CITEseq analysis of non-small-cell lung cancer lesions reveals an axis of immune cell activation associated with tumor antigen load and *TP53* mutations

Andrew M. Leader[1,2,3], John A. Grout[1,2,3], Christie Chang[1,4], Barbara Maier[1,2,3], Alexandra Tabachnikova[1,2,3], Laura Walker[1,4], Alona Lansky[1,2,3], Jessica LeBerichel[1,2,3], Naussica Malissen[1,2,3], Melanie Davila[1,4], Jerome Martin[1,2,3,5], Giuliana Magri[1,2,3,#], Kevin Tuballes[1,4], Zhen Zhao[6], Francesca Petralia[7,8], Robert Samstein[1,2,3,9], Natalie Roy D'Amore[10], Gavin Thurston[11], Alice Kamphorst[1,2,3], Andrea Wolf[12], Raja Flores[12], Pei Wang[7,8], Mary Beth Beasley[6], Helene Salmon[1,2,3,&], Adeeb H. Rahman[1,4,7], Thomas U. Marron[1,2,13], Ephraim Kenigsberg[1,7,8,$], Miriam Merad[1,2,3,4,$]

[1]The Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[2]The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[3]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[4]Human Immune Monitoring Center, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[5]Nantes Université, CHU Nantes, Inserm, Centre de Recherche en Transplantation et Immunologie, UMR 1064, F-44000 Nantes, France ; CHU Nantes, Nantes Université, Laboratoire d'Immunologie, F-44000 Nantes, France

[6]Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[7]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[8]Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[9]Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[10]Immuno-oncology Drug Discovery Unit, Millennium Pharmaceuticals, Inc. a wholly owned subsidiary of Takeda Pharmaceutical Company Limited.

[11]Department of Oncology & Angiogenesis, Regeneron Pharmaceuticals Inc., Tarrytown, NY, 10591, USA

[12]Department of Thoracic Surgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[13]Division of Hematology/Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Correspondence should be sent to: miriam.merad@mssm.edu, ephraim.kenigsberg@mssm.edu

[$]Authors contributed equally to this work

[#]Present address: Program for Inflammatory and Cardiovascular Disorders, Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), 08003 Barcelona, Spain

[&]Present address: INSERM U932, Institut Curie, 26 rue d'Ulm, 75005 Paris, France

SUMMARY

Immunotherapy is becoming a mainstay in the treatment of NSCLC. While tumor mutational burden (TMB) has been shown to correlate with response to immunotherapy, little is known about the relation of the baseline immune response with the tumor genotype. Here, we profiled 35 early stage NSCLC lesions using multiscale single cell sequencing. Unsupervised clustering identified in a subset of patients a key cellular module consisting of *PDCD1*+ *CXCL13*+ activated T cells, IgG+ plasma cells, and *SPP1*+ macrophages, referred to as the lung cancer activation module ($LCAM^{hi}$). Transcriptional data from two NSCLC cohorts confirmed a subset of patients with $LCAM^{hi}$ enrichment, which was independent of overall immune cell content. The $LCAM^{hi}$ module strongly correlated with TMB, expression of cancer testis antigens, and with *TP53* mutations in smokers and non-smokers. These data establish LCAM as a key mode of immune cell activation associated with high tumor antigen load and driver mutations.

INTRODUCTION

Lung cancer is the most common cause of cancer-related death[1], and the most common subgroup of lung cancer is non-small cell lung cancer (NSCLC)[2]. In recent years, immune checkpoint blockade (ICB) targeting the PD-1/PD-L1 axis has become first-line therapy for a majority of patients with metastatic and locally advanced disease[3]. Though ICB studies have achieved improved overall survival, fewer than half of patients achieve significant clinical benefit, though still may experience physical and financial toxicity. Biomarkers are lacking to determine optimal treatment regimens for patients, as our understanding of tumor-associated immune phenotypes and immune correlates of response to ICB remains incomplete.

While multiple studies have used single-cell assays to profile NSCLC tumor-infiltrating immune cells in comparison to patient-matched, non-involved lung (nLung)[4,5], blood[6], or both[7,8], or characterized tumor-infiltrating lymphocytes (TIL)[9-12], we continue to lack a comprehensive understanding of how immune cell phenotypes vary across patients. In particular, it remains unclear which immune cell populations and phenotypes are associated with robust, tumor-directed T cell responses and response to ICB, and how these features are connected to tumor-cell intrinsic characteristics such as tumor mutational burden (TMB)[13,14]. A deeper analysis is further required for uncovering the cell types and states associated with immunostimulatory versus immunoregulatory presentation of tumor-associated antigens, as well as parsing the tumor-related effects on tissue-resident and migratory innate cell types. Attempts to integrate these data across the innate and adaptive arms of the immune system are of crucial importance to optimizing rational design of immunotherapies. Furthermore, while response to ICB has been associated with specific patient groups, individual driver mutations, the degree of immune infiltrate, and TMB, the complex interplay between these factors remains poorly understood.

37 Here, we sought to define the molecular immune states induced in the tumor
38 microenvironment by profiling an expanded patient cohort compared to previous related studies[5,6]
39 via multiscale single-cell analyses. We integrated the results of single-cell RNA sequencing
40 (scRNAseq) of immune cells with cellular indexing of transcriptomes and epitopes by sequencing
41 (CITEseq)[15], a method allowing for combined scRNAseq and multiplexed single-cell surface
42 protein measurement. To further elucidate the TCR landscape across T cell phenotypes, we
43 analyzed these results together with joint scRNAseq/TCRseq. We revealed a pattern of inter-tumor
44 variability involving innate and adaptive immune responses which we validated in two bulk RNA
45 datasets, allowing us to detect an association with TMB and tumor driver mutations.
46
47 **RESULTS**
48
49 **Integrative analyses unify phenotypic mappings across substrates and datasets.** To probe
50 transcriptional states of immune cells in the lung cancer microenvironment, we set out to profile
51 cells from a cohort of untreated, early-stage NSCLC patients undergoing resection with curative
52 intent (Figure 1A). The cohort was diverse with respect to age, smoking status, sex, and
53 histological subtype (Figure 1B). We generated three datasets integrating antibody profiling of
54 surface marker proteins using CITEseq[15], scRNAseq, and TCRseq with single cell resolution
55 (Tables S1-S3). We performed CITEseq on matched tumor and non-involved lung (nLung) tissues
56 from 7 patients, in addition to performing scRNAseq of matched tumor and nLung in 28 additional
57 patients. Finally, to expand on our annotation of T cell clusters based on the distribution of clonally
58 expanded populations, we performed paired single-cell TCRseq and scRNAseq on T cells isolated
59 from 3 patients.

60 ScRNA profiles in tumor and nLung were clustered together using a batch-aware algorithm
61 we recently developed in order to combine data across patients, while accounting for batch-specific
62 background noise[16] (Figures 1C and S1A-C). To develop a general gene-expression model of
63 clusters representing different cell types and states, we relied only on 19 nLung and 22 tumor
64 samples processed with 10X Chromium 3' V2 chemistry. We then used this model to classify cells
65 from additional samples processed with different protocols or from different datasets showing
66 similar transcriptional profiles. (Figure S1B-D, methods). The RNA-based clustering identified
67 49 immune clusters within 6 compartments including subsets of T cells, B cells, plasma cells, mast
68 cells, plasmacytoid dendritic cells (pDC), and mononuclear phagocytes (MNP) consisting of
69 macrophages (MΦ), monocytes, putative monocyte-derived dendritic cells (MoDC), and
70 conventional dendritic cells (cDC; Figure 1C and S1E, F). Overall, 377,549 single cells from 35
71 tumors and 32 matched nLung samples from patients at Mount Sinai were classified into 6
72 compartments and 30 annotated transcriptional states. CITEseq data further confirmed cell
73 identities using well-established protein markers (Figure 1D). For example, annotation of pDC
74 was based on expression of transcripts associated with this lineage (*LILRA4, IRF8*; Figure 1C) and
75 high expression of known population-defining surface markers (CD123; Figure 1D). While cluster
76 frequencies varied widely among patients, clusters mapped between 590 and 23812 cells, and all
77 clusters included cells from multiple patients (Figure 1E and S1F).

78 We first compared the variability of samples from different regions within individual
79 tumors to the variability between different patients' tumors with respect to immune cell type
80 composition. To do this, we examined samples from a study that analyzed three regions per tumor
81 in 8 patients[5], mapping cells to the clusters produced with our expanded dataset. Clustering the
82 samples by correlation of cell type frequencies among immune cells demonstrated that samples

83 almost always clustered by patient (Figure S1G), and similarly, the Euclidean distances between
84 patient-matched samples of different tumor regions was strongly reduced compared to the
85 distances between samples from different patients (Figure S1H). Therefore, while the total level
86 of immune content may still vary regionally in and around tumors[17], these analyses demonstrated
87 that inter-tumor differences drive lung tumor immune variability in terms of the phenotypic
88 makeup of the immune cells that are present.

89 To understand whether the immune changes between tumor and nLung were distinct across
90 patients or, alternatively, globally similar, we estimated the immune diversity within tumor and
91 nLung using the Euclidean distances between the log-transformed cluster-frequencies. This
92 analysis indicated that nLung samples were significantly more homogeneous (Figure 1F; "nLung-
93 nLung distances") than tumor samples ("Tumor-Tumor distances"; $t=8.3$; $p<2.2e-16$). We further
94 compared distances among nLung and among tumor to the distances between nLung and tumor.
95 This analysis showed that the diversity between independent (unmatched) tumor and nLung
96 samples was larger than the diversity within tumor samples ($t=19.6$; $p<2.2e-16$) and nLung
97 samples ($t=24.6$; $p<2.2e-16$), suggesting that immune landscapes within the TME were
98 significantly changed compared to non-involved tissues, and that most tumors harbored many
99 conserved changes (Figure 1E-F).

100 We next sought to test if the differences between nLung and tumor could be observed in
101 an independent cohort. The cell type frequencies of 8 matched tumor-nLung pairs described in
102 ref.[5] indeed validated the distinct microenvironments we observed in our cohort (Figure 1G). This
103 result demonstrated that the observed tumor signatures were robust and reproducible, encouraging
104 us to further study the transcriptional states within it.

105

106 **The intratumoral dendritic cell compartment is characterized by expansion of monocyte-**
107 **derived DC.** We next investigated the heterogeneity with the myeloid compartment, given that
108 different myeloid cells have various important roles in generating or inhibiting tumor directed
109 immune responses, including antigen presentation, T cell co-stimulation, and shaping the cytokine
110 milieu within the TME[13]. We identified conventional DC1 (cDC1) expressing *IRF8*, *WDFY4*, and
111 *CLEC9A* transcripts (Figure 2A) as well as CD141 and CD26 surface markers (Figure 2B), and
112 cDC2 expressing high *CD1C* and *FCER1A* transcripts as well as CD1c and CD5 protein. We also
113 detected a DC cluster expressing *FSCN1* and *CCR7* transcripts and elevated HLADR, CD86, PD-
114 L1, and CD40 surface protein which we described as mature DC enriched in regulatory molecules
115 (mregDC) in great detail elsewhere[18]; in this study we found mregDC were correlated with tumor-
116 antigen uptake and thus help define antigen-charged DC[18]. This phenotype was also consistent
117 with an activated DC phenotype detected in lung and liver tumors by others[6,19]. We furthermore
118 identified clusters that expressed cDC2 markers such as CD1c and *CLEC10A*, but also expressed
119 high levels of monocyte and MΦ genes including *S100A8, S100A9, C1QA,* and *C1QB*, lacked
120 surface expression of the pre-DC surface marker CD5[20,21], and exhibited increased expression of
121 CD11b and CD14 (Figure 2A, B); we annotated such clusters as MoDC. Importantly, MoDC were
122 distinct from MΦ based on higher levels of CD1c surface protein in addition to their upregulation
123 of the DC2-like transcriptional signature (Figure 1C, clusters 52, 29, and 30). Overall, MoDC were
124 the most prevalent DC subtype and were increased in tumors compared to nLung, whereas
125 mregDC were the rarest (Figure 2C and S2A-C). As we had seen previously[7], the fraction of cDC1
126 were strongly reduced in tumors (Figure 2C).
127 Since the activation profile of mregDC is crucial for inducing tumor directed T cell
128 responses[18], we examined the mregDC distribution in tumors by multiplexed

129 immunohistochemical consecutive staining on a single slide (MICSSS)[22]. We stained for DC-
130 LAMP and PD-L1, as the transcripts of these genes (*LAMP3* and *CD274*, respectively) were highly
131 enriched in the mregDC cluster (Figure 2D). We found that mregDC expressing DC-LAMP and
132 PD-L1 accumulated in tertiary lymphoid structures (TLS) in close proximity to T cells (Figure
133 2E). CD3-negative areas of TLS, which are putatively analogous to lymph node B cell zones, were
134 frequently populated by MYH11+ follicular dendritic cells[23], a stromal cell type commonly found
135 in B cell zones (Figure 2F).

136 To better understand the relationship between MoDC and other MNP, we searched for
137 genes that were mutually exclusive among CD14 monocytes, cDC2, and MΦ (Figure S2D, Table
138 S4). Scoring MoDC using these gene lists in comparison with other MNP populations revealed
139 that MoDC were distinct from MΦ and CD14+ monocytes. Ordering cells within each of these
140 compartments by the expression of these distinct monocyte- and cDC2 gene programs
141 demonstrated anticorrelation of these gene sets among MoDC but not cDC2 (Figure 2G;
142 rho= -0.33, $p<2.2e-16$; rho=0.016, p=0.29, respectively), demonstrating that MoDC inhabited a
143 phenotypic spectrum between monocytes and cDC2-like cells. While some MΦ genes were
144 expressed in MoDC higher than in cDC2 cells, MoDC were distinct from MΦ based on higher
145 cDC2 gene expression and lower MΦ gene expression (score distributions are detailed in Figure
146 S2E).

147 To further uncover transcriptional programs that were variable among MoDC and cDC
148 without relying on specific cell classifications, we analyzed the covariance structure of variable
149 genes among all DC. This approach resulted in distinct sets of co-expressed genes (gene
150 "modules") that varied together across cells, independent of cluster assignments (Figure S2F, G,
151 Table S5). Gene module analysis across the DC compartment revealed upregulation in tumors of

152 multiple modules that were mainly restricted to MoDC and DC2 (Figure S2H, I). The gene
153 modules most upregulated in tumors compared to nlung included genes associated with glycolysis
154 (mod39) and cell cycle (mod38), which were mainly expressed in MoDC cluster 52 (Figure S2G-
155 I). Frequent upregulation of many monocyte- or MΦ-like modules (7, 3, 4, 6, 5, 37, 10) was
156 consistent with a higher frequency of MoDC compared to cDC in tumors.

157 We also identified a cDC2 module (mod34) which was enriched in tumor lesions compared
158 to nLung (Figure S2G, H) and included *CD1A* and *CD207*. These genes mark the lesional cells of
159 Langerhans cell histiocytosis (LCH), a myeloid inflammatory condition driven by enhanced ERK
160 activation[24]; we therefore referred to this module as "LCH-like". LCH cells produce many
161 inflammatory cytokines that promote the accumulation of Tregs and activated T cells in LCH
162 lesions[25]. Interestingly, *IL22RA2*, encoding the IL22 decoy receptor IL22-BP, was also included
163 in this module (Table S5). IL22 modulates epithelial cell growth and plays a role in tissue
164 protection through modulation of tissue inflammation and in promoting tumor growth through
165 induction of tissue repair[26]. Expression of the IL22 receptor (*IL22RA1*), meanwhile, negatively
166 correlated with survival in KRAS-mutated lung cancer lesions[27]. These genes were mainly induced
167 in the *bona fide* cDC2 cluster, but were also upregulated in MoDC (Figure 2H). Probing DC
168 expression in an independent scRNAseq dataset of NSCLC immune cells[5] confirmed upregulation
169 of these genes in tumor associated DC transcriptomes (Figure S2J).

170

171 **Tumors are dominated by monocyte-derived MΦ that are distinct from alveolar MΦ.**

172 While previous studies have demonstrated phenotypic differences between MΦ populating nLung
173 versus tumors[5,7], they have been limited in their ability to parse specific MΦ subpopulations with
174 potentially distinct ontogeny and function. Our data showed remarkable heterogeneity within the

175 MΦ compartment as demonstrated by the varying expression of classical marker genes among
176 clusters (Figures 3A and S3A). This level of resolution allowed the identification of alveolar MΦ
177 (AMΦ) clusters expressing *SERPINA1* and *PPARG* and a cluster expressing genes consistent with
178 interstitial MΦ (IMΦ), which thus far have only been defined to a limited extent in humans[28,29]. In
179 contrast to AMΦ that self-renew locally independent of blood precursors[30], IMΦ are thought to be
180 maintained by circulating monocyte pools even in steady state, albeit at lower rates of turnover
181 than in settings of overt inflammation. IMΦ lacked *PPARG* and expressed *MAF* family
182 transcription factors, *MERTK*, *CSF1R*, *LYVE1*, and *CX3CR1*[31]. CD14+ and CD16+ monocytes
183 were defined by the expression of *CD14* or *FCGR3A* respectively and the lack of MΦ markers
184 *MRC1*, *VSIG4*, and *SIGLEC1.* Other MΦ clusters expressed genes such as *MAFB*, *CEBPD*,
185 *FCGR2B*, and *CSF1R*, which are indicative of monocyte origin and shared by monocytes and
186 IMΦ; therefore, these clusters were annotated as MoMΦ. A remaining population of MΦ
187 expressed genes consistent with primary granule formation (*AZU1, ELANE, CTSG*) but distinct
188 from bone marrow progenitors due to lack of *MPO*, and also lacked elevation of neutrophil marker
189 genes[6] *CSF3R, LRG1, FFAR4,* and *VASP* compared to other myeloid cells (Figure S3A). This
190 cluster was referred to as *AZU1*+ MΦ.

191 Using a CITEseq panel of established immune surface markers, we validated the
192 transcription-based cluster annotations and associated new surface markers with the MΦ
193 subpopulations. For example, we found that CD10, not previously appreciated as a MΦ marker,
194 could distinguish AMΦ from other lung myeloid populations (Figures 3B and S3B). This staining
195 was consistent with RNA expression patterns (Figure S3A) and was verified by
196 immunohistochemical staining (IHC) of airspace-residing AMΦ in nLung (Figure S3C). MoMΦ
197 expressed higher levels of CD11c and CD14 than other MΦ populations, whereas IMΦ were

198 notably CD14+/HLADR+/CD11c^{int}/CD86-/CD10– (Figure 3B). Thus, CITEseq protein staining
199 confirmed the main MΦ subpopulations identified by the transcriptional classification and defined
200 potential sorting strategies (Figure S3B).

201 Gene module analysis across all monocyte and MΦ clusters (Figure S3D-G) revealed three
202 broad signatures, consistent with genes that were highly expressed in both AMΦ and MoMΦ
203 (module group I), MoMΦ or IMΦ (module group II), and monocytes (module group III; Figure
204 S3D). Individual modules could be identifiably associated with cell type annotations as well as
205 cell states reflecting, for example, interferon response (modules 32 and 19), heat shock genes
206 (module 49) cell cycle (module 42), HLA class-II expression (module 28), and glycolysis (module
207 47; Figure S3E). Examining the expression patterns of specific modules across MoMΦ clusters
208 led us to divide them into MoMΦ subtypes I-IV: MoMΦ-II clusters expressed the highest levels
209 of the tumor-enriched module 48, which was driven mainly by *SPP1* and also included IL-1
210 receptor antagonist *IL1RN*, and module 47 consisting of genes indicating a glycolytically active
211 state (*GAPDH, ENO1, LDHA, ALDOA, TPI1*), and lower levels than other MΦ of *C1Q* and HLA-
212 class-II transcripts (Figures 3C and S3E, G). MoMΦ-III clusters were enriched in module 24
213 (including *TREM2* and *LILRB4*) and module 25 (including *APOE* and *GPNMB*). MoMΦ-I and
214 MoMΦ-IV were less distinctive than the other MoMac subtypes, but each comprised their own
215 unique gene expression patterns. For example, MoMac-IV expressed the highest levels of module
216 27 which included *CTSS, CFD,* and *ALDH1A1* and also expressed some genes otherwise confined
217 to AMΦ (module 36), whereas MoMΦ-I was enriched in module 20 (including chemokine ligands
218 *CCL13* and *CCL2*) while MoMac IV was not. Together, these analyses identify multiple tumor
219 MoMΦ phenotypes with distinct metabolic and immunomodulatory gene programs that are
220 enriched in the tumor milieu and likely contribute to defining the tumor microenvironment.

221 Gene set scores based on mutually exclusive, differentially expressed genes among CD14+
222 monocytes, AMΦ, and MoMΦ (Figure S3H and Table S4) showed that AMΦ and MoMΦ were
223 each distinct from CD14+ monocytes (Figure 3D) but that MoMΦ expressed a gradient of the
224 CD14+ monocyte score (Figure 3E). Analysis of the gene expression patterns of hundreds of genes
225 within the scores supported the general trends (Figure 3F). MoMΦ clusters were also distinct from
226 the IMΦ cluster based on many transcripts and surface proteins (Figure 3A, B), although some
227 MoMΦ, especially those that were the most distant from AMΦ, shared some IMΦ genes such as
228 *CSF1R*, *FOLR2*, and *MERTK* (Figures 3A, F and S3A).

