Resolving the spatial and cellular architecture of lung adenocarcinoma by multi-region single-cell sequencing

Ansam Sinjab [1]*, Guangchun Han [2]*, Warapen Treekitkarnmongkol [1], Kieko Hara [1], Patrick Brennan [3], Minghao Dang [2], Dapeng Hao [2], Ruiping Wang [2], Enyu Dai [2], Hitoshi Dejima [1], Jiexin Zhang [4], Elena Bogatenkova [3], Beatriz Sanchez-Espiridion [1], Kyle Chang [5], Danielle R. Little [6], Samer Bazzi [7], Linh Tran [8], Kostyantyn Krysan [8], Carmen Behrens [9], Dzifa Duose [1], Edwin R. Parra [1], Maria Gabriela Raso [1], Luisa M. Solis [1], Junya Fukuoka [10], Jianjun Zhang [9], Boris Sepesi [10], Tina Cascone [9], Lauren Byers [9], Don L. Gibbons [9], Jichao Chen [6], Seyed Javad Moghaddam [6], Edwin J. Ostrin [11], Daniel G. Rosen [12], John V. Heymach [9], Paul Scheet [1,2,5], Steven Dubinett [8], Junya Fujimoto [1], Ignacio I. Wistuba [1], Christopher S. Stevenson [13], Avrum E. Spira [13,14], Linghua Wang [2#], Humam Kadara [1#]

Departments of Translational Molecular Pathology [1], Genomic Medicine [2], Pathology [3], Bioinformatics and Computer Biology [4], Epidemiology [5], Pulmonary Medicine [6], Thoracic, Head and Neck Medical Oncology [9], Cardiovascular and Thoracic Surgery [10] and General Internal Medicine [11], The University of Texas MD Anderson Cancer Center, Houston, TX, USA. Faculty of Arts and Sciences [7], University of Balamand, Koura, Lebanon. Department of Medicine [8], The University of California Los Angeles, Los Angeles, CA, USA. Department of Pathology [10], Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan. Department of Pathology and Immunology [12], Baylor College of Medicine, Houston, TX, USA. Lung Cancer Initiative at Johnson and Johnson [13], Boston, MA, USA. Section of Computational Biomedicine [14], Boston University, Boston, MA, USA.

*: equally contributing co-first authors

#: equally contributing senior co-corresponding authors

Running title: Spatial single-cell atlas of lung adenocarcinoma

Keywords: lung adenocarcinoma, single-cell RNA sequencing, tumor heterogeneity, spatial evolution, tumor microenvironment.

Financial support: Supported in part by research funding from Johnson and Johnson, National Cancer Institute (NCI) grants R01CA205608, 1U2CCA233238, University of Texas SPORE in Lung Cancer P50CA070907 and P50 core grant CA016672 (ATGC), and NIH grant 1S10OD024977, and the start-up research funds provided to L.W. by U.T. MD Anderson Cancer Center.

Correspondence to: Humam Kadara, Ph.D., Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center. Phone: +1-713-745-3186; Email: hkadara@mdanderson.org or Linghua Wang, M.D., Ph.D., Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center. Phone: +1-713-563-2293; Email: lwang22@mdanderson.org.

Disclosure of conflict of interest: CSS and AES are employees of Johnson and Johnson. HK receives research support from Johnson & Johnson. TC reports consulting fees from MedImmune/AstraZeneca and Bristol-Myers Squibb, advisory role fees from Bristol-Myers Squibb, MedImmune/AstraZeneca and EMD Serono, and clinical research funding to MD Anderson Cancer Center from Boehringer Ingelheim, MedImmune/AstraZeneca, EMD Serono, and Bristol-Myers Squibb.

Word count: 5,932 words

Number of main figures: 6

ABSTRACT

Little is known of the geospatial architecture of individual cell populations in lung adenocarcinoma (LUAD) evolution. Here, we perform single-cell RNA sequencing of 186,916 cells from 5 early-stage LUADs and 14 multi-region normal lung tissues of defined spatial proximities from the tumors. We show that cellular lineages, states, and transcriptomic features geospatially evolve across normal regions to LUADs. LUADs also exhibit pronounced intratumor cell heterogeneity within single sites and transcriptional lineage-plasticity programs. T regulatory cell phenotypes are increased in normal tissues with proximity to LUAD, in contrast to diminished signatures and fractions of cytotoxic CD8+ T cells, antigen-presenting macrophages and inflammatory dendritic cells. We further find that the LUAD ligand-receptor interactome harbors increased expression of epithelial CD24 which mediates pro-tumor phenotypes. These data provide a spatial atlas of LUAD evolution, and a resource for identification of targets for its treatment.

Statement of significance: The geospatial ecosystem of the peripheral lung and early-stage LUAD is not known. Our multi-region single-cell sequencing analyses unravel cell populations, states, and phenotypes in the spatial and ecological evolution of LUAD from the lung that comprise high-potential targets for early interception.

INTRODUCTION

Lung adenocarcinoma (LUAD) is the most common histological subtype of lung cancer and accounts for most cancer deaths (1,2). Over the past decade, annual low dose CT screening was endorsed in an effort to reduce lung cancer mortality (3). Since then, an increasing number of early-stage LUAD diagnoses has warranted the need for novel personalized early treatment strategies. This in turn heavily rests on improved understanding of molecular and cellular processes underlying early LUAD development.

Previous studies identified molecular alterations in histologically normal-appearing epithelial *fields* that are close to solid tumors including those of the lung and that are less prevalent or absent in relatively more distant (from the tumor) regions -- suggesting geospatial heterogeneity in the uninvolved lung that is pertinent to development of a nearby tumor (4). While these studies provided valuable insights into the *spatial* development of cancer from a particular niche in the lung, they have been mainly guided by bulk profiling approaches (4,5). It is now appreciated that editing of the immune microenvironment towards protumor phenotypes including escape of immune surveillance portends the underlying biology, development, and progression of LUAD (5). Yet, the interplay between individual immune cell populations and other cell subsets in geospatial development of early-stage LUAD is not known. Technologies that profile tissues at single-cell resolution have permitted delineating the molecular and cellular complexity of tumor ecosystems. Single-cell sequencing technologies were used to chart the immune microenvironment in metastasis and therapeutic response of advanced lung cancers to targeted therapies (6-9). Yet, the complex spatial evolution of heterogeneous cellular populations and their interactions, and as an early-stage LUAD develops from the peripheral lung, remains largely unresolved.

Here, we sought to discern the spatial atlas of the peripheral lung and early-stage LUAD at single-cell resolution to better understand the topological architecture of LUAD evolution. We performed deep scRNA-seq analysis of 19 spatial regions, including enriched epithelial populations, from 5 early-stage LUADs and 14 multi-region normal-

appearing lung tissues with differential and defined spatial proximities from the tumors. Our study unravels tumor evolutionary trajectories as well as geospatial evolution in cell populations and their expression signatures that portray how early-stage LUAD develops from the lung ecosystem.

RESULTS

Single-cell spatial landscape of early-stage LUAD

To begin to chart a comprehensive single-cell atlas of early-stage LUAD and the peripheral lung, we performed scRNA-seq on all cells from an early-stage LUAD (P1, **Supplementary Table S1**) as well as matched tumor-adjacent and relatively more distant normal lung tissues (**Fig. 1A**). Unsupervised clustering of 15,132 QC-passed cells revealed cell clusters representing 5 major cellular lineages, namely epithelial, endothelial, myeloid, lymphoid, and stromal cell subsets (**Fig. 1B, Supplementary Fig. S1**). Epithelial (*EPCAM*+) cell fractions were 3.7%, 5.4%, and 3.5% for tumor, tumor-adjacent and -distant normal samples, respectively, at an average of 4.2% and in line with previous studies ((8); **Fig. 1B-C**).

To increase the throughput and to better capture patterns of cellular heterogeneity based on distance from LUADs, in particular within the epithelial lineage, we performed separate scRNA-seq analysis of epithelial (*EPCAM*+) and non-epithelial (*EPCAM*-) cells enriched from early-stage LUADs of four additional early-stage LUAD patients (P2-P5, Methods, **Supplementary Fig. S1; Supplementary Table S1**), each with three matching normal lung tissues of defined spatial proximities to LUADs: tumor-adjacent, -intermediate and -distant (19 samples and 35 scRNA-seq libraries from P1-P5; **Fig. 1A**). The spatial locations of multi-region normal tissues were carefully defined with respect to the tumor edge (see Methods), such that the samples span a spatial continuum and, thus, enable interrogation of geospatial relationships among early-stage LUAD and the peripheral lung tissues. A total of 186,916 cells, uniformly derived from all patients and sequencing batches, were retained for subsequent analyses with a median

134 of 1,844 detected genes per cell (**Supplementary Fig. S2A-C; Supplementary Table**
135 **S1**). Cells clustered into the above-mentioned 5 distinct lineages (**Supplementary Fig.**
136 **S2D**), and clustering was deemed robust based on high ratio of overlapping cell
137 memberships with an independent clustering method (reciprocal PCA), and with
138 clusters of down-sampled cells (**Supplementary Fig. S3A-B**; Supplementary Methods).
139
140 We were able to profile samples markedly enriched with epithelial cells (37.6%, n =
141 70,030 epithelial cells, including 3,514 proliferating cells) in comparison to the unbiased
142 approach in P1 (4.2%) (**Fig. 1D**). Cells were adequately derived from all spatial samples
143 and their lineage compositions varied spatially across the LUAD and normal tissues
144 (**Fig. 1E**). Proliferating epithelial cells were highly enriched in LUADs (40%) compared
145 to normal tissues (P = 0.07; **Supplementary Fig. S2E**). We also noted higher
146 transcriptome complexity (number of genes) in EPCAM+ fractions compared to
147 EPCAM- fractions from P2-P5 (**Supplementary Fig. S2A** right; **Supplementary Table**
148 **S1**). This was also significantly evident among cells of P1 which were not subjected to
149 enrichment by EPCAM-based cell sorting (**Supplementary Fig. S4A-F**; $P < 0.01$).
150
151 Next, we analyzed hierarchical relationships among major cell lineages and found that
152 cells from all patient LUADs were transcriptomically distinct from their matched spatial
153 normal counterparts (**Fig. 1F,** see Methods). Overall, cells from adjacent normal
154 samples clustered more closely with those of the LUADs than with more distant normal
155 tissues (e.g., epithelial and myeloid lineages, **Fig. 1F** top dendrograms). This average
156 pattern of transcriptomic similarity between tumor and adjacent normal samples was
157 evident in specific patients (e.g., P2, **Fig. 1F** bottom dendrograms). We then further
158 classified lymphoid and myeloid lineages into major cell types (**Fig. 1G;**
159 **Supplementary Fig. S2F,** Methods). Analysis of spatial cell composition revealed
160 distinct topological gradients with greater tumor proximity, which were consistently
161 evident across patients, as well as increased fractions of B cells and decreased
162 abundance levels of NK cells specifically in the LUADs (**Fig. 1H-I; Supplementary**
163 **Table S2**). These observations highlight geospatial transcriptomic heterogeneity in
164 tumor microenvironment (TME) landscape of early-stage LUADs.

