional phenotype across
795 distinct human T cell lineages. *Cell Rep* **9**, 1075-1088 (2014).

796 41. Vandereyken, M., James, O.J. & Swamy, M. Mechanisms of activation of innate-like
797 intraepithelial T lymphocytes. *Mucosal Immunol* **13**, 721-731 (2020).

798 42. Li, H.*, et al.* Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated
799 Compartment within Human Melanoma. *Cell* **176**, 775-789 e718 (2019).

800 43. Yost, K.E.*, et al.* Clonal replacement of tumor-specific T cells following PD-1 blockade.
801 *Nat Med* **25**, 1251-1259 (2019).

802 44. Seo, W.*, et al.* Runx-mediated regulation of CCL5 via antagonizing two enhancers
803 influences immune cell function and anti-tumor immunity. *Nat Commun* **11**, 1562
804 (2020).

805 45. Oh, D.Y.*, et al.* Intratumoral CD4+ T Cells Mediate Anti-tumor Cytotoxicity in Human
806 Bladder Cancer. *Cell* **181**, 1612-1625.e1613 (2020).

807 46. Castellino, F.*, et al.* Chemokines enhance immunity by guiding naive CD8+ T cells to sites
808 of CD4+ T cell-dendritic cell interaction. *Nature* **440**, 890-895 (2006).

809 47. Gagliani, N.*, et al.* Coexpression of CD49b and LAG-3 identifies human and mouse T
810 regulatory type 1 cells. *Nat Med* **19**, 739-746 (2013).

811 48. Fan, X.*, et al.* CD49b defines functionally mature Treg cells that survey skin and vascular
812 tissues. *J Exp Med* **215**, 2796-2814 (2018).

813 49. Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the
814 transcription factor Foxp3. *Science* **299**, 1057-1061 (2003).

815 50. Lee, Y.*, et al.* Induction and molecular signature of pathogenic TH17 cells. *Nat Immunol*
816 **13**, 991-999 (2012).

817 51. Thornton, A.M. & Shevach, E.M. Helios: still behind the clouds. *Immunology* **158**, 161-
818 170 (2019).

819 52. Thornton, A.M.*, et al.* Expression of Helios, an Ikaros transcription factor family
820 member, differentiates thymic-derived from peripherally induced Foxp3+ T regulatory
821 cells. *J Immunol* **184**, 3433-3441 (2010).

822 53. Szurek, E.*, et al.* Differences in Expression Level of Helios and Neuropilin-1 Do Not
823 Distinguish Thymus-Derived from Extrathymically-Induced CD4+Foxp3+ Regulatory T
824 Cells. *PLoS One* **10**, e0141161 (2015).

825 54. Raffin, C.*, et al.* Human memory Helios- FOXP3+ regulatory T cells (Tregs) encompass
826 induced Tregs that express Aiolos and respond to IL-1beta by downregulating their
827 suppressor functions. *J Immunol* **191**, 4619-4627 (2013).

828 55. Pesenacker, A.M.*, et al.* CD161 defines the subset of FoxP3+ T cells capable of producing
829 proinflammatory cytokines. *Blood* **121**, 2647-2658 (2013).

830 56. Quintana, F.J.*, et al.* Aiolos promotes TH17 differentiation by directly silencing Il2
831 expression. *Nat Immunol* **13**, 770-777 (2012).

832 57. Gandhi, R.*, et al.* Activation of the aryl hydrocarbon receptor induces human type 1
833 regulatory T cell-like and Foxp3(+) regulatory T cells. *Nat Immunol* **11**, 846-853 (2010).

834 58. Kleinschek, M.A.*, et al.* Circulating and gut-resident human Th17 cells express CD161
835 and promote intestinal inflammation. *J Exp Med* **206**, 525-534 (2009).

836 59. Schneider, W.M., Chevillotte, M.D. & Rice, C.M. Interferon-Stimulated Genes: A Complex
837 Web of Host Defenses. *Annual Review of Immunology* **32**, 513-545 (2014).

838 60. Schoggins, J.W. & Rice, C.M. Interferon-stimulated genes and their antiviral effector
839 functions. *Curr Opin Virol* **1**, 519-525 (2011).

840 61. Cretney, E.*, et al.* The transcription factors Blimp-1 and IRF4 jointly control the
841 differentiation and function of effector regulatory T cells. *Nat Immunol* **12**, 304-311
842 (2011).

843 62. Sidwell, T.*, et al.* Attenuation of TCR-induced transcription by Bach2 controls regulatory
844 T cell differentiation and homeostasis. *Nat Commun* **11**, 252 (2020).