229 The predominant populations that increased in tumors were MoMΦ, while AMΦ were
230 strongly depleted from tumors and IMΦ frequencies were unchanged (Figure 3G). Monocyte
231 frequencies were also decreased, possibly reflecting their differentiation to MoMΦ or MoDC.
232 Given that individual MoMΦ-subsets changed between nLung and tumor to different extents, we
233 asked how these differences related to the underlying phenotypic heterogeneity within the MoMΦ
234 compartment, beyond signatures revealed by module analysis. Selecting for a set of highly
235 expressed transcripts encoding secreted factors demonstrated strong differences between MNP
236 subsets (Figure 3H). MoMΦ-II, the most tumor-enriched subset, expressed the highest levels of
237 inflammatory cytokines *TNF* and *IL6*, transcripts encoding the pleiotropic factor SPP1, a broad
238 collection of matrix metallopeptidases *MMP-7, -9*, and *-12*, as well as CCR2/5 ligands *CCL-2, -8*,
239 and *-7*. By comparison, other MoMΦ populations expressed less distinct secretory profiles.
240 Multiple MNP populations expressed the CXCR3-ligand chemotactic factors *CXCL-9,-10,-11*
241 including MoMΦ, MoDC, and mregDC, while these ligands were distinctly absent from AMΦ,
242 IMΦ, AZU1+ MΦ, monocytes, cDC1, and cDC2. MregDC, meanwhile, expressed distinct
243 cytokines and chemotactic factors associated with T cell engagement, including *IL12B*, *EBI3*,

244 *CCL17*, *CCL22*, and *CCL19*, which were expressed to a minimal or greatly reduced degree in
245 monocytes, MΦ, or MoDC.
246
247 **TCRs limited to tumors mark T cells with distinct phenotypic features.** CITEseq
248 characterization of T cells identified populations of CD8+ cells that were characterized by an NK-
249 like signature ($T_{NK\text{-}like}$), high expression of *GZMK* (T_{GZMK}), expression of genes related to tissue-
250 residence such as *ITGA1* transcript and CD103 and CD69 protein (CD8+ T_{rm}), and a cluster
251 consistent with activated T cells, expressing high levels of *IFNG, GZMB, LAG3*, *CXCL13*, and
252 *HAVCR2* transcripts, as well as high PD-1, ICOS, and CD39 protein ($T_{activated}$; Figure 4A, B). Other
253 clusters, which mostly consisted of CD4+ cells, could be separated into T_{reg}, T_{rm}, cells expressing
254 a profile consistent with either central memory or naïve cells ($T_{CM/Naïve\text{-}like\text{-}I}$; *TCF7*, *SELL*, *LEF1*,
255 *MAL,* and surface expression of CD127), and a group of clusters expressing both intermediate
256 levels of this signature as well as a tissue-residency signature ($T_{CM/Naïve\text{-}like\text{-}II}$). Cells within the
257 $T_{CM/Naïve\text{-}like}$ clusters did not otherwise segregate by signatures related to antigen experience, TCR
258 engagement, activation, or exhaustion state.
259 While clustering cells using their transcriptional profiles did not result in complete
260 separation of CD4+ and CD8+ cells, CITEseq allowed for the comparison of CD4+ versus CD8+
261 cells within otherwise transcriptionally similar groups. The $T_{activated}$ cluster could therefore be
262 separated into CD4+ and CD8+ components (15.5% and 74.8%, respectively). Differential
263 expression analysis between these subsets showed that, on average, CD4+ cells in this cluster
264 expressed increased levels of *CXCL13*, *CD40LG*, *BCL6*, and *IL21* (Figure S4A) consistent with a
265 phenotype similar to T-follicular-helper. We next asked whether we could use profiles of CD8+
266 and CD4+ cells within this cluster to classify CD4 and CD8 cells from samples lacking CITEseq

surface staining, which was not available for the majority of our dataset. Learning a signature-based classifier from a training set consisting of the $T_{activated}$ cells from 2 patients and testing this signature on the remaining patients with CITEseq staining demonstrated that transcriptional based classification guided by antibody signals was highly accurate (86% on test set; Figure S4B). This classification could further discriminate cells that uniquely expressed *CD8-A/B* transcripts or *CD4* transcripts across the remaining cells in the dataset (84% accuracy; Figure S4C). Applying this classification generally allowed for the separation of CD4 $T_{activated}$ from CD8 $T_{activated}$ across the dataset (Figure S4D). Similar to a recent report[32], independent quantification of these cells separately and comparing their frequencies demonstrated a high correlation across tumors (Figure S4E; rho=0.58, p=2.7e-4), so they were continued to be grouped for further analysis.

While $T_{activated}$ and T_{reg} were the most increased T cell populations in tumors compared to nLung (Figure 4C), another cluster, characterized by high expression of cell-cycle genes *MKI67* and *STMN1*, and surface expression of HLA-DR and CD38, was also significantly increased in tumors (T_{cycle}; Figure 4A-C). Other than expressing these hallmarks of proliferation, the T_{cycle} cluster was diverse with respect to RNA and protein expression (Figure 4A, B). Analyzing the cells comprising T_{cycle} by gene scores constructed from genes differentially expressed among the other clusters demonstrated that T_{cycle} is a mixture of multiple T cell phenotypes that share the cycling state (Figure S4F and Table S4). While tumors expressed overall higher frequencies of cycling T cells (Figures 4C and S4G), $T_{activated}$ and T_{reg} showed the highest frequencies of cycling cells compared to other phenotypes (Figure S4H).

To understand the clonal relationships among T cell phenotypes in tumor and nLung tissues, we performed paired scRNAseq and TCRseq using a nested PCR approach on paired tissues from 3 patients. Classification of the transcriptomes among the clusters and analysis of the

290 T cell repertoires among these phenotypes confirmed that cells mapping to the $T_{activated}$ cluster were
291 the most clonal population in tumors (Figure S4I). Furthermore, dividing clones into groups based
292 on their expansion in nLung or tumor determined the phenotypes of shared clones compared to
293 clones detected in either tissue specifically (Figures 4D, E,
294 S4J, K). In nLung samples, the phenotypic distribution of T cells with TCRs either shared with
295 tumor samples or only present only in nLung was similar (Figure 4Ei). In tumors, however, we
296 observed differences in phenotypic distributions between cells with shared versus tissue-specific
297 TCRs (Figure 4Eii). Specifically, no $T_{NK\text{-}like}$ cells in tumors had TCRs that were uniquely expanded
298 in tumors. Furthermore, the proportion of T_{cycle}, T_{reg}, and $T_{activated}$ among cells with TCRs uniquely
299 expanded in tumors were all markedly increased compared to their proportions among cells with
300 TCRs present in both tissues, and these relationships were not observed in nLung. By controlling
301 for the distribution of cells with shared TCRs in the tumor, we found that the clonal enrichment in
302 these populations was not simply due to enrichment of these phenotypes within tumor. Together,
303 the finding that $T_{activated}$ are enriched in clonally expanded and cycling T cells at the tumor suggests
304 that their accumulation is at least in part due to local clonal expansion.
305

306 **B cells and plasma cells are increased in tumors, but the B:plasma cell ratio is conserved**
307 **between tumor and nLung.**

308 B and plasma cells represented the most globally increased lineage among immune cells in tumors
309 compared to nLung across multiple datasets (Figure 1C-E); B cells were increased as a proportion
310 of immune cells by a median of 6.4-fold (IQR: 2.5-8.4), while plasma cells were similarly
311 increased by a median of 4.1-fold (IQR: 2.2-9.4). B cells and plasma cells were strongly distinct
312 both on the RNA and surface marker level (Figure 4F, G). Plasma cell clusters included rare IgD+

313 plasma cells, which were also the only CD38^{int} population. B cell frequencies overwhelmed plasma
314 cell frequencies with IgD+ and IgM+ plasma cells being the rarest, but lineage-normalized
315 frequencies were not different between nLung and Tumor (Figures 1E and S4L). B cells and
316 plasma cells were therefore found to increase in tumors without significant overall perturbation of
317 the B:plasma cell ratio or plasma cell isotype ratios.
318
319 **Ligand-receptor interactions identify potential drivers of an adaptive activation module.** In
320 order to identify links between cellular phenotypes that may drive patient diversity, we performed
321 correlation analyses across cell type frequencies in tumors normalized within lineage (Figure 5A).
322 Among the most highly correlated cell types were $T_{activated}$, IgG+ plasma cells, and MoMΦ-II; we
323 therefore called these cell types collectively the lung cancer activation module (LCAM). The cell
324 types that were most anticorrelated to this module included B cells, $T_{cm/naïve-II}$, AMΦ, resting cDC,
325 and AZU1+ MΦ. Sorting patients by these cell types revealed that patients could be broadly
326 grouped into those with high or low frequencies of LCAM cell types (Figure 5B). We called these
327 groups LCAM^{hi} and LCAM^{lo}, respectively. Including samples from external datasets in this
328 stratification supported the overall pattern (Figure 5B and S5A). This stratification was not
329 strongly associated with changes in lineage frequencies among total immune cells, and
330 accordingly, samples from both LCAM^{hi} and LCAM^{lo} groups generally displayed lineage-
331 population shifts in line with overall tumor versus nLung differences, such as decreased NK and
332 increased B lineage frequencies (Figure 5C). Therefore, while LCAM cell types included some of
333 the populations that were most enriched in tumor compared to nLung on average, the LCAM axis
334 was importantly not a reflection of tumor sample purity.

335 To identify tumor-specific immune dysregulation that may contribute to shaping the
336 LCAM^{hi} vs. LCAM^{lo} cellular organization among patients, we performed an unbiased analysis of
337 ligand-receptor pairs between immune subsets, leveraging a dataset of secreted ligands and their
338 experimentally validated receptors[16,33] (Figure S5B, C and Table S6), comparing differences in
339 ligand-receptor (LR) intensity scores[16] between LCAM^{hi} and LCAM^{lo} groups, as well as between
340 each group and their respective adjacent nLung tissues (Table S7). Overall, both LCAM^{hi} and
341 LCAM^{lo} patients demonstrated correlated modes of LR activation in tumors compared to nLung
342 (Figure 5D). In particular, tumors in both LCAM^{hi} patients and LCAM^{lo} patients exhibited strong
343 intensity scores between T-cell derived CXCL13 and B cell CXCR5, which is likely contributing
344 to the influx of B and plasma cells seen in tumors (Figure 5E). T cells in LCAM^{hi} but not LCAM^{lo}
345 patients also produced other factors in tumors but not nLung capable of stimulating B cells through
346 the IFNG-IFNGR1 axis and the BTLA-TNFRSF14 axis (Figure 5E). In addition, B cells from
347 LCAM^{hi} but not LCAM^{lo} patients highly expressed TNFSF9 (41BBL), which ligates TNFRSF9
348 (41BB), that we found highly expressed on $T_{activated}$ cells (Figure S5D), indicating B cells from
349 LCAM^{hi} patients participate in activation of T cells via TNFSF9-TNFRSF9 interaction.

350 In addition, we observed increased IFNG-IFNGR signaling between T cells and myeloid
351 cells in LCAM^{hi} patients (Figures 5F and S5E, F). Potentially in result, LCAM^{hi} patients displayed
352 higher activation of the CXCL9/10/11-CXCR3 axis between myeloid and T cells (Figure 5G-I).
353 Whereas MΦ and MoDC demonstrated many conserved ligands upregulated in both LCAM^{hi} and
354 LCAM^{lo} tumors such as IL10 and OSM (Figure 5G, H; note distribution of highlighted LR pairs
355 along the diagonal), tumor cDC shared few ligands between the two groups, and rather upregulated
356 CCL19 higher in LCAM^{hi} patients compared to CCL17 in LCAM^{lo} patients (Figure 5I); MoDC
357 also demonstrated the latter pattern, selectively expressing CCL17 in LCAM^{lo} patients (Figure

358 5H), suggesting that DC expression of CCL19 may be a unique feature of cDC necessary for
359 activation of T cells as well as induction of a humoral immune response. Overall, differences in
360 ligand-receptor intensity scores between LCAM^{hi} and LCAM^{lo} patients supported such a patient
361 stratification, and provide possible mechanistic insight into immune-cell crosstalk underlying the
362 development of the LCAM axis, including IFNg signaling as a major driver.
363
364 **Projection of bulk-transcriptomic data onto scRNA-derived signatures reveals the presence**
365 **of the LCAM^{hi} module in two independent LUAD datasets.**
366 To identify tumor-related correlates of the LCAM module, we aimed to analyze a larger patient
367 cohort in order to increase statistical power. Therefore, we implemented an unbiased method of
368 scoring bulk transcriptomic signatures along the LCAM axis[16,34]. Specifically, we identified genes
369 that were both differentially expressed between LCAM^{hi} and LCAM^{lo} tumor samples (Figure
370 S6A), and also highly specific to the cell types enriched or depleted in the LCAM^{hi} tumors (Figure
371 S6B, see methods). Using a published tabulation on estimated immune content of 512 lung
372 adenocarcinoma (LUAD) patients available from TCGA based on expression of immune genes of
373 all lineages[35], we saw as expected that scores generated with either gene set were highly correlated
374 with estimates of overall immune content (Figure S6C, D), but an ensemble LCAM score
375 computed by the difference of these scores (LCAM^{hi} score – LCAM^{lo} score) was not (Figure S6E).
376 As predicted by our scRNAseq data, when controlling for the immune content, we generally
377 observed negative correlations of LCAM^{hi} and LCAM^{lo} gene scores among tumors except for
378 samples with the 10% lowest immune content (Figure S6F, G), suggesting that the ensemble
379 LCAM score might measure a mode of immune activation that is independent of the overall
380 immune infiltration measured by immune content. We excluded the samples with the lowest 10%

381 immune content from further analysis because probing the immune signatures was likely less
382 informative within these samples. Sorting the patients by the ensemble LCAM score revealed the
383 presence of $LCAM^{hi}$ and $LCAM^{lo}$ patient groups within the cohort (Figure 6A).

384 To see whether a similar pattern was present in additional datasets, we probed the
385 independent CPTAC: LUAD cohort, consisting of 110 treatment-naive LUAD patients undergoing
386 surgical resection on whom bulk RNAseq and WES had been performed (Michael A. Gillette, et
387 al. *Cell*, In press). Similar to the TCGA cohort, sorting the patients by the ensemble LCAM score
388 revealed the presence of $LCAM^{hi}$ and $LCAM^{lo}$ patient groups (Figure S6H), further establishing
389 the prevalence of this cellular module in a subset of LUAD patients.

390

391 **LCAM immune response correlates with tumor-genotype and expression of tumor-antigens**
392 **in LUAD lesions**

393 While the anti-tumor immune response can be modulated by many tumor-intrinsic and tumor-
394 extrinsic factors, the tumor-infiltrating immune cells exist as part of a complex microenvironment
395 that includes many other stromal populations[13,36]. To ask whether the ensemble LCAM score is
396 associated with other non-tumor, non-immune stromal populations, we derived gene lists that were
397 specific for individual stromal populations identified in a public dataset of 8 NSCLC patients[5]
398 (Table S4), and used these genes to quantify enrichment of stromal populations in TCGA LUAD
399 data. The ensemble LCAM score correlated with a cancer-associated fibroblast (CAF) enrichment
400 score, anticorrelated with a normal fibroblast enrichment score, and strongly correlated with the
401 difference of these scores (Figure 6B-D). Meanwhile, it exhibited weak or absent correlations with
402 a tumor-associated blood endothelial cell (BEC) enrichment score, an nLung BEC enrichment
403 score, and a lymphatic endothelial score (Figure S6I). These data suggest an intimate link between

404 development of the LCAM cellular module and a CAF-like fibroblast phenotype, which should be
405 explored in further detail as CAF have been suggested to act as major regulators of TIL
406 function[36,37].

407 We hypothesized that variability in immune and stromal states captured by the LCAM and
408 CAF signatures could be associated with different tumor properties. While the ensemble LCAM
409 score demonstrated a small but significant increase in large tumors (t=2.60, p=0.01 between TNM
410 T-stage=T1 and T-stage>T1), we observed variable LCAM presence among tumors of all stages
411 (Figure S6J). Furthermore, while PD-L1 expression is the most commonly used biomarker guiding
412 ICB treatment, we also observed a weak correlation between the ensemble LCAM score and total
413 *CD274* expression (r=0.21, p=2.7e-5). TMB, meanwhile, has been demonstrated to be one of the
414 most robust predictors of checkpoint response[38], and is supported by the key mechanistic
415 hypothesis that tumors with many mutations more easily activate and are targeted by the immune
416 system via the generation of mutated peptides and damage-associated molecular patterns.
417 Strikingly, the data showed that the ensemble LCAM score was strongly correlated with TMB
418 both in TCGA (Figure 6E; r=0.47 p<2.2e-16) and in CPTAC (Figure S6K) (r=0.53 p=2.3e-9). By
419 comparison, other scores measuring overall immune content (Immune ESTIMATE[35]) or specific
420 aspects of immune state (T cell-inflamed gene expression profile (GEP) score[39,40]) had much
421 weaker associations with TMB (Figure S6L, M). Importantly, correlation with TMB was observed
422 broadly across $LCAM^{hi}$ genes expressed in multiple cell types, whereas conversely, anti-
423 correlation with TMB was also observed broadly across $LCAM^{lo}$ genes (Figure S6N). The
424 ensemble LCAM score correlated with TMB to similar extents among patients grouped within
425 each TNM T-stage (Figure S6O).

426 In LUAD cases, TMB is strongly associated with smoking history. Consistent with this
427 relationship, the ensemble LCAM score correlated with smoking pack-years (rho=0.23, p=4.4e-5).
428 Therefore, smoking history confounded the correlation we observed between TMB and the
429 ensemble LCAM score, suggesting that the immune signature could be only indirectly related to
430 mutations and specifically mutated neoantigens, but rather due to alternate modes of immune
431 dysregulation related to smoking exposure. To test this hypothesis, we stratified tumors by the
432 detection of the smoking-related mutational signature characterized by C>A de-aminations within
433 specific trinucleotide contexts[41,42]. This approach removed uncertainty related to unreliability of
434 patient-reported smoking statistics and missing clinical data. We observed that both tumors with
435 and without detection of this signature exhibited significant correlations between TMB and the
436 ensemble LCAM score (r=0.38; p=9.2e-5 in the undetected smoking signature group) despite
437 having clearly distinct distributions of TMB (Figure 6E), suggesting that this relationship was
438 independent of smoking-driven immunomodulation.

439 We then asked which additional features of the tumors may influence the ensemble LCAM
440 score beyond the effect caused by differences in TMB. To perform this analysis, we regressed the
441 ensemble LCAM score onto the LogTMB and correlated candidate variables with regression
442 residuals, which quantify the difference between the observed and expected LCAM scores based
443 on this relationship. For example, scores quantifying total predicted single-nucleotide-variant- or
444 Insertion/deletion-induced neoantigens did not correlate with these differences (Figure S6P, Q),
445 indicating that these neoantigen prediction scores did not provide more information regarding the
446 LCAM immune modulation than TMB alone. However, consistent with the hypothesis that
447 generation of tumor-associated antigens was the key mechanism connecting TMB to an LCAM
448 response, we found that a score quantifying total tumor associated but not tumor-specific cancer-

449 testis antigens (CTA) was correlated with the regression residuals (Figure 6F; r=0.16, p=3.4e-3),
450 suggesting that additional tumor-associated antigens beyond those directly caused by tumor
451 mutations may also contribute to induction of the LCAM response.

452 Most adenocarcinoma patients have at least one of a small number of common driver
453 mutations, including *KRAS, EGFR, STK11,* and *TP53*. Recently, it was shown that LUAD patients
454 responsive to immune checkpoint blockade frequently have tumors harboring *TP53* mutations, and
455 that *TP53* mutant status was associated with enrichment of CD8 T cells in the TME[43,44]. However,
456 immune-related effects of individual mutations have generally not been considered independently
457 given their correlation with TMB. Specifically, while *TP53*-mutant tumors had higher ensemble
458 LCAM scores compared to *TP53*-WT/(*EGFR* or *KRAS* or *STK11*)-mut tumors in both TCGA and
459 CPTAC datasets (Figures 6G and S6R), *TP53* was also most strongly associated with increased
460 TMB (Figures 6H and S6S). In order to statistically test whether these mutations were associated
461 with higher LCAM scores while controlling for TMB, we regressed the LCAM score onto the
462 LogTMB and asked whether any individual mutations were correlated with the regression
463 residuals. Interestingly, this analysis showed that *TP53*-mutant patients had higher LCAM scores
464 than expected by a model assuming only correlation with TMB (Figure 6I; p=1.4e-3). *KRAS*-
465 mutant patients, meanwhile, had lower LCAM scores than expected by this model (Figure 6J;
466 p=1.6e-4). There was no similar deviation seen in either *STK11*- or *EGFR*-mutant patients (Figure
467 S6T). Overall, projection of bulk signatures onto axes defined by variation in our scRNAseq cohort
468 suggested that expression of the LCAM cellular module is a marker of adaptive response against
469 mutated and ectopically-expressed tumor-associated antigens that is independent from the overall
470 level of immune infiltration.
471

472 **DISCUSSION**

473 The analysis of matched tumor and nLung tissues from 35 patients as described here provides the
474 largest unbiased single-cell map of the immune response of early-stage lung cancer lesions to date.
475 CITEseq analysis, combining phenotypic classifications based on surface protein expression with
476 transcriptomic profile, serves here to help unite high-dimensional models of cellular classification
477 and refine our understanding of the immune cellular landscape in disease lesions. By further
478 integrating tumor and nLung samples from public datasets, we demonstrated the robustness of the
479 reported signatures across platforms. Importantly, based on high levels of changes conserved
480 across tumor lesions, these data support the notion that common immunotherapy treatment
481 paradigms could be beneficial for large subsets of patients despite existing disease heterogeneity.