165

166 **Spatial diversity of lung epithelial cells and intratumoral heterogeneity in LUAD**

167 We next interrogated spatial epithelial features of the LUADs and multi-region normal
168 tissues. The 70,030 epithelial cells derived from all samples clustered into 10 distinct
169 epithelial cell types with high degree of robustness (**Fig. 2A**, **Supplementary Fig. S5A-**
170 **C**). These clusters represented alveolar type I (AT1; C2, *AGER*+), AT2 (C3; *SFTPC*+),
171 basal (C4; *KRT15*+), bronchioalveolar (C5; *SFTPC*+/*SCGB1A1*+), ciliated (C6; *PIFO*+)
172 and club/secretory (C7; *BPIFB1*+) cells (**Fig. 2A-B; Supplementary Table S3**). We also
173 identified the recently described and rare ionocytes (C8; *FOXI1*+/*CFTR*+; (10,11)),
174 bipotent alveolar progenitors (C1; (12)), and unique cell states such as proliferating
175 basal cells (C10; *TOP2A*+) (**Fig. 2A-B; Supplementary Table S3**).

176 We noted a malignant-enriched cluster (C9) with cells of mixed lineages (8) from all
177 patients, mostly their LUADs (**Fig. 2A-B**, **Supplementary Fig. S6A-B**). Interestingly,
178 few cells from the normal tissues were also found in the C9 cluster (**Fig. 2B**,
179 **Supplementary Fig. S6A**). To distinguish *bona fide* malignant cells from non-malignant
180 subsets, we employed a strategy that infers copy number variations (CNVs) from
181 scRNA-seq data in every epithelial cell and we generated a CNV score to quantify their
182 level of aneuploidy (13) (see Supplementary Methods). Cells in C9 exhibited overall
183 increased CNV scores as well as higher amplitudes of CNVs (**Fig. 2A** right), thereby
184 supporting the overall malignant assignment of this cluster. While cells from the LUADs
185 predominantly resided in C9, a fraction (29%) was transcriptomically and genotypically
186 (e.g., reduced copy number profiles) similar to basal cells (17%; **Supplementary Fig.**
187 **S6C; Supplementary Table S4**). Interestingly, basal cells were enriched in all LUADs
188 compared to their normal samples (P = 0.09; **Supplementary Fig. S6B and D;**
189 **Supplementary Table S5**). We also noted pronounced steady-state enrichment or
190 depletion of epithelial subsets with spatial proximity to the tumors (**Fig. 2C**). Relative to
191 cells from tumor-intermediate or -distant normal sites, cells from tumor-adjacent normal
192 tissues were, overall, more transcriptomically similar to (clustered closely with) those
193 from the LUADs (**Fig. 2D**).

194 Alveolar differentiation hierarchies have been shown to partake in lung tumor
195 development *in vivo* (12,14,15). In our cohort, alveolar cells with definitive lineage
196 features (e.g., AT1, AT2, and alveolar progenitors) were depleted in LUAD tissues (**Fig.**
197 **2C**), which prompted us to leverage the relatively large number of alveolar cells
198 sequenced to dissect potential alveolar differentiation trajectories. Pseudotemporal
199 ordering of alveolar cells revealed a developmental hierarchy that was initiated by AT2
200 cells and that followed a main trajectory of differentiation into AT1 cells
201 (**Supplementary Fig. S7A-B**) in close agreement with previous studies in mice
202 (12,14,15). Given the reported role of NOTCH in AT2-to-AT1 differentiation and alveolar
203 repair (16), we interrogated a NOTCH signaling score which we found to be increased
204 along the alveolar differentiation trajectory (**Supplementary Fig. S7C-D**).

205 To further investigate malignant programs, we performed subclustering of cells from
206 malignant-enriched cluster C9 (n = 10,667) while overlaying CNV scores, which
207 separated likely malignant cells from subsets derived from normal tissues (**Fig. 2E,**
208 **Supplementary Fig. S8A-E; Supplementary Table S6**). We noted low CNV scores
209 (**Supplementary Fig. 8A** and **Fig. 2E** right) in cells of the malignant-enriched cluster in
210 the LUAD of P2. We thus interrogated the presence of *KRAS* codon 12 driver mutations
211 which reside within genomic brackets captured in our 5' single-cell sequencing design
212 (see Supplementary Methods). Interestingly, among all P2 epithelial subsets, 29% of C9
213 cells (160 of 547) harbored the *KRAS* G12D mutation (**Fig. 2E** right **and 2F**) and were
214 exclusively derived from the tumor sample. Compared to *KRAS* wild-type cells from the
215 same LUAD, cells harboring *KRAS* G12D mutation exhibited distinctively high
216 expression of tumor markers (e.g., *CEACAM5*; **Supplementary Fig. S8B**), increased
217 expression of *MUC5AC* (**Fig. 2F**) and *LCN2* as well as reduced expression of *NKX2-1*
218 (**Supplementary Fig. S8B; Supplementary Table S7**), altogether suggestive of
219 mucinous differentiation (17,18) and in line with the histological (mucinous) pattern of
220 this tumor (**Supplementary Table S1**). These findings underscore spatial heterogeneity
221 patterns in *KRAS* driver mutation and cellular lineage that are likely unique to the
222 ecosystem of *KRAS*-mutant LUAD. Additional clustering of tumor-derived C9 cells by
223 patient underscored transcriptomic features that were shared between 2 or more
224 LUADs (e.g., increased *CEACAM5* or *CEACAM6*) (**Supplementary Fig. S8B-E**). We

225 also noted patient/LUAD-specific transcriptomic features such as enrichment of club
226 and secretory (P2) or AT2 (P5) canonical markers thus potentially signifying distinct
227 cells-of-origin among the LUADs (**Supplementary Fig. S8B and E**).

228 Unlike P2, C9 cells in P3 and P5 were almost exclusively derived from the LUAD
229 tissues (**Fig. 2E**). In P3 LUAD, we identified large-scale chromosomal alterations (**Fig.**
230 **2G** left, **Supplementary Fig. S9A; Supplementary Table S6**), based on which
231 unsupervised clustering analysis revealed 4 clusters with differential CNV profiles.
232 Among them, three clusters (C2, C3, and C4) exhibited pronounced CNVs that were
233 indicative of malignant cell features (**Fig. 2G** left). We found an additional CNV event
234 (i.e., gain of 1p) unique to cells of cluster C4 but not C2 or C3, possibly signifying a late
235 event in the evolutionary trajectory of P3 LUAD. We also observed a branched
236 trajectory that started with cells of C2 and C3 and comprised few "normal cells" with
237 club/secretory lineage, and that later branched into cells of CNV cluster C4 (**Fig. 2H** top)
238 -- suggesting that C4 evolved from C2 and C3 and that P3 LUAD perhaps originated
239 from club/secretory cells. This is consistent with increased expression of club/secretory
240 canonical markers (e.g., *TFF3, HP, MUC4*) in P3 tumor clones with also high CNV
241 scores (C0 and C2; **Supplementary Fig. S8C**). P5 LUAD comprised 7 distinct CNV
242 clusters (**Fig. 2G** right, **Supplementary Fig. S9B**). C1 harbored cells with increased
243 CNV events, including events with higher amplitude, while C4 clustered the closest to
244 C3 which mostly comprised non-malignant cells (**Fig. 2G** right). Pseudotime analysis
245 revealed a C4-to-C1 path which, in contrast to P3, was unbranched suggesting that C4
246 and C1 in P5 likely comprised malignant cells from early and late developmental states,
247 respectively (**Fig. 2H** bottom). Overall, hierarchical clustering analysis was consistent
248 with phylogenetic reconstruction of tumor clonal architecture using inferred large-scale
249 CNVs (r = 0.7 in P3; **Supplementary Fig**. **S10**). Our single-cell interrogation of a large
250 number of epithelial cells from multi-region tissues identified diverse epithelial identities,
251 malignant trajectories, as well as high-resolution intratumor cell heterogeneity.
252

253 **Lymphoid reprogramming towards a protumor microenvironment**

254 We further characterized lymphoid spatial dynamics (**Fig. 1H-I**) and states (n = 53,882
255 cells, see Methods). Following clustering, we found 10 transcriptomically distinct