845 63. Ghosh, S., Roy-Chowdhuri, S., Kang, K., Im, S.-H. & Rudra, D. The transcription factor
846 Foxp1 preserves integrity of an active Foxp3 locus in extrathymic Treg cells. *Nature*
847 *Communications* **9**, 4473 (2018).

848 64. Arpaia, N.*, et al.* A Distinct Function of Regulatory T Cells in Tissue Protection. *Cell* **162**,
849 1078-1089 (2015).

850 65. Wang, X.*, et al.* Preferential control of induced regulatory T cell homeostasis via a
851 Bim/Bcl-2 axis. *Cell Death & Disease* **3**, e270-e270 (2012).

852 66. Bindea, G.*, et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the
853 immune landscape in human cancer. *Immunity* **39**, 782-795 (2013).

854 67. Thome, J.J., *et al.* Spatial map of human T cell compartmentalization and maintenance
855 over decades of life. *Cell* **159**, 814-828 (2014).

856 68. Santegoets, S.J., *et al.* Tbet-positive regulatory T cells accumulate in oropharyngeal
857 cancers with ongoing tumor-specific type 1 T cell responses. *J Immunother Cancer* **7**, 14
858 (2019).

859 69. Arger, N.K., Machiraju, S., Allen, I.E., Woodruff, P.G. & Koth, L.L. T-bet Expression in
860 Peripheral Th17.0 Cells Is Associated With Pulmonary Function Changes in Sarcoidosis.
861 *Front Immunol* **11**, 1129 (2020).

862 70. Krejcik, J., *et al.* Daratumumab depletes CD38+ immune regulatory cells, promotes T-cell
863 expansion, and skews T-cell repertoire in multiple myeloma. *Blood* **128**, 384-394 (2016).

864 71. Stuart, T., *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902
865 e1821 (2019).

866 72. Hoadley, K.A., *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of
867 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304 e296 (2018).

868 73. Pereira, B., *et al.* The somatic mutation profiles of 2,433 breast cancers refines their
869 genomic and transcriptomic landscapes. *Nat Commun* **7**, 11479 (2016).
870