482 Among tumors, patients could, however, be stratified along a dominant LCAM axis that
483 was independent of overall immune infiltration or changes in proportions of immune lineages.
484 This axis was defined by a high level of IgG+ plasma cells, activated T cells that were clonally
485 enriched in the tumor and expressing a proliferation signature, and MoMac-II that expressed *SPP1*,
486 a glycolysis signature, and a set of inflammatory secreted factors; this module of cell types
487 anticorrelated with B cells, T cells with a Tcm/naïve-like phenotype, resting cDC, AMΦ, and MΦ
488 expressing *AZU1*. We therefore propose that LCAM^{hi} patients are undergoing a more vigorous
489 antigen-specific antitumor adaptive immune response, whereas LCAM^{lo} patients fail to mount an
490 adaptive response to such a degree. Unbiased ligand-receptor analyses showed that, while both
491 LCAM^{hi} and LCAM^{lo} tumors expressed similar patterns of ligand-receptor pairs among immune
492 cells compared to nLung, LCAM^{hi} status was specifically related to heightened *CXCL13*
493 expression by T cells, IFNg signaling from T cells to myeloid and B cells, and CXCL-9,10,11
494 signaling from myeloid cells; cDC meanwhile expressed more *CCL19* in LCAM^{hi} tumors

495 compared to more *CCL17* in LCAM^{lo} tumors. These factors likely served to modify the immune
496 response around a set of conserved changes compared to nLung observed in both LCAM^{hi} and
497 LCAM^{lo} tumors.

498 When analyzed in the context of broader datasets with paired bulk transcriptomics and
499 whole exome sequencing, an ensemble gene score learned from the LCAM^{hi} and LCAM^{lo} patients
500 and associated cell types strongly correlated with measures indicative of high levels of tumor-
501 associated antigens, namely TMB and a cancer testis antigen score. Interestingly, the LCAM score
502 was not correlated with the overall immune infiltrate, and was independent of the T-cell inflamed
503 gene expression profile score commonly used to reflect immune activation in tumor lesions[39,40].
504 While the ensemble LCAM score was correlated with smoking status and weakly correlated with
505 stage, the relationship with TMB remained even after controlling for these possible confounders.
506 Given that TMB has demonstrated predictive power with response to ICB response in NSCLC[14,38],
507 the relationship between TMB and the LCAM cellular module in treatment-naïve patients suggests
508 that this effect may be mediated via a conditioning of the immune system that exists prior to
509 treatment, and that measurement of this cellular module may provide a more direct indicator with
510 respect to the immune system's propensity for ICB response. Specifically, the fact that many
511 factors significantly influence the LCAM score, not just TMB, demonstrates how the immune
512 system integrates multiple types of signals to establish its set point.

513 Importantly, while previously reported immune signatures have been proposed to reflect
514 tumor cytolytic activity or T cell and IFNg-driven immune response in association with tumor
515 antigens and immune evasion modes[39,40,45], the LCAM axis presented here represents an integrated
516 assessment of the immune cellular organization, based on all immune cell types as defined by
517 scRNA across patients, likely arising as a direct response to tumor antigens.

518 An additional question of clinical interest relates to how different driver mutations affect
519 the conditioning of the immune system and ICB response. The analysis presented here shows that
520 the common LUAD driver mutations *EGFR* and *STK11* had little effect on the LCAM response
521 beyond that explained by their association with TMB. While *STK11* mutation status has been
522 shown to be the most prevalent genomic driver of primary resistance to ICB[44,46], there were no
523 patients with *STK11*-mutated tumors in our scRNAseq cohort, so this effect can therefore not be
524 addressed here. Meanwhile, compared to what was expected based on each tumor's TMB alone,
525 *TP53* mutation intensified the LCAM response and *KRAS* mutation blunted it. Interestingly, the
526 latter result is consistent with a recent report demonstrating that pharmacological blockade of
527 KRAS-G12C in preclinical studies resulted in a robust immune response and synergized with
528 anti-PD1 treatment[47]. The mechanisms of these effects remain to be seen, and may relate to the
529 expression of immunomodulatory factors by the tumor, or the re-shaping of the metabolic
530 microenvironment, for example. To elucidate such pathways, close study of the tumor on a broader
531 molecular scale, in conjunction with the immune cell composition and state, is necessary.

532 A further, surprising result from our bulk RNA analyses was that the LCAM axis was
533 highly consistent with a change in fibroblast phenotype based on signatures derived from
534 scRNAseq of NSCLC stromal clusters[5]. This association could suggest that the development of
535 the tumor fibroblast phenotype is in response to overwhelming immune activation that may be
536 instigated by an adaptive, antigen-specific response.

537 An important limitation of these findings relates to the site of initiation of the LCAM
538 response; while the LCAM cellular module consists of cells undergoing an apparent active immune
539 response, this study does not demonstrate the extent to which the module is instigated or
540 perpetuated *in situ* at the tumor lesion. Specifically, despite evidence of clonally expanding $T_{activated}$

541 cells, it remains unclear whether these lineages are primed *in situ* versus in tumor-draining lymph
542 nodes (TdLN). While understanding the timescale of the tumor specific response will always be
543 challenging due to variation in patient presentation timelines, it will nevertheless be important to
544 correlate the cell types and states present in the TdLN in order to determine whether the LCAM
545 response depends on lymph node priming, as well as to develop a deeper understanding of the
546 spatial dynamics of the LCAM cell types.

547 Overall, the model presented here identifies an immune activation signature, derived from
548 definitions of immune phenotypes defined by single-cell RNA and CITEseq, as an integrator of
549 tumor-associated antigen load and driver mutation status that is not related to overall immune
550 content. We believe that this axis, therefore, can serve as a more direct measure of antigen-specific,
551 anti-tumor immune activation compared to previously suggested immune readouts.

552

553 **METHODS**

554

555 **Human subjects**

556 Samples of tumor and non-involved lung were obtained from surgical specimens of patients
557 undergoing resection at Mount Sinai Hospital (New York, NY) after obtaining informed consent
558 in accordance with a protocol reviewed and approved by the Institutional Review Board at the
559 Icahn School of Medicine at Mount Sinai (IRB Human Subjects Electronic Research Applications
560 10-00472 and 10-00135) and in collaboration with the Biorepository and Department of Pathology.

561

562 **Tissue processing**

563 Tissues were rinsed in PBS, minced and incubated for 40 minutes at 37°C in Collagenase IV
564 0.25mg/ml, Collagenase D 200U/ml and DNAse I 0.1mg.ml (all Sigma). Cell suspensions were
565 then aspirated through a 18G needle ten times and strained through a 70-micron mesh prior to RBC
566 lysis. Cell suspensions were enriched for $CD45^+$ cells by either bead positive selection (Miltenyi)
567 per kit instructions or FACS sorting on a BD FACSAria flow sorter (as indicated in Table S1)
568 prior to processing for scRNAseq or CITEseq.

569

570 **ScRNA- and TCR-seq**

571 For each sample, 10,000 cells were loaded onto a 10X Chromium single-cell encapsulation chip
572 according to manufacturer instructions. Kit versions for each sample are indicated in Table S1.
573 Libraries were prepared according to manufacturer instructions. QC of cDNA and final libraries
574 was performed using CyberGreen qPCR library quantification assay. Sequencing was performed
575 on Illumina sequencers to a depth of at least 80 million reads per library.

576 TCRseq was performed using the Chromium Single Cell 5' VDJ kit, following
577 manufacturer's instructions. For patients 695 and 706, cells were subject to a CD2+ bead
578 enrichment (Miltenyi) instead of CD45+ enrichment prior to encapsulation.
579
580 **CITEseq**
581 For each sample, cell suspensions were split and barcoded using "hashing antibodies"[48] staining
582 beta-2-microglobulin and CD298 and conjugated to "hash-tag" oligonucleotides (HTOs). Hashed
583 samples were pooled and stained with CITEseq antibodies that had been purchased either from the
584 Biolegend TOTALseq catalog or conjugated using the Thunder-Link PLUS Oligo Conjugation kit
585 (Expedeon). Sample hashing schemes and CITEseq panels are detailed in Tables S1 and S2,
586 respectively. Stained cells were then encapsulated for single-cell reverse transcription using the
587 10X Chromium platform and libraries were prepared as previously described[15] with minor
588 modifications. Briefly, cDNA amplification was performed in the presence of 2pM of an antibody-
589 oligo specific primer to increase yield of antibody derived tags (ADTs). The amplified cDNA was
590 then separated by SPRI size selection into cDNA fractions containing mRNA derived cDNA
591 (>300bp) and ADT-derived cDNAs (<180bp), which were further purified by additional rounds of
592 SPRI selection. Independent sequencing libraries were generated from the mRNA and ADT cDNA
593 fractions, which were quantified, pooled and sequenced together on an Illumina Nextseq to a depth
594 of at least 80 million reads per gene expression library and 20 million reads per ADT library.
595
596 **MICSSS**
597 FFPE tissues were stained using multiplexed immunohistochemical consecutive staining on a
598 single slide as previously described[22]. Briefly, slides were baked at 37°C overnight, deparaffinized

599 in xylene, and rehydrated in decreasing concentrations of ethanol. Tissue sections were incubated
600 in citrate buffer (pH6 or 9) for antigen retrieval at 95°C for 30 minutes, followed by incubation in
601 3% hydrogen peroxide and in serum-free protein block solution (Dako, X0909) before adding
602 primary antibody for 1 hour at room temperature. After signal amplification using secondary
603 antibody conjugated to streptavidin-horseradish peroxidase and chromogenic revelation using 3-
604 amino-9-ethylcarbazole (AEC), slides were counterstained with hematoxylin, mounted with a
605 glycerol-based mounting medium and scanned for digital imaging (Pannoramic 250 Flash III
606 whole-slide scanner, 3DHISTECH). Then the same slides were successively bleached and re-
607 stained as previously described[22]. Primary antibodies were: anti-human CD10 (200103, R&D
608 systems), DC-Lamp (1010E1.01, Novus biologicals), pan-cytokeratin (AE1/AE3, Dako), PDPN
609 (D@-40, Ventana), CD163 (10D6, Novus Biologicals) and PD-L1 (E1L3N, Cell Signaling Tech).
610

611 **Analysis of Sequencing data**

612 Transcriptomic and TCR library reads were aligned to the GRCh38/84 reference genome and
613 quantified using Cellranger (v3.1.0). CITEseq ADT and CITEseq HTO reads were queried for
614 antibody- and cell-specific oligonucleotide sequence barcodes in the designated read positions,
615 including antibody sequences within a Hamming distance of 1 from the reference, using the
616 feature-indexing function of Cellranger. Resulting alignment statistics are reported in Table S3.
617 TCR data was aligned using Cellranger *vdj* function with default parameters.
618

619 **CITEseq processing and normalization**

620 Doublets were removed based on co-staining of distinct sample-barcoding ("hashing") antibodies
621 ([*maximum HTO counts*]/[2nd *most HTO counts*] < 5) and cell barcodes with few HTO counts

622 (*maximum HTO counts* < 10) were also excluded. Cells were then assigned to samples based on
623 their maximum staining HTO. HTO to sample associations are detailed in Table S1.
624 To normalize ADT counts across experimental batches given different CITEseq staining
625 panels and sequencing runs, we performed a quantile-normalization on the ADT count values for
626 each surface marker for the immune cells in each 10X encapsulation batch. To do this, the
627 geometric average of the quantile function was computed across batches

628 $$\overline{F_m{}^{-1}(p)} = \left(\prod_{b=1}^{N}[F_{m,b}^{-1}(p) + d]\right)^{\frac{1}{N}}$$

629 where $F_{m,b}^{-1}(p)$ is the quantile function, or inverse cumulative distribution function, for counts of
630 CITEseq marker *m* on immune cells in each of *N* 10X encapsulation batches *b* and regularization
631 factor *d*=1 ADT count, evaluated at quantile *p* in interval [0,1]. This geometric average quantile
632 function provided a reference function for a common mapping of cells based on their single-
633 channel, batch-specific staining quantile *p* to a normalized staining intensity. Of note, this
634 normalization method preserved the relationships between channels while constraining the
635 observed differences in staining across experiments within individual channels.
636

637 **Unsupervised batch-aware clustering analysis**

638 Immune cells from tumor and nLung samples were filtered for cell barcodes recording > 500 UMI,
639 with < 25% mitochondrial gene expression, and with less than defined thresholds of expression
640 for genes associated with red blood cells and with epithelial cells (Table S4). Cells were clustered
641 using an unsupervised batch-aware clustering method we have recently described[16] with minor
642 adjustments. This EM-like algorithm, which was also based on earlier studies[49,50], iteratively
643 updates both cluster assignments and sample-wise noise estimates until it converges, using a

644 multinomial mixture model capturing the transcriptional profiles of the different cell-states and
645 sample specific fractions of background noise. We clustered 19 nLung and 22 tumor samples
646 jointly and 46 additional tumor and nLung samples were mapped onto the final model as described
647 below.

648 The model definitions and estimation of model parameters were as described in ([16]).
649 Specifically, the probability of observing gene i in cell j is defined as:

650 $$p_{ji} = \frac{1}{Z}\left[K_{reg} + \left(1 - \eta_{b^j}\right) \cdot \alpha_{i,map^j} + \eta_{b^j} \cdot \beta_{i,b^j}\right]$$

651 Where map^j and b^j are assignments of cells j to cell-type and batch respectively; η_{b^j} is the fraction
652 of UMIs contributed by background noise; α_{i,map^j} is the probability that a molecule drawn from
653 celltype map^j is of gene i (assuming no background noise) β_{i,b^j} is the probability that a noise UMI
654 drawn from batch b^j will be of gene i, and K_{reg} is a small regularization constant.

655 We also used here the pseudo expectation-maximization (EM) algorithm[16] to infer the
656 model parameters with minor modifications: (1) training set size was 2000 instead of 1000 cells
657 and (2) the best clustering initiation was selected from 1000 instead of 10000 kmeans+ runs. For
658 this clustering we included barcodes with more than 800 UMIs and used $K_{reg_ds} = 0.2$; (P_1,P_2) =
659 $(0^{th},30^{th})$ percentiles; $K_{reg} = 5 \cdot 10^{-6}$; k=60. Genes with high variability between patients were
660 not used in the clustering. Those genes consisted of mitochondrial, stress, metallothionein genes,
661 immunoglobulin variable chain genes, HLA class I and II genes and 3 specific genes with
662 variable/noisy expression*: MALAT1, JCHAIN and XIST* (Table S4). Ribosomal genes were
663 excluded only from the k-means clustering (Step 2.D as described in ([16])). Samples used to
664 generate this model included only those that were enriched for CD45+ immune cells using bead
665 enrichment and were processed with the 10X Chromium V2 workflow.

666

667 **Integration of additional single-cell data**

668 The resulting clustering model was used to analyze additional data that was both generated in-
669 house or downloaded from public datasets. Single cells were mapped to clusters defined by the
670 previously generated model α. Similarly to the clustering iterations, this process associates single-
671 cells of a sample with multinomial probability vectors defined by the model and estimates the
672 noise fractions of the sample η_b by optimizing the likelihood function ([16]):

673 $$f(\eta_b) = \sum_j \sum_i U_{ij} \log\left(p_{ji}\right)$$

674 For p_{ji} as defined above, while α_{i,map^j} are updated using maximum likelihood.

675 Integrating inDrop data from ([6]) and 10X Chromium 5' data required addressing the
676 systematic differences[51] in gene capture present between these technologies and 10X Chromium
677 3' data that was used to develop the clustering model. Analysis of the differences in gene
678 expression between the technologies suggested that a multiplicative correction factor C_i per each
679 gene i could adjust for the capture efficiency differences. The following process was used to
680 estimate the correction parameters:

681 1. Map cells to the original cluster models, as above, assuming absent noise in order to prevent
682 the estimated noise term from being driven by error due to batch differences instead of true
683 noise.
684 2. Re-calculate models using the average expression of the mapped cells for each cluster to
685 form "data-based models" α^D.
686 3. Calculate a weight matrix W, that weights individual genes for each cluster. W is calculated
687 by

688 $$W_{i,j} = \max(\alpha_{i,j}, {\alpha^D}_{i,j}) + w_{reg}$$

689 for original cluster model matrix α, data-based cluster model α^D, gene i, cluster j, and
690 regularization constant $w_{reg} = 10^{-10}$. Since highly detected genes tend to dominate the
691 mapping results, it is important to account for genes that are highly detected in either the
692 original (10X Chromium V2) platform or the new platform

693 4. Construct a vector of gene-specific conversion factors that can operate between platforms:

694 $$C_i = \sum_j W_{i,j} \left([c + \alpha^D{}_{i,j}]/[c + \alpha_{i,j}]\right)$$

695 for regularization factor c = 10^{-6}.

696 5. Generate transformed cluster models $\alpha'_{i,j}$ by multiplying the original models by the
697 conversion vector and dividing by a normalization factor Z:

698 $$\alpha'_{i,j} = \frac{1}{Z} * C_i \, \alpha_{i,j}$$

699 6. Map cells to transformed models without fixing the noise.

700

701 Analysis of the gene expression profiles of the mapped cells in each cluster demonstrated
702 correspondence between the model and the mapped samples across the different technologies.
703 (Fig S1B).

704

705 **Analysis of public datasets**

706 Fastqs of scRNAseq data of tumors and nLung from 8 NSCLC patients[5] acquired using 10X
707 Chromium protocols was downloaded from ArrayExpress accessions E-MTAB-6149 and E-
708 MTAB-6653. Sequencing reads were re-aligned using Cellranger as described above . Single-cells
709 were mapped to clusters as described above. Tumor samples included 3 separate samples from the
710 core, middle, and edge of each tumor. Regional tumor samples were considered separately for the

711 intra- versus inter-patient variability analyses (Figure S1G, H). For remaining analyses, cell counts
712 of projected tumor samples were pooled by patient.

713 scRNAseq data of tumors from 7 NSCLC patients[6] acquired using inDrop was downloaded
714 from GEO accession GSE127465. Since neutrophils were not detected in 10X Chromium data,
715 cells that were annotated as neutrophils in the
716 *GSE127465_human_cell_metadata_54773x25.tsv.gz* file were excluded from analysis. Cells were
717 classified by projection as described above, using the modified procedure for inDrop data.

718 TCGA LUAD RNAseq data was downloaded using the *GDCquery* and *GDCdownload*
719 functions from the *TCGAbiolinks* R package. *GDCquery* options included *project="TCGA-*
720 *LUAD", data.category="Transcriptome Profiling", data.type="Gene Expression*
721 *Quantification", workflow.type="HTSeq – FPKM", experimental.strategy="RNA-Seq",* and
722 *legacy=F*. Whole exome sequencing data was downloaded using the *GDCquery_Maf* function
723 with arguments *tumor="TCGA-LUAD"* and *pipelines="mutect2"*. Clinical data was downloaded
724 using the *GDCquery_clinic* function with arguments *project="TCGA-LUAD"* and
725 *type="clinical"*.

726 Processed CPTAC lung adenocarcinoma data was downloaded from the CPTAC Data
727 Portal https://cptac-data-portal.georgetown.edu/cptacPublic/.

728

729 **Determination of sample-sample distances**

730 Sample-sample distances were computed as the Euclidean distance between vectors consisting of
731 the Log_{10}-transformed cell type frequencies, where frequencies were computed as a fraction of
732 total immune cells. A regularization factor of 10^{-3} was applied prior to applying the log-transform.

733

734 **Determination of myeloid cell type-specific gene scores**

735 Lists of mutually-exclusive genes were used to compare monocytes, cDC2, and MΦ in Figure 2,
736 and monocytes, AMΦ, and MoMΦ in Figure 3. For these analyses, genes were identified as
737 "mutually exclusive" if the average expression was at least 2x greater in a given population than
738 in the other comparison populations. To account for the large diversity of MoMΦ clusters, the
739 maximum average expression of each MoMΦ subtype was used instead of the overall average
740 expression. Resulting gene lists are presented in Table S4. Cells were scored according to the
741 resulting gene lists as the Log-transformed fraction of UMI belonging to the gene list. Histograms
742 were generated with the R function *density* using default parameters.