256 lymphoid cell types/states that were uniformly derived from all sequencing batches and
257 patients (**Fig. 3A, Supplementary Fig. S11A-B; Supplementary Table S8**). Lymphoid
258 clusters were overall spatially modulated by tumor proximity (**Fig. 3C**). Relative to
259 normal tissues, LUADs were heavily enriched with plasma *(SDC1+/MZB1+)*, B
260 (*CD19+/CD22+*), and T regulatory (Treg; *FOXP3+*) cells (**Fig. 3C, Supplementary Fig.**
261 **S11C**). With increasing tumor proximity, we noted a gradual decrease in NK cells
262 (*GNLY+*), innate lymphoid cells (ILCs), both *GZMA*-hi and *GNLY*-hi CD4+ cytotoxic T
263 lymphocytes (CTL; *CD40LG+, BATF+*), and *GNLY*-hi CD8+ CTLs, all of which were,
264 overall, depleted in the LUADs (**Fig. 3C, Supplementary Fig. S11C**).
265
266 We further performed subclustering analysis of CD8+ T cells which identified 3 robust
267 clusters: naïve, *GZMK*-hi, and *GNLY*-hi subpopulations with differentially expressed cell
268 state signatures (**Fig. 3D, Supplementary Fig. S12; Supplementary Tables S9 and**
269 **S10**)**.** Consistently, naïve CD8+ T cells showed high naïve and low cytotoxic T cell
270 scores and were composed of cells across all samples in the LUAD space (**Fig. 3D,**
271 **Supplementary Fig. S13A**). In contrast, *GNLY*-hi CD8+ CTLs exhibited high levels of
272 cytotoxicity genes (*TBX21, KLF3, FCGR3A, KLRG1*, *KLRB1*) and scores and were
273 depleted in the LUAD samples (**Fig. 3C-E**)**.** We observed a significant spatial pattern of
274 reduced cytotoxic activity in P3 and P4 (**Supplementary Fig. S13B** left and right,
275 respectively; $P < 0.05$)**.** Overall, CD8+ CTLs showed significant and spatially-modulated
276 reduction in cytotoxicity signature score ($P < 0.05$) and decreased expression of major
277 cytotoxic genes such as *NKG7* and *GNLY* ($P < 0.01$; **Fig. 3E**).
278
279 Spatial analysis of CD4+ T cell states (**Supplementary Tables S10 and S11**) showed
280 that LUAD tissues were specifically enriched with *FOXP3+* Tregs (**Fig. 3F-G;**
281 **Supplementary Table S12**) and the Treg signature scores were significantly and
282 spatially increased with proximity to all LUADs (**Supplementary Fig. S13C** left; $P <$
283 0.01) and in each of P3 and P5 (**Supplementary Fig. S13C** middle and right,
284 respectively; $P < 0.05$). The Tregs cells also expressed high levels of pro-tumor immune
285 checkpoints including *TIGIT, CTLA4, LAG3,* or *PDCD1* (**Fig. 3G**). The fraction of Tregs
286 co-expressing both *CTLA-4* and *TIGIT* immune checkpoints was progressively higher

287 along the spectrum of distant normal sites to more adjacent-to-tumor regions up to the
288 LUADs (**Fig. 3G** bottom). In contrast, we noted a reduction of cytotoxic CD4+ CTLs
289 characterized by high expression of *GZMA* ($P < 0.05$), or co-expression of *GZMA* and
290 *GZMH*, with increasing proximity to all LUADs (**Supplementary Fig. S13D-F**).
291
292 We further examined the spatial enrichment of LUADs with plasma and B cells (**Fig. 3C;**
293 **Supplementary Table S12**)**.** We found spatial changes in plasma cell isotype-
294 switching, such as increased *IGHA1/2* and decreased *IGHG1/3*, with increasing
295 proximity to P3 and P5 LUADs (**Supplementary Fig. S13 G-I; Supplementary Table**
296 **S13**). Smoker patients (P2 and P3) harbored strikingly high plasma cell fractions relative
297 to nonsmokers (P1 and P5) (**Supplementary Fig. S14A**). We confirmed the increased
298 plasma fractions in smoker LUADs following analysis of TCGA cohort ($P < 0.001$;
299 **Supplementary Fig**. **S14B**). We also identified 3 distinct subsets of B cell states
300 (**Supplementary Table S14**), including a LUAD-enriched subcluster (C0) with high
301 expression levels of *RAC2*+ and *ACTG*+ (**Supplementary Fig. S13J-L**), known to play
302 key roles in synapse formation in B cells (19). The B cell signature (C0) was
303 progressively increased across atypical adenomatous hyperplasias (AAH), the
304 preneoplastic precursors of LUAD, and invasive LUADs compared to matched normal
305 lung tissues (**Supplementary Fig. S13M**)**.** The B cell signature was also associated
306 with prolonged overall survival (OS) and progression free interval (PFI) in treatment-
307 naïve LUAD patients from TCGA ($P = 0.0001$, OS; $P = 0.02$, PFI; **Supplementary Fig**.
308 **S15A**) and in-house (MDACC; $P = 0.2$, OS; $P = 0.03$, PFI; **Supplementary Fig**. **S15B**)
309 cohorts. Our analyses identify spatial properties in lymphoid cell states that may
310 underlie pro-tumor immune remodeling in early-stage LUAD.
311
312 **Depletion of antigen presenting macrophages and inflammatory dendritic cells in**
313 **early-stage LUAD**
314 Spatial myeloid patterns (**Fig. 1H-I**) in LUAD space prompted us to further investigate
315 myeloid subsets and cellular states (**Fig. 4**). In total, 45,803 myeloid cells from all
316 sequencing batches and patients analyzed in this study (**Supplementary Fig. S16A-B**)
317 clustered into 13 distinct subsets: classical monocytes (*S100A8+, S100A9+*), non-

classical monocytes *(CDKN1C+*), mast cells *(MS4A2+)*, neutrophils (*IL1A+*), M2-like macrophages C1 (*TREM2+)*, M2-like macrophages C5 (*CD163+*), alveolar macrophages (*MARCO+)*, classical DC 1 (cDC1; *CLEC9A+*), cDC2 (*CLEC10A+*), plasmacytoid DC (pDC; *PLD4+*), other DCs (*CCL22+)* and proliferating myeloid cells *(TOP2A+)* (**Fig. 4A-B; Supplementary Tables S15 and S16**). Myeloid clusters were derived from tumor and all normal spatial samples with varying proportions (**Supplementary Fig. S16C**)**.** M2-like macrophages C5, monocytes (classical and non-classical), and mast cells were gradually depleted with increasing tumor proximity, whereas M2-like macrophages C1, proliferating myeloid subsets and cDC2 cells were steadily enriched in the tumors (**Fig. 4A and C, Supplementary Fig. S17A**)**.**

We next performed subclustering analysis of monocytes and macrophages (n = 27,664 cells) which identified 5 distinct subclusters and confirmed the unique enrichment of M2-like macrophages C1 in the LUAD tissues (**Fig. 4D, Supplementary Fig. S18**). Further, we found that C1 M2-like macrophages showed significantly diminished ($P < 0.001$) antigen presentation scores compared to C5 M2-like macrophages which were mainly enriched in normal samples (**Fig. 4D, Supplementary Fig. S17B; Supplementary Table S17**). We found markedly reduced expression levels of antigen presentation signature genes with increasing spatial proximity to the LUADs (**Fig. 4E-F, Supplementary Fig. S17C**). The spatial pattern of antigen presentation depletion was evident across M2-like macrophages combined from both clusters C1 and C5 and was statistically significant in 4 of the 5 LUAD patients (**Fig. 4G;** $P < 0.001$).

We examined gene and signature score differences between different subsets of DCs (n = 8,694). Spatial patterns were evident in cDC2 and pDC subsets (**Fig. 4H**). We observed differential expression of an inflammatory gene signature between the three cDC2 subclusters with C1, cells of which exhibited the lowest inflammatory scores, heavily enriched in the LUADs (**Fig. 4I-J; Supplementary Table S18**). Unsupervised subclustering of cDC2 cells identified three distinct cDC2 subsets characterized by differential expression of an inflammatory signature (highly enriched in C2) and MHC class II genes (enriched in C0/C1) previously shown to discriminate inflammatory from

349 non-inflammatory cDCs ((20); **Supplementary Fig. S17D**). Reduced expression of pro-
350 inflammatory genes and increased levels of anti-inflammatory features were evident
351 among cDC2 cells and along the continuum of normal-to-LUAD space (**Fig. 4K**). cDC2
352 subcluster with the highest inflammatory score (C2) was further characterized by an
353 inflammatory signature score that was spatially and significantly attenuated with
354 increasing tumor proximity ($P < 0.001$; **Fig. 4L**), and the fraction of C2 among cDC2
355 subset was overall under-represented in the LUADs ($P < 0.05$; **Fig. 4M**). Notably, the
356 inflammatory signature score significantly and progressively decreased along the
357 continuum from normal lung tissues, to matched premalignant AAHs and invasive
358 LUADs ($P < 0.01$; **Fig. 4N**) in sharp contrast to non-inflammatory DC expression
359 components (**Supplementary Fig. S17E**). We also studied pDC subsets and found
360 spatial enrichment of *FOS, FOSB,* and *JUN*, genes involved in regulating DC
361 immunogenicity (21), with increasing proximity to the tumors (**Supplementary Fig.
362 S17F-G**). We found that a relatively higher score of non-inflammatory to inflammatory
363 expression in cDC2 was associated with prolonged OS and PFI in TCGA LUAD cohort
364 ($P = 0.009$ and $P = 0.04$, respectively) with similar trends observed using an in-house
365 set (MDACC; $P = 0.2$ for OS; **Supplementary Fig. S19**). Altogether, these data
366 underscore spatial immune remodeling in early-stage LUAD that comprises loss of
367 antigen presentation by macrophage subsets and inflammatory phenotypes by DCs.
368
369 Stromal cells have been shown to impact diverse aspects of tumor immune
370 microenvironment and cancer progression (11,22). Intrigued by the enrichment of
371 stromal subsets in the LUADs (**Fig. 1H**), we further interrogated multiple stromal
372 populations in our dataset (**Supplementary Fig. S20A-C**), including tumor enrichment
373 of vascular smooth muscle cells, adventitial fibroblasts C5, and endothelial cell (EC)
374 venule clusters (**Supplementary Fig. S20D-E**). We pinpointed significantly differentially
375 expressed gene sets in the EC venule subpopulation and altered stromal signatures
376 (**Supplementary Fig. S20F**) that support the observed immune-related changes in the
377 LUAD space, and that are in line with previous observations (11). These comprised
378 tumor-specific activation of extracellular matrix reorganization, syndecan-2 pathway,