Fig. 1

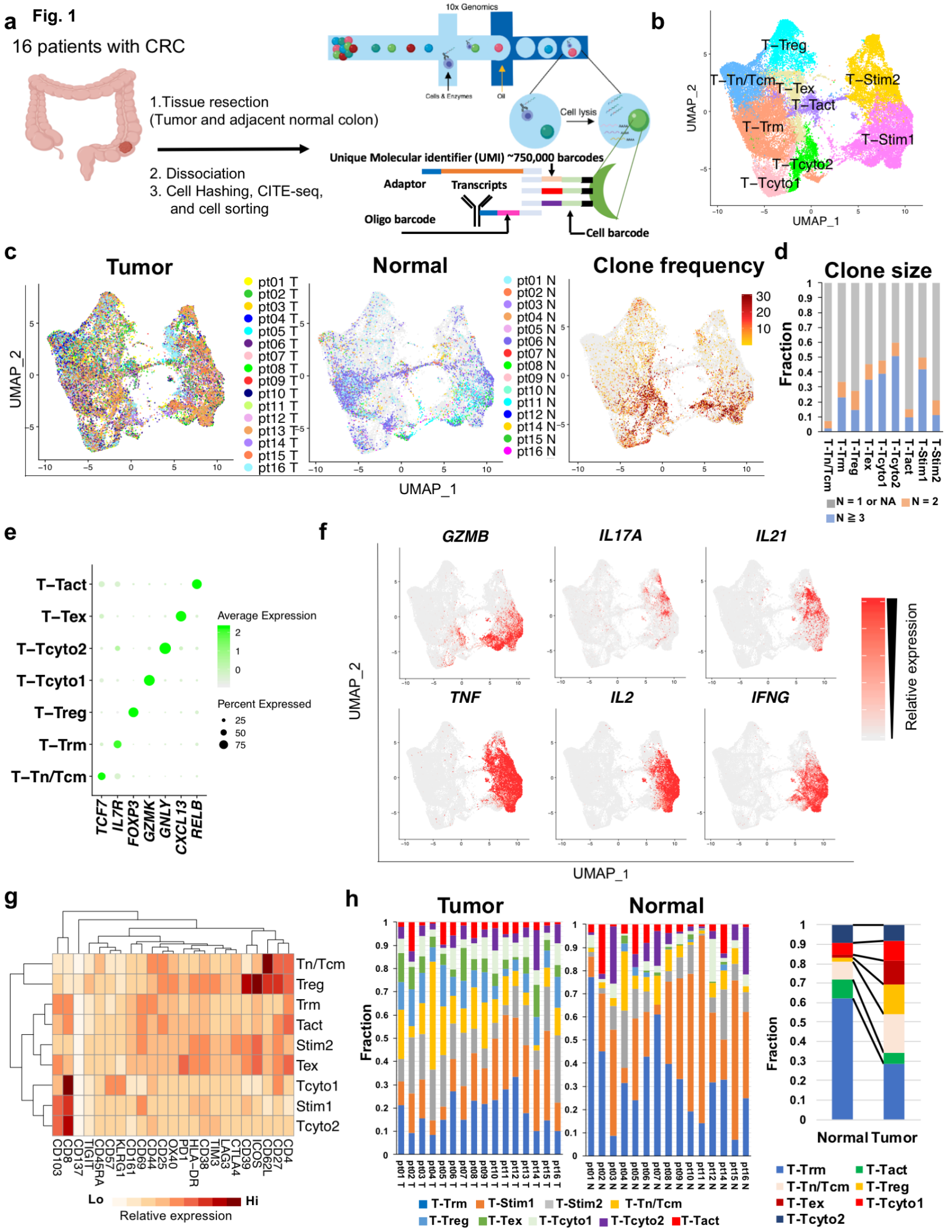
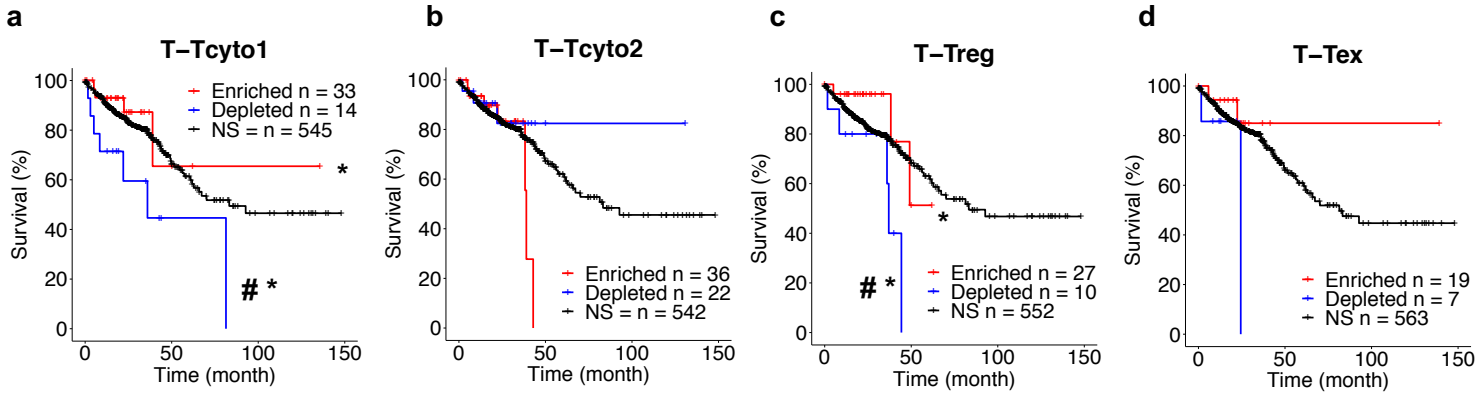