743

744 **Modules analyses**

745 Gene-gene correlation modules were generated using a similar method to that previously
746 described. Briefly, cells were downsampled to 2000 UMI prior to selecting a set of variable genes,
747 similar to the selection of genes in preparation for seeding the clustering[16]. The gene-gene
748 correlation matrix for this gene set was then computed for each sample over the cell population of
749 interest. Correlation matrices were averaged following a Fisher Z-transformation. Applying the
750 inverse transformation then resulted in the best-estimate correlation coefficients of gene-gene
751 interactions across the dataset. Genes were clustered into modules using complete linkage
752 hierarchical clustering over correlation distance. Histograms of module expression scores were
753 generated with the R function *density* using default parameters.

754

755 **Classification of CD4+ versus CD8+ $T_{activated}$ cells**

756 CITEseq staining on a subset of patients was used to build a gene-set-based classifier that could
757 use mRNA UMI data to discriminate CD4+ versus CD8+ cells within the $T_{activated}$ cluster. To
758 identify these gene sets, cells from 2 patients used as a training set were gated based on a Log_2FC
759 of raw ADT counts of raw CD4/CD8 > 1 and compared by differential expression. Genes were
760 filtered by expression > 10^{-4} and a Log_2FC > 1, and nonspecific or noise-related genes such as
761 those associated with cell-cycle, long-non-coding RNAs, heat shock proteins, immunoglobulin
762 genes, ribosomal proteins, *XIST*, and histone transcripts. Resulting gene lists are reported in Table
763 S4. Cells were scored based on the fraction of RNAs belonging to the resulting gene lists, and a
764 discrimination threshold for the ratio of the CD4 vs. CD8 gene lists was determined based on the
765 overall accuracy in discriminating between CITEseq-defined CD4+ vs. CD8+ cells in the training
766 set. This gene score discriminator was validated using cells from a test set comprised of cells from
767 4 additional patients analyzed by CITEseq (584, 593, 596, 630), and on cells with unique detection
768 of either *CD4* or at least one of (*CD8A, CD8B*).
769
770 **Analysis of cycling T cell cluster**
771 To analyze the phenotypic makeup of the cluster of T cells expressing cell-cycle genes, we
772 generated gene sets based on the other T cell phenotypes described here to score each cell within
773 the cluster. To do this, we pooled the cells of each other T cell phenotype to compute its average
774 expression. We then identified a gene list for each phenotype defined by expression > 1e-5 and
775 Log2FC > 0.25 compared to the maximum of the other phenotypes. From this list, we excluded
776 variable TCR genes, and other genes associated with noise or cell stress. The gene lists for the
777 $T_{naive/CM\text{-}like}$ cell types were grouped, since these phenotypes were very similar. Resulting gene lists
778 are reported in Table S4.

779 For each cell in the cycling cluster, we then computed the fraction of UMI belonging to
780 each gene signature after removing UMIs belonging to the list of genes associated with the cycling
781 cluster that was calculated as above. We performed spherical k-means clustering using the function
782 *skmeans()* in the *skmeans* R package on these signature fractions in order to group cells within the
783 cycling cluster according to phenotypic subtype by spherical k-means cluster.
784
785 **Single-cell TCRseq analysis**
786 Single T cells were grouped by clonotype according to their precise combination of α and β chains
787 present (uniquely defined by CDR3 sequence and V, D, and J gene usage), with the following
788 acceptations in order to filter for high quality singlets:
789 1. Cells with contigs encoding > 3 productive α and β chains were excluded as multiplets.
790 2. Cells with contigs encoding > 3 productive α and β chains that completely overlapped
791 with observed cells within the multiplets were also excluded as multiplets.
792 3. Remaining cells with 3 unique α and β chains that could be uniquely associated with
793 similar cells displaying 2 unique α and β chains were assumed to be clonally related, whereas cells
794 that could be similarly associated with multiple distinct sets of cells expressing 2 unique α and β
795 chains were excluded as doublets.
796 4. Cells in which a single TCR chain was observed were assumed to be clonally related to
797 any cells with 2 unique α and β chains to which they uniquely associated.
798 5. Remaining cells in which a single TCR chain was observed were excluded if they
799 matched ambiguously to multiple cells with 2- or 3-chains.
800 Clonality scores were computed for each T cell type in each patient as *1-Peilou's eveness*
801 over the set of unique TCRs as previously described[52].

802

803 **Ligand-receptor analysis**

804 Ligand-receptor intensity scores for a set of secreted ligands (ref([33]) and Table S6) were calculated
805 as previously reported[16]. Briefly, for each ligand-receptor interaction, for each source cell type and
806 each receiver cell type, the intensity score was equal to the product of ligand generation from the
807 source cell type relative to the total RNA with the expression of the receptor on the receiver cell
808 type. Scores were independently calculated for $LCAM^{hi}$ and $LCAM^{lo}$ patient sets in nLung and
809 Tumor tissues. To determine these patient sets, patients were sorted by the geometric mean of
810 lineage-normalized cellular frequencies of $LCAM^{hi}$ and $LCAM^{lo}$ cell types, and the top half of
811 patients were defined as $LCAM^{hi}$ with the bottom half defined as $LCAM^{lo}$. Only patients analyzed
812 using 10X Chromium V2 with immune cells purified with magnetic beads were used for this
813 analysis. The patients included in these groups were: $LCAM^{hi}$: (408, 403, 522, 371, 570, 714, 584,
814 377, 406, 564, 630, 578, 514); $LCAM^{lo}$: (571, 596, 393, 593, 626, 378, 370, 410, 572, 558, 581,
815 596, 729).

816

817 **Identification of $LCAM^{hi}$ and $LCAM^{lo}$ bulk-RNA gene signatures**

818 To define genes that could probe the presence of $LCAM^{hi}$ or $LCAM^{lo}$ cell types in bulk RNA
819 data, we adopted a similar strategy to that used previously for the projection of bulk data onto
820 signatures defined by cellular axes as measured with scRNA[16,34]. Cells were evenly sampled from
821 $LCAM^{hi}$ and $LCAM^{lo}$ patients (1409 cells per patient), and sampled cells were then pooled within
822 the groups. Differentially expressed genes ($FDR<10^{-3}$ and Log2FC > 1) were retained. Genes that
823 were expressed in the filtered epithelial cells > 2x higher than in immune cells on average were
824 removed. Among the remaining differentially expressed genes, those that were expressed on

825 average within any LCAM^{hi} or LCAM^{lo} subtype with Log2FC > 3 compared to the highest
826 expressing subtype in the opposite group were retained. These gene lists were further abbreviated
827 to include no more than 10 genes per cell type, in order to balance the number of genes coming
828 from any individual cell type. In order to increase the differential expression effect sizes observed,
829 only the most extreme 6 LCAM^{hi} and LCAM^{lo} patients processed with CD45+ bead enrichment
830 and 10X Chromium V2 were included in the differential expression analysis. These patients were
831 LCAM^{hi}: (408, 403, 714, 522, 371, 570), and LCAM^{lo}: (571, 596, 393, 593, 626, 378).
832

833 **Calculation of LCAM^{hi}, LCAM^{lo}, and ensemble LCAM scores in bulk-RNA datasets**

834 Bulk RNA expression datasets were log-transformed and z-scored. For each cell type associated
835 with LCAM^{hi} or LCAM^{lo}, the resulting z-scores of the associated genes were averaged and z-
836 scored. A summary average of these values was then computed across all the cell types associated
837 with either LCAM^{hi} or LCAM^{lo} cell types.
838

839 **Published statistics for TCGA Lung adenocarcinoma patients**

840 Estimates of total immune content present in each TCGA sample (ESTIMATE score)[35] were
841 download from https://bioinformatics.mdanderson.org/public-software/estimate/.

842 Scores associating mutational signatures[41] with individual TCGA samples were
843 downloaded from the mSignatureDB[42] website http://tardis.cgu.edu.tw/msignaturedb/. For the present
844 study, detection of Signature 4 was used to indicate presence of smoking-related mutations.

845 Counts of Indel Neoantigens, SNV Neoantigens, and CTA score in TCGA cases were
846 accessed from Table S1 of ref. ([53]).
847

848 **Generation of stromal cell type scores**

849 Fibroblast and endothelial cell count matrices from tumor and nLung of 8 NSCLC patients[5] were
850 downloaded from
851 https://gbiomed.kuleuven.be/english/research/50000622/laboratories/54213024/scRNAseq-
852 NSCLC. Previously-applied[5] cluster annotations were assumed, where endothelial cluster 6 was
853 defined as "lymphatics", and endothelial and fibroblast clusters were defined based on enrichment
854 in tumor or nLung: endothelial clusters 3 and 4 were pooled as "Tumor BEC", endothelial clusters
855 1 and 5 were pooled as "Normal BEC", fibroblast cluster 1 was defined as "Normal fibroblast",
856 and fibroblast clusters 1, 2, 3, 4, 5, and 7 were pooled as "CAF". For each of these cell types, gene
857 scores were defined based on a minimum average expression of 10^{-4} and a minimum fold-change
858 threshold of 4 compared to any other stromal cell type. Cell type gene-scores were defined in
859 TCGA lung adenocarcinoma using the average z-scored gene expression of each stromal gene list.
860

861 **DATA AVAILABILITY**

862 Human scRNAseq, TCRseq, and CITEseq data is available at GEO accession GSE154826.
863

864 **ACKNOWLEDGMENTS**

865 This work was supported by National Institutes of Health (NIH) grants 5T32CA078207 (to
866 A.M.L.). We thank A. Magen, P. Hamon, M. Casanova-Acebes for critical comments on the
867 manuscript; and the Mount Sinai flow cytometry core, Human Immune Monitoring Center and
868 Mount Sinai Biorepository for support. Research reported in this paper was supported by the Office
869 of Research Infrastructure of the National Institutes of Health under award numbers
870 S10OD018522 and S10OD026880. The content is solely the responsibility of the authors and does
871 not necessarily represent the official views of the National Institutes of Health. Data used in this

publication were generated by the TCGA Research Network, and the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). Research support was provided by Regeneron and Takeda. We recognize the patients and their families for their important contributions and sacrifices.

AUTHOR CONTRIBUTIONS

MM and EK conceived the project. AML, AR, and MM designed the experiments. AML, EK, and MM wrote the manuscript. AML, EK, and MD performed computational analysis. TM, MB, AW, and RF facilitated access to human samples. AML, JG, CC, BM, AT, LW, JL, NM, GM, and KT performed experiments. NRD and GT funded part of the study. AL conducted patient consents and facilitated regulatory items. JG, JM, GM, ZZ, FP, RS, AK, PW, HS, and TM provided further intellectual input.

DECLARATION OF INTERESTS

Research support for this work was provided by Regeneron and Takeda. The authors declare no other competing financial interests.

889 **FIGURE LEGENDS**

890

891 **Figure 1. scRNA- and CITE-seq establish the diversity of transcriptional states in the tumor**

892 **microenvironment.**

893 **A**, Study overview. Resected specimens of tumor tissue and non-involved lung (nLung) were

894 digested to single cell suspensions, enriched for CD45+ cells, and subjected to single cell assays

895 including CITEseq and TCRseq.

896 **B**, Clinical data of patients undergoing resection indicating summary pathological stage, smoking

897 history, histological diagnosis, and sex.

898 **C**, Expression of cell type marker genes across scRNAseq clusters of immune cells, grouped by

899 lineage annotation (MNP: mononuclear phagocyte; pDC: plasmacytoid dendritic cell). Heatmap

900 shows the number of unique molecular identifiers (UMI) per cell. Clusters are shown using an

901 even number of randomly selected cells from 7 matched tumor and nLung sample pairs who were

902 analyzed by CITEseq. Cells were downsampled to 2000 UMI/cell.

903 **D**, Expression of lineage-defining surface markers on single cells, as measured by CITEseq. Single

904 cells correspond directly to cells shown in (**C**). CITEseq count values were first quantile

905 normalized across patients, then row-normalized across cells in the heatmap.

906 **E**, Cells per cluster as a percent of total immune cells across 35 tumor and 32 matched nLung

907 samples. Clusters correspond directly to those shown in (**C**) and (**D**).

908 **F**, Box-plots of Euclidean distances between pairs of samples among nLung only (nLung-nLung),

909 tumor only (tumor-tumor), or between nLung samples and tumor samples (Tumor-nLung).

910 Distances between pairs of patient-matched samples were excluded from the Tumor-nLung

911 distribution to prevent confounding due to patient-specific effects. *** P < 0.001, Wilcoxon rank-
912 sum test.
913 **G**, Log-ratios between cell type frequencies in tumor and nLung. Clusters were grouped by cell
914 type annotation. Crosses represent error bars showing the mean ± SEM of Log_2FC estimates of
915 differences in cell type frequency between tumor and nLung using the cohort collected in the
916 present study (Mount Sinai; x-axis) or the cohort in ref. [5] (y-axis).
917
918 **Figure 2. Intratumoral DC comprise expanded MoDC and express an LCH-like signature.**
919 **A**, Expression of key genes discriminating scRNAseq clusters of DC, grouped by cell type
920 annotation (MoDC: monocyte-derived DC; cDC: classical DC). Heatmap shows the number of
921 UMI per cell. Clusters are shown using an even number of randomly selected cells from each,
922 drawing from patients who were analyzed by CITEseq with the DC panel shown in (**B**) (4 matched
923 tumor-nLung tissue pairs). Cells were downsampled to 2000 UMI/cell.
924 **B**, Expression of DC surface markers on single cells, as measured by CITEseq. Single cells
925 correspond directly to cells shown in (**A**). CITEseq count values were first quantile normalized
926 across patients, then row-normalized across cells in the heatmap.
927 **C**, Differences between tumor and nLung of DC frequencies normalized to total DC; *P<0.05,
928 **P<0.01, ***P,0.001 (Wilcoxon signed-rank test with Bonferroni correction; N=25 matched
929 tissue pairs with >50 DC observed in each tissue).
930 **D**, Barplots showing average expression of *LAMP3 (DC-LAMP)* and *CD274 (PD-L1)* across DC
931 clusters.
932 **E**, MICSSS imaging showing spatial distribution of DC-LAMP+/PD-L1+ DC in proximity to T
933 cells in a TLS.

934 **F**, Expression of follicular dendritic cell marker MYH11 in TLS in an adjacent section to that
935 shown in (**E**).
936 **G**, Expression among CD14+ monocytes and DC of monocyte, cDC2, and MΦ cell type specific
937 gene signatures (See Figure S2D, E). Heatmaps show expression of 20 genes from each score
938 among single-cells evenly sampled by cell type (left) and as corresponding summary scores. Cells
939 were ordered by the ratio of monocyte:cDC2 summary scores and were downsampled to 2000
940 UMI.
941 **H**, Boxplots showing average expression of LCH-like signature genes across DC populations in
942 distinct nLung and tumor samples.
943
944 **Figure 3. Tumors exclude AMΦ and exhibit a diversity of MoMΦ populations.**
945 **A**, Average cluster expression of lineage-defining monocyte and MΦ clusters based on literature
946 review, grouped by cell type annotation.
947 **B**, Expression of myeloid surface markers on single cells, as measured by CITEseq. Clusters are
948 shown using an even number of randomly selected cells from each, from patients who were
949 analyzed with the panel shown (4 matched tumor-nLung pairs). CITEseq count values were first
950 quantile normalized across patients, then row-normalized across cells in the heatmap.
951 **C**, Histograms of gene module scores per cell type (see also Figure S3D-J).
952 **D-F,** Expression among CD14+ monocytes, MoMΦ, and AMΦ of cell type specific gene scores.
953 Gene scores were generated based on sets of mutually exclusive, differentially expressed genes
954 among AMΦ, MoMΦ, and CD14+ monocytes (see Figure S3I). Cells are plotted by AMΦ and
955 MoMΦ score, and cell-annotations are indicated by colored dots or contour plots (**D**). Cells are

956 plotted on similar axes and colored by CD14+ monocyte score (**E**), or by expression of individual
957 genes (**F**).
958 **G**, Differences between tumor and nLung of lineage-normalized monocyte and MΦ frequencies;
959 *P<0.05; **P<0.01, ***P,0.001 (Wilcoxon signed-rank test with Bonferroni correction; N=32
960 matched tissue pairs).
961 **H**, Average cell type expression of secreted factors across MNP cell types.
962
963 **Figure 4. CITEseq and TCR analysis of the adaptive immune compartment.**
964 **A**, Expression of key genes discriminating scRNAseq clusters of T cells, grouped by cell type
965 annotation. Heatmap shows the number of UMI per cell. Clusters are shown using an even number
966 of randomly selected cells from each, drawing from patients who were analyzed by CITEseq with
967 the T cell panel shown in (**B)** (2 matched tumor-nLung tissue pairs). Cells were downsampled to
968 2000 UMI/cell.
969 **B**, Expression of T cell surface markers on single cells, as measured by CITEseq. Single cells
970 correspond directly to cells shown in (**A**). CITEseq count values were first quantile normalized
971 across patients, then row-normalized across cells in the heatmap.
972 **C**, Differences between tumor and nLung of population frequencies normalized by total NK and
973 T cells; *P<0.05; **P<0.01, ***P<0.001 (Wilcoxon signed-rank test with Bonferroni correction,
974 N=32 matched tissue pairs).
975 **D, E**, Phenotypic distribution of T cells among tissue-stratified clonotypes. Frequencies of unique
976 TCRs observed by scTCRseq in nLung (x-axis) or tumor in a representative patient (**D**). In (**E**),
977 cells were first grouped by TCR tissue tropism categories as defined in (**D**); for 3 patients, the
978 phenotypic makeup of the cells with unique TCRs, tissue-specific TCRs, or TCRs shared across

979 tissues is plotted for nLung (**i**) and tumor tissues (**ii**) is plotted as a percent of cells with similarly
980 tissue-distributed TCRs**.** Each patient is indicated by shape.
981 **F**, Expression of key genes discriminating scRNAseq clusters of B and plasma cells, grouped by
982 cell type annotation. Heatmap shows the number of UMI per cell. Clusters are shown using an
983 even number of randomly selected cells from each, drawing from patients who were analyzed by
984 CITEseq with the B cell panel shown in (**G**) (4 matched tumor-nLung tissue pairs). Cells were
985 downsampled to 2000 UMI/cell.
986 **G**, Expression of B and plasma cell surface markers on single cells, as measured by CITEseq.
987 Single cells correspond directly to cells shown in (**F**). CITEseq count values were first quantile
988 normalized across patients, then row-normalized across cells in the heatmap.
989
990 **Figure 5. Cell-cell interactions drive an axis of adaptive activation.**
991 **A**, Spearman correlation of cell type frequencies after normalization within lineage. Analysis
992 includes 23 tumors that were processed similarly using 10X Chromium V2 and CD45+ magnetic
993 bead enrichment.
994 **B**, Lineage-normalized cell type frequencies of $LCAM^{hi}$ and $LCAM^{lo}$ cell types among pooled
995 nLung and Tumor samples from Mount Sinai and refs. ([5,6]) (50 tumor patients with 40 matched
996 nLung samples). nLung samples are ordered to match the order of tumor samples based on
997 frequencies of LCAM celltypes.
998 **C**, Immune lineage frequencies of nLung and Tumor samples; with columns corresponding to
999 patient ordering in (**B**).
1000 **D-I**, Log2 Ratio of ligand-receptor (LR) intensity scores between tumor and nLung of $LCAM^{hi}$
1001 patients, ("LR ratio"; y-axis) and $LCAM^{lo}$ patients (x-axis). All interactions among T cells, B cells,

1002 MΦ, MoDC, cDC, and monocytes, colored by indication of significance (permutation test, **D**).
1003 Dashed diagonal line indicates unity.
1004 **E-I**, Showing same data as in (**D**), but highlighting in bold LR ratios for interactions between T
1005 cell ligands and B cell receptors (**E**), T cell ligands and cDC receptors (**F**), MΦ ligands and T cell
1006 receptors (**G**), MoDC ligands and T cell receptors (**H**), and cDC ligands and T cell receptors (**I**).
1007 Labelled interactions are plotted in red.
1008
1009 **Figure 6. Tumor features related to the LCAM immune response**.
1010 **A**, Normalized expression of $LCAM^{hi}$ and $LCAM^{lo}$ bulk-RNA signature genes, determined as
1011 shown in Figure S6A, B and as described in the methods, in TCGA lung adenocarcinoma
1012 dataset. Cell type association with sets of genes for each signature is shown. Patients are sorted
1013 along y-axis by ensemble LCAM score.
1014 **B-D**, Scatter plots of the ensemble LCAM score (y-axis) with signature scores based on genes
1015 that are specific for CAFs (**B**), normal fibroblasts (**C**), or the difference between these scores (**D**)
1016 in TCGA lung adenocarcinoma data. Stromal signatures are based on the stromal data reported in
1017 ref. [5].
1018 **E**, Scatter plot of LogTMB and ensemble LCAM score. Patients are divided into those with
1019 presence of a smoking-related mutational signature (black) and those without presence of the
1020 signature (red). Black and red lines indicate linear regression relationships computed over each
1021 group of patients independently (r=0.38; p=9.2e-5 in the undetected smoking signature group;
1022 r=0.34; p=1.1e-12 in the detected signature group).
1023 **F**, Scatter plot of Cancer testes antigen expression score (CTA score), as computed in ref. [53], and
1024 the residuals of the regression of the ensemble LCAM score on the LogTMB.