379 and neutrophil degranulation, as well as decreased JAK-STAT signaling and reduced
380 antigen-processing and cross-presentation (**Supplementary Fig. S20G-H**).
381
382 **Ligand-receptor cell-cell crosstalk in early-stage LUAD**
383 Crosstalk between tumor cells and elements in the TME is implicated in tumor
384 progression largely in part by mediating immunosuppressive phenotypes (23). We
385 utilized iTALK (24) to leverage signals from our scRNA-seq dataset and visualize
386 ligand-receptor (L-R)-mediated intercellular crosstalk (**Fig. 5A; Supplementary Table**
387 **S19**). Computational analysis and annotation were carried out using iTALK's built-in
388 database focusing on immune checkpoint-receptor pairs (n = 55) and cytokine-receptor
389 pairs (n = 327) (**Fig. 5A**). Overall, we found reduced overlap of L-R interactions between
390 the tumor and distant normal tissues compared to that between the tumor and more
391 proximal (adjacent, intermediate) regions (**Fig. 5B**). We identified altered cellular
392 interactions that were significantly and differentially modulated in LUADs versus their
393 respective spatial normal tissues (**Fig. 5C-F; Supplementary Fig. S21A and C**).
394 Among cytokine-receptor pairs, LUADs showed increased communication between
395 *CX3CL1+* tumor epithelial cells and DCs or macrophages expressing increased levels
396 of its cognate receptor *CX3CR1* (**Supplementary Fig. S21A and C; Supplementary**
397 **Table S20**). *CX3CR1* was increasingly expressed on macrophages and DCs but
398 decreased on CD8 T cells from LUADs (**Supplementary Fig. S21B and D**), in line with
399 previous reports demonstrating pro-tumor features for CX3CR1+ macrophages (25).
400
401 Notably, we identified increased interactions between immune checkpoint proteins
402 *CD24* or *LGALS9 (Galectin-9)* on tumor epithelial cells, and *SIGLEC10* on
403 macrophages or *HAVCR2 (TIM-3)* on DCs, respectively, and which were shared across
404 multiple patients (**Fig. 5C-F; Supplementary Table S20**). These interactions were
405 differentially enriched in tumors versus normal tissues at different distances from the
406 LUADs (**Fig. 5G-H; Supplementary Fig. S6A**). *CD24* and *LGALS9* expression levels
407 were overall increased in epithelial cells from LUAD tissues, particularly in cells of the
408 malignant-enriched cluster (**Fig. 5G-H, Supplementary Fig. S22A**). This pattern of
409 enrichment of *CD24* in cells of the malignant-enriched cluster was also observed in the

410 majority of patients analyzed (**Supplementary Fig. S22B**). Notably, *CD24* expression
411 was also markedly evident in B cells (**Supplementary Fig. S22A-B**) in line with
412 previous reports (26). These findings demonstrate that the early-stage LUAD ecosystem
413 harbors cell-cell communication that confers increased pro-tumor inflammatory and
414 immunosuppressive states.
415
416 Our findings above prompted us to validate *CD24* expression patterns using external
417 cohorts. We found progressively and markedly increased expression of the antigen
418 across normal lung tissues, AAHs, and LUADs (**Fig. 6A**; $P < 0.05$). *CD24* positively
419 correlated with expression of the epithelial marker *EPCAM* as well as with levels of pro-
420 tumor and immune suppressive features (*TIGIT, CTLA4, FOXP3, CCL19*)**,** in contrast to
421 negatively correlating with anti-tumor immune markers (*GZMB, GZMH, PRF1*) (**Fig. 6B,**
422 **Supplementary Fig. S23A**)**.** These observations were, for the most part, further
423 validated in LUADs and matched normal tissues from TCGA cohort (27) (**Fig. 6C-D,**
424 **Supplementary Fig. S23B**). Using an in-house cohort of early-stage LUADs analyzed
425 by targeted immune profiling (see Supplementary Methods), we found that relatively
426 higher *CD24* was associated with shortened OS and PFI (**Fig. 6E**; $P = 0.07$). Consistent
427 with the above, *CD24* expression positively correlated with that of *EPCAM* ($r = 0.31$)
428 and negatively correlated with both *PRF1* ($r = -0.37$) and immune cytotoxicity score
429 ((28), $r = -0.33$; **Fig. 6F**). Further, analysis of an independent in-house tissue microarray
430 of treatment-naïve LUADs revealed that CD24 immunohistochemical protein expression
431 was prevalent in tumor cells (**Fig**. **6G** left) and, in line with the above, was also
432 associated with reduced OS ($P = 0.04$) and PFI ($P = 0.007$) (**Fig**. **6G** right). We further
433 found in a syngeneic mouse model that loss of *Cd24a* by CRISPR mediated knockout
434 or its inhibition using neutralizing antibodies significantly reduced the growth of mouse
435 LUAD cells *in vivo* (Supplementary Methods**, Supplementary Fig. S24, Fig. 6H**). These
436 findings show that CD24, which we found to be at the core of an enriched cell-cell
437 interactome in early-stage LUAD, is associated with a pro-tumor immune contexture
438 and poor prognosis as well as promotes LUAD growth *in vivo*.
439
440

441 **DISCUSSION**

442

443 Molecular changes have been documented in the local niche of LUAD including loss-of-
444 heterozygosity in 3p and 9p, point mutations and tumor suppressor methylation (29,30).
445 Earlier work also underscored transcriptome profiles, somatic driver variants, as well as
446 genome-wide allelic imbalance events that are shared between lung cancer and
447 adjacent normal-appearing airway cells but that are absent in distant normal cells,
448 thereby pointing to putative drivers of lung oncogenesis (4,5). These earlier studies
449 have focused on understanding geospatial profiles by bulk profiling methods, thereby
450 inadvertently obscuring the individual contributions of epithelial and TME cues to lung
451 cancer pathogenesis. Our knowledge of the geospatial architecture of individual cell
452 populations in early-stage LUAD evolution remains poorly understood. By single-cell
453 interrogation of a unique multi-region sampling design with epithelial cell enrichment, we
454 here characterized spatial and ecological maps comprising various epithelial and non-
455 epithelial subsets that underlie emergence of early-stage LUAD from its local niche.

456

457 Multi-regional or spatial analyses have been employed to interrogate intratumor
458 heterogeneity (ITH) in solid tumors, including LUADs, in order to understand
459 evolutionary trajectories and therapy response (31). Our analyses showed that ITH is
460 evident at both the tumor epithelial cell and intra-site levels, i.e., within the same tumor
461 region. We applied an integrative approach to dissect ITH of likely malignant cells and
462 characterized cell clusters with differential transcriptomic profiles, evolutionary
463 trajectories, CNV burdens, and/or driver mutations. We also found “normal” cells in the
464 LUAD tissues themselves that are close in the inferred trajectory paths to specific
465 malignant-enriched subsets, possibly representing tumor cells-of-origin. Normal basal
466 cells (expressed *KRT5, KRT15, KRT17,* and *P63*) were consistently evident in the
467 LUAD samples, possibly corroborating previously reported basal differentiation
468 hierarchies in the normal lung (32,33). It is noteworthy that we found, in the normal-
469 appearing samples, cells with features of malignant-enriched subsets and
470 heterogeneous CNV profiles. Whether these cells comprise early LUAD precursors,

471 mutagenic clones that do not progress to malignancy, or putative molecular field
472 cancerization remains to be investigated (34).
473
474 LUADs exhibit remarkable inter-patient heterogeneity in histological differentiation
475 patterns, driver alterations and inferred tumor cells-of-origin (35). In our cohort, we
476 identified malignant gene expression features that varied between the patients and that
477 pointed to likely distinct tumor cells-of-origin and/or LUAD histopathological patterns.
478 We pinpointed G12D mutations in *KRAS* to a unique subset of the malignant-enriched
479 cluster within P2 LUAD. These cells had overall low CNVs, perhaps reminiscent of
480 findings in LUADs driven by strong driver genes *in vivo* (36). These cells exhibited
481 increased expression of genes associated with *KRAS*-mutant cancer such as *LCN2* (a
482 marker of inflammation, (17)), and reduced expression of the lineage-specific oncogene
483 *NKX2-1* (18), in line with expression patterns in mucinous *KRAS*-mutant LUADs (e.g.,
484 P2). Compared to other C9 (malignant-enriched cluster) cells from the same LUAD
485 (P2), *KRAS*-mutant cells showed reduced expression of airway lineage-specific genes
486 (e.g., *SCGB3A1*, *SFTPB*) in accordance with recent reports (8). Our findings also allude
487 to the possibility that tumor cell lineage plasticity may occur at an early-stage of *KRAS*-
488 mutant LUAD carcinogenesis – a supposition that warrants exploration of a larger and
489 more diverse repertoire of *KRAS*-mutant cells. Of note, we could not characterize the
490 full spectrum of ITH, clonality, and inter-patient heterogeneity given the limited number
491 of patients profiled and the use of scRNA-seq which is not ideal for examining tumor
492 clonality or genomic alterations. Nonetheless, our in-depth analysis of a relatively large
493 number of epithelial cells unveiled different characteristics (airway lineages, malignant
494 programs, potential tumor cells-of-origin, and cellular ITH) of the epithelial architecture
495 of early-stage LUAD. These insights could be further extended to profile a larger and
496 more diverse array of LUADs using anticipated advances enabling simultaneous
497 scRNA-seq and scDNA-seq of the same cell.
498
499 Earlier studies have shown that immunosuppressive T regulatory cells are crucial for
500 immune evasion in lung cancer (37,38). We found Tregs co-expressing both *TIGIT* and
501 *CTLA-4* immune checkpoints and that were progressively enriched with increasing