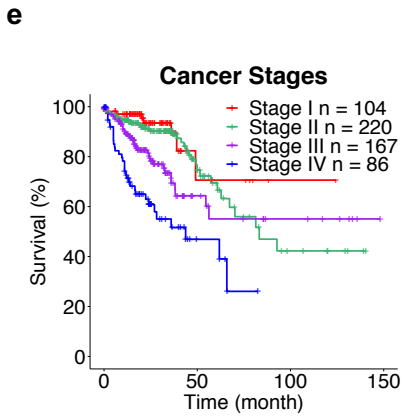
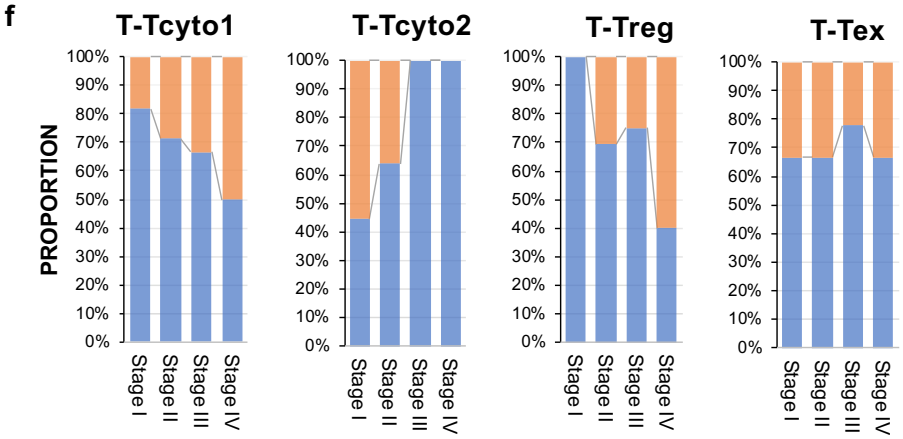