1025 **G and H**, Boxplots showing either the ensemble LCAM score (**G**), or TMB (**H**) among TCGA
1026 lung adenocarcinoma patients, divided by combinations of driver mutations.
1027 **I and J**, Histograms of residuals of the regression of the ensemble LCAM score on the LogTMB,
1028 with patients stratified by *TP53* (**I**) or *KRAS* (**J**) mutational status (Two-sided t-test).
1029

1030 **Figure S1. Integration of scRNA samples and datasets for common cell type analysis.**
1031 **A**, Comparison of per-sample estimated noise levels in the training set of cells used for clustering
1032 and model formation (x-axis) compared to the per-sample estimated noise in a withheld test set of
1033 cells that were mapped to the model clusters by probabilistic projection.
1034 **B and C**, Illustration of how incorporating a fit noise component improves the concordance
1035 between predicted expression and of cells mapped to the T_{reg} cluster and observed expression. Y-
1036 axis shows the predicted expression of T_{regs} in individual samples without accounting for noise (**B**)
1037 or accounting for noise (**C**), against the observed average expression (X-axis)**.** Genes were color-
1038 coded by the ratio between the observed expression and the model without accounting for noise.
1039 Estimation of the noise component is detailed in the methods.
1040 **D**, Per-sample estimated noise levels in 10X chromium V2 samples that were used for clustering,
1041 10X chromium V2 samples that were analyzed by projection onto the clustering model and not
1042 used in the clustering, 10X chromium 5' samples that were analyzed by projection, and samples
1043 from external datasets[5,6] that were analyzed by projection.
1044 **E**, Boxplots showing the distribution of UMI per cell in each cluster.
1045 **F**, Barplots showing number of cells in Mount Sinai dataset mapping to each cluster.

1046 **G**, Heatmap showing row-normalized cell type frequencies in a public dataset[5] with samples
1047 spanning 3 regions each in a cohort of 8 NSCLC patients. Samples are clustered by spearman
1048 correlation distance. Sample names are colored by patient.
1049 **H**, Box plots of Euclidean distances based on log-transformed cluster frequencies between samples
1050 of different patients or from the same patient, as in (**G**), from ref. [5]. *** $P < 0.001$, Wilcoxon rank-
1051 sum test.
1052
1053 **Figure S2. Module analysis of DC.**
1054 **A,** Barplots showing total number of cells mapped to each individual DC cluster in the Mount
1055 Sinai cohort.
1056 **B**, Boxplots showing number of cells mapped to each individual DC cluster per tumor sample in
1057 the Mount Sinai cohort.
1058 **C**, Differences between tumor and nLung of DC frequencies normalized to total MNP; *P<0.05,
1059 **P<0.01, ***P,0.001 (Wilcoxon signed-rank test with Bonferroni correction; N=26 matched
1060 tissue pairs with >250 MNP observed in each tissue).
1061 **D**, Log_2FC and expression level distributions of gene sets that are mutually exclusively expressed
1062 in CD14+ monocytes, MΦ, and cDC2 (See Figure 2G).
1063 **E.** Histograms of cDC2 and MΦ scores, using gene lists generated as shown in (**D**).
1064 **F-I**, Gene module analysis of DC clusters. Correlation of gene module expression across all DC
1065 (**F**), five example genes from each module, ranked by correlation to the other genes in the module
1066 and colored by total expression in DC (**G**), boxplots showing Log_2FC of module expression among
1067 all DC between patient matched tumor and nLung samples (**H**), and normalized average cluster
1068 expression of modules (**I**).

1069 **J**, Boxplots showing expression of LCH-like signature genes across DC populations in distinct
1070 nLung and tumor samples from ref. [5].
1071
1072 **Figure S3. Diversity of nlung and tumor-infiltrating MΦ populations.**
1073 **A**, Expression of key genes discriminating scRNAseq clusters of monocytes and MΦ, grouped by
1074 cell type annotation. Heatmap shows the number of UMI per cell. Clusters are shown using an
1075 even number of randomly selected cells from each, drawing from 35 tumor and 32 nLung samples.
1076 Cells were downsampled to 2000 UMI/cell.
1077 **B**, Scatter plots showing normalized CITEseq CD10 and CD206 surface marker counts on AMΦ,
1078 MoMΦ, and CD14+ monocytes in nLung of a representative patient.
1079 **C**, IHC of CD10 staining AMΦ in the airspaces of nLung tissue.
1080 **D-G**, Gene module analysis of monocyte and MΦ clusters. Correlation of gene module expression
1081 across all monocytes and MΦ (**D**). Module groups illustrate groups of correlated modules which
1082 are expressed most specifically on AMΦ, MoMΦ, and monocytes (see **G**). Five example genes
1083 from each module, ranked by correlation to the other genes in the module and colored by total
1084 expression in monocytes and MΦ (**E**), boxplots showing Log_2FC of module expression among all
1085 monocytes and MΦ between patient matched tumor and nLung samples (**F**), and normalized
1086 average cluster expression of modules (**G**).
1087 **H**, Log_2FC and expression level of gene sets that are mutually exclusively expressed in CD14+
1088 monocytes, AMΦ, and MoMΦ (See Figure 3D-F).
1089
1090 **Figure S4. Phenotypic dissection of activated, cycling, and clonally expanded T cells.**

1091 **A**, Differential expression within the $T_{activated}$ cluster of cells staining for CD4 versus CD8 by
1092 CITEseq (y-axis) vs. average $T_{activated}$ expression (x-axis).
1093 **B**, Classification of $T_{activated}$ cells as CD4+ or CD8+ based on the ratio of CD4-related or CD8-
1094 related gene signatures learned from cells of 2 patients (training set; open circles) and validated on
1095 cells of 4 additional validation patients (test set; black dots). Red line indicates gene ratio threshold
1096 learned from the training set. Only cells where the CITEseq CD4:CD8 count ratio is >2 or <1/2
1097 are considered.
1098 **C**, Validation of CD4/CD8 classification scheme shown in (**B**) for cells without CITEseq staining.
1099 Cells were considered to be CD4+ or CD8+ based on unique RNA detection of either *CD4* (blue
1100 points) or at least one *CD8A* or *CD8B* transcript (green points). The discriminant line is equivalent
1101 to the gene ratio threshold learned from CITEseq data, shown in (**B**).
1102 **D**, Expression of key genes in CD4-related and CD8-related gene signatures for discriminating
1103 CD4+ and CD8+ activated T cells. Cells are sorted by ratio of these gene signatures, and the line
1104 is drawn to indicate the cells discriminated based on the threshold in panel (**B**). Heatmap shows
1105 the number of UMI per cell. Cells represent $T_{activated}$ cells from 35 tumors, and were downsampled
1106 to 2000 UMI/cell.
1107 **E**, Frequency of CD8+ or CD4+ $T_{activated}$ cells across 35 patients, as determined by gene signature
1108 scores learned from CITEseq (as in **A-D**).
1109 **F-H**, Spherical k-means sub-clustering on cell type scores of cells within the cycling T cell cluster
1110 18 based on gene scores generated from other T cell clusters. Heatmap of single-cell expression of
1111 cell type scores, grouped by sub-cluster (**F**), number of cells in each sub-cluster (**G**; nLung shown
1112 in blue, tumor in brown; lines dividing bars horizontally discriminate groups of cells from distinct

1113 patients), and the frequency of cycling T cells of each T cell phenotype (**H**; data points represent
1114 samples with at least 50 cells of the given phenotype).
1115 **I**, TCR clonality score of phenotypic groups in nLung (blue) and tumor (brown). Dots represent
1116 individual samples with at least 30 cells of indicated phenotype. N=3 patients with tumor-nLung
1117 pairs.
1118 **J**, Number of cells within each TCR category, determined as in Figure 5D, in matched nLung and
1119 tumor samples of 3 patients, each patient indicated by shape.
1120 **K**, Number of unique TCRs represented in each TCR category, determined as in Figure 5D.
1121 **L**, Differences between tumor and nLung of lineage-normalized B and plasma cell type
1122 frequencies. All comparisons were not significant (P>0.05, Wilcoxon signed-rank test, N=32
1123 matched tissue pairs).
1124
1125 **Figure S5. Ligand-receptor intaractions in LCAM^{hi} and LCAM^{lo} tumors.**
1126 **A**, Lineage-normalized cell type frequencies of all cell types among pooled nLung and Tumor
1127 samples from Mount Sinai and refs. [5,6] (50 tumor patients with 40 matched nLung samples).
1128 **B and C,** Column-normalized expression of highly expressed secreted ligands (**B**) and associated
1129 receptors (**C**) across all immune cell types, connected by lines linking ligands to receptors.
1130 Connectors are colored by association with LCAM^{hi} patients (purple), LCAM^{lo} patients (green),
1131 or all tumors (orange).
1132 **D-F**, Log2 Ratio of ligand-receptor (LR) intensity scores between tumor and nLung of LCAM^{hi}
1133 patients, ("LR ratio"; y-axis) and LCAM^{lo} patients (x-axis) as in Figure 5D, highlighting in bold
1134 LR ratios for interactions between B cell ligands and T cell receptors (**D**), T cell ligands and MoDC
1135 receptors (**E**), and T cell ligands and MΦ receptors (**F**). Labelled interactions are plotted in red.

1136

1137 **Figure S6. Projection of bulk RNA samples onto signatures defined by the LCAM scRNA**

1138 **axis.**

1139 **A and B**, Derivation of the $LCAM^{hi}$ and $LCAM^{lo}$ gene signatures for scoring bulk RNA samples.

1140 Identification of differentially-expressed genes between averaged scRNAseq samples of $LCAM^{hi}$

1141 and $LCAM^{lo}$ patients (**A**), and identification of differentially expressed genes that are specific to

1142 genes in the $LCAM^{hi}$ or $LCAM^{lo}$ cell types (**B**).

1143 **C-E**, Scatter plots of immune ESTIMATE score[35] with the $LCAM^{hi}$ signature score (**C**), the

1144 $LCAM^{lo}$ signature score (**D**), or the difference between the $LCAM^{hi}$ and $LCAM^{lo}$ signature scores

1145 (i.e. the ensemble LCAM score; **E**).

1146 **F**, Spearman correlation of the $LCAM^{hi}$ and $LCAM^{lo}$ signature scores among the deciles of

1147 immune content. Error bars represent the 95% confidence interval around the estimate of the

1148 spearman correlation.

1149 **G**, Scatter plots of the $LCAM^{hi}$ and $LCAM^{lo}$ signature scores, showing the 1st (black), 3rd (green),

1150 and 10th (red) deciles of immune content. Labelled trend lines are shown for other deciles.

1151 **H**, Normalized expression of $LCAM^{hi}$ and $LCAM^{lo}$ bulk-RNA signature genes in the CPTAC lung

1152 adenocarcinoma dataset.

1153 **I**, Scatter plots of the ensemble LCAM score (y-axis) with signature scores based on genes that

1154 are specific for tumor blood endothelial cells (BEC; left), normal BEC (center), and lymphatic

1155 endothelial cells. Stromal signatures are based on the stromal data reported in ref. [5].

1156 **J**, Boxplots showing ensemble LCAM scores among TCGA lung adenocarcinoma patients by

1157 TNM T-stage.

1158 **K**, Scatter plot of LogTMB and ensemble LCAM score in CTPAC lung adenocarcinoma patients,
1159 with linear regression line.
1160 **L and M**, Scatter plots of the LogTMB and immune ESTIMATE score[35] (**L**) and the T-cell
1161 inflamed gene expression profile (GEP[39,40]; **M**) in TCGA lung adenocarcinoma patients.
1162 **N**, Correlation between individual genes comprising the $LCAM^{hi}$ and $LCAM^{lo}$ bulk gene
1163 signatures and LogTMB in TCGA lung adenocarcinoma patients.
1164 **O**, Scatter plots of LogTMB and the LCAM ensemble score for patients by T-stage.
1165 **P and Q**, Scatter plots of the number of indel-induced neoantigens (**P**) and SNV-induced
1166 neoantigens (**Q**) as computed in ref. [53], and the residuals of the regression of the ensemble LCAM
1167 score on the LogTMB.
1168 **R and S**, Boxplots showing either the ensemble LCAM score (**G**), or TMB (**H**) among CPTAC
1169 lung adenocarcinoma patients, divided by combinations of mutated driver mutations.
1170 **T and U**, Histograms of residuals of the regression of the ensemble LCAM score on the LogTMB,
1171 with patients stratified by *STK11* (**T**) or *EGFR* (**U**) mutational status (Two-sided t-test).
1172
1173 **SUPPLEMENTAL TABLES**
1174 **Table S1.** Sample table, with information about patient, tissue, 10X loading, and QC metrics
1175 **Table S2.** CITEseq panels used
1176 **Table S3.** QC table of GEX, HTO, and ADT libraries
1177 **Table S4.** Gene lists used in paper
1178 **Table S5.** Gene modules
1179 **Table S6.** Ligand-receptor pairs used in the analysis
1180 **Table S7.** Ligand-receptor statistics
1181

1182 **REFERENCES**

1183 1. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J Clin* **69**, 7-34 (2019).

1184 2. Travis, W.D., Brambilla, E., Burke, A.P., Marx, A. & Nicholson, A.G. Introduction to The 2015
1185 World Health Organization Classification of Tumors of the Lung, Pleura, Thymus, and Heart. *J*
1186 *Thorac Oncol* **10**, 1240-1242 (2015).

1187 3. Remon, J.*, et al.* Immune Checkpoint Inhibitors in Thoracic Malignancies: Review of the Existing
1188 Evidence by an IASLC Expert Panel and Recommendations. *Journal of Thoracic Oncology* (2020).

1189 4. Laughney, A.M.*, et al.* Regenerative lineages and immune-mediated pruning in lung cancer
1190 metastasis. *Nat Med* **26**, 259-269 (2020).

1191 5. Lambrechts, D.*, et al.* Phenotype molding of stromal cells in the lung tumor microenvironment.
1192 *Nat Med* **24**, 1277-1289 (2018).

1193 6. Zilionis, R.*, et al.* Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals
1194 Conserved Myeloid Populations across Individuals and Species. *Immunity* **50**, 1317-1334 e1310
1195 (2019).

1196 7. Lavin, Y.*, et al.* Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell
1197 Analyses. *Cell* **169**, 750-765.e717 (2017).

1198 8. Kargl, J.*, et al.* Neutrophils dominate the immune cell composition in non-small cell lung cancer.
1199 *Nature Communications* **8**, 14381 (2017).

1200 9. Guo, X.*, et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell
1201 sequencing. *Nature Medicine* **24**, 978-985 (2018).

1202 10. Thommen, D.S.*, et al.* A transcriptionally and functionally distinct PD-1+ CD8+ T cell pool with
1203 predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nature Medicine* **24**,
1204 994-1004 (2018).

1205 11. Li, H.*, et al.* Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated
1206 Compartment within Human Melanoma. *Cell* **176**, 775-789 e718 (2019).

1207 12. Wu, T.D.*, et al.* Peripheral T cell expansion predicts tumour infiltration and clinical response.
1208 *Nature* **579**, 274-278 (2020).

1209 13. Binnewies, M.*, et al.* Understanding the tumor immune microenvironment (TIME) for effective
1210 therapy. *Nat Med* **24**, 541-550 (2018).

1211 14. Rizvi, N.A.*, et al.* Mutational landscape determines sensitivity to PD-1 blockade in non–small cell
1212 lung cancer. *Science* **348**, 124 (2015).

1213 15. Stoeckius, M.*, et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nature*
1214 *methods* **14**, 865-868 (2017).

1215 16. Martin, J.C., *et al.* Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular
1216 Module Associated with Resistance to Anti-TNF Therapy. *Cell* **178**, 1493-1508.e1420 (2019).

1217 17. Rosenthal, R., *et al.* Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**,
1218 479-485 (2019).

1219 18. Maier, B., *et al.* A conserved dendritic-cell regulatory program limits antitumour immunity. *Nature*
1220 **580**, 257-262 (2020).

1221 19. Zhang, Q., *et al.* Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma.
1222 *Cell* **179**, 829-845.e820 (2019).

1223 20. See, P., *et al.* Mapping the human DC lineage through the integration of high-dimensional
1224 techniques. *Science* **356**(2017).

1225 21. Dutertre, C.A., *et al.* Single-Cell Analysis of Human Mononuclear Phagocytes Reveals Subset-
1226 Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. *Immunity* **51**, 573-589
1227 e578 (2019).

1228 22. Remark, R., *et al.* In-depth tissue profiling using multiplexed immunohistochemical consecutive
1229 staining on single slide. *Sci Immunol* **1**, aaf6925 (2016).

1230 23. Ioannidis, I. & Laurini, J.A. Use of Smooth Muscle Myosin Heavy Chain as an Effective Marker
1231 of Follicular Dendritic Cells. *Appl Immunohistochem Mol Morphol* **27**, 48-53 (2019).

1232 24. Allen, C.E., Merad, M. & McClain, K.L. Langerhans-Cell Histiocytosis. *New England Journal of*
1233 *Medicine* **379**, 856-868 (2018).

1234 25. Senechal, B., *et al.* Expansion of regulatory T cells in patients with Langerhans cell histiocytosis.
1235 *PLoS Med* **4**, e253 (2007).

1236 26. Dudakov, J.A., Hanash, A.M. & van den Brink, M.R. Interleukin-22: immunobiology and
1237 pathology. *Annu Rev Immunol* **33**, 747-785 (2015).

1238 27. Khosravi, N., *et al.* IL22 Promotes Kras Mutant Lung Cancer by Induction of a Pro-Tumor Immune
1239 Response and Protection of Stemness Properties. *Cancer Immunology Research*,
1240 canimm.0655.2017 (2018).

1241 28. Chakarov, S., *et al.* Two distinct interstitial macrophage populations coexist across tissues in
1242 specific subtissular niches. *Science* **363**(2019).

1243 29. Leach, S.M., *et al.* Human and mouse transcriptome profiling identifies cross-species homology in
1244 pulmonary and lymph node mononuclear phagocytes. *bioRxiv*, 2020.2004.2030.070839 (2020).

1245 30. Hashimoto, D., *et al.* Tissue-Resident Macrophages Self-Maintain Locally throughout Adult Life
1246 with Minimal Contribution from Circulating Monocytes. *Immunity* **38**, 792-804 (2013).

1247 31. Gibbings, S.L., *et al.* Three Unique Interstitial Macrophages in the Murine Lung at Steady State.
1248 *Am J Respir Cell Mol Biol* **57**, 66-76 (2017).

1249 32. Singh, D., *et al.* CD4+ follicular helper-like T cells are key players in anti-tumor immunity. *bioRxiv*
1250 (2020).

1251 33. Ramilowski, J.A., *et al.* A draft network of ligand–receptor-mediated multicellular signalling in
1252 human. *Nature Communications* **6**, 7866 (2015).

1253 34. Neftel, C., *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma.
1254 *Cell* **178**, 835-849.e821 (2019).

1255 35. Yoshihara, K., *et al.* Inferring tumour purity and stromal and immune cell admixture from
1256 expression data. *Nature Communications* **4**, 2612 (2013).

1257 36. Salmon, H., Remark, R., Gnjatic, S. & Merad, M. Host tissue determinants of tumour immunity.
1258 *Nat Rev Cancer* **19**, 215-227 (2019).

1259 37. Salmon, H., *et al.* Matrix architecture defines the preferential localization and migration of T cells
1260 into the stroma of human lung tumors. *J Clin Invest* **122**, 899-910 (2012).

1261 38. Samstein, R.M., *et al.* Tumor mutational load predicts survival after immunotherapy across
1262 multiple cancer types. *Nat Genet* **51**, 202-206 (2019).

1263 39. Ayers, M., *et al.* IFN-γ–related mRNA profile predicts clinical response to PD-1 blockade. *The*
1264 *Journal of Clinical Investigation* **127**, 2930-2940 (2017).

1265 40. Cristescu, R., *et al.* Pan-tumor genomic biomarkers for PD-1 checkpoint blockade–based
1266 immunotherapy. *Science* **362**, eaar3593 (2018).

1267 41. Alexandrov, L.B., *et al.* Mutational signatures associated with tobacco smoking in human cancer.
1268 *Science* **354**, 618 (2016).

1269 42. Huang, P.-J., *et al.* mSignatureDB: a database for deciphering mutational signatures in human
1270 cancers. *Nucleic Acids Research* **46**, D964-D970 (2017).