502 geospatial proximity to the LUADs – suggesting a value in combinatorial targeting of
503 multiple checkpoints for immunotherapy of early-stage LUAD (39). scRNA-seq analysis
504 of our limited patient cohort also pinpointed important attributes to other lymphoid
505 populations. We identified B cell signatures that are spatially enriched in the LUADs,
506 progressively increased along the course of normal to preneoplasia and invasive LUAD
507 and associated with prolonged survival. These data suggest important yet unexplored
508 roles for B cell phenotypes in immune evolution of LUAD (40). We also noted strikingly
509 high plasma cell fractions among LUADs from smokers, suggesting that plasma cells
510 may play important roles in the pathogenesis of smoking-associated LUAD and its
511 microenvironment (41). Our analysis also pointed to mechanisms by which the myeloid
512 immune microenvironment permits LUAD pathogenesis. Tumor-specific M2-like
513 macrophages displayed diminished antigen presentation, whereby expression levels of
514 MHC genes as well as genes involved in peptide transport and loading (*TAP1, TAP2,*
515 *TAPBP;* (42)) were markedly reduced. We also found DC subclusters and signatures
516 that were recently reported in healthy donors and in cancer (20,43). These included a
517 shift from an inflammatory to a non-inflammatory DC state across normal lung to
518 premalignant AAH up to LUAD and that was associated with prolonged survival in
519 patients -- thus pointing to immune cues that could perhaps be harnessed to manipulate
520 the immunogenicity of tumors (43). We also identified an inflammatory tumor-depleted
521 cDC cluster with increased expression of scavengers such as *CD36*. Dendritic CD36
522 permits acquisition and presentation of cell surface antigens (44), and the precise
523 effects of its stark absence in a LUAD-specific cDC subset the tumor immune
524 microenvironment warrant further investigation. Nevertheless, these findings suggest
525 that unique cDC subsets play critical roles in LUAD pathogenesis and, thus, could be
526 potential targets for immune-based interception.
527
528 Deciphering ligand-receptor mediated interactions can elucidate cell-to-cell
529 communication implicated during carcinogenesis, tumor-immune co-evolution and
530 immune reprogramming, and thus, may help identify potential immune therapeutic
531 targets (23). In this study, we applied the iTALK tool (24) developed by our group and
532 performed a deep analysis of cellular interaction networks. We identified significant

533 immune checkpoint- (e.g., *CD24*, *Galectin-9*, or *TIM-3)* and cytokine- (e.g., *CX3CL1)*
534 receptor interactions whose enrichment or depletion in the LUAD space signified a
535 highly pro-tumorigenic milieu. These findings are in accordance with the CD24–Siglec-
536 10 interaction and subsequent "do not eat me" signal recently highlighted in breast
537 cancer (26). Additionally, our findings on CD24 including prominent expression in tumor
538 epithelium, association with pro-tumor immune phenotypes and reduced survival, and
539 functional role in LUAD *in vivo*, suggest that CD24 may be a viable target for treatment
540 of early-stage LUAD. Overall, our interactome analyses implicate potential targets as
541 culprits in LUAD pathogenesis.
542
543 In summary, our results provide a spatial atlas of early-stage LUAD and its nearby and
544 distant lung ecosystem. This atlas comprises high cellular heterogeneity as well as
545 spatial dynamics in cell populations, cell states and their transcriptomic features that
546 underlie LUAD evolution from the peripheral lung ecosystem. Our extensive
547 transcriptomic dataset of lung epithelial and immune cells, other populations such as
548 stromal and endothelial subsets, as well as of tumor-pertinent cell-cell interactions,
549 constitutes a valuable resource to functionally interrogate LUAD trajectories at high
550 resolution and generate strategies for its early treatment. Also, our study's multi-region
551 sampling design in conjunction with single-cell analysis could help address specific
552 questions in early malignant and immune biology of other solid tumors.
553
554
555 **METHODS**
556 *Additional description of methods can be found in the Supplementary Data file.*
557
558 **Multi-region sampling of surgically resected LUADs and normal lung tissues**
559 Patients undergoing surgical resection for primary early-stage lung adenocarcinoma (I-
560 IIIA) were carefully selected for derivation of multi-region samples for single cell analysis
561 (**Supplementary Table S1**). All patients were evaluated at the University of Texas MD
562 Anderson Cancer Center and had provided informed consents under approved
563 institutional review board protocols. Immediately following surgery, resected tissues

564 were processed by an experienced pathologist assistant (PB). One side of the
565 specimen was documented and measured, followed by tumor margin identification.
566 Based on the placement of the tumor within the specimen, incisions were made at
567 defined collection sites in one direction along the length of the specimen and spanning
568 the entire lobe: tumor-adjacent and -distant normal parenchyma at 0.5 cm from the
569 tumor edge and from the periphery of the overall specimen/lobe, respectively. An
570 additional tumor-intermediate normal tissue was selected for P2-5 that ranged between
571 3-5 cm from the edge of the tumor. Sample collection was initiated at normal lung
572 tissues that are farthest from the tumor moving inwards towards the tumor to minimize
573 cross-contamination during collection.
574
575 **scRNA-seq analysis**
576 Tumor and spatial normal parenchyma tissues (n = 19 samples) were immediately
577 transported on ice for mincing and enzymatic digestion (Supplementary Methods). Cells
578 were sorted (by FACS) for viable singlets cells (and also EPCAM+/- fractions from P2-
579 P5), followed by processing for scRNA-seq library construction using 10X Genomics
580 and sequencing using Novaseq6000 platforms. Single-cell analyses were performed
581 using available computational framework. Raw scRNA-seq data were pre-processed,
582 demultiplexed, aligned to human GRCh38 reference and feature-barcodes generated
583 using CellRanger (10X Genomics, version 3.0.2). Details of quality control including
584 quality check, data filtering, as well as identification and removal of cellular debris,
585 doublets and multiplets are found in Supplementary Methods. Following quality filtering,
586 a total of 186,916 cells were retained for downstream analysis. Raw unique molecular
587 identifier (UMI) counts were log normalized and used for principal component analysis
588 using Seurat (45). Batch effects were statistically assessed using k-BET (46) and
589 corrected by Harmony (47) (Supplementary Methods), followed by unsupervised
590 clustering analysis using Seurat v3 (45). Uniform Manifold Approximation and Projection
591 (UMAP) (48) was used for visualization of clusters. Clustering robustness was
592 examined for major clusters and subclusters (see Supplementary Methods), and all
593 subclustering analyses were performed separately for each compartment. For
594 unsupervised clustering analysis, we applied single-cell consensus clustering (SC3;

595 (49)) approach using default parameters and independently of cell lineage annotation.
596 Differentially expressed genes (DEGs) for each cell cluster were identified using the
597 *FindAllMarkers* function in Seurat R package. We defined cell types and cluster
598 functional states by integrating the enrichment of canonical marker genes, top-ranked
599 DEGs in each cell cluster, and the global cluster distribution.
600 To study hierarchical relationships among cell types, pairwise Spearman correlations
601 were calculated from average expression levels (Seurat function *AverageExpression*),
602 based on which Euclidean distances were calculated. Monocle 2 (version 2.10.1; (50))
603 was applied to construct trajectories. Likely malignant cells were distinguished from
604 non-malignant subsets based on information integrated from multiple sources including
605 cluster distribution of the cells, gene expression, as well as presence of mutations
606 (*KRAS* in P2, see Supplementary Methods) and CNVs. CNVs were inferred using
607 inferCNV tool (13) with NK cells as the control. To evaluate the robustness of CNV
608 inference, T and B cells were also used as controls. To identify significant ligand-
609 receptor pairs among major cell lineages, the top 30% of most highly expressed genes
610 were included in the analysis. Significant cellular interactions were identified using
611 iTALK as previously described (24). For ligand-receptor annotation, the iTALK built-in
612 ligand-receptor database was used (https://github.com/Coolgenome/iTALK).
613
614 **Statistical analysis**
615 All statistical analyses were performed using R package v3.6.0. Pseudo-bulk gene
616 expression values for defined cell clusters were calculated by taking mean expression
617 of each gene across all cells in a specific cluster. Pearson's correlation analysis was
618 used to identify genes significantly correlated with *CD24* expression. Log-rank tests and
619 Kaplan-Meier plots were used for survival analysis. All statistical significance testing
620 was two-sided, and results were considered statistically significant at $P < 0.05$. The
621 Benjamini-Hochberg method was applied to control the false discovery rate (FDR) in
622 multiple comparisons and to calculate adjusted P – value (q-values).
623
624
625

626 **Data Availability**

627 All sequencing data generated in this study are deposited in the European Genome-
628 phenome Archive (EGAS00001005021).

629 **REFERENCES**

630

631 1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA: a cancer journal for
632 clinicians **2018**;68(1):7-30 doi 10.3322/caac.21442.
633 2. Goldstraw P, Ball D, Jett JR, Le Chevalier T, Lim E, Nicholson AG*, et al.* Non-
634 small-cell lung cancer. Lancet (London, England) **2011**;378(9804):1727-40 doi
635 10.1016/s0140-6736(10)62101-0.
636 3. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM*, et al.*
637 Reduced lung-cancer mortality with low-dose computed tomographic screening.
638 The New England journal of medicine **2011**;365(5):395-409 doi
639 10.1056/NEJMoa1102873.
640 4. Abbosh C, Venkatesan S, Janes SM, Fitzgerald RC, Swanton C. Evolutionary
641 dynamics in pre-invasive neoplasia. Current opinion in systems biology **2017**;2:1-
642 8 doi 10.1016/j.coisb.2017.02.009.
643 5. Saab S, Zalzale H, Rahal Z, Khalifeh Y, Sinjab A, Kadara H. Insights Into Lung
644 Cancer Immune-Based Biology, Prevention, and Treatment. Front Immunol
645 **2020**;11:159 doi 10.3389/fimmu.2020.00159.
646 6. Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL*, et al.* Therapy-
647 Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA
648 Sequencing. Cell **2020** doi https://doi.org/10.1016/j.cell.2020.07.017.
649 7. Lavin Y, Kobayashi S, Leader A, Amir ED, Elefant N, Bigenwald C*, et al.* Innate
650 Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell
651 Analyses. Cell **2017**;169(4):750-65.e17 doi 10.1016/j.cell.2017.04.014.
652 8. Laughney AM, Hu J, Campbell NR, Bakhoum SF, Setty M, Lavallée V-P*, et al.*
653 Regenerative lineages and immune-mediated pruning in lung cancer metastasis.
654 Nature Medicine **2020**;26(2):259-69 doi 10.1038/s41591-019-0750-6.
655 9. Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW*, et al.* Single-cell RNA
656 sequencing demonstrates the molecular and cellular reprogramming of
657 metastatic lung adenocarcinoma. Nature Communications **2020**;11(1):2285 doi
658 10.1038/s41467-020-16164-1.

