Disrupted priming within draining lymph nodes drives immune quiescence in gastric cancer

Sohrab Salehi[1,2,3*], Emily E. Stroobant[4*], Hannah Lees[1,2**], Ya-Hui Lin[5**], Shoji Shimada[6**], Miseker Abate[4,7**], Matthew J. Zatzman[1,2], Nicholas Ceglia[1,2], Samuel S. Freeman[1,2], Monika Laszkowska[8], Steven Maron[9], Andrew McPherson[1,2], Nicole Rusk[1,2], Eliyahu Havasov[1,2], Harrison Drebin[10], Ping Gu[9], Laura H. Tang[11], Yelena Y. Janjigian[9], Ruslan Soldatov[1,2], Ronan Chaligne[12], Sohrab P. Shah[1,2***], Vivian E. Strong[4,7***], Santosha A. Vardhana[5***]

*These authors contributed equally

** These authors contributed equally

***Co-corresponding

[1]The Halvorsen Center for Computational Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[2]Computational Oncology, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[3]Irving Institute for Cancer Dynamics, Columbia University, New York, NY, USA

[4]Gastric and Mixed Tumor Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[5]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[6]Digestive Disease Center, Showa University, Northern Yokohama Hospital, Yokohama, Japan

[7]Department of Surgery, New York-Presbyterian Hospital/Weill Cornell Medicine, New York, NY, USA

26 [8]Gastroenterology, Hepatology, and Nutrition Service, Department of Medicine, Memorial Sloan
27 Kettering Cancer Center, New York, NY, USA
28 [9]Gastrointestinal Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer
29 Center, New York, NY, USA
30 [10]Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New
31 York, NY, USA
32 [11]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA
33 [12]Single-cell Analytics Innovation Laboratory, Memorial Sloan Kettering Cancer Center, New
34 York, NY, USA
35 Correspondence: shahs3@mskcc.org, strongv@mskcc.org, vardhans@mskcc.org

Abstract

The gastric mucosa is characterized by continuous innate immune surveillance and inflammatory signaling, yet a high proportion of gastric carcinomas (GCs) are recalcitrant to immune-directed therapies. The mechanisms by which GCs evade adaptive immune surveillance within the highly antigenic microenvironment of the gastric mucosa remains unknown. To address this, we collected patient-matched tumor tissue, distant normal tissue, metastasis, and draining lymph nodes to generate a large-scale single-cell immune profiling dataset from 64 patients (n=179 samples, >150,000 cells). From single cell analysis, we identified two distinct sources of impaired tumor surveillance within tumor draining lymph nodes. First, we observed that a significant fraction of tumor draining lymph nodes had undergone cytokine-driven reprogramming, leading to reduced dendritic cell homing and limited T cell priming. Second, T cells undergoing successful activation exhibited limited expansion and constrained differentiation, marked by expression of the quiescence-associated transcription factor Kruppel-like Factor 2 (*KLF2*). Overexpression of *KLF2* in primary T cells limited both their differentiation and cytotoxic capacity. These findings implicate both impaired T cell priming and *KLF2*-dependent T cell quiescence in limiting T cell immunity in gastric adenocarcinoma. We suggest these findings represent an emerging model for immune silencing in tumors developing from tissues with chronic inflammation.

55 **Introduction**

56 Tumor growth and progression require evasion of immune surveillance; yet paradoxically,
57 chronic inflammation is a well-established risk factor for multiple cancers. This dichotomy is
58 evident in gastric adenocarcinoma, a malignancy with rising global incidence and poor outcomes,
59 particularly once lymphatic or distant spread occurs[1]. The role of chronic inflammation in gastric
60 carcinogenesis was first established with the classification of *Helicobacter pylori* as a carcinogen
61 in 1994[2]. Subsequent associations with microbiota and autoimmune conditions further
62 positioned chronic inflammation as a critical promotion factor of gastric carcinogenesis[3-6]. As
63 most gastric cancers, and particularly those without evidence of microsatellite instability, are
64 immunologically silent with sparse infiltration of cytotoxic CD8+ T lymphocytes (TILs)[7], we
65 sought to determine if chronic inflammation disrupts CD8+ T cell-dependent immune
66 surveillance. We recently found that intra-tumoral T cell abundance in gastric cancer correlates
67 with effective T cell priming in tumor-draining lymph nodes[8]. To test if chronic inflammation
68 impairs adaptive immune surveillance by disrupting T cell priming within draining lymph nodes,
69 we developed a surgical collection protocol enabling paired profiling of immune cells from gastric
70 tumors and matched tumor-draining lymph nodes. We then performed an integrated analysis of
71 single-cell RNA sequencing, T-cell receptor (TCR) sequencing and cell-surface epitope
72 assessment (CITE-Seq)[9] on the matched samples and validated the resultant phenotypes with
73 functional assays. Our findings establish two complementary mechanisms by which sustained
74 innate inflammation disrupts anti-tumor T cell immunity: by i) reducing the frequency of
75 successful dendritic cell-T cell priming events within draining lymph nodes and ii) stabilizing
76 transcriptional programs that reinforce T cell quiescence.

77

78 **Patient cohort, sample collection and single cell data generation**

79 To characterize adaptive immune networks during both the priming and effector stages of the
80 immune response, we developed a tissue collection protocol from patients with gastric

81 adenocarcinoma, emphasizing tumor resections paired with matched draining lymph nodes
82 (dLN) (**Fig. 1a**). 64 patients with gastric adenocarcinoma were included, broadly distributed
83 across gastric cancer stages, subtypes and established covariates (**Extended Data Table 1, 2**).
84 This included at least eight patients from each tumor subtype by Lauren classification[10], and
85 representation from all TCGA subtypes[11] (**Fig. 1b**). Importantly, PD-L1 expression by combined
86 positive score (CPS)[12] spanned the full distribution (0-100), indicating representation across a
87 broad range of immunogenicity by conventional metrics. Median follow-up was 2.5 years after
88 initial specimen collection. In total, 179 fresh tissue samples (74 tumor samples, 56 distant
89 normal samples, 36 draining lymph nodes, and 13 metastasis samples) were collected from
90 diagnostic laparoscopy (DL) with esophagogastroduodenoscopy (EGD) or gastrectomy for
91 gastric adenocarcinoma by a specialist gastric cancer surgeon (**Methods**) (**Fig. 1a,b**). All
92 samples were dissociated into single cell suspensions and viably cryopreserved within 2.5 hours
93 of resection.
94
95 **Distinct single cell immune profiles of gastric tumors and matched draining lymph nodes**
96 We first analyzed immune cell subsets from 152,981 CD45+ cells passing quality control (QC)
97 filtering (**Methods, Extended Data Table 3**). Unsupervised graph-based clustering identified six
98 main CD45+ cell types marked by canonical identifying genes (**Fig. 2a, Extended Fig. 1a**):
99 98,274 T/NK cells, 23,560 B cells, 19,774 plasma cells, 8,876 myeloid cells, 1,687 mast cells,
100 and 810 plasmacytoid dendritic cells (DCs). Further phenotypic subsets of CD8+ T cells (**Fig.**
101 **2b**), CD4+ T cells (**Fig. 2c**), and myeloid cells/DCs (**Fig. 2d**) were determined using a
102 combination of known marker genes and surface proteins (**Extended Data Table 4**). Classical
103 CD8+ T cell phenotypes including naïve, early activated, late activated, effector, central memory,
104 proliferating, early exhausted, progenitor exhausted, terminally exhausted, effector memory,
105 heat shock protein (HSP) associated and effector memory with expression of CD45RA (EMRA)
106 subsets were identified, as well as CD8+ T cells marked by expression of *NEAT1* (**Extended**

107 **Fig. 1b**). CD4+ T cell phenotypes included naïve, central memory, effector memory, follicular
108 helper, cytotoxic, HSP-associated, proliferating, and Th17 cells, as well as a *NEAT1*-expressing
109 cluster and two subsets of regulatory T cells distinguished by *CCR8* expression (**Extended Fig.**
110 **1c**). Proliferating CD4+ and CD8+ T cells were also identified (**Extended Fig. 1d**). Within the
111 myeloid compartment, we identified mast cell and dendritic cell subsets including cDC1, cDC2,
112 migratory DCs, and plasmacytoid DCs. We observed two macrophage populations,
113 distinguished by either NLRP3+ inflammasome activation or expression of complement proteins
114 (**Extended Fig. 1e**).
115
116 Given the repertoire of immune cell phenotypes, we next quantified their relative distributions,
117 comparing tumors to their matched dLNs. Lymph node and tumor samples were largely non-
118 overlapping across all three cell compartments (CD8+, CD4+ and myeloid). CD8+ T cells were
119 the more abundant immune cell type within tumors, whereas CD4+ T cells dominated both
120 tumors and dLN (**Fig. 2a, Extended Fig. 1f**). Amongst CD8+ T cells, dLN primarily contained
121 naïve and central memory cells while tumors contained primarily activated, effector memory, and
122 exhausted T cells (**Fig. 2b**). CD4+ T cells within dLN were primarily naïve while tumors
123 contained a mix of regulatory, Th17, and cytotoxic CD4 T cells (**Fig. 2c**). Finally, within the
124 myeloid compartment, lymph nodes contained predominantly cDC2s and plasmacytoid DCs,
125 while tumors were marked by NLRP3 macrophages and mast cells (**Fig. 2d**). Heat maps for the
126 distribution of immune cells in normal and metastasis samples are in **Extended Fig. 2**. Thus,
127 gastric tumors and draining lymph nodes comprise a non-overlapping immune network within
128 which the distribution of cell types may enable or impair anti-tumor immune surveillance.
129
130 **Patient-specific immune networks defined by cytokine-driven cell-cell communication**
131 Cytokines play an essential role in regulating the survival, trafficking, and function of immune
132 cells across tissues[13]. Given the distinct immune cell composition of tumors and draining LN as

133 well as the established role of chronic inflammation in gastric carcinogenesis, we set out to
134 quantitatively measure cytokine-driven interactions between immune cells within gastric tumors
135 and draining lymph nodes. Our goal was to phenotype LN and tumors based on their cytokine-
136 immune crosstalk. Using a recently generated cytokine dictionary of gene expression changes
137 in individual immune cell types isolated from mice treated with a library of 86 distinct cytokines[14]
138 (**Fig. 3a**, **Methods**), we quantified activation of gene expression programs associated with
139 cytokine exposure in each cell type from our cohort. Briefly, after mapping cell types identified in
140 **Fig. 2** to their murine counterparts, we computed cytokine enrichment in each cell type using an
141 over-representation analysis at the sample level (**Methods**). We applied a newly developed
142 Bayesian sparse tensor factorization approach (FRACTAL - Factor and Receiver Analysis of
143 Cytokines in Tumor And Lymph nodes), effectively modeling variation in cytokine-driven
144 expression programs across samples. This resulted in five robustly enumerated factors, with
145 model weights in both 'receiver' cell-type and 'cytokine' activation dimensions (**Extended Fig.
146 3a**). Factor 0 was strongly associated with cDC2 activation, and moderately with pDCs. Factor
147 1 was strongly associated with Treg activation, and (γδ) T cells. Factor 2 was associated with
148 CD4 T cells and Tregs, Factor 3 with macrophages, and Factor 4 with CD8+ T cells, NK cells,
149 and B cells. Factors 0 and 2 were enriched in the dLN, while Factors 3 and 4 were enriched in
150 the tumor; this is consistent with dendritic cell and CD4+ T cell abundance within lymph nodes,
151 and macrophage and CD8+ T cell abundance within tumors. Regarding cytokines, IL-1a was
152 enriched in Factor 2 while IL-15 and TNF were enriched in Factor 3. IL-1 alarmin was the highest
153 activated cytokine family in Factors 1, 2, and 3, while Factors 0 and 4 were more balanced with
154 the highest activity from the IL-2 family. Growth factors and receptor antagonists were the least
155 activated cytokine families across all Factors.
156
157 We next stratified dLN and tumor samples by factor enrichment, yielding two clusters of cytokine-
158 driven immune networks in dLN, and 3 clusters in tumors (**Methods**) (**Fig. 3b**). One dLN cluster

159 (“canonical” LN) was defined by crosstalk between classical dendritic cells and T cells.
160 Consistent with intact LN priming events, central memory CD8+ T cells, follicular helper CD4+ T
161 cells, and type 2 and plasmacytoid dendritic cells were prevalent in this cluster (**Fig. 3c**,
162 **Extended Fig. 3b**). The second dLN cluster (“disrupted” LN) showed a decrease in cytokine-
163 driven transcriptional activity, with macrophages as the dominant cytokine-responsive cell type
164 (**Fig. 3c**, **Extended Fig. 3b**). Notably this cluster harbored an abundance of CD8+ and CD4+ T
165 cells with a naïve phenotype (**Extended Fig. 3b**) and macrophages with an NLRP3+ phenotype
166 (**Fig. 3c**). Given the relative dominance of macrophage-driven crosstalk and the absence of
167 dendritic cells needed for classical T cell priming, we labeled LN within this cluster as “disrupted”.
168 Tumor clusters were similarly interpreted based on cytokine-driven immune cell interactions,
169 with “T cell engaged” dominated by high-strength T cell interactions, activated and exhausted
170 CD8+ T cells, and Th17 CD4+ T cells, “inflamed myeloid” with primarily NLRP3+ macrophage
171 driven interactions, and “immune silent” with minimal immune cell infiltration or cytokine-driven
172 interactions (**Fig. 3c, Extended Fig. 3b**).
173
174 We next asked which ligand-receptor interactions contributed most strongly to these cytokine-
175 driven phenotypes. We inferred intercellular ligand-receptor (L-R) interactions, with a focus on
176 cytokine-driven L-R interactions[15,16] (**Methods**). Inference of interactions was consistent with
177 dendritic cells as the dominant target of cytokine-driven L-R interactions within canonical LN,
178 integrating signals from several immune cell subsets. Analogously, macrophages were the
179 dominant source of cytokine-driven L-R interactions, broadly targeting immune cell subsets
180 within disrupted LN (**Fig. 3d**). Transforming growth factor beta (*TGFB1*) was the dominant
181 source of cytokine-mediated crosstalk within canonical LN, whereas several innate inflammatory
182 cytokines, including IL-1 (*IL1A, IL1B*) and TNF, dominated cytokine-mediated crosstalk in
183 disrupted LN samples (**Fig. 3d**).
184

185 We then assessed whether paired tumors and their dLN showed co-varying cytokine-driven
186 reprogramming. The two dLN types led to significantly different tumor type distributions (p-value
187 = 0.04, chi-squared test); canonical dLN were more often paired with T-cell engaged tumors,
188 while disrupted dLN were more often paired with either immune silent or inflamed myeloid tumors
189 (**Fig. 3e**). Individual patient tumor and dLN types are listed in **Extended Data Table 5**. There
190 was no clear correlation of tumor genomic features with either tumor or dLN type (**Extended Fig.**
191 **3c**). Interestingly, there was an uneven distribution amongst traditional tumor types, with diffuse-
192 type tumors, which carry the poorest prognosis amongst gastric tumor subtypes[17], pairing
193 exclusively with disrupted dLN (**Fig. 3f**). As T cell priming occurs in the lymph node, these results
194 suggest that altered priming within dLN contributes to the lack of intratumoral T cell surveillance
195 in these tumor subtypes.
196
197 **Reduced T cell priming in disrupted tumor draining lymph nodes**
198 Linked cytokine-driven reprogramming between tumor and matched dLN suggests that chronic
199 inflammation-driven loss of dendritic cell homing to disrupted dLN might impair adaptive immune
200 responses. To investigate this, we next focused on the impact of canonical and disrupted dLN
201 on CD8+ T cell phenotypic trajectories. We generated a CD8+ T cell trajectory manifold
202 (**Methods**) and mapped T cells according to their transcriptional phenotypes (**Fig. 4a**). T cells
203 from dLN and tumors had limited phenotypic overlap (**Fig. 4b**). A computational fate analysis
204 (**Methods**) revealed 6 primary terminal states for naïve CD8+ T cells, including activated,
205 effector, EMRA, Neat1+, and proliferating fates, as well as a cell fate in which T cells remained
206 in a naïve-like state (**Fig. 4c**). Using single-cell TCR analysis, we confirmed that clonally
207 expanded T cells were mostly, though not exclusively, in the tumor (**Extended Fig. 4a**). The
208 tumor-resident clonally expanded T cells were comprised of a wide phenotypic range of effector
209 and exhausted states, while dLN expanded T cells were primarily in a central memory state
210 (**Extended Fig. 4a**).