1271 43. Dong, Z.-Y., *et al.* Potential Predictive Value of TP53 and KRAS Mutation Status for Response to
1272 PD-1 Blockade Immunotherapy in Lung Adenocarcinoma. *Clinical Cancer Research*,
1273 clincanres.2554.2016 (2017).

1274 44. Biton, J., *et al.* TP53, STK11, and EGFR mutations predict tumor immune profile and the response
1275 to anti–PD-1 in lung adenocarcinoma. *Clinical Cancer Research* **24**, 5710-5723 (2018).

1276 45. Rooney, Michael S., Shukla, Sachet A., Wu, Catherine J., Getz, G. & Hacohen, N. Molecular and
1277 Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* **160**, 48-61
1278 (2015).

1279 46. Skoulidis, F., *et al.* STK11/LKB1 mutations and PD-1 inhibitor resistance in KRAS-mutant lung
1280 adenocarcinoma. *Cancer discovery* **8**, 822-835 (2018).

1281 47. Canon, J., *et al.* The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity.
1282 *Nature* **575**, 217-223 (2019).

1283 48. Stoeckius, M., *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet
1284 detection for single cell genomics. *Genome Biol* **19**, 224 (2018).

1285 49. Paul, F., *et al.* Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors.
1286 *Cell* **163**, 1663-1677 (2015).

1287 50. Jaitin, D.A., *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues
1288 into cell types. *Science* **343**, 776-779 (2014).

1289 51. Ding, J., *et al.* Systematic comparison of single-cell and single-nucleus RNA-sequencing methods.
1290 *Nature Biotechnology* (2020).

1291 52. Kirsch, I., Vignali, M. & Robins, H. T-cell receptor profiling in cancer. *Mol Oncol* **9**, 2063-2070
1292 (2015).

1293 53. Thorsson, V., *et al.* The Immune Landscape of Cancer. *Immunity* **48**, 812-830.e814 (2018).

1294

Figure 1

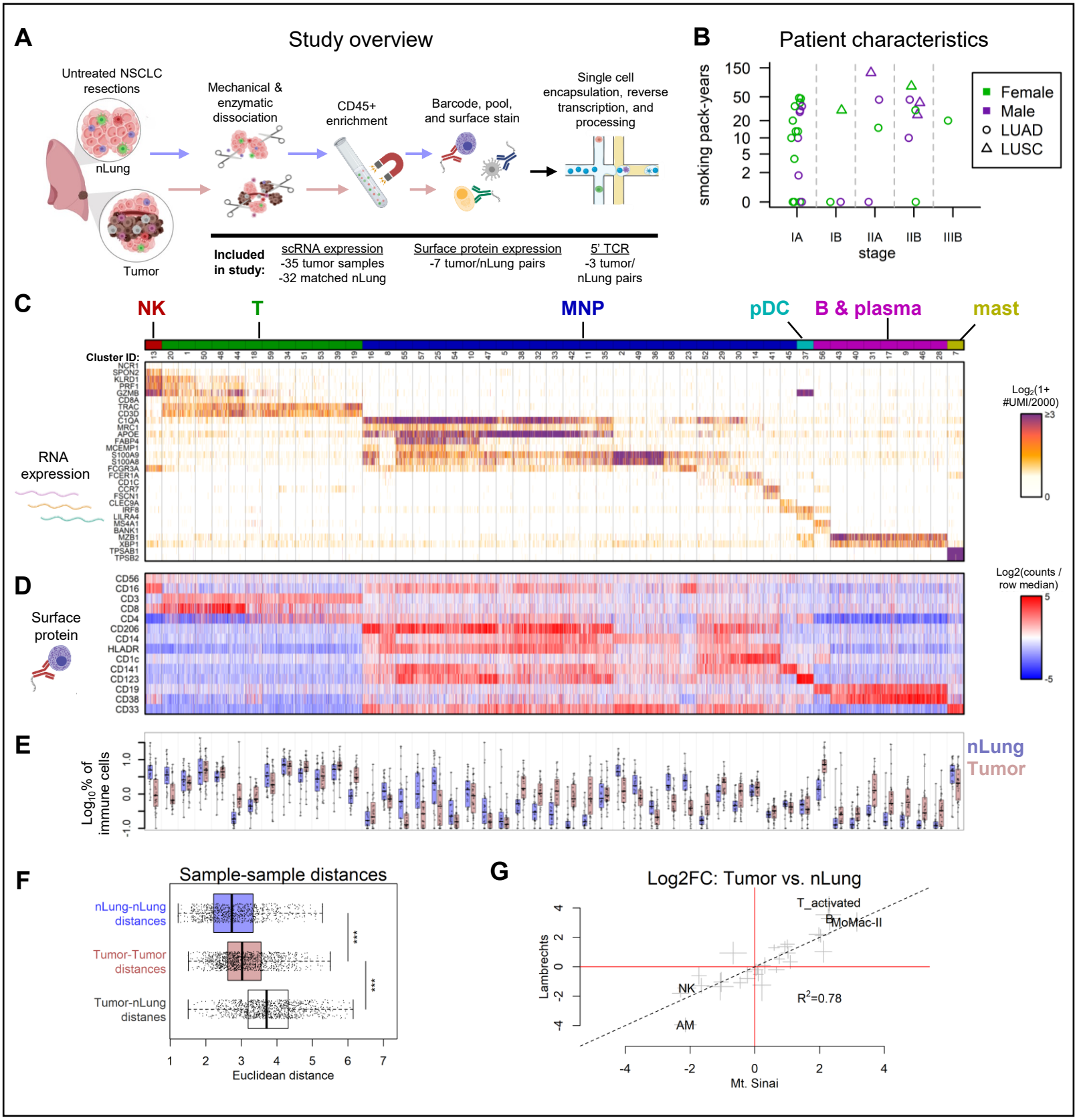

A
Study overview
Untreated NSCLC resections
nLung
Tumor
Mechanical & enzymatic dissociation
CD45+ enrichment
Barcode, pool, and surface stain
Single cell encapsulation, reverse transcription, and processing
Included in study:
scRNA expression
-35 tumor samples
-32 matched nLung
Surface protein expression
-7 tumor/nLung pairs
5' TCR
-3 tumor/ nLung pairs
B
Patient characteristics
smoking pack-years
stage
IA
IB
IIA
IIB
IIIB
Female
Male
LUAD
LUSC
C
NK
T
MNP
pDC
B & plasma
mast
Cluster ID:
RNA expression
Log2(1+ #UMI/2000)
D
Surface protein
Log2(counts / row median)
E
Log10 % of immune cells
nLung
Tumor
F
Sample-sample distances
nLung-nLung distances
Tumor-Tumor distances
Tumor-nLung distanes
Euclidean distance
G
Log2FC: Tumor vs. nLung
T_activated
B
MoMac-II
NK
AM
R2=0.78
Lambrechts
Mt. Sinai

Figure S1

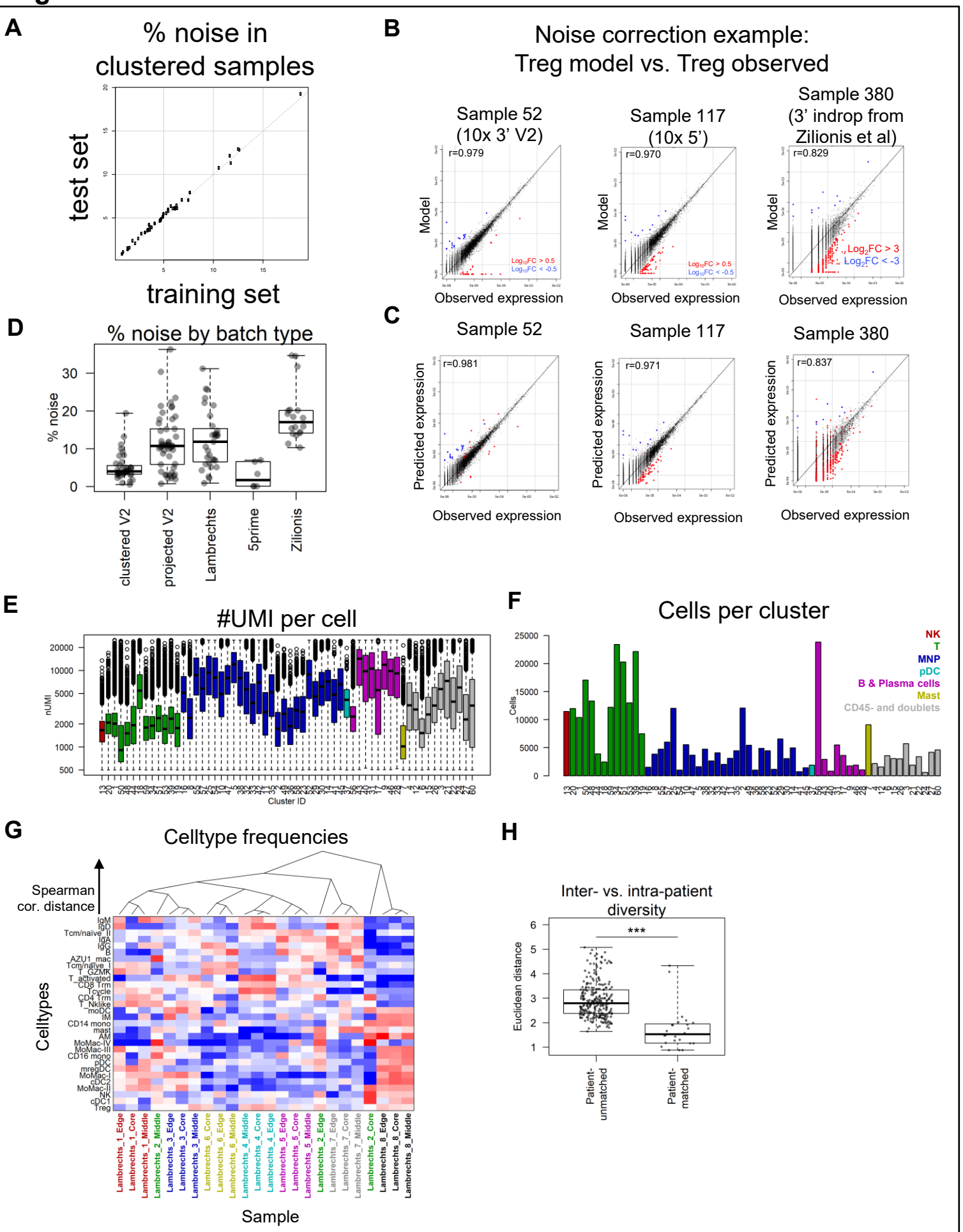

A
% noise in clustered samples
test set
training set
B
Noise correction example: Treg model vs. Treg observed
Sample 52 (10x 3' V2)
Sample 117 (10x 5')
Sample 380 (3' indrop from Zilionis et al)
r=0.979
r=0.970
r=0.829
Model
Observed expression
Log10FC > 0.5
Log10FC < -0.5
Log2FC > 3
Log2FC < -3
C
Sample 52
Sample 117
Sample 380
r=0.981
r=0.971
r=0.837
Predicted expression
Observed expression
D
% noise by batch type
% noise
clustered V2
projected V2
Lambrechts
5prime
Zilionis
E
#UMI per cell
nUMI
Cluster ID
F
Cells per cluster
Cells
NK
T
MNP
pDC
B & Plasma cells
Mast
CD45- and doublets
G
Celltype frequencies
Spearman cor. distance
Celltypes
Sample
H
Inter- vs. intra-patient diversity
Euclidean distance
***
Patient-unmatched
Patient-matched

Figure 2

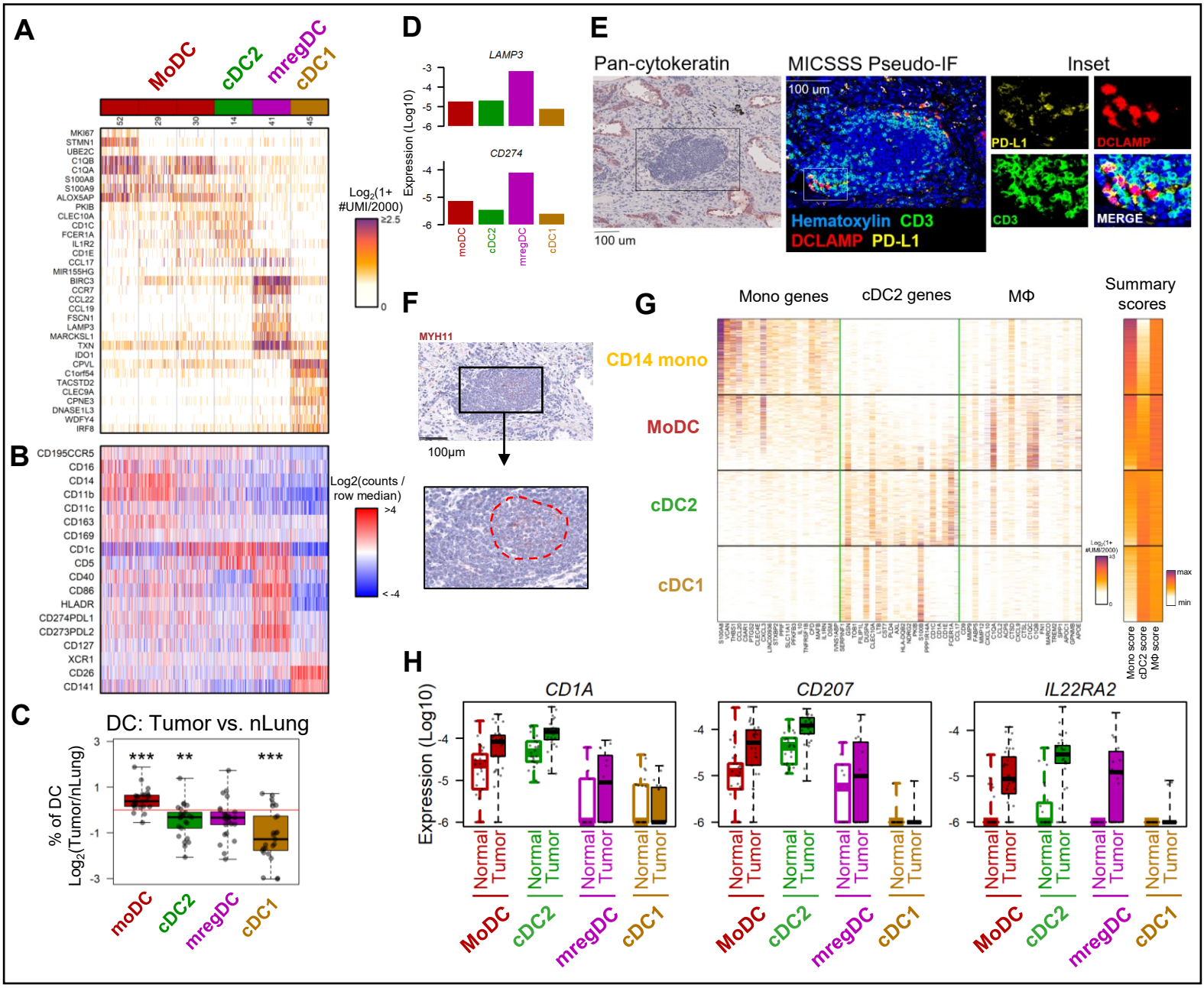

A
MoDC
cDC2
mregDC
cDC1
Log2(1+ #UMI/2000)
B
Log2(counts / row median)
C
DC: Tumor vs. nLung
% of DC
Log2(Tumor/nLung)
D
LAMP3
CD274
Expression (Log10)
E
Pan-cytokeratin
MICSSS Pseudo-IF
Inset
100 um
Hematoxylin CD3
DCLAMP PD-L1
PD-L1
DCLAMP
CD3
MERGE
F
MYH11
100µm
G
Mono genes
cDC2 genes
MΦ
Summary scores
CD14 mono
MoDC
cDC2
cDC1
Mono score
cDC2 score
MΦ score
H
CD1A
CD207
IL22RA2
Expression (Log10)
Normal
Tumor

Figure S2

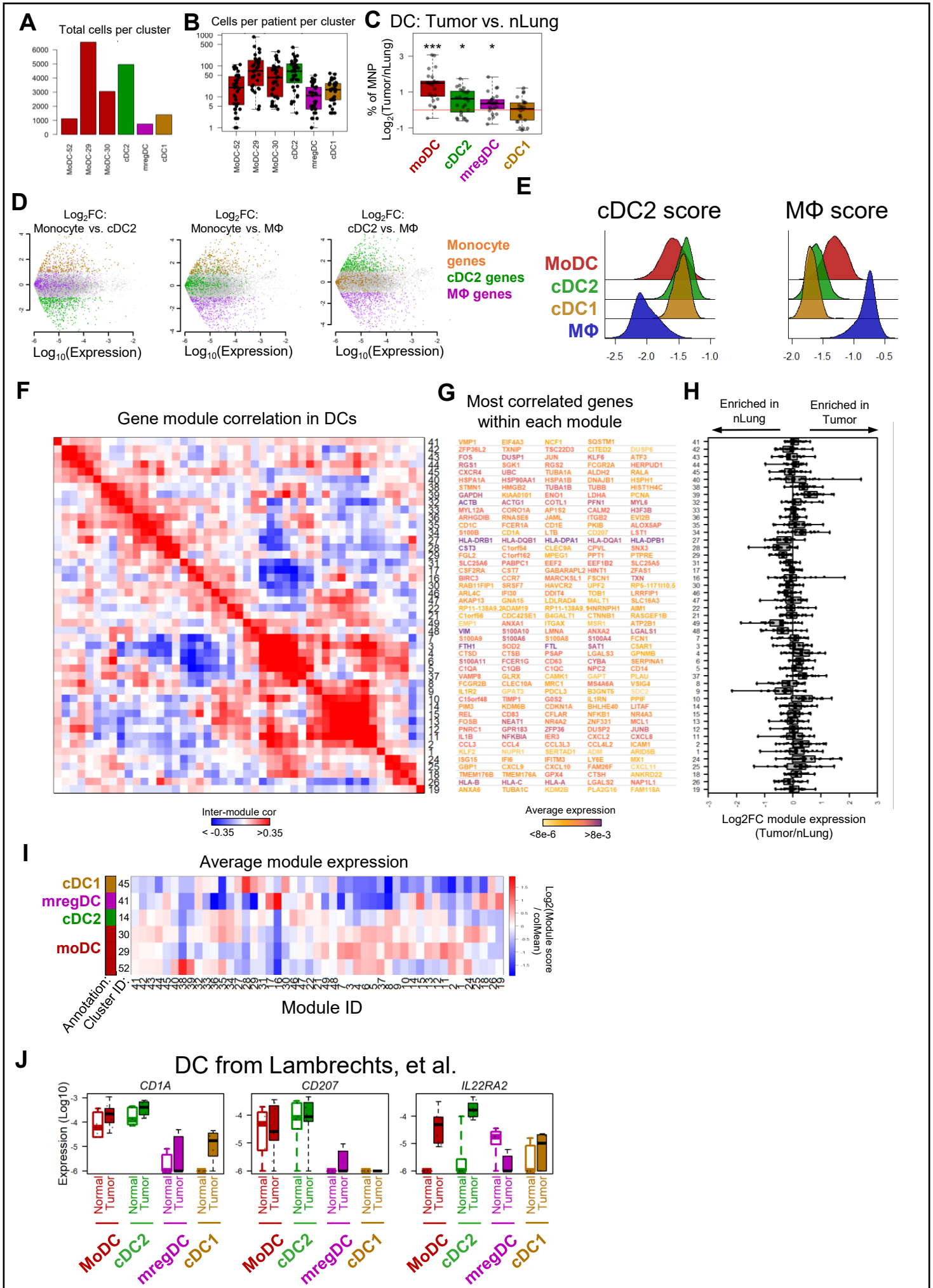

A
Total cells per cluster
MoDC-52
MoDC-29
MoDC-30
cDC2
mregDC
cDC1
B
Cells per patient per cluster
C
DC: Tumor vs. nLung
% of MNP
Log2(Tumor/nLung)
moDC
cDC2
mregDC
cDC1
D
Log2FC: Monocyte vs. cDC2
Log2FC: Monocyte vs. MΦ
Log2FC: cDC2 vs. MΦ
Log10(Expression)
Monocyte genes
cDC2 genes
MΦ genes
E
cDC2 score
MΦ score
MoDC
cDC2
cDC1
MΦ
F
Gene module correlation in DCs
Inter-module cor
< -0.35
>0.35
G
Most correlated genes within each module
Average expression
<8e-6
>8e-3
H
Enriched in nLung
Enriched in Tumor
Log2FC module expression (Tumor/nLung)
I
Average module expression
cDC1
mregDC
cDC2
moDC
Annotation:
Cluster ID:
Module ID
Log2(Module score / colMean)
J
DC from Lambrechts, et al.
CD1A
CD207
IL22RA2
Expression (Log10)
Normal
Tumor
MoDC
cDC2
mregDC
cDC1