659 10. Plasschaert LW, Žilionis R, Choo-Wing R, Savova V, Knehr J, Roma G*, et al.* A
660 single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary
661 ionocyte. Nature **2018**;560(7718):377-81 doi 10.1038/s41586-018-0394-6.
662 11. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O*, et al.*
663 Phenotype molding of stromal cells in the lung tumor microenvironment. Nat Med
664 **2018**;24(8):1277-89 doi 10.1038/s41591-018-0096-5.
665 12. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH*, et al.*
666 Reconstructing lineage hierarchies of the distal lung epithelium using single-cell
667 RNA-seq. Nature **2014**;509(7500):371-5 doi 10.1038/nature13173.
668 13. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K*, et al.*
669 Single-cell RNA-seq supports a developmental hierarchy in human
670 oligodendroglioma. Nature **2016**;539(7628):309-13 doi 10.1038/nature20123.
671 14. Little DR, Gerner-Mauro KN, Flodby P, Crandall ED, Borok Z, Akiyama H*, et al.*
672 Transcriptional control of lung alveolar type 1 cell development and maintenance
673 by NK homeobox 2-1. Proc Natl Acad Sci U S A **2019**;116(41):20545-55 doi
674 10.1073/pnas.1906663116.
675 15. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in
676 lung development, renewal and cancer. Nature **2014**;507(7491):190-4 doi
677 10.1038/nature12930.
678 16. Finn J, Sottoriva K, Pajcini KV, Kitajewski JK, Chen C, Zhang W*, et al.* Dlk1-
679 Mediated Temporal Regulation of Notch Signaling Is Required for Differentiation
680 of Alveolar Type II to Type I Cells during Repair. Cell Rep **2019**;26(11):2942-
681 54.e5 doi 10.1016/j.celrep.2019.02.046.
682 17. Treekitkarnmongkol W, Hassane M, Sinjab A, Chang K, Hara K, Rahal Z*, et al.*
683 Augmented Lipocalin-2 is Associated with COPD and Counteracts Lung
684 Adenocarcinoma Development. Am J Respir Crit Care Med **2020** doi
685 10.1164/rccm.202004-1079OC.
686 18. Snyder EL, Watanabe H, Magendantz M, Hoersch S, Chen TA, Wang DG*, et al.*
687 Nkx2-1 represses a latent gastric differentiation program in lung
688 adenocarcinoma. Mol Cell **2013**;50(2):185-99 doi 10.1016/j.molcel.2013.02.018.

689 19. Arana E, Vehlow A, Harwood NE, Vigorito E, Henderson R, Turner M*, et al.*
690 Activation of the small GTPase Rac2 via the B cell receptor regulates B cell
691 adhesion and immunological-synapse formation. Immunity **2008**;28(1):88-99 doi
692 10.1016/j.immuni.2007.12.003.
693 20. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J*, et al.*
694 Single-cell RNA-seq reveals new types of human blood dendritic cells,
695 monocytes, and progenitors. Science (New York, NY) **2017**;356(6335) doi
696 10.1126/science.aah4573.
697 21. Mann J, Oakley F, Johnson PW, Mann DA. CD40 induces interleukin-6 gene
698 transcription in dendritic cells: regulation by TRAF2, AP-1, NF-kappa B, AND
699 CBF1. The Journal of biological chemistry **2002**;277(19):17125-38 doi
700 10.1074/jbc.M109250200.
701 22. Bremnes RM, Dønnem T, Al-Saad S, Al-Shibli K, Andersen S, Sirera R*, et al.*
702 The role of tumor stroma in cancer progression and prognosis: emphasis on
703 carcinoma-associated fibroblasts and non-small cell lung cancer. J Thorac Oncol
704 **2011**;6(1):209-17 doi 10.1097/JTO.0b013e3181f8a1bd.
705 23. Maman S, Witz IP. A history of exploring cancer in context. Nat Rev Cancer
706 **2018**;18(6):359-76 doi 10.1038/s41568-018-0006-7.
707 24. Wang Y, Wang R, Zhang S, Song S, Jiang C, Han G*, et al.* iTALK: an R Package
708 to Characterize and Illustrate Intercellular Communication. bioRxiv **2019**:507871
709 doi 10.1101/507871.
710 25. Schmall A, Al-Tamari HM, Herold S, Kampschulte M, Weigert A, Wietelmann A*,*
711 *et al.* Macrophage and cancer cell cross-talk via CCR2 and CX3CR1 is a
712 fundamental mechanism driving lung cancer. Am J Respir Crit Care Med
713 **2015**;191(4):437-47 doi 10.1164/rccm.201406-1137OC.
714 26. Barkal AA, Brewer RE, Markovic M, Kowarsky M, Barkal SA, Zaro BW*, et al.*
715 CD24 signalling through macrophage Siglec-10 is a target for cancer
716 immunotherapy. Nature **2019**;572(7769):392-6 doi 10.1038/s41586-019-1456-0.
717 27. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung
718 adenocarcinoma. Nature **2014**;511(7511):543-50 doi 10.1038/nature13385.

719 28. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic
720 properties of tumors associated with local immune cytolytic activity. Cell
721 **2015**;160(1-2):48-61 doi 10.1016/j.cell.2014.12.033.
722 29. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA*, et*
723 *al.* Aberrant promoter methylation in bronchial epithelium and sputum from
724 current and former smokers. Cancer Res **2002**;62(8):2370-7.
725 30. Powell CA, Klares S, O'Connor G, Brody JS. Loss of heterozygosity in epithelial
726 cells obtained by bronchial brushing: clinical utility in lung cancer. Clin Cancer
727 Res **1999**;5(8):2025-34.
728 31. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J*, et al.* Intratumor
729 heterogeneity in localized lung adenocarcinomas delineated by multiregion
730 sequencing. Science **2014**;346(6206):256-9 doi 10.1126/science.1256930.
731 32. Morrisey EE. Basal Cells in Lung Development and Repair. Dev Cell
732 **2018**;44(6):653-4 doi 10.1016/j.devcel.2018.03.004.
733 33. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, Birket SE*, et al.* A revised
734 airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature
735 **2018**;560(7718):319-24 doi 10.1038/s41586-018-0393-7.
736 34. Sinjab A, Han G, Wang L, Kadara H. Field Carcinogenesis in Cancer Evolution:
737 What the Cell Is Going On? Cancer Research **2020**;80(22):4888 doi
738 10.1158/0008-5472.CAN-20-1956.
739 35. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong K-K. Non-small-cell lung
740 cancers: a heterogeneous set of diseases. Nature reviews Cancer
741 **2014**;14(8):535-46 doi 10.1038/nrc3775.
742 36. McFadden DG, Politi K, Bhutkar A, Chen FK, Song X, Pirun M*, et al.* Mutational
743 landscape of EGFR-, MYC-, and Kras-driven genetically engineered mouse
744 models of lung adenocarcinoma. Proceedings of the National Academy of
745 Sciences of the United States of America **2016**;113(42):E6409-E17 doi
746 10.1073/pnas.1613601113.
747 37. Quezada SA, Peggs KS, Simpson TR, Allison JP. Shifting the equilibrium in
748 cancer immunoediting: from tumor tolerance to eradication. Immunol Rev
749 **2011**;241(1):104-18 doi 10.1111/j.1600-065X.2011.01007.x.

750 38. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q*, et al.* Global
751 characterization of T cells in non-small-cell lung cancer by single-cell
752 sequencing. Nat Med **2018**;24(7):978-85 doi 10.1038/s41591-018-0045-3.
753 39. Cascone T, William WN, Weissferdt A, Lin HY, Leung CH, Carter BW*, et al.*
754 Neoadjuvant nivolumab (N) or nivolumab plus ipilimumab (NI) for resectable non-
755 small cell lung cancer (NSCLC): Clinical and correlative results from the
756 NEOSTAR study. Journal of Clinical Oncology **2019**;37(15_suppl):8504- doi
757 10.1200/JCO.2019.37.15_suppl.8504.
758 40. Wang S-s, Liu W, Ly D, Xu H, Qu L, Zhang L. Tumor-infiltrating B cells: their role
759 and application in anti-tumor immunity in lung cancer. Cellular & Molecular
760 Immunology **2019**;16(1):6-18 doi 10.1038/s41423-018-0027-x.
761 41. Alisoltani A, Qiu X, Jaroszewski L, Sedova M, Li Z, Godzik A. Large scale
762 analysis of smoking-induced changes in the tumor immune microenvironment.
763 bioRxiv **2020**:2020.03.06.981225 doi 10.1101/2020.03.06.981225.
764 42. Seliger B, Maeurer MJ, Ferrone S. TAP off--tumors on. Immunol Today
765 **1997**;18(6):292-9 doi 10.1016/s0167-5699(97)01052-9.
766 43. Wculek SK, Cueto FJ, Mujal AM, Melero I, Krummel MF, Sancho D. Dendritic
767 cells in cancer immunology and immunotherapy. Nature Reviews Immunology
768 **2020**;20(1):7-24 doi 10.1038/s41577-019-0210-z.
769 44. Perry JSA, Russler-Germain EV, Zhou YW, Purtha W, Cooper ML, Choi J*, et al.*
770 Transfer of Cell-Surface Antigens by Scavenger Receptor CD36 Promotes
771 Thymic Regulatory T Cell Receptor Repertoire Development and Allo-tolerance.
772 Immunity **2018**;48(5):923-36.e4 doi 10.1016/j.immuni.2018.04.007.
773 45. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of
774 single-cell gene expression data. Nat Biotechnol **2015**;33(5):495-502 doi
775 10.1038/nbt.3192.
776 46. Büttner M, Miao Z, Wolf FA, Teichmann SA, Theis FJ. A test metric for assessing
777 single-cell RNA-seq batch correction. Nature methods **2019**;16(1):43-9 doi
778 10.1038/s41592-018-0254-1.

779 47. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K*, et al.* Fast,
780 sensitive and accurate integration of single-cell data with Harmony. Nature
781 Methods **2019**;16(12):1289-96 doi 10.1038/s41592-019-0619-0.
782 48. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG*, et al.*
783 Dimensionality reduction for visualizing single-cell data using UMAP. Nat
784 Biotechnol **2018** doi 10.1038/nbt.4314.
785 49. Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T*, et al.* SC3:
786 consensus clustering of single-cell RNA-seq data. Nature Methods
787 **2017**;14(5):483-6 doi 10.1038/nmeth.4236.
788 50. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ*, et al.* The single-cell
789 transcriptional landscape of mammalian organogenesis. Nature
790 **2019**;566(7745):496-502 doi 10.1038/s41586-019-0969-x.