211

212 We next focused on comparing T cell differentiation trajectories from canonical and disrupted
213 LN priming. In canonical LN-tumor pairs, dLN CD8+ T cells were primarily comprised of a central
214 memory phenotype and were found in all six terminal states within the tumor. In sharp contrast,
215 T cells from disrupted LN-tumor pairs were restricted to undifferentiated naïve-like states within
216 the dLN, and their tumor resident trajectories were substantially more restricted (**Fig. 4d**). We
217 attributed this constrained differentiation directly to altered priming within the dLN, as TCR
218 analysis revealed that CD8+ T cells within disrupted LN displayed reduced clonal expansion and
219 a reduced phenotypic entropy (**Fig. 4e**), with the majority of CD8+ T cells occupying either a
220 naïve or EMRA phenotypic state (**Fig. 4f**).

221

222 Curiously, a fraction of phenotypically naïve CD8+ T cells from disrupted dLN downregulated
223 surface expression of CD45RA, which typically occurs in the context of antigen experience[18,19]
224 (**Fig. 4g**), despite being transcriptionally similar to naïve cells (**Fig. 4h**). We therefore stratified
225 phenotypically naïve T cells based on CD45 isoform expression (CD45RA vs. CD45RO,
226 included in our CITE-seq antibody-derived tags (ADT) panel). Phenotypically naïve T cells
227 negative for CD45RA and positive for CD45RO ('RA^-RO^+') expressed markers consistent with
228 antigen experience on the protein level, including downregulation of L-selectin and CD3[20,21], as
229 well as upregulation of Fas[22], consistent with being recently activated. In contrast, T cells
230 negative for CD45RA and CD45RO ('RA^-RO^-') largely overlapped with CD45RA+, CD45RO-
231 canonically naïve T cells, with sustained expression of L-selectin and CD3 (**Fig. 4i**). Trajectory
232 analysis of these 3 phenotypically naïve T cell subsets along with early activated T cells
233 confirmed that RA^-RO^- T cells were similar to RA^+RO^- naïve T cells, whereas RA^-RO^+ cells were
234 more similar to early activated T cells (**Fig 4j, Extended Fig. 4b**). We therefore combined double
235 negative RA^-RO^- T cells with RA^+RO^- true naïve cells and RA^-RO^+ cells with early activated T
236 cells for downstream analysis. We found that compared to canonical dLN, T cells within

237 disrupted dLN were enriched in either a naïve or EMRA phenotype, indicating that inflammatory
238 reprogramming of disrupted dLN is sufficient to limit T cell priming and downstream tumor
239 surveillance (**Extended Fig. 4c**).
240

241 **KLF2 suppresses effector T cell differentiation in gastric tumor-draining LN**

242 Despite showing increased clonal expansion relative to T cells from disrupted LN, T cells from
243 canonical LN also showed evidence of phenotypic constraint, with the majority of clonally
244 expanded T cells within dLN remaining in a central memory state (**Extended Fig. 5a**). We
245 therefore sought to identify transcriptional regulators that might restrict differentiation and egress
246 of CD8+ T cells from dLN. We identified differentially expressed transcription factors in clonally
247 expanded CD8+ T cells compared with clonally expanded T cells from tumors and found
248 Kruppel-like Factor-2 (*KLF2*)[23] as the most differentially upregulated transcription factor in
249 clonally expanded T cells retained within dLN (**Fig. 5a**). Notably, within both dLN and tumors,
250 *KLF2* expression correlated strongly with expression of genes linked to quiescence (*LEF1,*
251 *TCF7*) and LN retention (*S1PR1, S1PR4, CCR7, SELL*) and inversely with genes linked to tissue
252 infiltration (*ITGAE*, *CXCR6*) and effector function (*IRF4, GZMB*) (**Fig 5b**). Moreover, as *KLF2*
253 expression inversely correlated with clone size amongst expanded clones within the LN and
254 tumor, we inferred that *KLF2* expression may be a regulator of quiescence and LN retention
255 while restricting T cell expansion, differentiation, and LN egress (**Fig. 5c**).
256

257 To directly assess the consequences of *KLF2* stabilization on T cell activation and differentiation,
258 we transduced activated primary human T cells with either lentivirus encoding *KLF2* or a vector
259 control (**Extended Fig. 5b**) and determine the impact of *KLF2* overexpression on T cell
260 phenotypes using a combination of flow cytometry and combined single-cell RNA and ATAC
261 sequencing (DOGMA-Seq)[24]. *In vitro* activated T cells differentiated into several T cell states
262 marked by distinct transcriptional modules, including proliferating, effector, effector memory, and

263 central memory T cells, with a minority of T cells occupying innate-like, activated, or intermediate
264 states (**Fig 5d, Extended Fig. 5c**). Naïve cells were not present as T cells had been previously
265 activated to facilitate lentiviral transduction. Consistent with *KLF2* expression and accessibility
266 marking a central memory phenotype[25], we found that overexpression of *KLF2* in primary T cells
267 led to an increased abundance of CD8+ T cells with a central memory phenotype and a
268 decreased abundance of T cells with an effector memory phenotype (**Fig. 5e,f**). To determine
269 the mechanism by which *KLF2* alters T cell differentiation, we identified differentially accessible
270 motifs in central memory versus effector memory T cells using ChromVar[26]. We found that
271 CCCTC-binding factor (*CTCF*) and Interferon Regulatory Factor (*IRF*) motif accessibility were
272 associated with effector memory T cells, whereas *FOXO1* and T-box (*TBX*) family members
273 were associated with central memory T cells (**Fig. 5g**). Measurement of promoter-centered
274 chromatin accessibility confirmed that promoter-associated accessibility of both *CTCF* and *IRF*
275 transcription factors was reduced in *KLF2*-overexpressing cells (**Fig. 5h**). As both *CTCF* and
276 *IRF* transcription factors are known to be required for effector CD8+ T cell differentiation[27-32], we
277 measured both effector T cell differentiation and cytotoxic function and found them to be reduced
278 in *KLF2* overexpressing T cells (**Fig 5i, Extended Fig. 5d**).
279
280 Finally, to identify potential mechanisms by which *KLF2* might limit T cell differentiation and
281 tumor infiltration, we identified putative *KLF2* target genes by asking which genes containing a
282 KLF-binding motif[33] (CCACGCCC) were differentially expressed in *KLF2* versus vector
283 overexpressing CD8+ T cells. *KLF2* overexpression was associated with reduced expression of
284 target genes associated with effector memory T cells, such as *TOX*, *TOX2*, and *ITGAE* and
285 increased expression of the GTPase activating proteins (GAP) *RAPGAP2* and *RASA3* (**Fig. 5j**).
286 *RAPGAP2* and *RASA3* bind and stimulate the catalytic activity of the Rap1 GTPase[34-36]. Rap1
287 has been shown to be required for the TCR-dependent conformational change in Lymphocyte
288 function-associated antigen-1 (LFA-1) that enables interaction with its ligands, otherwise known

289 as ‘inside out’ signaling. This led us to hypothesize that *KLF2*-driven expression of Rap1 GAPs
290 may limit T cell priming by disrupting LFA-1 activation. Consistent with this hypothesis, surface
291 expression of both components of the LFA-1 heterodimer (CD11a and CD18) was reduced in
292 naïve and central memory T cells from disrupted LN; moreover, *KLF2* expression *in vitro* was
293 sufficient to lower expression levels of CD11a and CD18 when compared to our empty vector
294 control (**Fig. 5k**). To directly test the impact of LFA-1 activity on T cell priming and differentiation,
295 we activated OT-I TCR transgenic T cells with cognate peptide across a range of antigen dose
296 and affinities using altered peptide ligands of the canonical OT-I TCR cognate peptide SIINFEKL
297 (**Extended Fig. 6a**)[37]**.** Addition of LFA-1 blocking antibodies[38] during T cell priming by peptide-
298 pulsed DCs increased the threshold for activation, proliferation, and effector function across
299 peptide doses and affinities, providing a mechanism by which *KLF2* stabilization influences both
300 T cell priming and effector differentiation in GC (**Extended Fig. 6b,c**).
301
302 **Dynamic state switching of cytokine driven dLN types in a remarkable patient**
303 Lymph nodes are typically only surgically resected during gastrectomy, as a fundamental part of
304 staging. However, for one specific patient (GC028) a biopsy of a visible draining lymph node
305 was collected during diagnostic laparoscopy (DL) for evaluation of metastatic gastric cancer as
306 well as during surgical resection after achieving a near complete response to neoadjuvant
307 chemotherapy (FOLFOX)[39] (**Fig. 6a**). This unique pairing of tumor and draining lymph node both
308 before and after neoadjuvant chemotherapy allowed us to investigate whether cytokine driven
309 dLN types remain fixed in the context of therapy. Classification of both pre-treatment and post-
310 treatment samples using FRACTAL revealed a disrupted dLN prior to treatment and a canonical
311 dLN at the time of resection (**Fig. 6b**). Accordingly, the post-treatment dLN showed a decreased
312 abundance of C1QC+ macrophages and an increased abundance of plasmacytoid DCs (**Fig.
313 6c, Extended Fig. 7a**).
314

315 Clonotype tracking between the dLN and tumor pairs revealed that shared clones were present
316 between all 4 samples (**Extended Fig. 7b**). Phenotypes of shared CD8+ T cell TCR clonotypes
317 changed post-treatment, with an increase in the fraction of progenitor Texh in the tumor and CM
318 in the dLN (**Extended Fig. 7c**). The post-treatment canonical dLN also had both a larger number
319 of expanded clones and a greater diversity of expanded T cell phenotypes from the dLN to its
320 paired tumor (**Fig. 6d**). We examined overall *KLF2* expression in CD8+ T cells in the pre-therapy
321 disrupted dLN and the post-therapy canonical dLN and found that *KLF2* expression was
322 significantly higher in the disrupted dLN (pre) than the canonical dLN (post) (**Fig. 6e**). *KLF2* was
323 also the second most transcriptionally decreased gene between these pre-treatment and post-
324 treatment dLN (**Extended Fig. 7d**). Notably, expression of both LFA-1 components (CD11a and
325 CD18) was also increased in the canonical dLN (post-treatment) (**Extended Fig. 7e**). These
326 observations imply that the disrupted and canonical types we report above are not fixed in
327 disease evolution. Rather, changes in *KLF2* expression coupled with cytokine driven dLN
328 dynamics suggest therapeutic intervention can modulate lymph node priming in gastric cancer.
329 The clinical and prognostic significance of this single patient remains unclear, however we note
330 that this patient remains in complete remission 2.7 years after diagnosis despite presenting with
331 metastatic disease.
332

333 **Discussion**

334 Our study sought to understand the impact of chronic inflammation on lymph node priming
335 events during gastric tumorigenesis. Through multi-modal single cell investigation of gastric
336 tumors and matched draining lymph nodes, we provide here a novel mechanistic explanation for
337 how persistent innate inflammation can impair anti-tumor T cell immunity. Our findings are
338 consistent with reduced dendritic cell accumulation within disrupted draining lymph nodes,
339 preventing exposure of putatively tumor-reactive naïve T cells within the lymph node to tumor-
340 derived neoantigens. While diffuse-type GC had exclusively disrupted dLN, we identified this LN

341 phenotype across every molecular GC subtype, indicating that LN disruption suppresses
342 immune surveillance in GC across histologic and molecular classifications. These findings
343 nominate DC homing and T cell priming as potential therapeutic targets that cross previously
344 reported GC stratifications.
345
346 We also identify a second, previously unappreciated checkpoint on T cell surveillance in gastric
347 cancer. Activated T cells within the draining LN largely fail to differentiate into effectors, instead
348 becoming central memory T cells that remain in the LN and express high levels of *KLF2*. This
349 may be in part due to reduced integrin activity increasing the threshold for T cell activation during
350 naïve T cell priming events within dLN. These findings are supported by recent evidence
351 demonstrating a role for *KLF2* in preventing T cell exhaustion[40]. By suppressing the exhaustion
352 program, *KLF2* may paradoxically reduce the clinical efficacy of immune checkpoint inhibitors
353 that target inhibitory receptors preferentially expressed on exhausted T cells, including PD-1,
354 CTLA-4, and LAG-3. Future studies to determine if *KLF2* expression in T cells is associated
355 with lack of immune checkpoint blockade (ICB) response or whether targeting either *KLF2* or
356 *KLF2*-dependent integrin activation can enhance sensitivity to ICB would have immediate clinical
357 relevance.
358
359 Our findings provide insights into how reciprocal balances between innate and adaptive immune
360 responses maintain homeostasis in an inflammatory microenvironment, and how an imbalance
361 between these responses contributes to human disease. The extent to which the T cell
362 phenotypes observed in this study apply to other tumor types remains unknown. However, this
363 unique dataset, derived from prospective surgical sampling of paired tumors and draining LN,
364 may serve as a valuable reference to understand whether lymph node priming is disrupted in
365 other cancers. Blunting of the adaptive immune response during chronic inflammation may
366 represent a physiologic mechanism to preserve mucosal tolerance and barrier integrity;

367 consistent with this hypothesis, T cell specific *KLF2* expression is required to maintain
368 commensal tolerance and prevent inflammatory bowel disease[41]. However, adaptive immune
369 suppression has been reported in other chronic inflammatory conditions, including sepsis-
370 associated immune paralysis[42]. Potential targets such as *KLF2* may enable T cells to escape
371 from this maladaptive state. Furthermore, our anecdotal evidence that dLN can change with
372 intervention supports tumor draining lymph nodes as a viable focus for therapeutic development.
373 We report that a patient with pre- and post-treatment tissue collections demonstrated phenotypic
374 T cell state transitions linked to a remarkable clinical response. This indicates that the non-
375 canonical immune states we describe in this study are potentially reversible with therapeutic
376 intervention. Characterizing the prevalence and determinants of such immune state transitions
377 will be central to future translational efforts aimed at enhancing lymph node priming in gastric
378 cancer.

379 **Figure Legends**

380

381 **Figure 1**. **Patient cohort and sample collection.** (a) Schematic of surgical sample collection
382 from gastric cancer patients and processing for combined single cell RNA-seq, TCR-seq and
383 surface epitope analysis. (b) Swimmer's plot showing individual patients with samples collected,
384 tumor characteristics, and simplified patient timelines. Samples are organized by those with
385 paired tumor and dLN collections, and 'additional sample pairs' without both tumor and dLN.

386

387 **Figure 2: Single cell characterization of immune cells from primary tumor, distal normal,**
388 **draining lymph nodes, and metastases in GC.** Uniform manifold approximation and
389 projections **(**UMAP) depicting (a) broad CD45+ cell subsets (left), density heatmap of CD45+
390 cells from dLN (center) and tumors (right), (b) CD8+ T cell subsets (left), density heatmap of
391 CD8+ cells from dLN (center) and tumors (right), (c) CD4+ T cell subsets (left), density heatmap
392 of CD4+ cells from dLN (center) and tumors (right), and (d) myeloid cell subsets (left), density
393 heatmap of myeloid cells from dLN (center) and tumors (right).