Figure 3

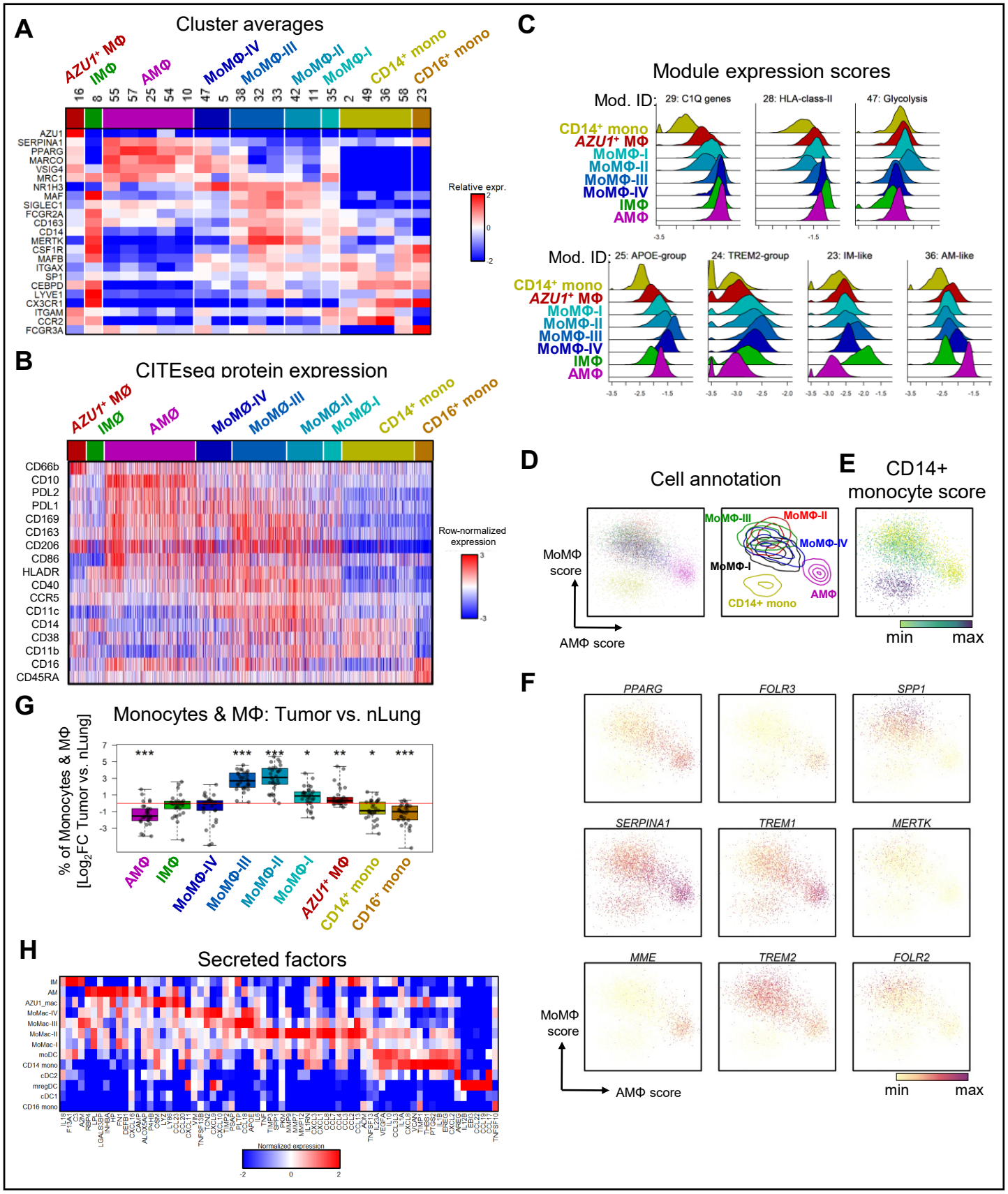

A
Cluster averages
AZU1+ MΦ
IMΦ
AMΦ
MoMΦ-IV
MoMΦ-III
MoMΦ-II
MoMΦ-I
CD14+ mono
CD16+ mono
16 8 55 57 25 54 10 47 5 38 32 33 42 11 35 2 49 36 58 23
AZU1
SERPINA1
PPARG
MARCO
VSIG4
MRC1
NR1H3
MAF
SIGLEC1
FCGR2A
CD163
CD14
MERTK
CSF1R
MAFB
ITGAX
SP1
CEBPD
LYVE1
CX3CR1
ITGAM
CCR2
FCGR3A
Relative expr.
2
0
-2
B
CITEseq protein expression
AZU1+ MΦ
IMΦ
AMΦ
MoMΦ-IV
MoMΦ-III
MoMΦ-II
MoMΦ-I
CD14+ mono
CD16+ mono
CD66b
CD10
PDL2
PDL1
CD169
CD163
CD206
CD86
HLADR
CD40
CCR5
CD11c
CD14
CD38
CD11b
CD16
CD45RA
Row-normalized expression
3
-3
C
Module expression scores
Mod. ID:
29: C1Q genes
28: HLA-class-II
47: Glycolysis
CD14+ mono
AZU1+ MΦ
MoMΦ-I
MoMΦ-II
MoMΦ-III
MoMΦ-IV
IMΦ
AMΦ
-3.5
-1.5
Mod. ID:
25: APOE-group
24: TREM2-group
23: IM-like
36: AM-like
-3.5 -2.5 -1.5
-3.5 -3.0 -2.5 -2.0
-3.5 -3.0 -2.5 -2.0 -1.5
-3.5 -2.5
16
8
55
57
25
54
10
47
5
38
32
33
42
11
35
2
49
36
58
23
D
Cell annotation
MoMΦ score
AMΦ score
MoMΦ-III
MoMΦ-II
MoMΦ-IV
MoMΦ-I
CD14+ mono
AMΦ
E
CD14+ monocyte score
min
max
F
PPARG
FOLR3
SPP1
SERPINA1
TREM1
MERTK
MME
TREM2
FOLR2
MoMΦ score
AMΦ score
min
max
G
Monocytes & MΦ: Tumor vs. nLung
% of Monocytes & MΦ [Log2FC Tumor vs. nLung]
7
5
3
1
-1
-3
***
***
***
*
**
*
***
AMΦ
IMΦ
MoMΦ-IV
MoMΦ-III
MoMΦ-II
MoMΦ-I
AZU1+ MΦ
CD14+ mono
CD16+ mono
H
Secreted factors
IM
AM
AZU1_mac
MoMac-IV
MoMac-III
MoMac-II
MoMac-I
moDC
CD14 mono
cDC2
mregDC
cDC1
CD16 mono
Normalized expression
-2
0
2

Figure S3

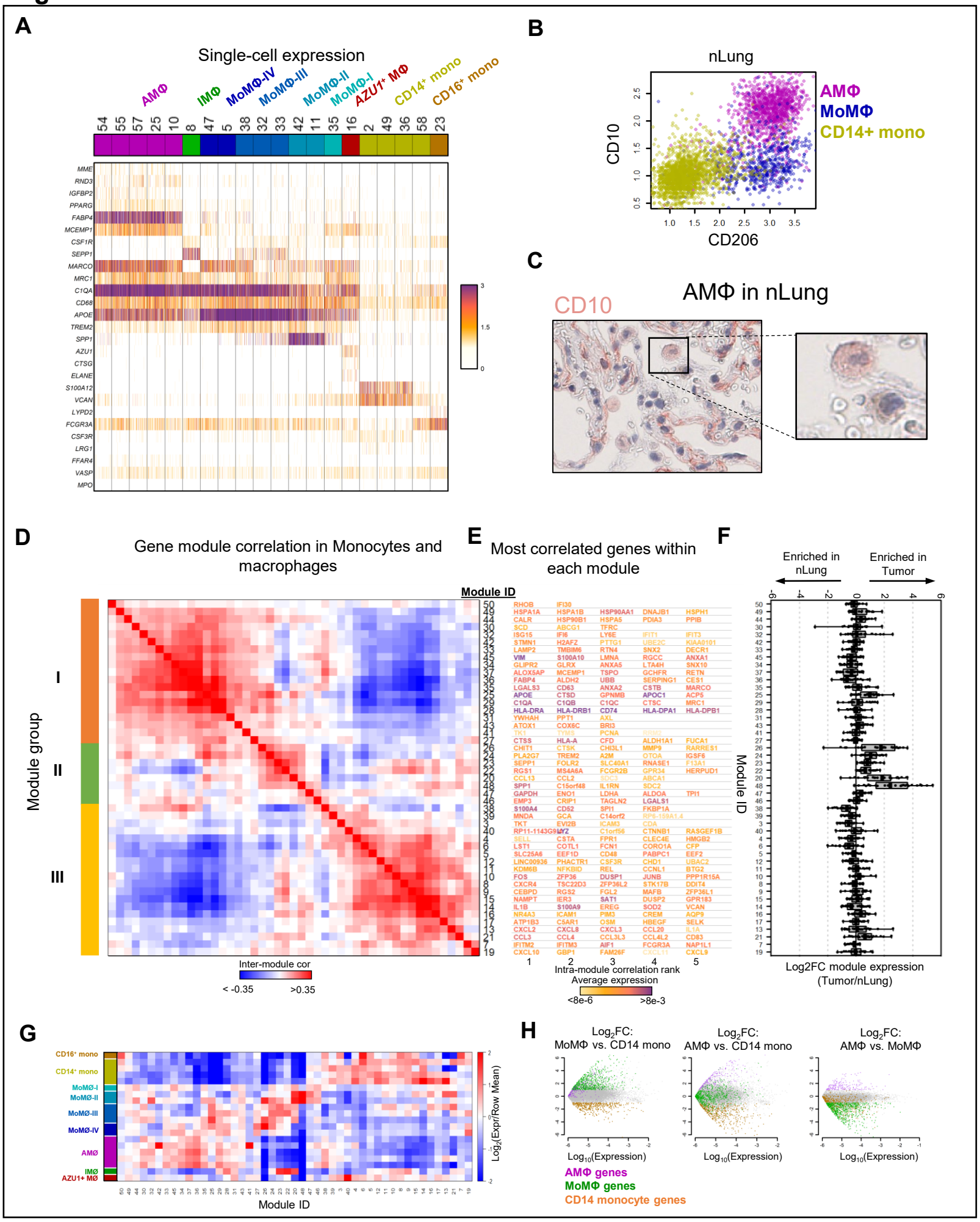

A
Single-cell expression
AMΦ
IMΦ
MoMΦ-IV
MoMΦ-III
MoMΦ-II
MoMΦ-I
AZU1+ MΦ
CD14+ mono
CD16+ mono
54 55 57 25 10 8 47 5 38 32 33 42 11 35 16 2 49 36 58 23
MME
RND3
IGFBP2
PPARG
FABP4
MCEMP1
CSF1R
SEPP1
MARCO
MRC1
C1QA
CD68
APOE
TREM2
SPP1
AZU1
CTSG
ELANE
S100A12
VCAN
LYPD2
FCGR3A
CSF3R
LRG1
FFAR4
VASP
MPO
B
nLung
AMΦ
MoMΦ
CD14+ mono
CD10
CD206
C
AMΦ in nLung
CD10
D
Gene module correlation in Monocytes and macrophages
Module group
I
II
III
Inter-module cor
< -0.35
>0.35
E
Most correlated genes within each module
Module ID
Intra-module correlation rank
Average expression
<8e-6
>8e-3
F
Enriched in nLung
Enriched in Tumor
Module ID
Log2FC module expression (Tumor/nLung)
G
CD16+ mono
CD14+ mono
MoMØ-I
MoMØ-II
MoMØ-III
MoMØ-IV
AMØ
IMØ
AZU1+ MØ
Log2(Expr/Row Mean)
Module ID
H
Log2FC: MoMΦ vs. CD14 mono
Log2FC: AMΦ vs. CD14 mono
Log2FC: AMΦ vs. MoMΦ
Log10(Expression)
AMΦ genes
MoMΦ genes
CD14 monocyte genes

Figure 4

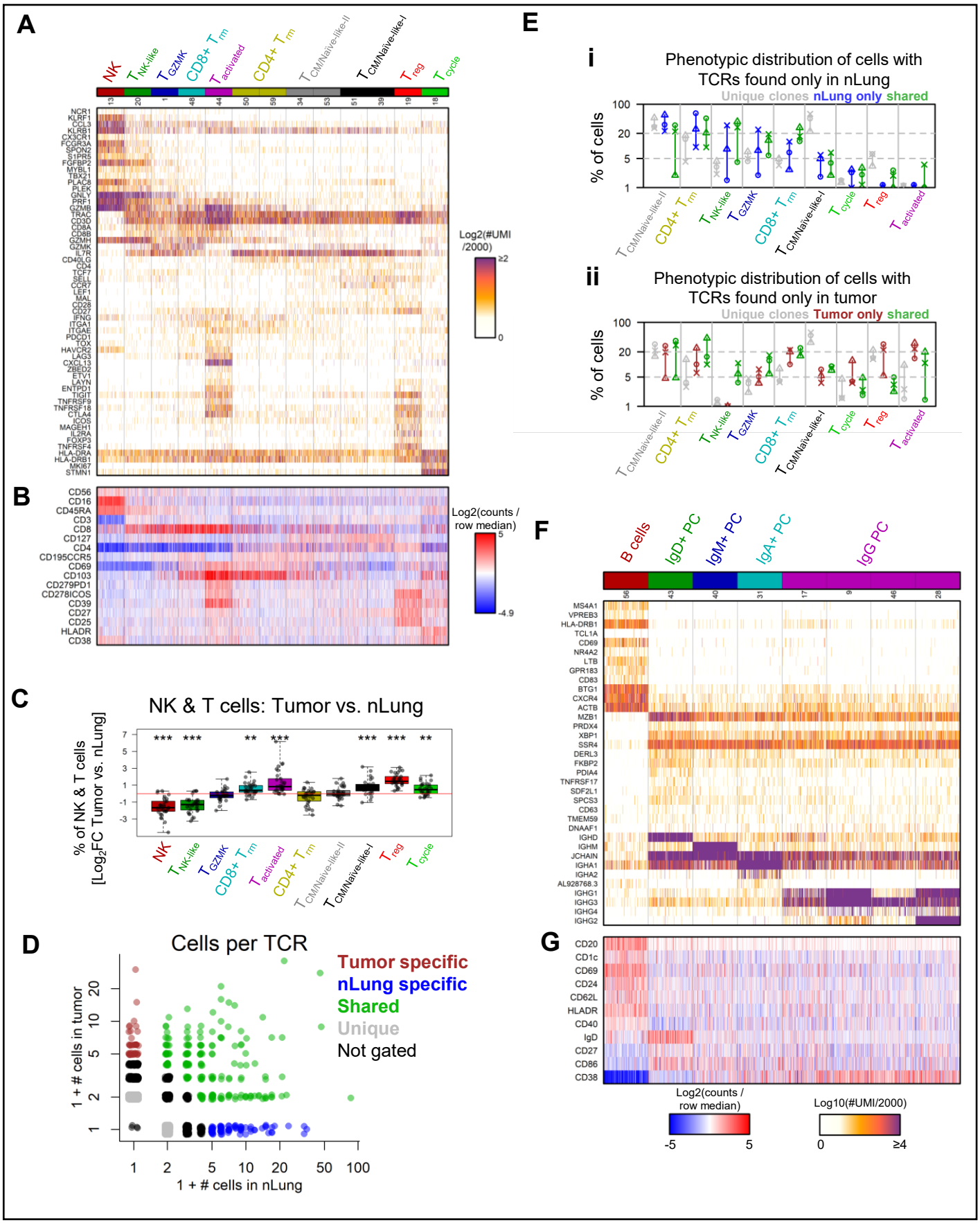

A
NK
T NK-like
T GZMK
CD8+ T rm
T activated
CD4+ T rm
T CM/Naïve-like-II
T CM/Naïve-like-I
T reg
T cycle
Log2(#UMI /2000)
B
Log2(counts / row median)
C
NK & T cells: Tumor vs. nLung
% of NK & T cells [Log2FC Tumor vs. nLung]
D
Cells per TCR
Tumor specific
nLung specific
Shared
Unique
Not gated
1 + # cells in tumor
1 + # cells in nLung
E
i
Phenotypic distribution of cells with TCRs found only in nLung
Unique clones nLung only shared
% of cells
ii
Phenotypic distribution of cells with TCRs found only in tumor
Unique clones Tumor only shared
F
B cells
IgD+ PC
IgM+ PC
IgA+ PC
IgG PC
G
Log2(counts / row median)
Log10(#UMI/2000)

Figure S4

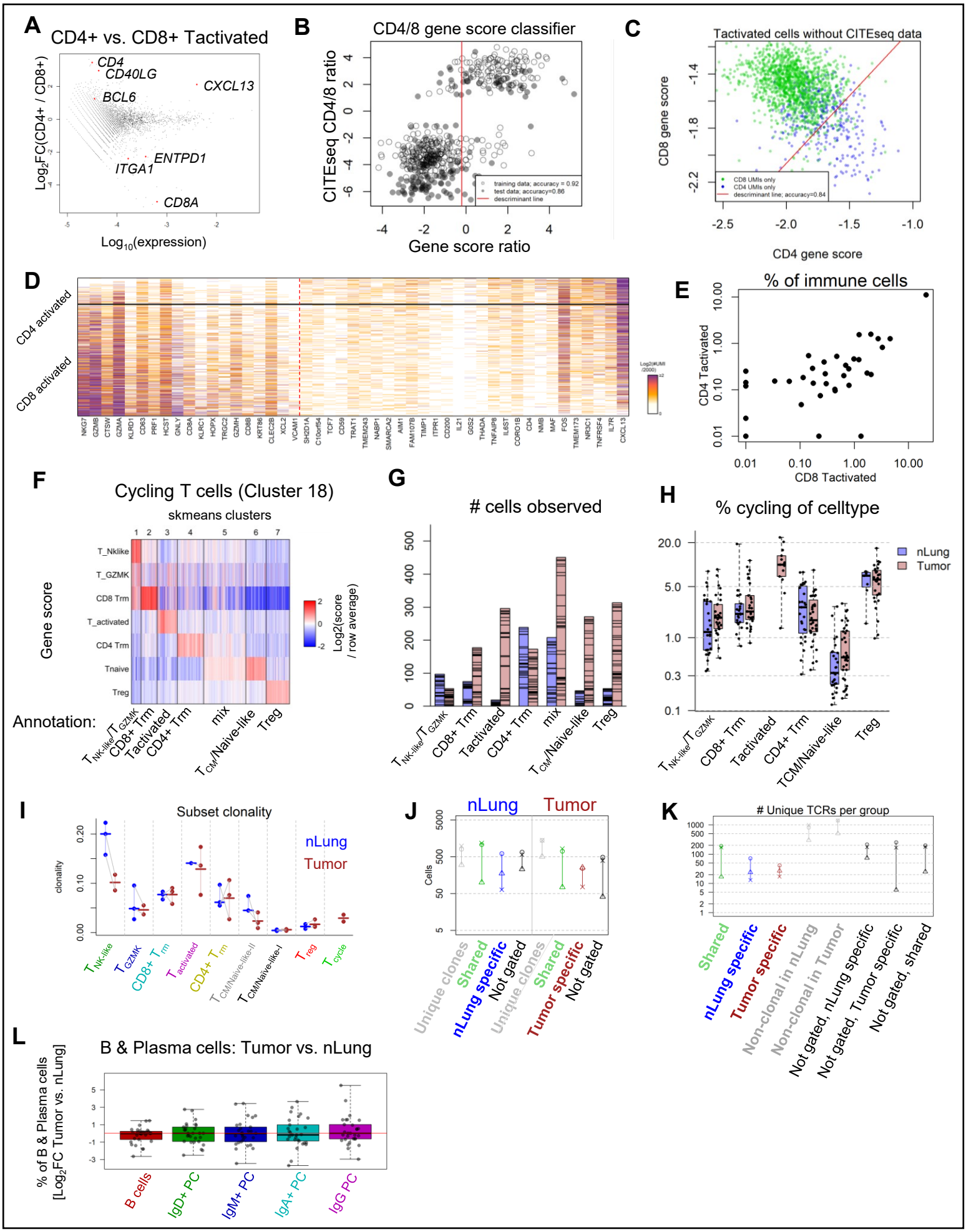

A
CD4+ vs. CD8+ Tactivated
CD4
CD40LG
CXCL13
BCL6
ENTPD1
ITGA1
CD8A
Log2FC(CD4+ / CD8+)
Log10(expression)
B
CD4/8 gene score classifier
CITEseq CD4/8 ratio
Gene score ratio
training data; accuracy = 0.92
test data; accuracy=0.86
descriminant line
C
Tactivated cells without CITEseq data
CD8 gene score
CD4 gene score
CD8 UMIs only
CD4 UMIs only
descriminant line; accuracy=0.84
D
CD4 activated
CD8 activated
E
% of immune cells
CD4 Tactivated
CD8 Tactivated
F
Cycling T cells (Cluster 18)
skmeans clusters
Gene score
T_Nklike
T_GZMK
CD8 Trm
T_activated
CD4 Trm
Tnaive
Treg
Log2(score / row average)
Annotation:
TNK-like/TGZMK
CD8+ Trm
Tactivated
CD4+ Trm
mix
TCM/Naive-like
Treg
G
cells observed
H
% cycling of celltype
nLung
Tumor
I
Subset clonality
clonality
nLung
Tumor
TNK-like
TGZMK
CD8+ Trm
Tactivated
CD4+ Trm
TCM/Naive-like-II
TCM/Naive-like-I
Treg
Tcycle
J
nLung
Tumor
Cells
Unique clones
Shared
nLung specific
Not gated
Unique clones
Shared
Tumor specific
Not gated
K
Unique TCRs per group
Shared
nLung specific
Tumor specific
Non-clonal in nLung
Non-clonal in Tumor
Not gated, nLung specific
Not gated, Tumor specific
Not gated, shared
L
B & Plasma cells: Tumor vs. nLung
% of B & Plasma cells
[Log2FC Tumor vs. nLung]
B cells
IgD+ PC
IgM+ PC
IgA+ PC
IgG PC