791

792 **FIGURE LEGENDS**

793

794 **Figure 1. Dissecting early-stage LUAD and the peripheral lung ecosystem by**
795 **multi-region single-cell RNA sequencing. A**, Workflow showing multi-region sampling
796 strategy of 5 LUADs and 14 spatially defined normal lung tissues for analysis by
797 scRNA-seq. Dis, distant normal; Int, intermediate normal; Adj, adjacent normal; LUAD,
798 tumor tissue. **B**, Uniform manifold approximation and projection (UMAP) embedding of
799 cells from tumor, adjacent normal and distant normal samples of patient one (P1). Cells
800 are colored by their inferred cell types. **C**, Cell composition in absolute cell numbers
801 (stacked bar plots) and relative fractions (pie charts) in each spatial sample derived
802 from P1. **D**, UMAP view of cells from all 5 patients, including EPCAM+ and EPCAM-
803 pre-enriched cells from P2-P5. Colors represent assigned major cell types. Proliferating,
804 proliferating cells. **E**, UMAP view of cell types and their fractions (stacked bar plot) by
805 spatial samples. Clusters in embedded squares represent proliferating cells. Colors
806 represent assigned cell types as in panel D. **F**, Dendrograms showing hierarchical
807 relationships of cells among the spatial samples based on the computed Euclidean
808 distance using transcriptomic features. Dendrograms are shown for 5 major cell types
809 (from left to right), for all patients together (top) and by patient (bottom). **G**, Same UMAP
810 as in panel D, with further subclustering of lymphoid and myeloid cells. Colors
811 correspond to the cell type annotation in panel H for EPCAM- cells. **H-I**, Line plot
812 showing changes in the relative fractions among the EPCAM negative subsets across
813 spatial samples for all patients together (H) and by patient (pie charts, I). Stacked bar
814 plots in panel I show absolute cell numbers of the fractions by patient and spatial
815 sample**.**

816

817 **Figure 2. Epithelial lineage diversity and intratumoral heterogeneity in the spatial**
818 **ecosystem of early-stage LUAD. A**, UMAP visualization of all EPCAM+ cells from P1-
819 P5 colored (from left to right) by their assigned cell types, spatial samples, and inferred
820 copy number variation (CNV) scores. **B**, Heatmap of major lineage marker genes for
821 EPCAM+ cell clusters (C1-10 as shown in panel A, left), with corresponding bar blots
822 outlining fraction by spatial sample. **C**, Area plot showing changes in EPCAM+ subset

823 fractions across spatial samples. **D**, Hierarchical relationships of 3 representative
824 subsets of epithelial cells (from top to bottom) among the spatial samples based on the
825 computed Euclidean distance using transcriptomic features (left), and corresponding
826 heatmaps quantifying similarity levels among spatial samples (right). Similarity score is
827 defined as one minus the Euclidean distance. **E**, UMAP plots of cells in the malignant-
828 enriched cluster C9 (panel A), colored by their corresponding patient origin (left), spatial
829 sample (middle), and CNV score (right). The zoom in view of the right panel shows
830 *KRAS* G12D mutant cells in P2. **F**, Fraction of cells carrying *KRAS* G12D mutation (left
831 bar plot), with numbers indicating the absolute cell numbers, as well as expression
832 levels of *KRAS* (violin plot, top right) and *MUC5AC* (violin plot, bottom right), within cells
833 of each epithelial lineage cluster of P2. **G**, Unsupervised clustering of CNV profiles
834 inferred from scRNA-seq data from patient P3 (left) and P5 (right) tumor samples using
835 NK cells as control and demonstrating intratumoral heterogeneity in CNV profiles.
836 Chromosomal amplifications (red) and deletions (blue) are inferred for all 22
837 chromosomes (color bars on the top). Each row represents a single cell, with
838 corresponding cell type annotated on the right (same as in panel A). **H**, Potential
839 developmental trajectories for EPCAM+ cells from P3 (top) and P5 (bottom) inferred by
840 Monocle 3 analysis. Cells on the tree are colored by pseudotime (dotted boxes) and
841 CNV clusters.
842
843 **Figure 3. Spatial reprogramming of lymphoid subsets towards protumor**
844 **phenotypes in early-stage LUAD. A**, UMAP visualization of lymphoid cell subsets
845 from P1-P5 colored by cell lineage (left) and spatial sample (right). CTL, cytotoxic T
846 lymphocyte; Treg, T regulatory cell; ILC, innate lymphoid cell; NK, natural killer cell. **B**,
847 Bubble plot showing the expression of lineage markers. Both the fraction of cells
848 expressing signature genes (indicated by the size of the circle) as well as their scaled
849 expression levels (indicated by the color of the circle) are shown. **C**, Changes in the
850 abundance of lymphoid cell lineages and cellular states across the LUADs and spatial
851 normal samples. Embedded pie charts show the contribution of each spatial sample to
852 the indicated cell subtype/state. **D**, UMAP plots of CD8+ T lymphocytes colored by cell
853 states (top), spatial sample (middle), and cytotoxic score (bottom). The heatmap on the

854 right shows normalized expression of marker genes for defined CD8 T cell subsets.
855 Each column represents a cell. Top annotation tracks indicate (from top to bottom) cell
856 states, naïve T cell scores and cytotoxic scores calculated using curated gene
857 signatures, and the corresponding spatial sample of each cell. **E**, Depletion of CD8+
858 GNLY-hi CTLs in the tumor microenvironment (TME) of LUADs. Bar plot (top left) and
859 boxplot (top middle) showing percentage of CD8+ GNLY-hi CTLs among total CD8+
860 cells from all patients across the spatial sample. Each circle in the boxplot represents a
861 patient sample. *P* – value was calculated using Kruskal-Wallis test. Cytotoxicity
862 signature score (violin plot, top right) of CD8+ CTLs across spatial samples (*, $P < 0.05$;
863 ***, $P < 0.001$). *P* – values were calculated using Wilcoxon rank sum test. The
864 percentage of CD8+ CTLs expressing cytotoxic signature genes (indicated by the size
865 of the circle) and their scaled expression levels (indicated by the color of the circle)
866 across the LUADs and spatial normal lung samples (bubble plot, bottom left).
867 Expression levels of *NKG7* and *GNLY* in CD8+ CTLs across the spatial samples (violin
868 plots, bottom right; **, $P < 0.01$; ***, $P < 0.001$). *P* – values were calculated using
869 Wilcoxon rank sum test. **F**, UMAP plots of CD4+ T lymphocytes colored by cell states
870 (top), spatial sample (middle), and Treg signature score (bottom). The heatmap on the
871 right shows normalized expression of marker genes for CD4+ T cells grouped by
872 defined subcluster. Each column represents a cell. Top annotation tracks indicate (from
873 top to bottom) cell states, Treg signature score, cytotoxic scores, and naïve T cell score
874 calculated using curated gene signatures, and the corresponding spatial sample of each
875 cell. **G**, Enrichment of CD4+ T regulatory cells (Treg) in the TME of LUADs. Bar plot (top
876 left) and boxplot (top middle) showing percentage of CD4+ Tregs among total CD4+
877 cells from all patients across the spatial samples. Each circle in the boxplot represents a
878 patient sample. *P* – value was calculated using Kruskal-Wallis test. Percentage of CD4+
879 Tregs expressing inhibitory immune checkpoint genes (indicated by the size of the
880 circle) and their scaled expression levels (indicated by the color of the circle, color
881 assignment same as panel E) across the spatial samples (bubble plot, top right).
882 Frequency of CD4+ Treg cells co-expressing *CTLA4* and *TIGIT* immune checkpoints
883 across the spatial samples (scatter plots, bottom). The fractions of *CTLA4*+*TIGIT*+
884 Tregs are labeled on each plot.