394

395 **Figure 3**. **Patient stratification by distinct cytokine driven immune networks in tumors and**
396 **their draining lymph nodes.** (a) Workflow of FRACTAL - Factor and Receiver Analysis of
397 Cytokines in Tumor And Lymph nodes our sample level Bayesian sparse tensor decomposition
398 method. (b) Clustering of individual patient samples from dLN (left) and tumors (right) based on
399 FRACTAL decompostion from (a). (c) Summary of enriched cytokine-cell type interactions in
400 dLN and tumor subtypes with UMAPs showing the fraction of myeloid cell type content and
401 paired myeloid cell density. (d) Significant cytokine ligand-receptor interactions identified using
402 CellChat, with accompanying normalized gene expression. Interaction probability strength is
403 shown relative to the disrupted lymph node samples, with blue arrows indicating interactions
404 more probable in the canonical and red arrows more probable in the disrupted lymph node

405 samples. (e) Correlation of cytokine interaction-dependent dLN and tumor subtypes in patients
406 with paired samples demonstrates a significant association between dLN and tumor types as
407 indicated (p-value = 0.04, chi-squared test) (f) Distribution of Lauren and molecular subtypes
408 across paired dLN-Tumor networks.
409
410 **Figure 4**. **Reduced T cell priming in disrupted tumor draining lymph nodes.** (a-d) Trajectory
411 manifold of CD8+ T cell subsets across dLN and tumor, showing distribution of (a) CD8+ T cell
412 subsets, (b) distribution within dLN and tumors, (c) fate probabilities, and (d) density mapping of
413 cells from 'disrupted-immune silent' pairs and 'canonical-T cell engaged' pairs. (e) Fraction of
414 shared expanded (clonotypes >20) CD8+ T cell clones between dLN and Tumor (left) and
415 average phenotype entropy in CD8+ T cells (right) (**Methods**, section *Phenotype entropy*),
416 stratified by canonical dLN versus disrupted dLN. (f) Difference in mean cell fractions of CD8+
417 T cell subsets between disrupted dLN and canonical dLN. (g) Surface CD45 isoform expression
418 distribution on CD8+ T cells from disrupted dLN with a transcriptionally naïve phenotype. (h)
419 Cells from (g) projected on the trajectory manifold from (a), coded by CD45 isoform expression.
420 (i) Expression of genes and proteins associated with T cell activation on phenotypically naïve T
421 cells from (g,h), stratified by surface expression of CD45 isoforms. (j) Density heatmaps of
422 subtypes of naïve cells from (g-i), projected onto the CD8+ trajectory manifold from (a).
423
424 **Figure 5**. **KLF2 suppresses effector T cell differentiation.** (a) Differentially expressed
425 transcription factors expressed in non-naïve CD8+ T cells from tumors (up) vs. dLN (down). (b)
426 Genes correlated with the KLF2 expression in non-naïve CD8+ T cells across tumors and dLN
427 (c) Correlation of KLF2 expression on non-naïve CD8+ T cells across tumors and dLN with
428 clonotype size. (d) UMAP of activated human CD8+ T cells transduced with either empty vector
429 or a KLF2 overexpression vector. (e) Density plot showing relative distribution of CD8+ T cell
430 phenotypes in primary human CD8+ T cells transduced with either empty vector (left) or KLF2

431 overexpressing vector (right). (f) Mean fraction of T cell subsets in primary human CD8+ T cells
432 transduced with either empty vector or KLF2 overexpressing vector as indicated. (g)
433 Transcription factor motifs with top accessibility scores in central memory (right) compared with
434 effector memory (left) T cell subsets. (h) Volcano plot demonstrating accessibility of CTCF/L
435 (left) or IRF4/8 (right) motifs, centered around the transcriptional start site, in human CD8+ T
436 cells transduced with empty vector or KLF2 overexpression vector. (i) Expression of CXCR6 and
437 PD-1 (left) and intracellular accumulation of IFN-g and TNF following restimulation with PMA and
438 ionomycin in CD4+ and CD8+ T cells transduced with empty vector or KLF2 overexpression
439 vector. *ADT* antibody-derived tags (j) Relative expression of genes (KLF2 overexpression vector
440 vs. Empty vector) with an overlapping KLF2 motif (k) Differential surface expression of the LFA-
441 1 subunits CD11a and CD18 in indicated CD8+ T cell subsets from dLN, stratified by canonical
442 vs. disrupted dLN phenotype (left) and in primary human T cells transduced with empty vector
443 or KLF2 overexpression vector.
444
445 **Figure 6**. **Cytokine driven dLN types are not fixed.** (a) Patient timeline with tumor and LN
446 collection pre- and post-neoadjuvant chemotherapy treatment. (b) UMAP projections of myeloid
447 cells (left) and CD8+ T cells (right) for patient GC028 across both pre-treatment and post-
448 treatment conditions, colored by either cell subset or distribution in tumors and dLN. (c) Heat-
449 encoded density embeddings of myeloid and CD8+ T cell compartments prior to and following
450 neoadjuvant treatment, stratified by timepoint and anatomical site. (d) Alluvial plots showing
451 CD8+ T cells with shared TCRs across dLN and tumor, both pre- and post-treatment (e) KLF2
452 expression in CD8+ T cells within dLN of patient GC028, pre- and post-treatment.
453
454
455 **Extended Figure Legends**
456

457 **Extended Figure 1**. **Marker genes for major CD45+ cell compartments.** The expression of
458 marker genes in each major CD45+ cell compartment (a), CD8+ T cells (b), CD4+ T cells (c),
459 cycling cells in CD4+ and CD8+ T cells (d) and myeloid cells (e). (f) Fraction of cells in the major
460 CD45+ cell compartments within tumors and dLN as indicated.
461
462 **Extended Figure 2**. **Single cell characterization of immune cells from normal tissue and**
463 **metastatic sites in GC.** UMAPs depicting (a) broad CD45+ cell subsets across tissue sites (left),
464 density heatmap of CD45+ cells from distant normal tissue (center) and metastases (right), (b)
465 CD8+ T cell subsets across tissue sites (left), density heatmap of CD8+ cells from distant normal
466 tissue (center) and metastases (right), (c) CD4+ T cell subsets across tissues (left), density
467 heatmap of CD4+ cells from distant normal tissue (center) and metastases (right), and (d)
468 myeloid cell subsets across tissue sites (left), density heatmap of myeloid cells from distant
469 normal tissue (center) and tumors (right).
470
471
472 **Extended Figure 3**. **Patient stratification by distinct cytokine driven immune networks in**
473 **tumors and their draining lymph nodes.** (a) Posterior mean for latent sample by factor (left),
474 cell-type by factor (center), and cytokine by factor (right) tensor slices. For visualization purposes,
475 the cytokine by factor slide heatmap includes only cytokines with an average factor activation
476 score of more than 0.2. (b) UMAPs showing the fraction of CD8+ T cell type content and paired
477 CD8+ T cell density (left) and fraction of CD4 and myeloid cell subsets in each cytokine
478 interaction-driven immune network cluster. (c) Oncoplot of tumor mutation[43], separated by either
479 dLN or tumor subtype.
480
481 **Extended Figure 4**. **CD8+ T cell expansion.** (a) CD8+ T cell clonotype size distribution in dLN
482 and Tumor (left) and cell phenotypes for expanded TCR clonotypes (right). (b) Average distance

483 of naïve like T cells from non-naïve T cell phenotypes indicating closer proximity of RA-RO+
484 naïve T cells to early activated T cells (**Methods**). (c) Difference in mean cell fractions of CD8+
485 T cell subsets between disrupted dLN and canonical dLN following recategorization of naïve-like
486 CD8+ T cells (RA^+RO^- naïve and RA^-RO^- naïve into 'naïve merged', RA^-RO^+ naïve and early
487 activated into 'early activated merged').
488
489 **Extended Figure 5**. **KLF2 suppresses effector T cell differentiation.** (a) Fraction of CD8+ T
490 cell phenotypes in expanded TCR clonotypes in each cytokine-driven immune network cluster.
491 (b) Western blot showing KLF2 and Histone H3 (loading control) protein abundance in primary
492 activated human T cells transduced with either empty vector control or KLF2 overexpression
493 vector. (c) Marker genes and their expression in CD8+ T cell phenotypes reported in empty
494 vector and KLF2 overexpression vector-transduced cells from **Fig. 5d**. (d) Flow cytometry gating
495 for **Fig. 5i**.
496
497 **Extended Figure 6. LFA-1 blockade attenuates T cell priming and effector differentiation.**
498 (a) Experimental design to assess the contribution of LFA-1 to DC-mediated T cell priming and
499 effector differentiation. (b) Flow cytometry scatter plots and gating of live, $B220^-NK1.1^-$
500 $TCRb^+CD8^+$ T cells to assess T cell activation ($CD69^+CD25^+$), proliferation ($CD44^+$CellTrace
501 $Violet^{low}$), effector differentiation ($CX3CR1^-CXCR6^+$) and cytokine production ($TNF^+IFN\text{-}g^+$). (c)
502 Quantitation of impact of peptide dose and affinity on CD8+ T cell activation, proliferation,
503 effector differentiation, and cytokine production as described in (b). Each **table entry** represents
504 the **average** of three independent biological replicates for each peptide affinity and dose.
505 *p<0.05 **p<0.01 ***p<0.001
506
507 **Extended Figure 7**. **Cytokine driven dLN types are not fixed** (a) Fraction of CD8+ T (left) and
508 myeloid cell type subsets (right) in patient GC0028, stratified by tissue (dLN and tumor) as well

509 as timepoint (pre- and post-treatment) (b) Abundance of TCR clonotypes that are shared across
510 time and anatomical sites (left) and their projection onto the patient GC0028-specific CD8+ T
511 cell UMAP (right) (c) Alluvial plots showing phenotypes of CD8+ T cells with shared TCR pre-
512 and post-treatment in both dLN (left) and tumor (right) samples. (d) Differentially expressed
513 genes in the dLN pre-treatment (up) vs post-treatment (down). (e) Expression of CD11a and
514 CD18 in CD8+ T cells from patient GC0028 dLN pre- and post-treatment.
515

516 **EXTENDED DATA TABLES**

517 **Extended Data Table 1**: Individual patient and tumor characteristics

518 **Extended Data Table 2**: Patient demographics

519 **Extended Data Table 3**: Data subsets and cell counts

520 **Extended Data Table 4:** Top 50 DEGs for cell type cluster annotation for CD45+, CD8, CD4
521 and Myeloid subsets

522 **Extended Data Table 5**: Patients with dLN and tumor IREA types
523
524

525 **END NOTES**

526 **Acknowledgements**

527 This research was supported in part by the institution's Cancer Center Support Grant (P30
528 CA008748) from the NIH, and the Cycle for Survival-Equinox Grant (MA, EES, VES, SAV). SAV
529 was supported by a Damon Runyon Clinical Investigator Award. SS is supported by an NCI
530 Pathway to Independence Award (5 K99 CA277562-02). Additional support was provided by the
531 Halvorsen Center for Computational Oncology. SPS holds the Nicholls-Biondi Endowed Chair
532 in Computational Oncology. This work was additionally supported by research funding from
533 Bristol Myers Squibb. Figure 1a and Extended Figure 6a were made using Biorender.com. We
534 thank the MSKCC Single Cell Analysis and Innovation Lab (SAIL) for their thoughtful insights

and assistance, Anthony Verdone for assistance with data management, and Monique Du Plessis for assistance with Figure preparation. This work uses resources from the High-Performance Computing Group at Memorial Sloan Kettering Cancer Center.

Author Contributions

EES, SS, SS, MA, HL, VES, SPS, and SAV conceived the project and designed all experiments. SS, MA, HD, VES and EES performed all sample acquisition. EES and SS assembled all metadata and reviewed clinical characteristics. YHL, SS and EES performed all tissue processing. YHL and EES performed all spectral cytometry assays. YHL and EES performed all laboratory experiments. SS, HL, MZ, and NC performed all computational and statistical analyses. EES, SS, HL, SPS, VES and SAV wrote the manuscript; all authors reviewed the manuscript.

Competing Interests Statement

VES has received speaking honoraria from Merck Pharmaceuticals and Astra Zeneca. SAV is an advisor for Generate Bio and for the Institute for Follicular Lymphoma Innovation (I-FLI), previously served as a consultant for Koch Disruptive Technologies and has received research funding from Bristol Meyers Squibb. SPS reports grants/contracts from AstraZeneca and Bristol Meyers Squibb and a leadership role in the Genome Canada Scientific Advisory Council. ML is supported by a grant from the NIDDK of the NIH (K08 DK125876). YJ has received research funding from Astellas, AstraZeneca, Arcus Biosciences, Bayer, Bristol-Myers Squibb, Cycle for Survival, Department of Defense, Eli Lilly, Fred’s Team, Genentech/Roche, Inspirna, Merck, NCI, Stand Up 2 Cancer, and Transcenta. YJ is on the advisory board/consults for Abbvie, Alphasights, Amerisource Bergen, Ask-Gene Pharma, Inc., Arcus Biosciences, Astellas, Astra

Zeneca, Basilea Pharmaceutica, Bayer, Boehringer Ingelheim, Bristol-Myers Squibb, Clinical Care Options, Daiichi-Sankyo, eChina Health, Ed Med Resources (OncInfo), Eisai, Eli Lilly, Geneos Therapeutics, GlaxoSmithKline, Guardant Health, Inc., H.C. Wainwright & Co., Health Advances, HMP Global, Imedex, Imugene, Inspirna, Lynx Health, Mashup Media LLC, Master Clinician Alliance, Merck, Merck Serono, Mersana Therapeutics, Michael J. Hennessy Associates, Oncology News, Paradigm Medical Communications, PeerMD, PeerView Institute, Pfizer, Physician's Education Resource, LLC, Research to Practice, Sanofi Genzyme, Seagen, Silverback Therapeutics, Suzhou Liangyihui Network Technology Co., Ltd, Talem Health, TotalCME, and Zymeworks Inc.. YJ also discloses stock options in Inspirna and Veda Life Sciences, Inc.. No other authors have relationships with outside entities to disclose.

Data and Code Availability

Supplementary Information is available for this paper. Scripts used in the analysis, including the implementation for FRACTAL is available online at https://github.mskcc.org/shahcompbio/gac_code.

REFERENCES

1 *SEER*Explorer: An interactive website for SEER cancer statistics [Internet].* <https://seer.cancer.gov/statistics-network/explorer/> (2025 Apr 16).

2 Schistosomes, liver flukes and Helicobacter pylori. *IARC Monogr Eval Carcinog Risks Hum* **61**, 1-241 (1994).

3 Fu, K. *et al.* Streptococcus anginosus promotes gastric inflammation, atrophy, and tumorigenesis in mice. *Cell* **187**, 882-896 e817 (2024). https://doi.org/10.1016/j.cell.2024.01.004

4 Norcross, J. W., Monroe, S. E. & Griffin, B. G. The development of gastric carcinoma in pernicious anemia. *Ann Intern Med* **37**, 338-343 (1952). https://doi.org/10.7326/0003-4819-37-2-338

5 Sasaki, H. *et al.* Presence of Streptococcus anginosus DNA in esophageal cancer, dysplasia of esophagus, and gastric cancer. *Cancer Res* **58**, 2991-2995 (1998).

6 Zamcheck, N., Grable, E., Ley, A. & Norman, L. Occurrence of gastric cancer among patients with pernicious anemia at the Boston City Hospital. *N Engl J Med* **252**, 1103-1110 (1955). https://doi.org/10.1056/NEJM195506302522601

7 Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61 (2015). https://doi.org/10.1016/j.cell.2014.12.033

8 Abate, M. *et al.* Host Tissue Factors Predict Immune Surveillance and Therapeutic Outcomes in Gastric Cancer. *Cancer Immunol Res* **13**, 591-601 (2025). https://doi.org/10.1158/2326-6066.CIR-23-0563

9 Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865-868 (2017). https://doi.org/10.1038/nmeth.4380

10 Lauren, P. The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification. *Acta Pathol Microbiol Scand* **64**, 31-49 (1965). https://doi.org/10.1111/apm.1965.64.1.31

11 Cancer Genome Atlas Research, N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202-209 (2014). https://doi.org/10.1038/nature13480

12 Bellmunt, J. *et al.* Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. *N Engl J Med* **376**, 1015-1026 (2017). https://doi.org/10.1056/NEJMoa1613683

13 Dranoff, G. Cytokines in cancer pathogenesis and cancer therapy. *Nat Rev Cancer* **4**, 11-22 (2004). https://doi.org/10.1038/nrc1252

14 Cui, A. *et al.* Dictionary of immune responses to cytokines at single-cell resolution. *Nature* **625**, 377-384 (2024). https://doi.org/10.1038/s41586-023-06816-9

15 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021). https://doi.org/10.1038/s41467-021-21246-9

16 Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. *Nat Protoc* **20**, 180-219 (2025). https://doi.org/10.1038/s41596-024-01045-4

17 Nakauchi, M. *et al.* Outcomes of Neoadjuvant Chemotherapy for Clinical Stages 2 and 3 Gastric Cancer Patients: Analysis of Timing and Site of Recurrence. *Ann Surg Oncol* **28**, 4829-4838 (2021). https://doi.org/10.1245/s10434-021-09624-5

18 Akbar, A. N., Terry, L., Timms, A., Beverley, P. C. & Janossy, G. Loss of CD45R and gain of UCHL1 reactivity is a feature of primed T cells. *J Immunol* **140**, 2171-2178 (1988).