Figure 5

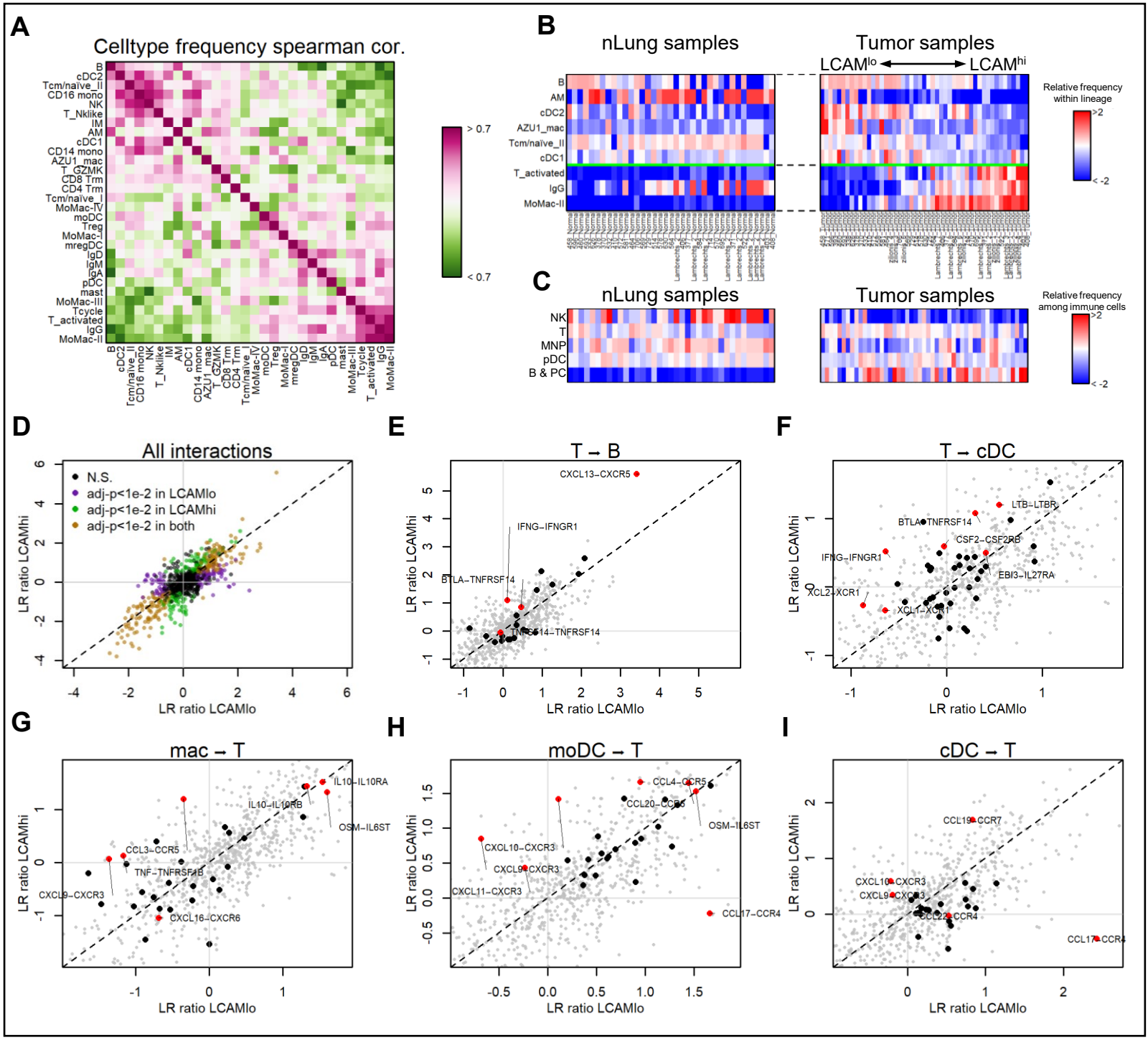

A
Celltype frequency spearman cor.
B
nLung samples
Tumor samples
LCAMlo
LCAMhi
Relative frequency within lineage
C
nLung samples
Tumor samples
Relative frequency among immune cells
D
All interactions
N.S.
adj-p<1e-2 in LCAMlo
adj-p<1e-2 in LCAMhi
adj-p<1e-2 in both
LR ratio LCAMhi
LR ratio LCAMlo
E
T → B
CXCL13–CXCR5
IFNG–IFNGR1
BTLA–TNFRSF14
F
T → cDC
LTB–LTBR
BTLA–TNFRSF14
CSF2–CSF2RB
IFNG–IFNGR1
EBI3–IL27RA
XCL2–XCR1
XCL1–XCR1
G
mac → T
IL10–IL10RA
IL10–IL10RB
OSM–IL6ST
CCL3–CCR5
TNF–TNFRSF1B
CXCL9–CXCR3
CXCL16–CXCR6
H
moDC → T
CCL4–CCR5
CCL20–CCR6
OSM–IL6ST
CXCL10–CXCR3
CXCL9–CXCR3
CXCL11–CXCR3
CCL17–CCR4
I
cDC → T
CCL19–CCR7
CXCL10–CXCR3
CXCL9–CXCR3
CCL22–CCR4
CCL17–CCR4

Figure S5

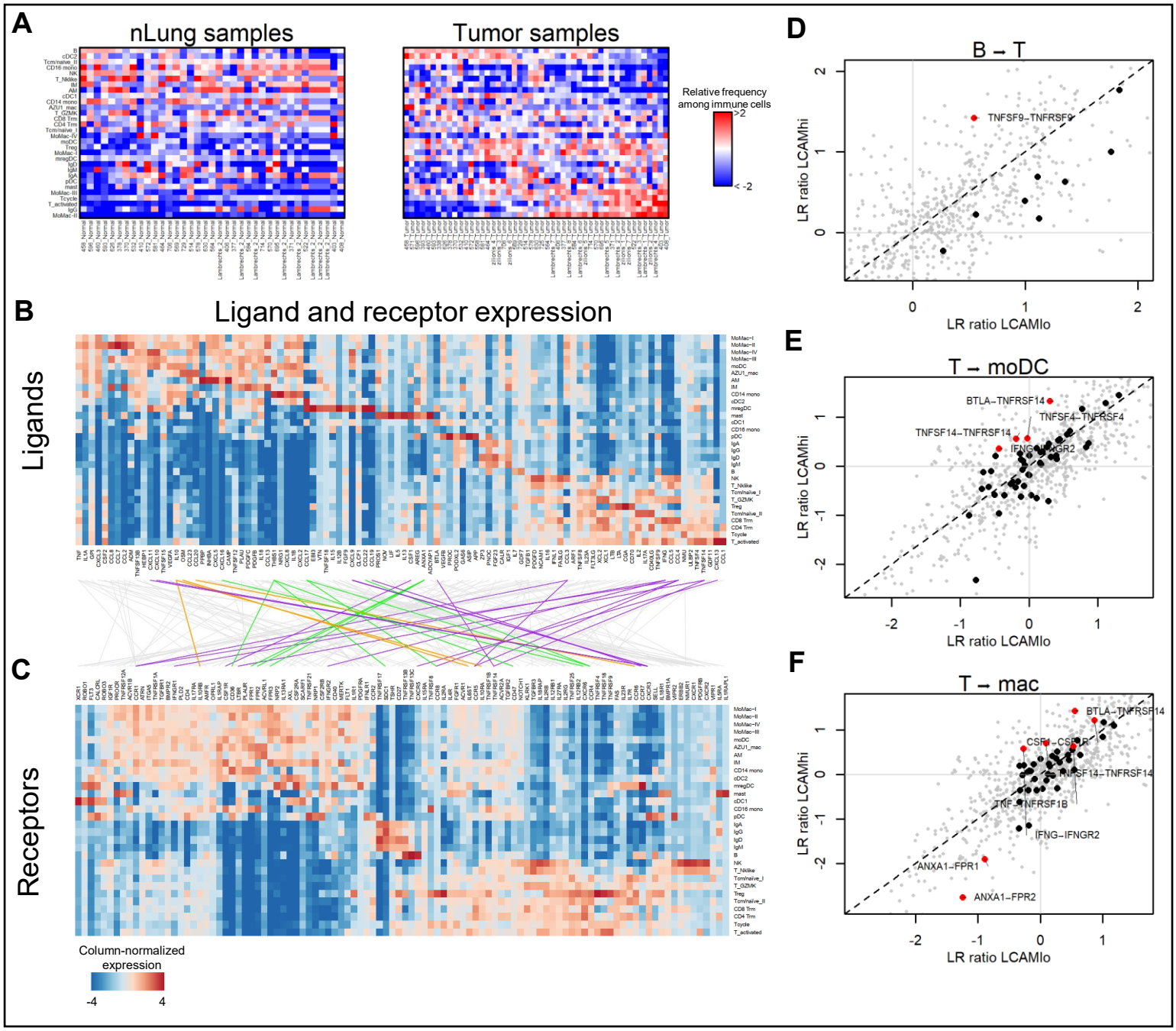

A
nLung samples
Tumor samples
Relative frequency among immune cells
>2
< -2
B
Ligand and receptor expression
Ligands
C
Receptors
Column-normalized expression
-4
4
D
B → T
TNFSF9–TNFRSF9
LR ratio LCAMhi
LR ratio LCAMlo
E
T → moDC
BTLA–TNFRSF14
TNFSF4–TNFRSF4
TNFSF14–TNFRSF14
IFNG–IFNGR2
F
T → mac
BTLA–TNFRSF14
CSF1–CSF1R
TNFSF14–TNFRSF14
TNF–TNFRSF1B
IFNG–IFNGR2
ANXA1–FPR1
ANXA1–FPR2

Figure 6

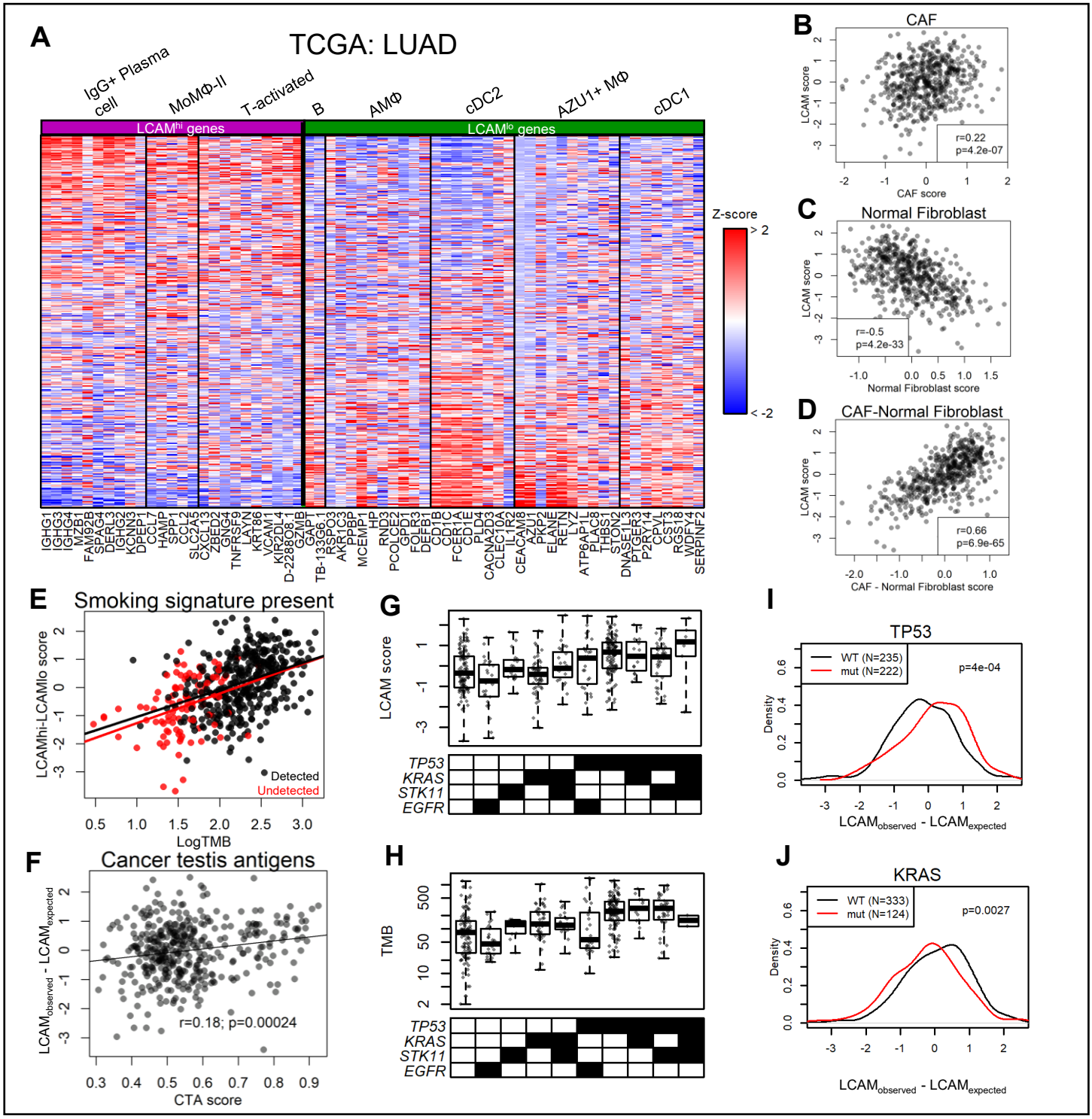
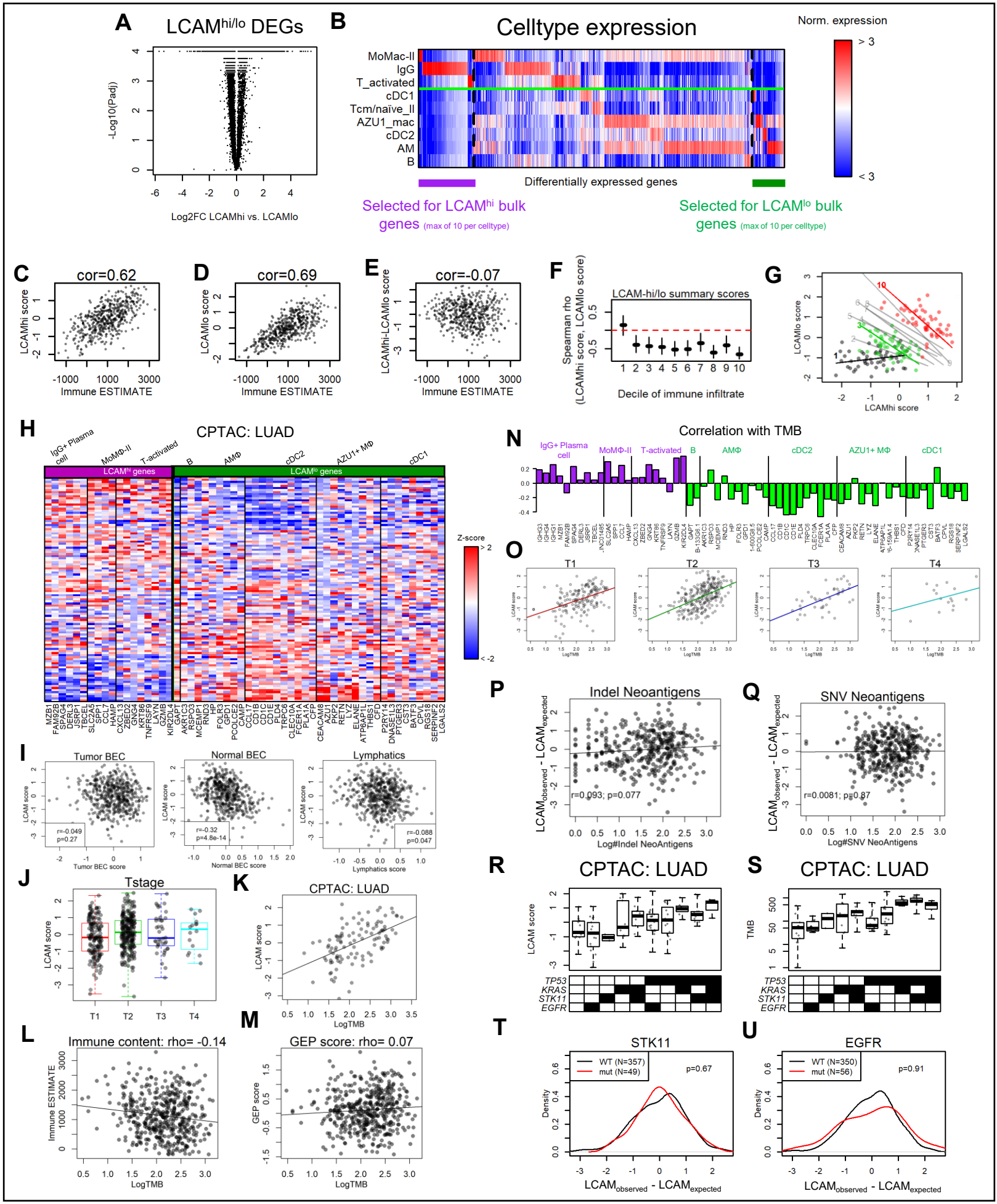

A
TCGA: LUAD
IgG+ Plasma cell
MoMΦ-II
T-activated
B
AMΦ
cDC2
AZU1+ MΦ
cDC1
LCAMhi genes
LCAMlo genes
Z-score
> 2
< -2
B
CAF
LCAM score
CAF score
r=0.22
p=4.2e-07
C
Normal Fibroblast
Normal Fibroblast score
r=-0.5
p=4.2e-33
D
CAF-Normal Fibroblast
CAF - Normal Fibroblast score
r=0.66
p=6.9e-65
E
Smoking signature present
LCAMhi-LCAMlo score
LogTMB
Detected
Undetected
F
Cancer testis antigens
CTA score
r=0.18; p=0.00024
G
LCAM score
TP53
KRAS
STK11
EGFR
H
TMB
I
TP53
WT (N=235)
mut (N=222)
p=4e-04
Density
J
KRAS
WT (N=333)
mut (N=124)
p=0.0027

Figure S6


A
LCAMhi/lo DEGs
-Log10(Padj)
Log2FC LCAMhi vs. LCAMlo
B
Celltype expression
Norm. expression
> 3
< 3
MoMac-II
IgG
T_activated
cDC1
Tcm/naïve_II
AZU1_mac
cDC2
AM
B
Differentially expressed genes
Selected for LCAMhi bulk genes (max of 10 per celltype)
Selected for LCAMlo bulk genes (max of 10 per celltype)
C
cor=0.62
LCAMhi score
Immune ESTIMATE
D
cor=0.69
LCAMlo score
Immune ESTIMATE
E
cor=-0.07
LCAMhi-LCAMlo score
Immune ESTIMATE
F
LCAM-hi/lo summary scores
Spearman rho (LCAMhi score, LCAMlo score)
Decile of immune infiltrate
G
LCAMlo score
LCAMhi score
H
CPTAC: LUAD
IgG+ Plasma cell
MoMΦ-II
T-activated
B
AMΦ
cDC2
AZU1+ MΦ
cDC1
LCAMhi genes
LCAMlo genes
Z-score
> 2
< -2
N
Correlation with TMB
O
T1
T2
T3
T4
LogTMB
I
Tumor BEC
Normal BEC
Lymphatics
r=-0.049
p=0.27
r=-0.32
p=4.8e-14
r=-0.088
p=0.047
Tumor BEC score
Normal BEC score
Lymphatics score
P
Indel Neoantigens
r=0.093; p=0.077
Log#Indel NeoAntigens
Q
SNV Neoantigens
r=0.0081; p=0.87
Log#SNV NeoAntigens
J
Tstage
LCAM score
K
CPTAC: LUAD
LogTMB
R
CPTAC: LUAD
S
CPTAC: LUAD
TMB
TP53
KRAS
STK11
EGFR
L
Immune content: rho= -0.14
Immune ESTIMATE
M
GEP score: rho= 0.07
GEP score
T
STK11
WT (N=357)
mut (N=49)
p=0.67
Density
U
EGFR
WT (N=350)
mut (N=56)
p=0.91