a
16 patients with CRC
1.Tissue resection
(Tumor and adjacent normal colon)
2. Dissociation
3. Cell Hashing, CITE-seq,
and cell sorting
10x Genomics
Cells & Enzymes
Oil
Cell lysis
Unique Molecular identifier (UMI) ~750,000 barcodes
Adaptor
Transcripts
Oligo barcode
Cell barcode
b
T–Treg
T–Tn/Tcm
T–Tex
T–Tact
T–Stim2
T–Trm
T–Stim1
T–Tcyto2
T–Tcyto1
UMAP_2
UMAP_1
c
Tumor
Normal
Clone frequency
pt01 T
pt02 T
pt03 T
pt04 T
pt05 T
pt06 T
pt07 T
pt08 T
pt09 T
pt10 T
pt11 T
pt12 T
pt13 T
pt14 T
pt15 T
pt16 T
pt01 N
pt02 N
pt03 N
pt04 N
pt05 N
pt06 N
pt07 N
pt08 N
pt09 N
pt10 N
pt11 N
pt12 N
pt14 N
pt15 N
pt16 N
30
20
10
UMAP_2
UMAP_1
d
Clone size
Fraction
T-Tn/Tcm
T-Trm
T-Treg
T-Tex
T-Tcyto1
T-Tcyto2
T-Tact
T-Stim1
T-Stim2
N = 1 or NA
N = 2
N ≧ 3
e
T–Tact
T–Tex
T–Tcyto2
T–Tcyto1
T–Treg
T–Trm
T–Tn/Tcm
TCF7
IL7R
FOXP3
GZMK
GNLY
CXCL13
RELB
Average Expression
Percent Expressed
f
GZMB
IL17A
IL21
TNF
IL2
IFNG
Relative expression
UMAP_2
UMAP_1
g
Tn/Tcm
Treg
Trm
Tact
Stim2
Tex
Tcyto1
Stim1
Tcyto2
CD103
CD8
CD137
TIGIT
CD45RA
CD57
KLRG1
CD161
CD69
CD44
CD25
OX40
PD1
HLA–DR
CD38
TIM3
LAG3
CTLA4
CD39
ICOS
CD62L
CD27
CD4
Lo
Hi
Relative expression
h
Tumor
Normal
Fraction
T-Trm
T-Stim1
T-Stim2
T-Tn/Tcm
T-Treg
T-Tex
T-Tcyto1
T-Tcyto2
T-Tact
Normal
Tumor
T-Trm
T-Tact
T-Tn/Tcm
T-Treg
T-Tex
T-Tcyto1
T-Tcyto2