19 Serra, H. M., Krowka, J. F., Ledbetter, J. A. & Pilarski, L. M. Loss of CD45R (Lp220) represents a post-thymic T cell differentiation event. *J Immunol* **140**, 1435-1441 (1988).
20 Tedder, T. F., Penta, A. C., Levine, H. B. & Freedman, A. S. Expression of the human leukocyte adhesion molecule, LAM1. Identity with the TQ1 and Leu-8 differentiation antigens. *J Immunol* **144**, 532-540 (1990).
21 Telerman, A. *et al.* Internalization of human T lymphocyte receptors. *Eur J Immunol* **17**, 991-997 (1987). https://doi.org/10.1002/eji.1830170715
22 Miyawaki, T. *et al.* Differential expression of apoptosis-related Fas antigen on lymphocyte subpopulations in human peripheral blood. *J Immunol* **149**, 3753-3758 (1992).
23 Kuo, C. T., Veselits, M. L. & Leiden, J. M. LKLF: A transcriptional regulator of single-positive T cell quiescence and survival. *Science* **277**, 1986-1990 (1997). https://doi.org/10.1126/science.277.5334.1986
24 Mimitou, E. P. *et al.* Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat Biotechnol* **39**, 1246-1258 (2021). https://doi.org/10.1038/s41587-021-00927-2
25 Schober, S. L. *et al.* Expression of the transcription factor lung Kruppel-like factor is regulated by cytokines and correlates with survival of memory T cells in vitro and in vivo. *J Immunol* **163**, 3662-3667 (1999).
26 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978 (2017). https://doi.org/10.1038/nmeth.4401
27 Liu, J. *et al.* CTCF mediates CD8+ effector differentiation through dynamic redistribution and genomic reorganization. *J Exp Med* **220** (2023). https://doi.org/10.1084/jem.20221288
28 Raczkowski, F. *et al.* The transcription factor Interferon Regulatory Factor 4 is required for the generation of protective effector CD8+ T cells. *Proc Natl Acad Sci U S A* **110**, 15019-15024 (2013). https://doi.org/10.1073/pnas.1309378110
29 Yao, S. *et al.* Interferon regulatory factor 4 sustains CD8(+) T cell expansion and effector differentiation. *Immunity* **39**, 833-845 (2013). https://doi.org/10.1016/j.immuni.2013.10.007
30 Man, K. *et al.* Transcription Factor IRF4 Promotes CD8(+) T Cell Exhaustion and Limits the Development of Memory-like T Cells during Chronic Infection. *Immunity* **47**, 1129-1141 e1125 (2017). https://doi.org/10.1016/j.immuni.2017.11.021
31 Harberts, A. *et al.* Interferon regulatory factor 4 controls effector functions of CD8(+) memory T cells. *Proc Natl Acad Sci U S A* **118** (2021). https://doi.org/10.1073/pnas.2014553118
32 Miyagawa, F. *et al.* Interferon regulatory factor 8 integrates T-cell receptor and cytokine-signaling pathways and drives effector differentiation of CD8 T cells. *Proc Natl Acad Sci U S A* **109**, 12123-12128 (2012). https://doi.org/10.1073/pnas.1201453109
33 Miller, I. J. & Bieker, J. J. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins. *Mol Cell Biol* **13**, 2776-2786 (1993). https://doi.org/10.1128/mcb.13.5.2776-2786.1993
34 Schultess, J., Danielewski, O. & Smolenski, A. P. Rap1GAP2 is a new GTPase-activating protein of Rap1 expressed in human platelets. *Blood* **105**, 3185-3192 (2005). https://doi.org/10.1182/blood-2004-09-3605

35 Molina-Ortiz, P. *et al.* Rasa3 controls megakaryocyte Rap1 activation, integrin signaling and differentiation into proplatelet. *PLoS Genet* **10**, e1004420 (2014). https://doi.org/10.1371/journal.pgen.1004420

36 Stefanini, L. *et al.* RASA3 is a critical inhibitor of RAP1-dependent platelet activation. *J Clin Invest* **125**, 1419-1432 (2015). https://doi.org/10.1172/JCI77993

37 Zehn, D., Lee, S. Y. & Bevan, M. J. Complete but curtailed T-cell response to very low-affinity antigen. *Nature* **458**, 211-214 (2009). https://doi.org/10.1038/nature07657

38 Zumwalde, N. A., Domae, E., Mescher, M. F. & Shimizu, Y. ICAM-1-dependent homotypic aggregates regulate CD8 T cell effector function and differentiation during T cell activation. *J Immunol* **191**, 3681-3693 (2013). https://doi.org/10.4049/jimmunol.1201954

39 Wang, X. *et al.* A phase II study of a modified FOLFOX6 regimen as neoadjuvant chemotherapy for locally advanced gastric cancer. *Br J Cancer* **114**, 1326-1333 (2016). https://doi.org/10.1038/bjc.2016.126

40 Fagerberg, E. *et al.* KLF2 maintains lineage fidelity and suppresses CD8 T cell exhaustion during acute LCMV infection. *Science* **387**, eadn2337 (2025). https://doi.org/10.1126/science.adn2337

41 Shao, T. Y. *et al.* Kruppel-like factor 2+ CD4 T cells avert microbiota-induced intestinal inflammation. *Cell Rep* **42**, 113323 (2023). https://doi.org/10.1016/j.celrep.2023.113323

42 Schwandt, T. *et al.* Expression of type I interferon by splenic macrophages suppresses adaptive immunity during sepsis. *EMBO J* **31**, 201-213 (2012). https://doi.org/10.1038/emboj.2011.380

43 Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* **23**, 703-713 (2017). https://doi.org/10.1038/nm.4333

METHODS

Patient Inclusion

Under Institutional Review Board approval (IRB 20-025) fresh tissue was collected from gastric tumors, distant normal gastric mucosa, draining gastric lymph nodes (dLN), and metastases from patients with gastric adenocarcinoma at Memorial Sloan Kettering Cancer Center (MSK). Patients had either institutional 6-107 and/or 12-245 tissue consent or specific IRB approval for tissue use after patient death. IRB approval was obtained for patients under 18. Tissues were collected from patients undergoing either gastric resection or diagnostic laparoscopy (DL) with esophagogastroduodenoscopy (EGD) as part of treatment or staging/diagnosis. Gastric resections were either open or robotic assisted. Biopsies taken during EGD were removed using cold endoscopic forceps, and biopsies taken during DL were removed using cold laparoscopic graspers. Pathology was reviewed by institutional gastrointestinal pathologists and all tumors were pathologically confirmed to be gastric adenocarcinoma. The 8th edition of the AJCC staging system was used by convention for pathologic TNM staging.[1] Clinical information for patients was collected prospectively in a clinical gastric database. Lauren classification[2] and TCGA classifications[3] were combined to describe patients with intestinal, diffuse, mixed, MSI-H and EBV tumors. MSI-H or EBV designation superseded Lauren classification. Lauren classification was determined by pathology report. MSI-H status was determined by immunohistochemistry (IHC) of MLH1, MSH2, MSH6, and PMS2 as reported by pathology or by tumor mutation burden on institutional genomic testing (MSK-IMPACT).[4] EBV positivity was determined by IHC on pathology report. HER2 status was reported as negative, positive, or NA (not available) as a summary of HER2 IHC and fluorescence in situ hybridization (FISH) analysis for equivocal IHC results from pathology reports. PD-L1 scores were reported as CPS from 0-100, with any scores over 100 reduced to 100. For patients who received systemic therapy prior to resection, tumor response grade (TRG) was reported as called in the institutional pathology report from 0-3

28 according to the NCCN system.[5] *H. pylori* history was determined by a previous positive IHC
29 test or a documentation of previous *H. pylori* infection or treatment in a patient's chart.
30 Treatments were grouped as chemotherapy, immunotherapy, trastuzumab, radiation, or other
31 and were categorized as prior to or following specimen collection. Survival, follow-up, and
32 recurrence were collected up to January 17, 2025. Outcomes for two patients were not included
33 due to their enrollment in a clinical trial after specimen collection. Disaggregated data could not
34 be publicly shared for **7** patients due to IRB restrictions.
35
36 **Sample Collection and Processing**
37 As mentioned above, tissue was collected in the operating room from clinically indicated gastric
38 cancer resections or DLs with EGDs. During robotic gastric resections, Indocyanin Green (ICG)
39 was injected intra-operatively to mark the tumor bed and highlight tumor-draining lymph nodes
40 as previously described.[6] During open operations, tumor draining lymph nodes were determined
41 by the operating surgeon according to tumor location and lymph node station. For gastrectomies,
42 within 3 minutes of resection, 0.5-1.0 cm^3 of tumor, distant normal gastric mucosa (defined as
43 macroscopically normal appearing to the surgeon and not directly adjacent to the tumor), and
44 LNs (identified to be draining nodes by ICG or station by operating surgeon) were isolated, cut
45 into ~1 mm^3 pieces and placed into 15mL conical tubes containing phosphate buffered saline
46 (PBS). These samples were placed on ice and immediately transported to the laboratory. During
47 DL with EGD tumor and distant normal samples were collected during EGD with endoscopic
48 cold forceps, 4 bites of tumor were collected then immediately suspended in PBS, as were 3
49 bites of distant normal gastric mucosa with the forceps cleaned in-between collections. If
50 peritoneal or liver metastases were seen on DL with sufficient tissue for pathologic diagnosis, a
51 biopsy of 0.5 cm^3 of metastasis was collected and immediately transferred to a 15 mL conical
52 tube with PBS. These samples were placed on ice and immediately transported to the laboratory

53 in the same manner as for gastric resections. All samples were digested using a GentleMACS
54 Dissociator in 2.2 ml or 4.7 mL (based on tissue size) of R-10 solution (RPMI, 10% fetal bovine
55 serum, 10mM HEPES, 1% penicillin-streptomycin, 1% L-glutamine) and Miltenyi enzymes
56 (Miltenyi Biotec Tumor Dissociation Kit, Human) at 37°C using GentleMACS program TK3. The
57 digested tissue was then filtered through a 100 µm cell strainer and washed with cold R-10.
58 Samples were centrifuged at 500 *g* x 5 minutes at 4°C and supernatant was aspirated. Pelleted
59 cells were resuspended in 1mL of ammonium-chloride-potassium (ACK) lysing buffer for 3
60 minutes at room temperature and then washed with 5 mL of R-10. Samples were centrifuged
61 again at 500 *g* x 5 minutes at 4°C and supernatant was aspirated, the cell pellet was
62 resuspended in 1mL of RPMI plus 2 uL of 20 mg/mL DNase and incubated at room temperature
63 for 5 minutes. After incubation 4 mL of R10 was added and cells were counted. Cells were spun
64 again at 500 *g* x 5 minutes at 4°C and resuspended in 1 million/mL aliquots in BAMBANKER
65 Serum-Free Cell Freezing Medium (Wako Chemicals) for cryopreservation at -80°C.
66

67 **Single cell immune profiling**

68 Samples were batched for multiplexed combined 3' single-cell RNA, 5' T-cell receptor (TCR),
69 and cell-surface epitope assessment (CITE-seq) (10x Genomics.) Selected samples in batches
70 of no more than 30 samples, were thawed from -80°C in a 37°C water bath with gentle agitation,
71 transferred to a 15 mL conical tube and then slowly diluted with 9mL of RPMI. Samples were
72 then centrifuged at 200 rpm for 10 min at 4°C and the supernatant was aspirated. Each sample
73 was resuspended in 50 uL of 1:400 dilution in PBS of Ghost Dye 510 (Tonbo) and transferred to
74 a 96-well V-bottom plate, then incubated at room temperature in the dark for 10 minutes.
75 Samples were washed with 100 uL of PBS, centrifuged at 700 *g* for 2 minutes at 4°C, and
76 supernatant was discarded. Each sample was then resuspended in 25 uL of Human TruStain
77 FcX (Biolegand) diluted 1:20 in PBS and incubated for 10 minutes in the dark on ice. Then an

78 antibody pool of TotalSeq-C Human Universal Cocktail V1.0 (Biolegend), TotalSeqC anti-human
79 Hashtag Antibodies (Biolegend) was added, and samples were incubated for 30 minutes on ice.
80 Samples were washed with 100 uL of PBS, centrifuged at 700 *g* for 2 minutes at 4°C and
81 supernatant was discarded. Cells were then resuspended in 200 uL of 0.5% bovine serum
82 albumin (BSA) in PBS and transferred to a flow tube with strainer cap. Cells were flow sorted for
83 live cells and then pooled and counted. All samples were concentrated to a volume of 50 uL and
84 then delivered to the Single Cell Analytics Innovation Lab (SAIL) for library construction and
85 sequencing.
86
87 Single-cell RNA-Seq of FACS-sorted cell suspensions was performed on Chromium instrument
88 (10x genomics) following the user manual for 5′ v2 chemistry. In brief, FACS-sorted cells were
89 washed once with PBS containing 2% bovine serum albumin (BSA) and resuspended in PBS
90 containing 2% BSA to a final concentration of 700–1,200 cells per µL. The viability of cells was
91 usually above 70%, as confirmed with 0.2% (w/v) Trypan Blue staining (Countess II). Between
92 10,000 and 30,000 cells were targeted for each sample, where several samples were
93 multiplexed together on one lane of 10x Chromium and specific extracellular protein epitope
94 characterized (using Antibody Derived Tag - ADT) according to the cell hashing protocol.[7]
95
96 The scRNA-seq and scTCR-seq libraries were prepared using the 10x Single Cell Immune
97 Profiling Solution Kit, according to the manufacturer's instructions. Briefly, amplified cDNA was
98 used for both 5′ gene expression library construction and TCR enrichment. For gene expression
99 library construction, amplified cDNA was fragmented and end-repaired, double-sided size-
100 selected with SPRIselect beads, PCR-amplified with sample indexing primers (98 °C for 45 s;
101 14–16 cycles of 98 °C for 20 s, 54 °C for 30 s, 72 °C for 20 s; 72 °C for 1 min), and double-sided
102 size-selected with SPRIselect beads.

103 For TCR library construction, TCR transcripts were enriched from 2 µl of amplified cDNA by PCR
104 (primer sets 1 and 2: 98 °C for 45 s; 10 cycles of 98 °C for 20 s, 67 °C for 30 s, 72 °C for 1 min;
105 72 °C for 1 min). Following TCR enrichment, enriched PCR product was fragmented and end-
106 repaired, size-selected with SPRIselect beads, PCR-amplified with sample-indexing primers
107 (98 °C for 45 s; 9 cycles of 98 °C for 20 s, 54 °C for 30 s, 72 °C for 20 s; 72 °C for 1 min), and
108 size-selected with SPRIselect beads.
109 Final libraries (GEX, TCR and HTO/CITE) were sequenced on Illumina NovaSeq S4 or X+
110 platform (R1 – 26 cycles, i7 – 10 cycles, i5 – 10 cycles, R2 – 90 cycles).
111
112 **Demultiplexing**
113 HTO sequencing data were aligned to the HTO barcodes, and UMIs were counted for each cell
114 using CITE-seq-Count. Using a two-component K-means algorithm, we partitioned logged HTO
115 counts into two distributions: background noise (lower mean) and positive tags (larger mean).
116 Each droplet was then assigned to its source sample based on tags in the positive signal
117 component. We classified droplets with multiple assignments as doublets and those with a single
118 assignment as singlets. This analysis was performed using SHARP v0.1.1
119 (https://github.com/hisplan/sharp)[8].
120
121 **Single-Cell Analysis**
122 We used a hierarchical supervised approach to assign phenotypes to cells as follows. After QC
123 and pre-processing, we used unsupervised graph-based clustering to define the main CD45+
124 cell types. This yielded 6 separate clusters, namely T/NK, B, Plasma, Mast, Myeloid, and
125 Plasmacytoid DC (Figure 2a), annotated using known marker genes (Extended Fig. 1,
126 Supplementary Table 4).
127
128 **Preprocessing**

129 Raw binary base call (BCL) files were demultiplexed into FASTQ files using Cell Ranger. Gene
130 expression libraries were aligned to the GRCh38 reference genome and quantified using
131 `cellranger count` (v7.1.0). TCR V(D)J libraries were processed with `cellranger vdj`
132 (v6.1.2). Cell barcodes were selected using Cell Ranger's cell-calling algorithm EmptyDrops,
133 and the resulting whitelist was used to filter ADT data. The ADT data was then quantified using
134 the SHARP pipeline, which builds on CITE-seq-Count to count cell barcodes and UMIs. Finally,
135 these matrices—gene expression, ADT, and TCR—were loaded into individual AnnData objects
136 for downstream single-cell analysis.
137 Cells with fewer than 200 genes and more than 40% mitochondrial content were removed and
138 samples with fewer than 100 cells were removed. Doublet cells were identified using
139 scDblFinder[9] and excluded. Raw counts were normalized by library size and log-transformed
140 with a pseudo-count of 1 using Scanpy.[10] The top 2000 highly variable genes were selected
141 following the removal of ribosomal, mitochondrial and T cell receptor genes using the
142 `highly_variable_genes` function with the `flavor` option set to `seurat` and sample as
143 `batch_key`. Principal component analysis (PCA) was performed retaining the top 50 PCs.
144 `Harmonypy`[11] (v0.0.9) was used to perform batch correction to control for inter-patient variation.
145 The batch corrected embeddings were used for leiden clustering and creating the UMAP[12]
146 projection for visualization. Clusters for the broad cell lineages were annotated using marker
147 genes, derived from Wilcoxon rank-sum tests comparing each cluster to all others. The 8
148 annotated clusters were: B (*MS4A1, CD19, CD79A*), T/NK (*CD3D, CD3E, NCAM1*), epithelial
149 (*EPCAM, KRT19, KRT8*), endothelial (*PECAM1, VWF, CDH5*), fibroblasts (*COL1A1, FAP,*
150 *PDGFRA*), myeloid (*CD14, LYZ, CSF1R*), mast (*KIT, TPSB2, FCER1A, CPA3, HDC, MS4A2*),
151 and plasma (*MZB1, SDC1, JCHAIN*). Using these annotated cell type clusters, the data were
152 subset to CD45+ cells and cells with more than 10% mitochondrial content were removed. Highly

153 variable gene selection, PCA, batch correction, leiden clustering and UMAP computation were
154 performed as outlined above.
155
156 **Cell-type annotations**
157 CD45+ cells were annotated using a combination of gene expression and protein marker
158 intensity. Cells were grouped into major compartments, namely T/NK, myeloid, B, plasma, mast
159 and plasmacytoid DC cells (**Supplementary Table 4**). T cells were annotated using cite-seq
160 protein CD3D, with custom intensity thresholds to filter for CD4+ and CD8+ T cells using CD8A
161 and CD4 proteins, respectively. For CD4, CD8 and myeloid cells, clustering analysis was
162 performed and the UMAP re-computed for each subset. Differentially expressed genes between
163 clusters were evaluated alongside protein expression (e.g. CD45RA and CD45RO) to annotate
164 13 CD8, 11 CD4 and 7 myeloid subclusters (**Supplementary Table 4**).
165
166 **Gaussian Kernel Density Estimates**
167 We estimated cell density across the UMAP embedding using Scanpy's
168 `sc.tl.embedding_density` with default parameters, grouping cells by the respective
169 covariate. For each group, we visualized the resulting Gaussian kernel density using
170 `sc.pl.embedding_density`.
171
172 **Tensor factorization for Cytokine Activation using IREA**
173 Our objective was to identify intercellular cytokine driven signaling bridging tumors and tumor-
174 draining lymph nodes. To do so, we computed a transcriptional impact for each of a list of 86
175 cytokines in each cell type per sample. To account for the variation of cell types and cytokines
176 in our samples, we used a Bayesian Poisson tensor factorization algorithm shown to be robust
177 to sparse datasets.[13] The algorithm infers sample-level representations in terms of activities of

latent multi-cytokine factors, which represent a group of cytokines that coordinately impact multiple cell types across samples. For each anatomical site of origin, we grouped samples based on their factor activation scores, using a nearest-neighbor graph community detection algorithm.

Concretely, top up-regulated genes in each cell type for each cytokine were acquired from a recently published cytokine dictionary (hereafter called IREA genes.)[14] For each sample, and each cytokine, the overlap between the IREA genes and the DE genes in each cell type was used to compute an overrepresentation analysis and p-values were computed via the hypergeometric test and corrected for multiple testing using false discovery rate (FDR). An enrichment score was computed as the negative log10 of the adjusted p-values per (sample, cell-type, cytokine) tuple. A sample by cytokine score was computed as the sum of all cytokine-cell type interactions. The sample by cell type by cytokine enrichment score, computed thus, was used as the input to the tensor factorization algorithm. Let x_{ijc} denote the enrichment of cytokine c, in cell type j, in sample i, then we assume the following decomposition: $x_{ijc} \approx \widehat{x_{ijc}} = \sum_{k=1}^{K} A_{ik} B_{jk} C_{ck}$ where $A \in R_{+}^{N \times K}$, $B \in R_{+}^{J \times K}$, $C \in R_{+}^{D \times K}$ are latent non-negative matrices, representing sample by factor, cell type by factor, and cytokine by factor scores respectively. To encourage sparsity in the latent factors, we placed Gamma priors on each factor $A_{ik} \sim \text{Gamma}(\alpha, \beta)$, $B_{jk} \sim \text{Gamma}(\alpha, \beta)$, and $C_{ck} \sim \text{Gamma}(\alpha, \beta)$, and adopted a Poisson likelihood $x_{ijc} \sim \text{Poisson}(. \mid \widehat{x_{ijc}})$. The Poisson likelihood favors zero values for small rates, and its closed-form probability mass function accommodates non-negative real-valued inputs.[15] The target of inference is the posterior distribution on the latent matrices A, B, C, that is $P(A, B, C \mid X, \lambda)$. We used variational inference to infer the latent variables.[16]

Cell type-Cytokine Activation Graph Construction

202 Let ψ_{ij} denote the activity of cytokine i in cell type j. We used $\psi \in R_{+}^{J \times C}$ as a weighted
203 adjacency matrix for a bipartite graph where nodes are the set of cytokines and cell types, and
204 edges denote the inferred activity level of a source cytokine in a receiver cell type. To compute
205 ψ, we averaged over factors to compute sample-specific (cell-type, cytokine) pairings, for each
206 sample, per cell type per cytokine $\psi_{ij} = \sum_{k=1}^{K} A_{ik} B_{jk} C_{:K} \in R^{M}$. For each group of samples G, we
207 computed a cell type-cytokine activation matrix as $\psi_{Gj} \coloneqq \sum_{i \in G} \psi_{ij}$. Similarly, for a set of
208 cytokines H, we reported a summary ψ_{GH} as $\psi_{GH} \coloneqq \sum_{i \in G, j \in H} \psi_{ij}$.
209

210 **Cytokine type groupings**

211 We grouped into biologically relevant cytokine families as below:

212 - Class 2 cytokines: IL-10 and all the interferons (k, g, e, b, a1)
213 - Common beta-chain cytokines: IL3, IL5, GM-CSF
214 - Colony stimulating factors: G-CSF, M-CSF, Flt3l
215 - Hormones: Prolactin, Adiponectin
216 - IL-1/Alarmin family: TSLP, IL1a, Il1b, IL18, IL33, IL36a
217 - IL-12 family: IL12, IL23, IL27
218 - IL-17 family: IL17E
219 - IL-2 family: IL2, IL7, IL15, IL21
220 - IL-6 family: IL11, IL31, LIF, Cardiotrophin, Neuropoietin, Oncostatin M
221 - Th2-like: IL4, IL9, IL13
222 - TNF family: TNF, TL1A, CD40L, BAFF

223

224 **Clustering of samples based on IREA scores**

We grouped samples that belong to each anatomical site separately. Briefly, given the sample by factor matrix A , we selected rows corresponding to samples from the dLN (tumor), formed the nearest neighbor graph and computed the leiden clustering with resolution 0.5. For tumor samples, this initially yielded 4 clusters. For interpretability, we merged the two that were high in T cell interactions.

Receptor Ligand activation analysis

We inferred cell-cell interactions using CellChat,[17] calculating communication probabilities for each cell type pair and associated significance using permutation testing for disrupted and canonical lymph node samples separately. We merged these CellChat objects to perform comparisons between the two groups and to identify significantly differing ligand-receptor interaction probabilities. We then filtered these results to cytokine interactions. To further evaluate the evidence for these predicted interactions, we compared the gene expression for each significant ligand or receptor highlighted by CellChat. This was performed by pooling all cells from each lymph node group and comparing gene expression using a Wilcoxon rank-sum test with Benjamini–Hochberg correction for multiple comparisons. This strategy allowed us to include all samples (particularly valuable for myeloid cell types due to the relatively low cell counts for this cell lineage). Gene expression after pseudobulking by sample was assessed and only genes that showed a consistent trend between the pseudobulked and pooled data were retained in our results.

CD8+ T cell Trajectory Analysis

For trajectory analysis, we used Palantir and CellRank. First, we selected CD8+ T cells derived from dLN and tumor samples based on prior annotations. We then identified the top 2,000 highly variable genes using the Scanpy function `sc.pp.highly_variable_genes`. Next, we

250 computed the top 50 principal components using `sc.tl.pca` and performed batch correction
251 via Harmony integration across individual patient identifiers using
252 `sc.external.pp.harmony_integrate`. We computed a diffusion map based on these
253 Harmony-corrected PCA embeddings using `sc.tl.diffmap` to capture global structure and
254 lineage relationships. We initialized the pseudotime trajectory by selecting a root cell
255 representative of a biologically meaningful starting point—specifically, a cluster of phenotypically
256 naïve T cells identified using our annotations. We then used Palantir's
257 `palantir.core.run_palantir` function to infer cell fate potential and generate pseudotime
258 values, employing 1,000 waypoints. Finally, we use CellRank's
259 `cellrank.tl.terminal_states` and `cellrank.tl.lineages` functions, leveraging
260 Palantir-derived pseudotime and diffusion-based transition probabilities to identify terminal
261 states and major cellular lineages. We executed CellRank with 1,000 permutations to identify
262 lineage drivers and associated genes.
263
264 **Phenotype entropy**
265 We first compute the fraction of CD8+ T cells per sample. We then compute the Shannon entropy
266 as $H = \sum_{k=1}^{K} p_k \log(p_k)$.
267
268 **Identifying naïve cell subpopulations**
269 We used Gaussian Mixture Models (GMM) to identify subpopulations of naïve cells based on
270 the protein expression of CD45 isoforms. Briefly, we selected all cells that had a naïve phenotype
271 based on their transcription profile. We used Bayesian information criterion to select the number
272 of components for the GMM. We used the from `sklearn.mixture.GaussianMixture` with
273 a full covariance, `max_iter=10000`, and 100 random restarts.
274

275 **Distance between T Cell Phenotypes**

276 To compute the distance between different T cell phenotype clusters, we used the geometric
277 median computed Weiszfeld's algorithm.[18] Geometric median of a set of n points $\{X_i\}$, is defined
278 as the point y that minimizes the sum of distances from all poits in the set, i.e., $y \coloneqq$
279 $\text{argmin}_{y \in R^D} \sum_{i=1}^{n} D(X_i - y)$, where D is the L_2 norm (Euclidean distance). We used the
280 diffusion map representation of cells as input. We computed the geometric median for each
281 phenotype group the computed the pairwise distance of geometric medians of each phenotype
282 cluster from all others.

283

284 **Tumor vs dLN Transcription Factor Comparison**

285 To find which transcription factors (TF) affect expansion in CD8+ T cells, we compared the
286 expression of TFs in CD8+ T cells, excluding the naïve phenotype, derived from Tumor to
287 draining lymph node. We obtained a list of human TFs[19] and ranked their expression using the
288 Wilcoxon rank-sum test.

289

290 **KLF2 Correlated Genes**

291 We used regularized linear regression to find genes predictive of KLF2's expression in a sample
292 specific manner. Gene expression counts were normalized for library-size, log-transformed, and
293 standardized such that the mean and the variance of each gene are zero and one respectively.
294 Using linear regression with an L2 penalty (i.e., ridge regression), we estimated a per sample
295 coefficient vector of per gene association with KLF2 expression. We used the collection of these
296 coefficients to compute a mean coefficient and Z-score per gene.

297

298 **KLF2 Gene Cloning**

299 The plasmid PpyCAG-KLF2-IB[20] was a gift from Austin Smith (Addgene #60441), which

300 contained human *KLF2*. PCR amplification of *KLF2* was performed, using primers (in table
301 below) and Q5® High-Fidelity DNA Polymerase (New England Biolabs) according to
302 manufacturer's instructions. A restriction digest was performed of both source plasmid PpyCAG-
303 KLF2-IB and destination plasmid pTRIP-SFFV-mTagBFP-P2A (Addgene #102585), using
304 BamHI-HF (New England Biolabs) and SalI-HF (New England Biolabs) with rCutSmart™ Buffer
305 (New England Biolabs) according to manufacturer's instructions, gel isolated and purified.
306 Construct assembly was then performed using Gibson Assembly® Master Mix (New England
307 Biolabs) according to manufacturer's instructions. The constructed plasmid was then
308 transformed into NEB 5-alpha competent *E. coli* cells (New England Biolabs) using 1uL of
309 plasmid DNA. Mini-prep using Qiagen Mini Kit (Qiagen) was performed for 3 individual clones
310 per manufacturer's instructions, and then Sanger sequencing was used to confirm the
311 appropriate sequence for clone 2 which was subsequently used for all experiments.
312
313 **Cloning primer sequences**

Description	Name	Sequence
KLF2 Overexpression	NL65_KLF2.F	gagaaccctggacctgcgatggcgctgagtgaacccatcc
	NL65_KLF2.R	cgaggtcgacactagtgctacatgtgccgtttcatgtgc
KLF2 Sequencing	SFFV-mTagBFP-F	caaacgccaagaccacatatag

314
315
316 **KLF2 Overexpression**
317 Lentivirus was produced using 293T cells as follows: briefly, 293T cells were transiently co-
318 transfected with viral packaging (pCG-gag-pol), and envelope (pCG-VSVg) constructs along with
319 the expression vector (pTRIP-SFFV-mTagBFP-P2A-Empty or pTRIP-SFFV-mTagBFP-P2A-
320 KLF2) using 100 mg/mL polyethylenimine (Polysciences Inc). Virus-containing supernatant was

321 collected beginning 48 hours after transfection and filtered with a 0.45μm syringe, then aliquoted
322 and stored at -80°C until use. Primary human T-cell activation was performed as follows: Buffy
323 coats containing PBMCs were obtained from New York Blood Center. PBMCs were isolated
324 using Lymphoprep density gradient-based centrifugation. Subsequently, T cells were isolated
325 using a Dynabeads Untouched Human T Cell kit according to manufacturer's instructions. T cells
326 were activated in T cell media (RPMI-1640 supplemented with 10% heat-inactivated FBS, 2 mM
327 *L*-glutamine, and 100 U/mL Penicillin-Streptomycin) supplemented with recombinant IL-2
328 (Peprotech, 5 ng/mL). Cells were plated onto non-tissue culture-treated 6-well plates that had
329 been coated with anti-CD3 (Biolegend, 5 μg/mL) and anti-CD28 antibody (Biolegend, 2 μg/mL)
330 in PBS. 24 hours after activation, 100x concentrated virus was diluted 1:50 in T cell media (40
331 uL virus in 160 uL T cell media, see above), then 200 uL was added dropwise to each well and
332 plates were shaken gently. 24 hours later, cells were passaged into tissue culture-treated 6-well
333 plates containing fresh T cell media supplemented with IL-2. 5 days following initial stimulation,
334 cells were either harvested directly for spectral cytometry-based phenotyping, DOGMA-Seq,
335 assessment of KLF2 by Western blot, or restimulated to interrogate intracellular cytokine
336 production as described below.
337
338 **Protein extraction and Western Blot**
339 Cell lysis was performed using RIPA buffer (Cell Signaling) containing 1% SDS as well as
340 protease and phosphatase inhibitors. Following lysis, cells were sonicated (Bioruptor) and
341 centrifuged and supernatants were analyzed for protein concentration using BCA. 40 ug of each
342 lysate were run on a 4-12% Bis-Tris Gel (Thermo Fisher) and transferred to nitrocellulose
343 membranes. Membranes were blocked using 5% milk and then blotted with the following
344 antibodies: anti-KLF2 (see table below) and anti-H3 (see table below) in 3% BSA overnight at
345 4C on a rocker. Membranes were washed and then incubated with HRP-conjugated secondary

346 antibodies at 1:5000 dilution for 1 hour at room temperature in 5% milk. Membranes were then
347 washed and imaged using a Bio-Rad ChemiDoc Imaging System.
348

Target Name	Host	Company	Catalog #	Dilution
KLF2	Mouse	Novus Biologicals	NBP2-61812	1:1000
Histone H3	Rabbit	Cell Signaling	9715	1:1000
Anti-mouse IgG, HRP-linked Antibody	Horse	Cell Signaling	7076	1:5000
Amersham ECL Rabbit IgG, HRP-linked whole Ab	Donkey	Cytiva	NA934	1:5000

349

350 **Spectral Cytometry**

351 For phenotyping, cells were incubated with Fc block (Human TruStain FcX, Biolegend, cat.
352 422302) at a 1:20 dilution for 30 minutes at 4°C. Viability staining was performed for 30 minutes
353 at 4°C using Ghost Dye 510 Violet (cat. 59863) diluted 1:800 in PBS. Cells were subsequently
354 washed and stained for surface markers in FACS buffer containing 1X Brilliant Buffer (cat.
355 566349) for 30 minutes at 4°C, then washed twice and resuspended in 1% BSA in PBS. Data
356 was acquired using a Cytek Aurora spectral cytometer.
357

Staining type	Marker	Fluorophore	Catalog #	Clone	Dilution
Phenotype, Surface	PD-1	APC	BioLegend, 329908	EH12.2H7	1:100
	CCR7	BB700	BD Biosciences, 566437	3D12	1:50
	CD45RA	BUV395	BD Biosciences, 740298	HI100	1:400
	CD8	BUV496	BD Biosciences, 612943	RPA-T8	1:800
	CD38	BUV661	BD Biosciences, 612969	HIT2	1:200
	HLA-DR	BV786	BD Biosciences, 564041	G46-6	1:200
	CD4	BV570	BioLegend, 300534	RPA-T4	1:800

	CD28	BUV805	BD Biosciences, 742037	CD28.2 (RUO)	1:100
	CD69	BV605	BioLegend, 310937	FN50	1:200
	Tim-3	BV650	BioLegend, 345028	F38-2E2	1:50
	CD103	PE-594	BioLegend, 350223	Ber-ACT8	1:100
	CD39	PE-810	BioLegend, 328245	A1	1:100
	CXCR3	A700	BD Biosciences, 561320	1C6/CXCR3	1:100
	CXCR6	BV421	BioLegend, 356013	K041E5	1:100
Phenotype, Intracellular	TCF-1	A488	Cell Signaling, 6444S	C63D9	1:100
	TOX	PE	MACS Miltenyi Biotec, 130-120-716	REA473	1: 25
	CTLA4	BV711	BioLegend, 369632	BNI3	1: 200

358

359 For interrogation of cytokine production, PMA and ionomycin were added to wells to a final
360 concentration of 50 ng/mL and 500 ng/mL, respectively. 1 hour after addition of PMA and
361 ionomycin, Brefeldin A was added at a 1:1000 dilution and cells were incubated for another 3
362 hours, after which they were harvested and stained with surface antibodies as described above
363 using the panel below. Fixation was performed using an eBioscience FoxP3
364 Fixation/Permeabilization kit and stained overnight with intracellular antibodies diluted in
365 permeabilization buffer overnight at 4C. Data was acquired using a Cytek Aurora spectral
366 cytometer.