Fig. 2





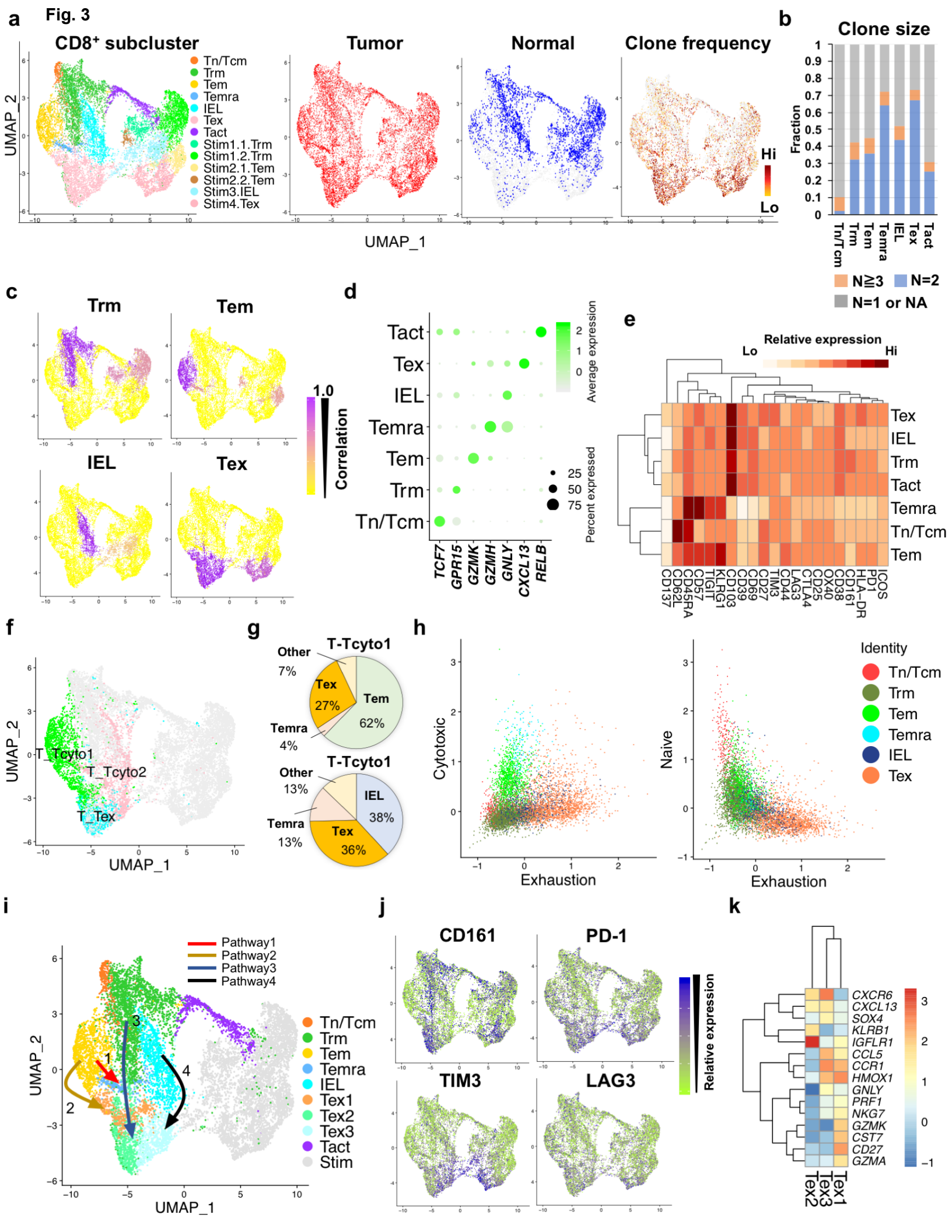


Depleted

Enriched

Average stage	T-Tcyto1	T-Tcyto2	T-Treg	T-Tex
Enriched	2.23	2.43	2.27	2.31
Depleted	2.57	1.64	2.90	2.29

Fig. 3


a
CD8+ subcluster
Tumor
Normal
Clone frequency
Tn/Tcm
Trm
Tem
Temra
IEL
Tex
Tact
Stim1.1.Trm
Stim1.2.Trm
Stim2.1.Tem
Stim2.2.Tem
Stim3.IEL
Stim4.Tex
UMAP_2
UMAP_1
Hi
Lo
b
Clone size
Fraction
N≧3
N=2
N=1 or NA
c
Trm
Tem
IEL
Tex
1.0
Correlation
d
Tact
Tex
IEL
Temra
Tem
Trm
Tn/Tcm
TCF7
GPR15
GZMK
GZMH
GNLY
CXCL13
RELB
Average expression
Percent expressed
e
Relative expression
Lo
Hi
CD137
CD62L
CD45RA
CD57
TIGIT
KLRG1
CD103
CD39
CD69
CD27
TIM3
CD44
LAG3
CTLA4
CD25
OX40
CD38
CD161
HLA-DR
PD1
ICOS
f
T_Tcyto1
T_Tcyto2
T_Tex
UMAP_2
UMAP_1
g
T-Tcyto1
Other
7%
Tex
27%
Tem
62%
Temra
4%
T-Tcyto1
Other
13%
IEL
38%
Temra
13%
Tex
36%
h
Cytotoxic
Exhaustion
Naive
Exhaustion
Identity
Tn/Tcm
Trm
Tem
Temra
IEL
Tex
i
Pathway1
Pathway2
Pathway3
Pathway4
Tn/Tcm
Trm
Tem
Temra
IEL
Tex1
Tex2
Tex3
Tact
Stim
UMAP_2
UMAP_1
j
CD161
PD-1
TIM3
LAG3
Relative expression
k
CXCR6
CXCL13
SOX4
KLRB1
IGFLR1
CCL5
CCR1
HMOX1
GNLY
PRF1
NKG7
GZMK
CST7
CD27
GZMA
Tex1
Tex3
Tex2