367

Staining type	Marker	Fluorophore	Catalog #	Clone	Dilution
Restimulation, Surface	PD-1	APC	BioLegend, 329908	EH12.2H7	1: 100
	CCR7	BB700	BD Biosciences, 566437	3D12	1: 50
	CD45RA	BUV395	BD Biosciences, 740298	HI100	1: 400

	CD8	BUV496	BD Biosciences, 612943	RPA-T8	1: 800
	CD38	BUV661	BD Biosciences, 612969	HIT2	1: 200
	HLA-DR	BV786	BD Biosciences, 564041	G46-6	1: 200
	CD4	BV570	BioLegend, 300534	RPA-T4	1: 800
	CD28	BUV805	BD Biosciences, 742037	CD28.2	1: 100
	CD69	BV605	BioLegend, 310937	FN50	1: 200
	Tim-3	BV650	BioLegend, 345028	F38-2E2	1: 50
	CD103	PE-594	BioLegend, 350223	Ber-ACT8	1: 100
	CD39	PE-810	BioLegend, 328245	A1	1: 100
Restimulation, Intracellular	GzmB	A700	BD Biosciences, 560213	GB11	1: 1600
	GzmK	FITC	BioLegend, 370508	GM26E7	1: 800
	IL-2	APC-F750	BioLegend, 500351	MQ1-17H12	1: 400
	IFN-g	BV421	BioLegend, 502531	4S.B3	1: 800
	TNF	PE	BioLegend, 502908	MAb11	1: 800

DOGMA-Seq

Primary human T cells were isolated and cultured as described above ('**KLF2 Overexpression**'). On day 5 following initial activation, T cells were collected and stained with a viability dye followed by surface staining with anti-CD4 (RPA-T4, BioLegend 300534, 1:100), anti-CD8 (RPA-T8, BD Biosciences 612914, 1:1200), TotalSeq-A Human Universal Cocktail, and TSA-compatible hashtag oligos for multiplexing. Live BFP+ $CD4^+$ and $CD8^+$ T cells from both KLF2 overexpressing and empty vector conditions were flow sorted and delivered to SAIL for sequencing. Single Cell Multiome ATAC + Gene Expression was performed with 10x genomics system using Chromium Next GEM Single Cell Multiome Reagent Kit. We followed User Guide with the difference that single cell suspension was prepared as described in Mimitou et al (DOGMA-seq protocol with Digitonin buffer).[21] Around 10,000 cells were targeted for each sample. Samples were multiplexed together on one lane of 10x Chromium (using Hash Tag

381 Oligonucleotides - HTO) and specific extracellular protein epitope characterized (using Antibody
382 Derived Tag - ADT) following previously published protocol.[7] Next, cells were treated with a
383 digitonin lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 mM MgCl2, 0.01% DIG and 2 U
384 μl−1 of RNase inhibitor) for 5 min on ice, followed by adding 1 ml of chilled DIG wash buffer
385 (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 3 mM MgCl2 and 1 U μl−1 of RNAse inhibitor) and
386 inversion before centrifugation at 500g for 5 min at 4 °C. The supernatant was discarded, and
387 cells were resuspended in DIG wash buffer, followed by counting using trypan blue and DAPI
388 on a Countess II FL Automated Cell Counter. Then cells were processed according to the
389 Chromium Next GEM Single Cell Multiome ATAC + Gene Expression user guide. Final libraries
390 were sequenced on Illumina NovaSeq X+ platform following 10x genomics recommendations.
391
392 For sample demultiplexing, HTO sequencing data were aligned to the HTO barcodes, and UMIs
393 were counted for each cell using CITE-seq-Count. Using a two-component K-means algorithm,
394 we partitioned logged HTO counts into two distributions: background noise (lower mean) and
395 positive tags (larger mean). Each droplet was then assigned to its source sample based on tags
396 in the positive signal component. We classified droplets with multiple assignments as doublets
397 and those with a single assignment as singlets. This analysis was performed using SHARP
398 v0.1.1 (https://github.com/hisplan/sharp)[8].
399
400 **Cell typing of DOGMA-seq**
401 The KLF2 over-expression and empty vector were demultiplexed as above in '**DOGMA-Seq**'.
402 We processed single-cell RNA-, ATAC- and CITE-seq (DOGMA-seq) data using the Seurat (v.
403 5.0.3) and Signac packages (v.1.13.0). A total of 9,616 cells were sequenced, 6,487 in the KLF2
404 over-expression and 3,129 in the empty vector conditions. We loaded RNA and ATAC-seq data
405 into a single Seurat object. We computed quality control metrics—including nucleosome signal

406 and transcription start site (TSS) enrichment and filtered cells based on RNA/ATAC count
407 thresholds, nucleosome signal, and TSS enrichment to exclude low-quality cells. Next, we
408 established a unified peak set by merging and filtering called peaks, removing non-standard
409 chromosomes and blacklist regions. We quantified ATAC counts within these peaks using
410 `FeatureMatrix` and added them to a new chromatin assay in the Seurat object. We then split
411 the CD4+ and CD8+ T cells based on the CD4 and CD8A protein expression. To annotate cell
412 phenotypes, we normalized the RNA expression counts with `SCTransform` and reduced
413 dimensions using `RunPCA`. We processed ATAC data with term frequency-inverse document
414 frequency (`RunTFIDF`) and `RunSVD`. We then integrated RNA and ATAC modalities using
415 `FindMultiModalNeighbors` and `FindClusters` on the resulting joint weighted nearest-
416 neighbor graph to identify clusters. We used `FindAllMarkers` on the SCT assay and
417 separately on the ADT assay to report marker genes and proteins per cluster. We used these
418 markers to annotate the T cell phenotypes.
419

420 **Motif Accessibility Analysis**

421 We used the `RunChromVAR` function with default parameters from the Signac package to
422 compute transcription factor (motif activities from single-cell chromatin accessibility data. Briefly,
423 it quantifies how accessible each cell is at genomic regions associated with specific TF motifs
424 by comparing observed accessibility to background expectations. ChromVAR generates bias-
425 corrected, normalized activity scores (z-scores) for each motif in each cell, enabling the
426 identification of TFs that are potentially driving differences in chromatin states across cell
427 populations or experimental conditions. With the accessibility scores at hand, we used the
428 `FindMarkers` function to compute the differentially accessible markers per cell type (with
429 parameters `mean.fxn = rowMeans` and `fc.name = "avg_diff"`).
430

431 **KLF2 Targets**

432 We first extracted KLF2's motif's binding sites that overlapped with any identified peak. We used
433 the motif MA1515.2.[22] We then filtered for regions that were within 1000bp of a protein coding
434 gene's promoter to identify putative target genes. Among these genes, we reported those with a
435 log-fold-change (between the KLF2 over expression versus the empty vector conditions) of
436 above 1.0 or below –1.0 and a gene expression value of 0.1 in at least one of the conditions.

437

438 **LFA-1 dependent T cell priming assay**

439 *Cell Isolation*

440 OT-I T cells and dendritic cells were isolated as follows: for DC isolation, spleens from C57/Bl6
441 mice were minced and incubated in digestion buffer (200X DNase, 50X Liberase, RPMI) for 30
442 minutes at 37 degrees. During this incubation period, spleens from OT-I mice were isolated,
443 mashed over a 40 µm strainer. Subsequently, spleens were washed with RPMI and treated with
444 hypotonic ACK lysis buffer to lyse red blood cells. RBC lysis was quenched with 10x excess
445 PBS and centrifuged. OT-I T cells and DCs were then enriched using T-cell isolation
446 (Dynabeads™ Untouched™ Mouse T Cells Kit) and DC enrichment kits (Dynabeads™ Mouse
447 DC Enrichment kit), respectively. These are both negative selection kits in which cell
448 suspensions were first incubated with antibodies for 20 minutes at 4 degrees, subsequently
449 incubated with magnetic beads for 15 minutes with gentle tilting and rotation, and finally placed
450 on a magnet after which suspensions containing OT-I T cell and dendritic cell-enriched
451 populations were transferred to a fresh tube for downstream assays.

452

453 *T cell-DC coculture*

454 After isolation, DCs were resuspended in complete RPMI-1640 media containing 10% FBS, 2
455 mM L-glutamine and Pen/Strep and incubated at 37 degrees for 1 hour with gentle rotation in
456 the presence of escalating doses of SIINFEKL (N4), SIIQFEKL (Q4), SIITFEKL (T4), or

457 SIIVFEKL (V4) ranging from 0.1, 1, 10, and 100 nM. Cells were subsequently washed and
458 resuspended at 1 million cells/mL.
459 During this period, OT-I T cells were labeled with CellTrace Violet by resuspending in PBS at 1
460 million cells/mL, warming to 37 degrees, mixing 1:1 with CellTrace Violet diluted 1:1000 in pre-
461 warmed PBS and incubating in a 37-degree water bath for 20 minutes. Labeling was quenched
462 by adding 5x volume of RPMI+++. Labeled T cells were counted, centrifuged and resuspended
463 in complete T cell media (RPMI+++ containing 50 uM BME and 10 ng/mL IL-2) at 200K/mL, with
464 either isotype control (rat IgG2a kappa, BioXCell) or anti-LFA-1 (M17/4, BioXCell) at a final
465 concentration of 5 ug/mL. 500 uL containing 100K T cells were added to each well of a 96 well
466 Polypropylene Deep Well Plate (Corning) into which 5 uL DCs (5,000 DC total) were added. The
467 plate was then incubated at 37 degrees for 4 days.
468 For interrogation of cytokine production, PMA and ionomycin were added to wells to a final
469 concentration of 50 ng/mL and 500 ng/mL, respectively. 1 hour after addition of PMA and
470 ionomycin, Brefeldin A was added at a 1:1000 dilution and cells were incubated for another 3
471 hours, after which they were harvested and stained with surface antibodies as described above
472 using the panel below. Fixation was performed using an eBioscience FoxP3
473 Fixation/Permeabilization kit and stained overnight with intracellular antibodies diluted in
474 permeabilization buffer overnight at 4C. Data was acquired using a Cytek Aurora spectral
475 cytometer.

Staining type	**Marker**	**Fluorophore**	**Catalog #**	**Clone**	**Dilution**
Restimulation, Surface	B220	BUV496	BD Biosciences 612950	RA3-6B2	1:800
	TCRb	BUV805	BD Biosciences 748405	H57-597	1:100
	NK1.1	BUV615	BD Biosciences 751111	PK136	1:50
	CD4	BUV563	BD Biosciences 612923	GK1.5	1:800
	CD8	BV650	Biolegend 100741	53-6.7	1:800
	CD62L	BV570	Biolegend 104433	ME-14	1:200
	CD44	BUV737	BD Biosciences 612799	IM7	1:400

	PD-1	APC-Cy7	Biolegend 135224	29F.1A12	1:200
	CD39	BUV395	BD Biosciences 567264	Y23-1185	1:800
	CD25	PE	Biolegend 102008	PC61	1:200
	CD127	PE-Cy5	Biolegend 135015	A7R34	1:200
	CD69	PerCP-Cy5.5	Biolegend 104522	H1.2F3	1:200
	CX3CR1	APC	Biolegend 149008	SA011F11	1:400
	CXCR6	PE-Dazzle594	Biolegend 151117	SA051D1	1:200
Restimulation, Intracellular	IL-2	BV711	Biolegend 503837	JES6-5H4	1:800
	TNF	FITC	Biolegend 506304	MP6-XT22	1:800
	IFN-g	PE-Cy7	Biolegend 505826	XMG1.2	1:1600
	Granzyme B	Alexa 700	Biolegend 372222	QA16A02	1:3200

Methods Reference Section:

1 Amin, M., Edge, S., Greene, F. & et al. *AJCC Cancer Staging Manual*. 8th edn, (Springer, 2017).

2 Lauren, P. The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification. *Acta Pathol Microbiol Scand* **64**, 31-49 (1965). https://doi.org/10.1111/apm.1965.64.1.31

3 Cancer Genome Atlas Research, N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202-209 (2014). https://doi.org/10.1038/nature13480

4 Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* **23**, 703-713 (2017). https://doi.org/10.1038/nm.4333

5 NCCN®. *NCCN Guidelines® Gastric Cancer 2.2023*, <https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1434> (2023).

6 Herrera-Almario, G., Patane, M., Sarkaria, I. & Strong, V. E. Initial report of near-infrared fluorescence imaging as an intraoperative adjunct for lymph node harvesting during robot-assisted laparoscopic gastrectomy. *Journal of Surgical Oncology* **113**, 768-770 (2016). https://doi.org/10.1002/jso.24226

7 Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biology* **19** (2018). https://doi.org/10.1186/s13059-018-1603-1

8 Wan, S., Kim, J. & Won, K. J. SHARP: hyperfast and accurate processing of single-cell RNA-seq data via ensemble random projection. *Genome Res* **30**, 205-213 (2020). https://doi.org/10.1101/gr.254557.119

9 Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Research* **10**, 979 (2022). https://doi.org/10.12688/f1000research.73600.2

10 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19** (2018). https://doi.org/10.1186/s13059-017-1382-0

11 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289-1296 (2019). https://doi.org/10.1038/s41592-019-0619-0

12 Leland McInnes, J. H., James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. https://doi.org/10.48550/arxiv.1802.03426

13 Schein, A., Paisley, J., Blei, D. M. & Wallach, H. Bayesian Poisson Tensor Factorization for Inferring Multilateral Relations from Sparse Dyadic Event Counts. *Kdd'15: Proceedings of the 21st Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 1045-1054 (2015). https://doi.org/10.1145/2783258.2783414

14 Cui, A. *et al.* Dictionary of immune responses to cytokines at single-cell resolution. *Nature* (2023). https://doi.org/10.1038/s41586-023-06816-9

15 Salehi, S., Nazaret, A., Shah, S. P. & Blei, D. M. Population Priors for Matrix Factorization. *Transactions on Machine Learning Research* (2024).

16 Blei, D. M., Alp, K. & and McAuliffe, J. D. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association* **112**, 859-877 (2017). https://doi.org/10.1080/01621459.2017.1285773

17 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nature Communications* **12** (2021). https://doi.org/10.1038/s41467-021-21246-9

18 Chandrasekaran, R. & Tamir, A. Open questions concerning Weiszfeld's algorithm for the Fermat-Weber location problem. *Mathematical Programming* **44**, 293-295 (1989). https://doi.org/10.1007/bf01587094

19 Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665 (2018). https://doi.org/10.1016/j.cell.2018.01.029

20 Takashima, Y. *et al.* Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell* **158**, 1254-1269 (2014). https://doi.org/10.1016/j.cell.2014.08.029

21 Mimitou, E. P. *et al.* Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nature Biotechnology* **39**, 1246-1258 (2021). https://doi.org/10.1038/s41587-021-00927-2

22 Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **52**, D174-D182 (2024). https://doi.org/10.1093/nar/gkad1059

Figure 1

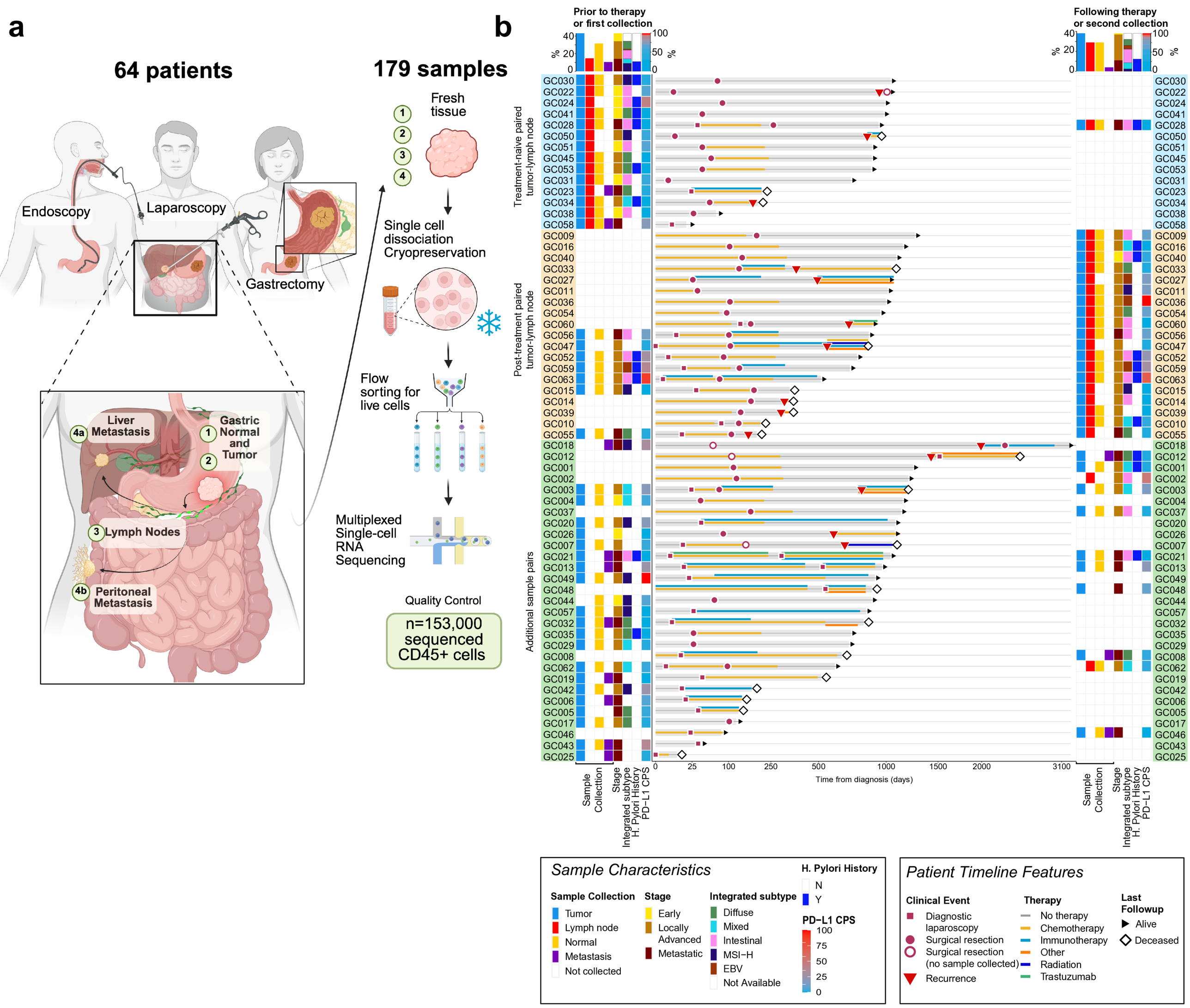

a
64 patients
Endoscopy
Laparoscopy
Gastrectomy
Liver Metastasis
Gastric Normal and Tumor
Lymph Nodes
Peritoneal Metastasis
179 samples
Fresh tissue
Single cell dissociation Cryopreservation
Flow sorting for live cells
Multiplexed Single-cell RNA Sequencing
Quality Control
n=153,000 sequenced CD45+ cells
b
Prior to therapy or first collection
Following therapy or second collection
Treatment-naive paired tumor-lymph node
Post-treatment paired tumor-lymph node
Additional sample pairs
Time from diagnosis (days)
Sample Collection
Stage
Integrated subtype
H. Pylori History
PD-L1 CPS
Sample Characteristics
Sample Collection
Tumor
Lymph node
Normal
Metastasis
Not collected
Stage
Early
Locally Advanced
Metastatic
Integrated subtype
Diffuse
Mixed
Intestinal
MSI-H
EBV
Not Available
H. Pylori History
N
Y
PD-L1 CPS
Patient Timeline Features
Clinical Event
Diagnostic laparoscopy
Surgical resection
Surgical resection (no sample collected)
Recurrence
Therapy
No therapy
Chemotherapy
Immunotherapy
Other
Radiation
Trastuzumab
Last Followup
Alive
Deceased

Figure 2

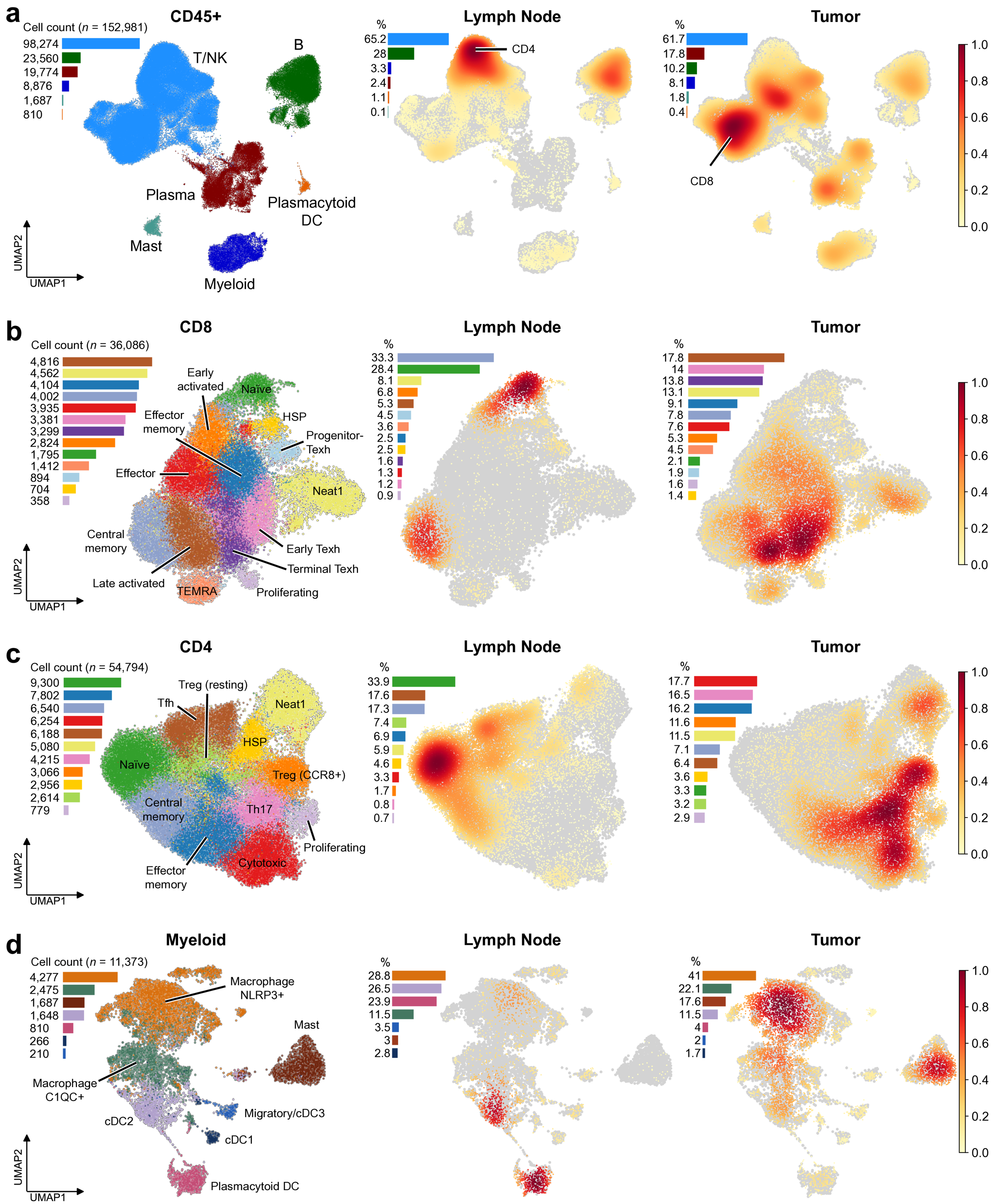

a
CD45+
Cell count (n = 152,981)
98,274
23,560
19,774
8,876
1,687
810
T/NK
B
Plasma
Plasmacytoid DC
Mast
Myeloid
UMAP2
UMAP1
Lymph Node
%
65.2
28
3.3
2.4
1.1
0.1
CD4
Tumor
%
61.7
17.8
10.2
8.1
1.8
0.4
CD8
1.0
0.8
0.6
0.4
0.2
0.0
b
CD8
Cell count (n = 36,086)
4,816
4,562
4,104
4,002
3,935
3,381
3,299
2,824
1,795
1,412
894
704
358
Early activated
Naïve
Effector memory
HSP
Progenitor-Texh
Effector
Neat1
Central memory
Early Texh
Terminal Texh
Late activated
TEMRA
Proliferating
UMAP2
UMAP1
Lymph Node
%
33.3
28.4
8.1
6.8
5.3
4.5
3.6
2.5
2.5
1.6
1.3
1.2
0.9
Tumor
%
17.8
14
13.8
13.1
9.1
7.8
7.6
5.3
4.5
2.1
1.9
1.6
1.4
1.0
0.8
0.6
0.4
0.2
0.0
c
CD4
Cell count (n = 54,794)
9,300
7,802
6,540
6,254
6,188
5,080
4,215
3,066
2,956
2,614
779
Treg (resting)
Tfh
Neat1
HSP
Naïve
Treg (CCR8+)
Central memory
Th17
Proliferating
Effector memory
Cytotoxic
UMAP2
UMAP1
Lymph Node
%
33.9
17.6
17.3
7.4
6.9
5.9
4.6
3.3
1.7
0.8
0.7
Tumor
%
17.7
16.5
16.2
11.6
11.5
7.1
6.4
3.6
3.3
3.2
2.9
1.0
0.8
0.6
0.4
0.2
0.0
d
Myeloid
Cell count (n = 11,373)
4,277
2,475
1,687
1,648
810
266
210
Macrophage NLRP3+
Mast
Macrophage C1QC+
Migratory/cDC3
cDC2
cDC1
Plasmacytoid DC
UMAP2
UMAP1
Lymph Node
%
28.8
26.5
23.9
11.5
3.5
3
2.8
Tumor
%
41
22.1
17.6
11.5
4
2
1.7
1.0
0.8
0.6
0.4
0.2
0.0

Figure 3

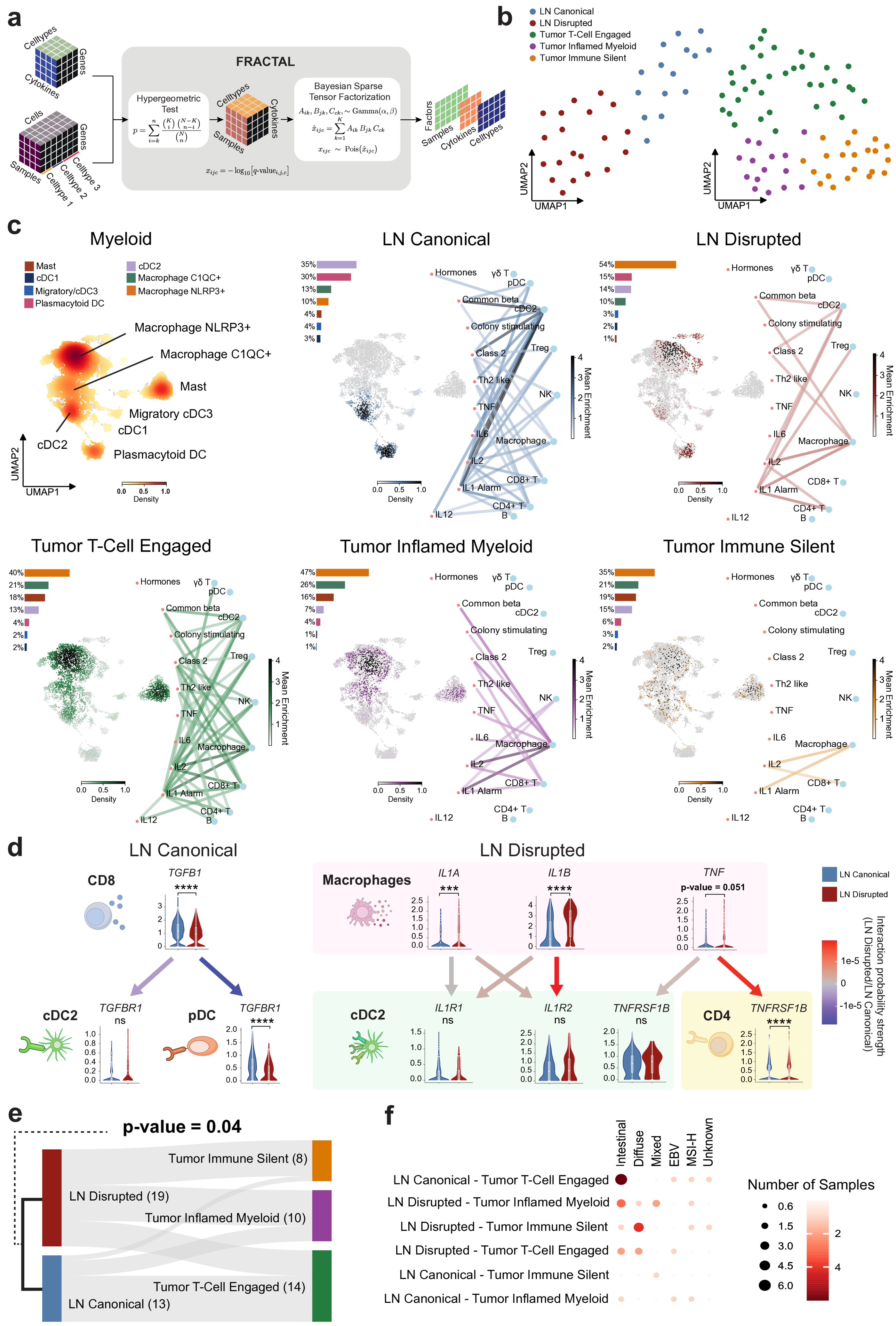

a
Celltypes
Genes
Cytokines
Cells
Genes
Samples
Celltype 1
Celltype 2
Celltype 3
FRACTAL
Hypergeometric Test
Bayesian Sparse Tensor Factorization
Factors
Samples
Cytokines
Celltypes
b
LN Canonical
LN Disrupted
Tumor T-Cell Engaged
Tumor Inflamed Myeloid
Tumor Immune Silent
UMAP2
UMAP1
c
Myeloid
LN Canonical
LN Disrupted
Mast
cDC1
Migratory/cDC3
Plasmacytoid DC
cDC2
Macrophage C1QC+
Macrophage NLRP3+
Tumor T-Cell Engaged
Tumor Inflamed Myeloid
Tumor Immune Silent
Hormones
γδ T
pDC
Common beta
cDC2
Colony stimulating
Class 2
Treg
Th2 like
NK
TNF
IL6
Macrophage
IL2
CD8+ T
IL1 Alarm
CD4+ T
B
IL12
Density
Mean Enrichment
d
LN Canonical
LN Disrupted
CD8
TGFB1
cDC2
TGFBR1
pDC
TGFBR1
Macrophages
IL1A
IL1B
TNF
p-value = 0.051
cDC2
IL1R1
IL1R2
TNFRSF1B
CD4
TNFRSF1B
LN Canonical
LN Disrupted
Interaction probability strength (LN Disrupted/LN Canonical)
e
p-value = 0.04
Tumor Immune Silent (8)
LN Disrupted (19)
Tumor Inflamed Myeloid (10)
Tumor T-Cell Engaged (14)
LN Canonical (13)
f
Intestinal
Diffuse
Mixed
EBV
MSI-H
Unknown
LN Canonical - Tumor T-Cell Engaged
LN Disrupted - Tumor Inflamed Myeloid
LN Disrupted - Tumor Immune Silent
LN Disrupted - Tumor T-Cell Engaged
LN Canonical - Tumor Immune Silent
LN Canonical - Tumor Inflamed Myeloid
Number of Samples