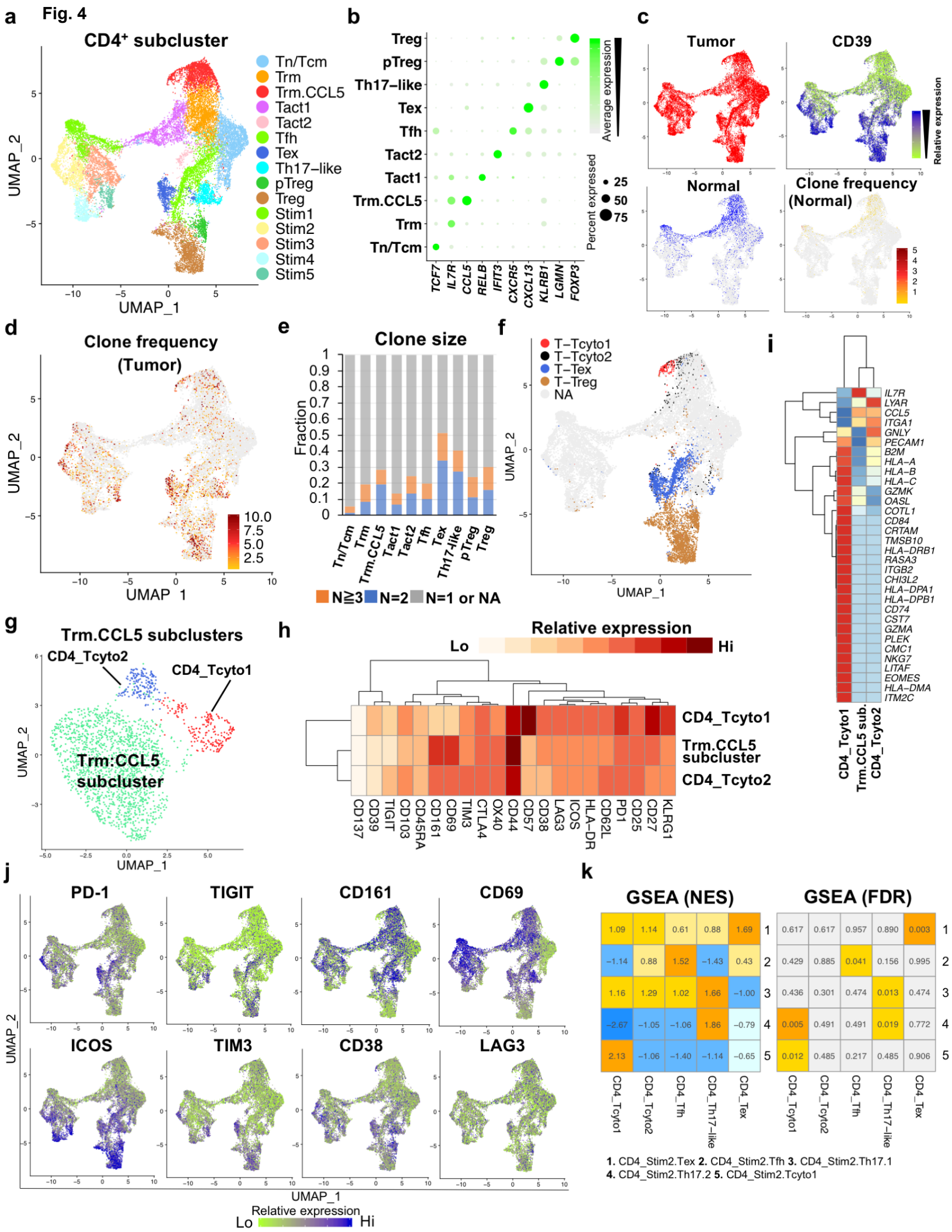


Fig. 4
a
CD4+ subcluster
Tn/Tcm
Trm
Trm.CCL5
Tact1
Tact2
Tfh
Tex
Th17–like
pTreg
Treg
Stim1
Stim2
Stim3
Stim4
Stim5
UMAP_1
UMAP_2
b
Treg
pTreg
Th17–like
Tex
Tfh
Tact2
Tact1
Trm.CCL5
Trm
Tn/Tcm
TCF7
IL7R
CCL5
RELB
IFIT3
CXCR5
CXCL13
KLRB1
LGMN
FOXP3
Average expression
Percent expressed
25
50
75
c
Tumor
CD39
Relative expression
Normal
Clone frequency (Normal)
d
Clone frequency (Tumor)
10.0
7.5
5.0
2.5
UMAP_1
UMAP_2
e
Clone size
Fraction
Tn/Tcm
Trm
Trm.CCL5
Tact1
Tact2
Tfh
Tex
Th17–like
pTreg
Treg
N≧3
N=2
N=1 or NA
f
T–Tcyto1
T–Tcyto2
T–Tex
T–Treg
NA
UMAP_1
UMAP_2
i
IL7R
LYAR
CCL5
ITGA1
GNLY
PECAM1
B2M
HLA–A
HLA–B
HLA–C
GZMK
OASL
COTL1
CD84
CRTAM
TMSB10
HLA–DRB1
RASA3
ITGB2
CHI3L2
HLA–DPA1
HLA–DPB1
CD74
CST7
GZMA
PLEK
CMC1
NKG7
LITAF
EOMES
HLA–DMA
ITM2C
CD4_Tcyto1
Trm.CCL5 sub.
CD4_Tcyto2
g
Trm.CCL5 subclusters
CD4_Tcyto2
CD4_Tcyto1
Trm:CCL5 subcluster
UMAP_1
UMAP_2
h
Relative expression
Lo
Hi
CD4_Tcyto1
Trm.CCL5 subcluster
CD4_Tcyto2
CD137
CD39
TIGIT
CD103
CD45RA
CD161
CD69
TIM3
CTLA4
OX40
CD44
CD57
CD38
LAG3
ICOS
HLA–DR
CD62L
PD1
CD25
CD27
KLRG1
j
PD-1
TIGIT
CD161
CD69
ICOS
TIM3
CD38
LAG3
UMAP_1
UMAP_2
Relative expression
Lo
Hi
k
GSEA (NES)
1.09 1.14 0.61 0.88 1.69 1
−1.14 0.88 1.52 −1.43 0.43 2
1.16 1.29 1.02 1.66 −1.00 3
−2.67 −1.05 −1.06 1.86 −0.79 4
2.13 −1.06 −1.40 −1.14 −0.65 5
CD4_Tcyto1
CD4_Tcyto2
CD4_Tfh
CD4_Th17–like
CD4_Tex
GSEA (FDR)
0.617 0.617 0.957 0.890 0.003 1
0.429 0.885 0.041 0.156 0.995 2
0.436 0.301 0.474 0.013 0.474 3
0.005 0.491 0.491 0.019 0.772 4
0.012 0.485 0.217 0.485 0.906 5
CD4_Tcyto1
CD4_Tcyto2
CD4_Tfh
CD4_Th17–like
CD4_Tex
1. CD4_Stim2.Tex 2. CD4_Stim2.Tfh 3. CD4_Stim2.Th17.1
4. CD4_Stim2.Th17.2 5. CD4_Stim2.Tcyto1