Figure 4

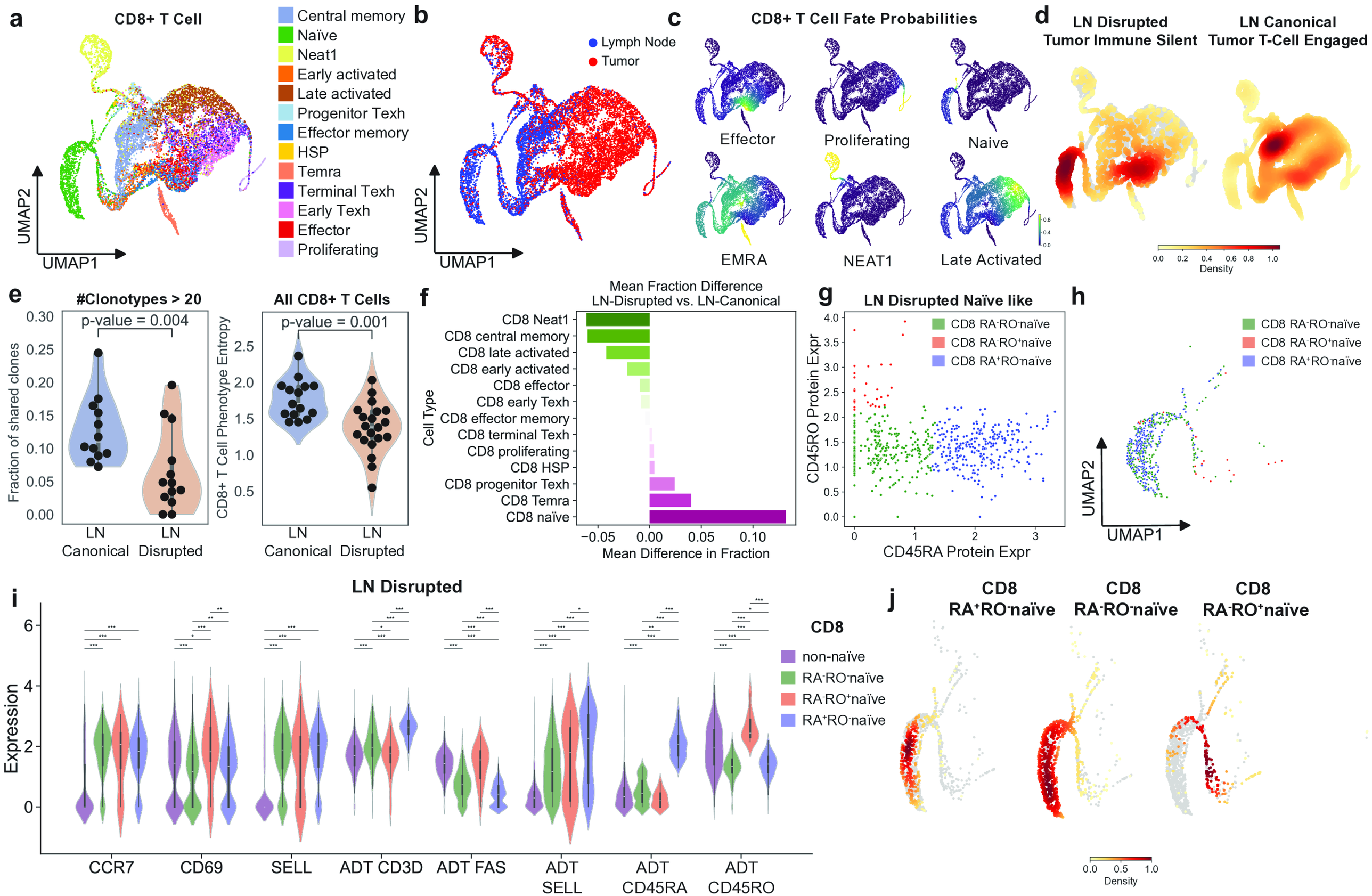

a
CD8+ T Cell
Central memory
Naïve
Neat1
Early activated
Late activated
Progenitor Texh
Effector memory
HSP
Temra
Terminal Texh
Early Texh
Effector
Proliferating
UMAP2
UMAP1
b
Lymph Node
Tumor
UMAP2
UMAP1
c
CD8+ T Cell Fate Probabilities
Effector
Proliferating
Naive
EMRA
NEAT1
Late Activated
d
LN Disrupted
Tumor Immune Silent
LN Canonical
Tumor T-Cell Engaged
Density
e
#Clonotypes > 20
p-value = 0.004
Fraction of shared clones
LN Canonical
LN Disrupted
All CD8+ T Cells
p-value = 0.001
CD8+ T Cell Phenotype Entropy
LN Canonical
LN Disrupted
f
Mean Fraction Difference
LN-Disrupted vs. LN-Canonical
CD8 Neat1
CD8 central memory
CD8 late activated
CD8 early activated
CD8 effector
CD8 early Texh
CD8 effector memory
CD8 terminal Texh
CD8 proliferating
CD8 HSP
CD8 progenitor Texh
CD8 Temra
CD8 naïve
Cell Type
Mean Difference in Fraction
g
LN Disrupted Naïve like
CD8 RA-RO-naïve
CD8 RA-RO+naïve
CD8 RA+RO-naïve
CD45RO Protein Expr
CD45RA Protein Expr
h
CD8 RA-RO-naïve
CD8 RA-RO+naïve
CD8 RA+RO-naïve
UMAP2
UMAP1
i
LN Disrupted
CD8
non-naïve
RA-RO-naïve
RA-RO+naïve
RA+RO-naïve
Expression
CCR7
CD69
SELL
ADT CD3D
ADT FAS
ADT SELL
ADT CD45RA
ADT CD45RO
j
CD8 RA+RO-naïve
CD8 RA-RO-naïve
CD8 RA-RO+naïve
Density

Figure 5

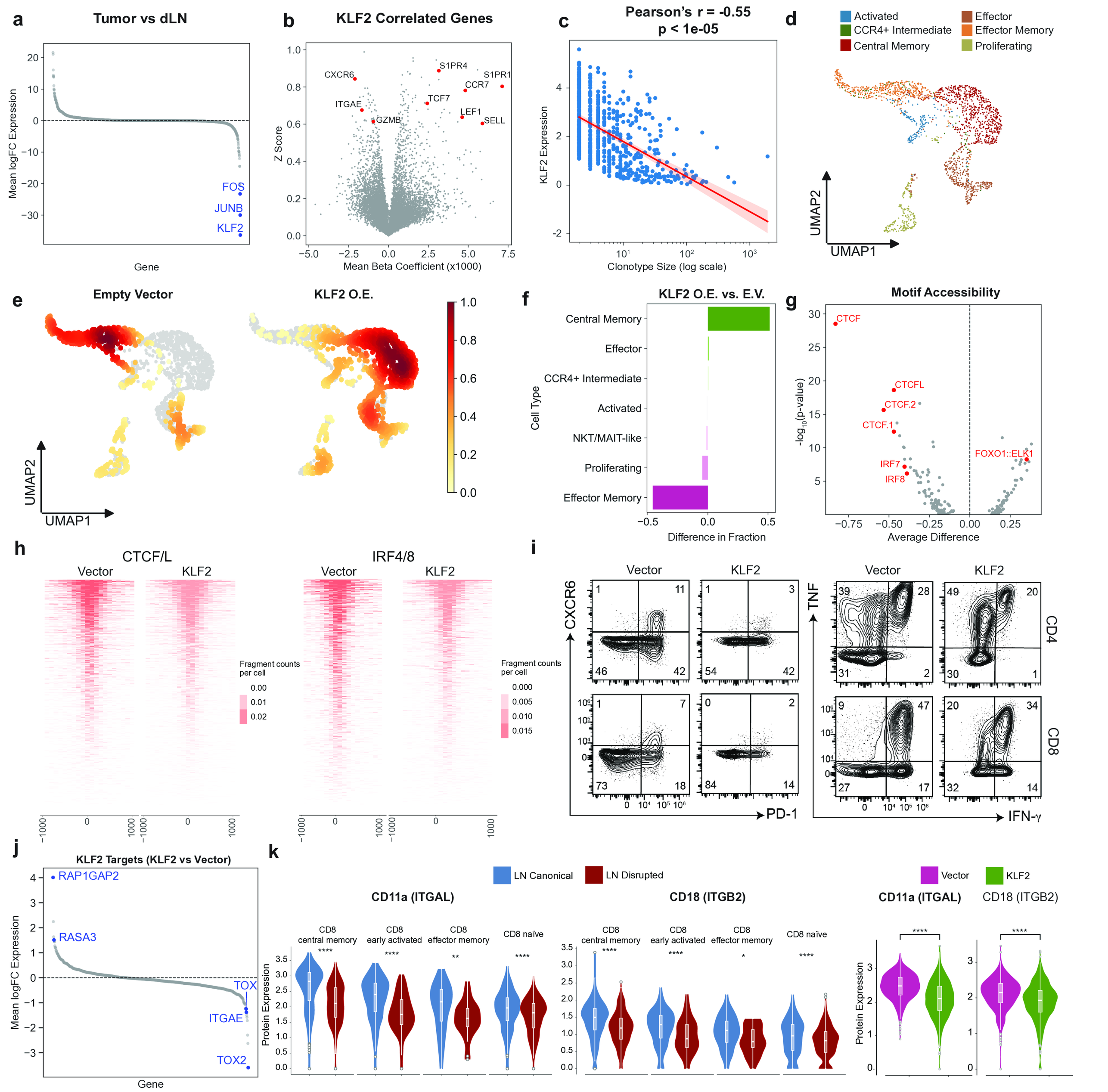
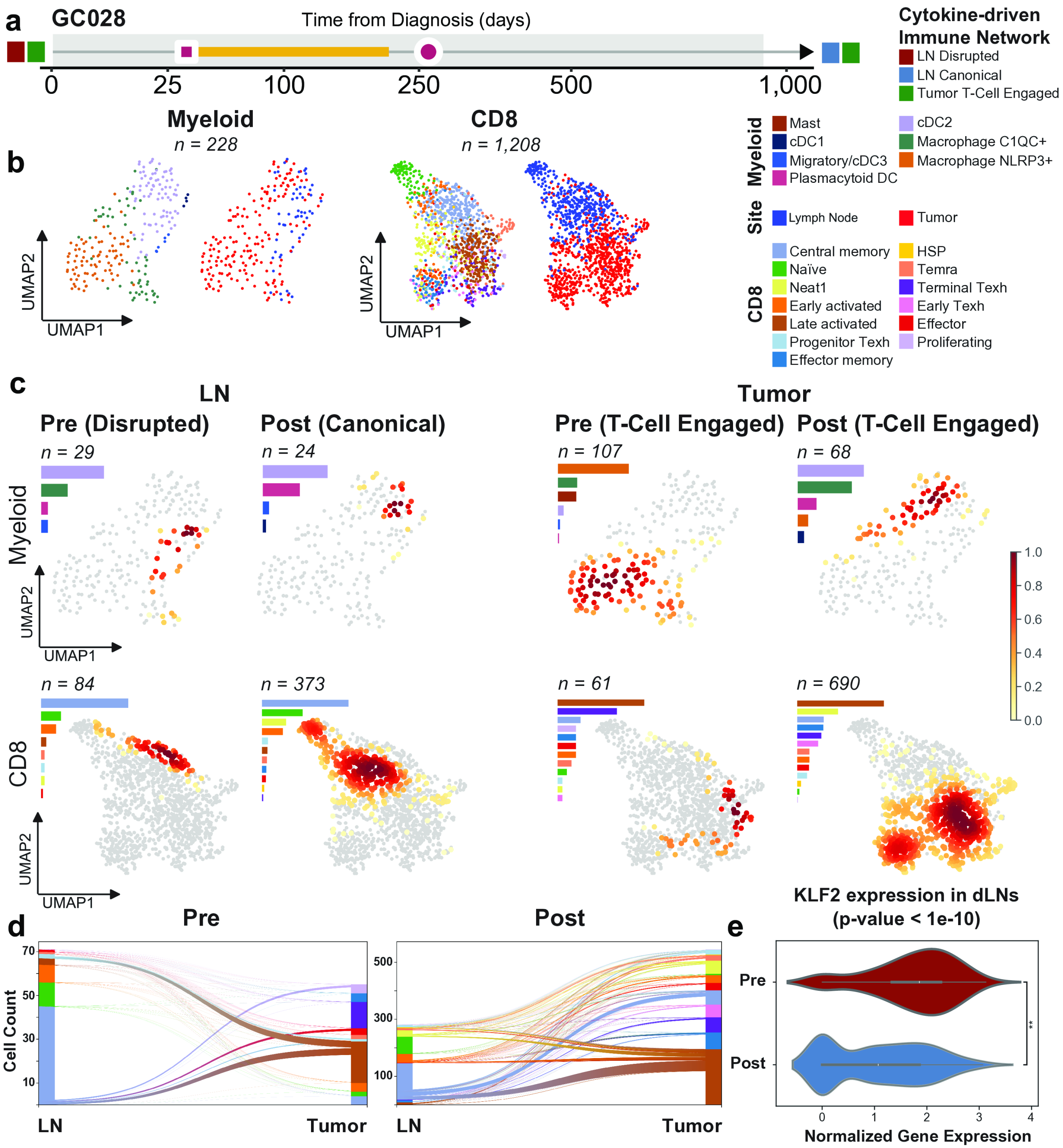

a
Tumor vs dLN
Mean logFC Expression
Gene
FOS
JUNB
KLF2
b
KLF2 Correlated Genes
Z Score
Mean Beta Coefficient (x1000)
CXCR6
ITGAE
GZMB
S1PR4
S1PR1
CCR7
TCF7
LEF1
SELL
c
Pearson's r = -0.55
p < 1e-05
KLF2 Expression
Clonotype Size (log scale)
d
Activated
CCR4+ Intermediate
Central Memory
Effector
Effector Memory
Proliferating
UMAP2
UMAP1
e
Empty Vector
KLF2 O.E.
UMAP2
UMAP1
f
KLF2 O.E. vs. E.V.
Cell Type
Central Memory
Effector
CCR4+ Intermediate
Activated
NKT/MAIT-like
Proliferating
Effector Memory
Difference in Fraction
g
Motif Accessibility
CTCF
CTCFL
CTCF.2
CTCF.1
IRF7
IRF8
FOXO1::ELK1
Average Difference
h
CTCF/L
IRF4/8
Vector
KLF2
Fragment counts per cell
i
Vector
KLF2
CXCR6
PD-1
TNF
IFN-γ
CD4
CD8
j
KLF2 Targets (KLF2 vs Vector)
RAP1GAP2
RASA3
TOX
ITGAE
TOX2
Mean logFC Expression
Gene
k
LN Canonical
LN Disrupted
CD11a (ITGAL)
CD18 (ITGB2)
CD8 central memory
CD8 early activated
CD8 effector memory
CD8 naïve
Protein Expression
Vector
KLF2

Figure 6


a
GC028
Time from Diagnosis (days)
0
25
100
250
500
1,000
Cytokine-driven Immune Network
LN Disrupted
LN Canonical
Tumor T-Cell Engaged
b
Myeloid
n = 228
CD8
n = 1,208
UMAP2
UMAP1
Myeloid
Mast
cDC1
Migratory/cDC3
Plasmacytoid DC
cDC2
Macrophage C1QC+
Macrophage NLRP3+
Site
Lymph Node
Tumor
CD8
Central memory
Naïve
Neat1
Early activated
Late activated
Progenitor Texh
Effector memory
HSP
Temra
Terminal Texh
Early Texh
Effector
Proliferating
c
LN
Tumor
Pre (Disrupted)
Post (Canonical)
Pre (T-Cell Engaged)
Post (T-Cell Engaged)
n = 29
n = 24
n = 107
n = 68
Myeloid
n = 84
n = 373
n = 61
n = 690
CD8
1.0
0.8
0.6
0.4
0.2
0.0
d
Pre
Post
Cell Count
70
50
30
10
500
300
100
LN
Tumor
e
KLF2 expression in dLNs
(p-value < 1e-10)
Pre
Post
**
0
1
2
3
4
Normalized Gene Expression