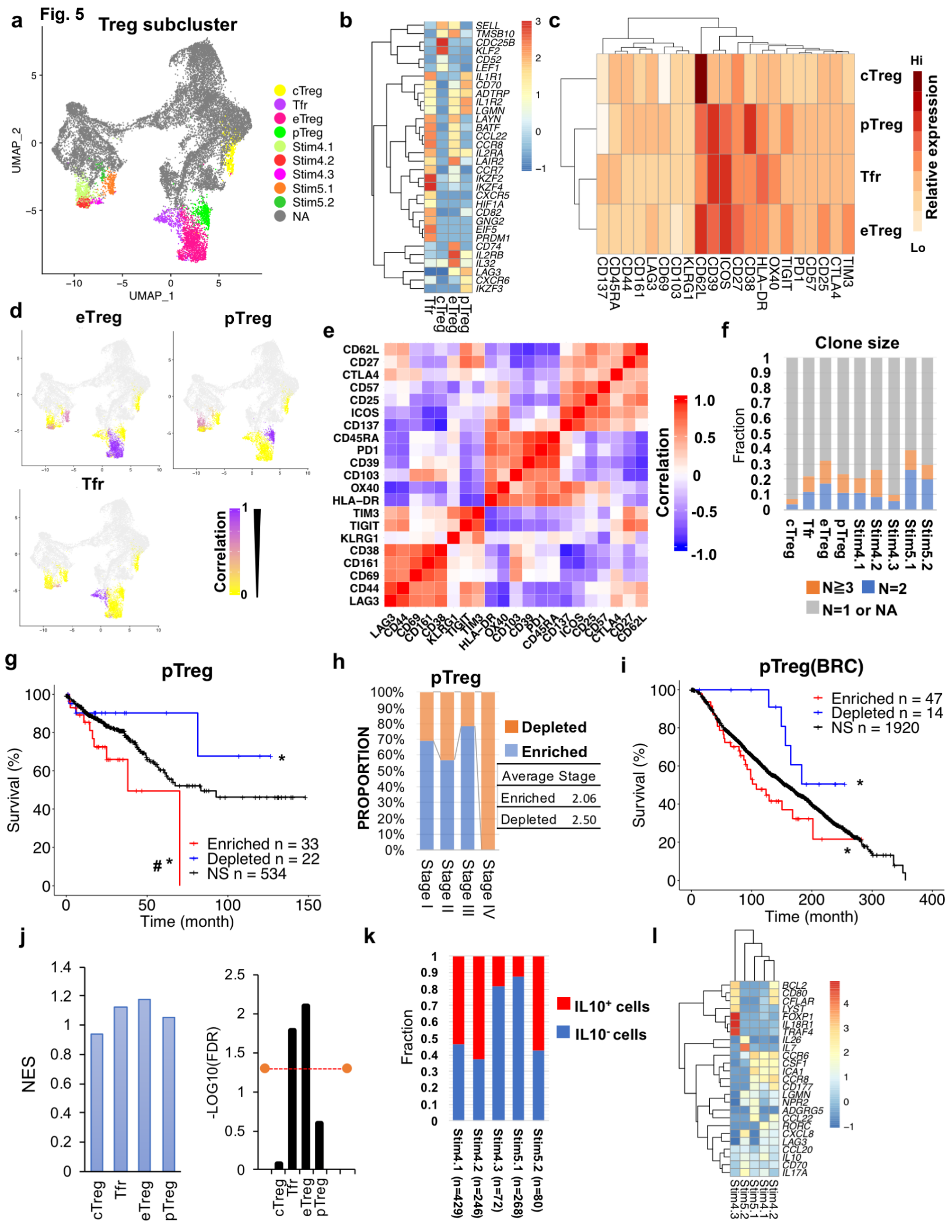


Fig. 5
a
Treg subcluster
UMAP_2
UMAP_1
cTreg
Tfr
eTreg
pTreg
Stim4.1
Stim4.2
Stim4.3
Stim5.1
Stim5.2
NA
b
SELL
TMSB10
CDC25B
KLF2
CD52
LEF1
IL1R1
CD70
ADTRP
IL1R2
LGMN
LAYN
BATF
CCL22
CCR8
IL2RA
LAIR2
CCR7
IKZF2
IKZF4
CXCR5
HIF1A
CD82
GNG2
EIF5
PRDM1
CD74
IL2RB
IL32
LAG3
CXCR6
IKZF3
c
Hi
Lo
Relative expression
CD137
CD45RA
CD44
CD161
LAG3
CD69
CD103
KLRG1
CD62L
CD39
ICOS
CD27
CD38
HLA-DR
OX40
TIGIT
PD1
CD57
CD25
CTLA4
TIM3
d
eTreg
pTreg
Tfr
Correlation
e
CD62L
CD27
CTLA4
CD57
CD25
ICOS
CD137
CD45RA
PD1
CD39
CD103
OX40
HLA-DR
TIM3
TIGIT
KLRG1
CD38
CD161
CD69
CD44
LAG3
Correlation
f
Clone size
Fraction
cTreg
Tfr
eTreg
pTreg
Stim4.1
Stim4.2
Stim4.3
Stim5.1
Stim5.2
N≧3
N=2
N=1 or NA
g
pTreg
Survival (%)
Time (month)
Enriched n = 33
Depleted n = 22
NS n = 534
h
pTreg
PROPORTION
Depleted
Enriched
Average Stage
Enriched 2.06
Depleted 2.50
Stage I
Stage II
Stage III
Stage IV
i
pTreg(BRC)
Survival (%)
Time (month)
Enriched n = 47
Depleted n = 14
NS n = 1920
j
NES
-LOG10(FDR)
cTreg
Tfr
eTreg
pTreg
k
Fraction
IL10+ cells
IL10- cells
Stim4.1 (n=429)
Stim4.2 (n=246)
Stim4.3 (n=72)
Stim5.1 (n=268)
Stim5.2 (n=80)
l
BCL2
CD80
CFLAR
LYST
FOXP1
IL18R1
TRAF4
IL26
IL7
CCR6
CSF1
ICA1
CCR8
CD177
LGMN
NPR2
ADGRG5
CCL22
RORC
CXCL8
LAG3
CCL20
IL10
CD70
IL17A
Stim4.3
Stim5.2
Stim5.1
Stim4.1
Stim4.2