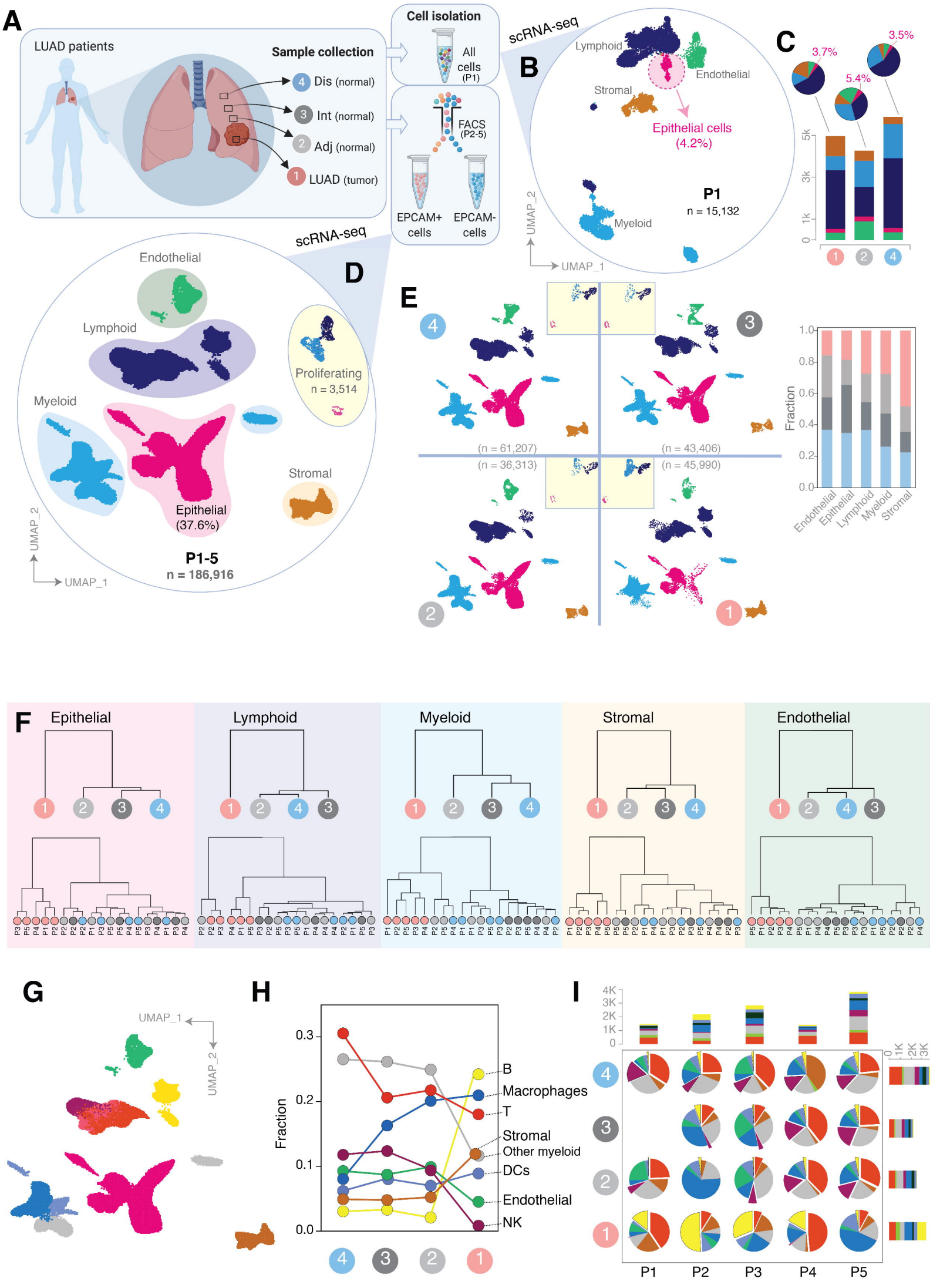
885

886 **Figure 4. Reduced signatures of antigen presentation and inflammatory dendritic**
887 **cells in the microenvironment of early-stage LUAD. A**, UMAP visualization of
888 myeloid cell lineages colored by cell type/state (left) and the spatial samples (right).
889 Mac, macrophages; Mono, monocytes; DC, dendritic cell; cDC, classical dendritic cell;
890 pDC, plasmacytoid dendritic cell. **B**, Bubble plot showing the percentage of myeloid
891 cells expressing lineage specific marker genes (indicated by the size of the circle) as
892 well as their scaled expression levels (indicated by the color of the circle). **C**, Changes
893 in the abundance of myeloid cell subsets across the LUADs and spatial normal lung
894 samples. Embedded pie charts show the contribution of each spatial sample to the
895 indicated cell subtype/state. **D**, UMAP plot of monocyte and macrophage
896 subpopulations, color coded by cell type/state (left), spatial sample (middle), and
897 antigen presentation score (right). **E**, Percentage of M2-like macrophages cluster 1
898 expressing antigen presentation genes (indicated by the size of the circle) and their
899 scaled expression levels (indicated by the color of the circle) across the spatial samples
900 (bubble plot). **F**, Ridge plots showing the distribution of MHC class I and MHC class II
901 gene expression densities in M2-like macrophages cluster 1, and across LUADs and
902 spatial normal lung samples. **G**, Violin plots showing the antigen presentation score in
903 M2-like macrophages (clusters 1 and 5) across LUADs and spatial normal lung samples
904 for all patients together (left) and within patients (right) (***, $P < 0.001$, N.S, $P > 0.05$). P
905 – values were calculated by Wilcoxon rank sum test. **H**, UMAP plots of dendritic cells,
906 color coded by cell state (left) and spatial sample (right). **I**, UMAP plots showing
907 unsupervised subclustering of cDC2 cells colored by cluster ID (top left), spatial sample
908 (top right) and the computed inflammatory signature score (bottom). **J**, Heatmap
909 showing normalized expression of marker genes of cDC2 cell subsets. The top
910 annotation tracks indicate (from top to bottom) the inflammatory signature scores,
911 spatial tissue of origin, and cDC2 cell clusters. **K**, Bubble plot showing the percentage of
912 cDC2 cells expressing inflammatory and non-inflammatory signature genes (indicated
913 by the size of the circle) as well as their scaled expression levels (indicated by the color
914 of the circle) in the LUADs and spatial normal lung tissues. **L**, Violin plot showing
915 expression of inflammatory signature score in cells from cDC2 cluster C2 in the TME of

916 LUADs. *P* – value was calculated using Kruskal-Wallis test (***, *P* < 0.001). **M**, Boxplot
917 showing fraction of cDC2 C2 cells among total cDC2 cells and across the LUADs and
918 spatial normal lung tissues. Individual circles correspond to patient samples. *P* – value
919 was calculated using Kruskal-Wallis test. **N**, Violin plot showing the inflammatory
920 signature scores in cDC2 C2 cells across LUADs and spatial normal lung samples. **O,**
921 boxplot showing the inflammatory signature score in normal lung (NL), in premalignant
922 atypical adenomatous hyperplasia (AAH) and in LUAD from an independent cohort (**,
923 *P* < 0.01; ***, *P* < 0.001; N.S, *P* > 0.05 of the Wilcoxon rank sum test).
924
925 **Figure 5. Enriched ligand-receptor cell-cell communication networks between**
926 **LUADs and their immune microenvironment**. **A**, Computational analysis workflow of
927 cell-cell communication using iTALK to identify, from a database of curated ligand-
928 receptor (L-R) pairs, the highly expressed immune checkpoint- and cytokine-receptor
929 pairs, that are significantly and differentially altered (i.e., interactions lost or gained)
930 between LUADs and spatial normal lung tissues. **B**, Heatmaps showing the overlap
931 (quantified by Jaccard index) of predicted ligand-receptor based interactions among
932 individual LUADs and their corresponding spatially distributed normal lung tissues. **C-F**,
933 Representative circos plots showing details of immune checkpoint-mediated L-R pairs
934 compared between each of the LUADs of patients 2 (C), 3 (D), 4 (E) and 5 (F), and
935 selected matching spatial normal lung samples. **G-H**, Violin plots showing expression of
936 the ligand and receptor genes (selected from panels C-F) involving immune checkpoints
937 and showing spatial gain-of-interaction patterns as highlighted in circus plots for patient
938 2 (panel G left), 3 (panel G right), 4 (panel H left) and 5 (panel H right).
939
940 **Figure 6. Pro-tumor phenotypes associated with augmented CD24 in LUAD**. **A,**
941 Boxplot showing *CD24* expression levels in an independent cohort of normal lung
942 tissues (NL), premalignant atypical adenomatous hyperplasias (AAH) and LUADs
943 assessed using the Nanostring immune Counter panel (see Supplementary Methods)
944 (*, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001 of the Wilcoxon rank sum test). **B**, Scatterplots
945 using Pearson correlation coefficients between levels of *CD24* with *EPCAM* and *PRF1*
946 in the NL, AAH, and LUAD samples. **C**, Boxplot depicting *CD24* expression levels in

947 LUADs and matched normal lung (NL) tissues from TCGA LUAD cohort (***, $P < 0.001$
948 of the Wilcoxon rank sum test). **D**, Scatter plots showing correlation of expression using
949 Pearson's correlation coefficients between *CD24* and *EPCAM* or *PRF1*. **E,** Overall
950 survival (OS) and progression-free interval (PFI) in a subset of early-stage LUAD
951 patients analyzed by targeted immune profiling (MDACC cohort) and dichotomized by
952 median *CD24* mRNA expression (*CD24* low n = 28 and high n= 28). Survival analysis
953 was performed using Kaplan–Meier estimates and two-sided log-rank tests. **F,** Scatter
954 plots showing correlation of expression using Pearson's correlation coefficients between
955 *CD24* and *EPCAM*, *PRF1* or cytotoxicity score in MDACC cohort. Cytotoxicity score
956 was calculated as the square root of the product of *GZMB* and *PRF1* expression. **G,**
957 Representative images showing relatively high (top) and low (bottom)
958 immunohistochemical CD24 staining (left). Scale bar = 100 μm. OS (middle) and PFI
959 (right) analysis in the early-stage LUAD TMA. Patients were dichotomized based on
960 median CD24 protein expression (CD24 low n = 83 and high n = 83). The analysis was
961 performed with the Kaplan–Meier estimates and two-sided log-rank tests. **H**, *In vivo*
962 growth of MDA-F471 cells subcutaneously implanted into syngeneic mice. Lengths and
963 widths of tumors were measured twice per week for 3 weeks and tumor volumes were
964 calculated according to the formula (length x width^2)/2, and tumor growth was plotted as
965 mean ± SEM. Mice in the left panel were implanted with MDA-F471 sgCt or MDA-F471
966 sg*Cd24a* cells following FACS-sorting for high or low expression of CD24 surface
967 protein, respectively. Mice in the right panel were implanted with parental MDA-F471
968 cells and treated with either control IgG or anti-CD24 antibody at the indicated
969 timepoints (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; unpaired Student's *t* test).

Figure 1


A
LUAD patients
Sample collection
4 Dis (normal)
3 Int (normal)
2 Adj (normal)
1 LUAD (tumor)
Cell isolation
All cells (P1)
FACS (P2-5)
EPCAM+ cells
EPCAM- cells
scRNA-seq
B
Lymphoid
Endothelial
Stromal
Epithelial cells (4.2%)
Myeloid
P1
n = 15,132
UMAP_2
UMAP_1
C
3.7%
5.4%
3.5%
5k
3k
1k
0
1
2
4
D
scRNA-seq
Endothelial
Lymphoid
Proliferating
n = 3,514
Myeloid
Stromal
Epithelial (37.6%)
P1-5
n = 186,916
UMAP_2
UMAP_1
E
4
3
2
1
(n = 61,207)
(n = 43,406)
(n = 36,313)
(n = 45,990)
Fraction
1.0
0.8
0.6
0.4
0.2
0.0
Endothelial
Epithelial
Lymphoid
Myeloid
Stromal
F
Epithelial
Lymphoid
Myeloid
Stromal
Endothelial
G
UMAP_1
UMAP_2
H
Fraction
0.3
0.2
0.1
0.0
B
Macrophages
T
Stromal
Other myeloid
DCs
Endothelial
NK
4
3
2
1
I
4K
3K
2K
1K
0
P1
P2
P3
P4
P5

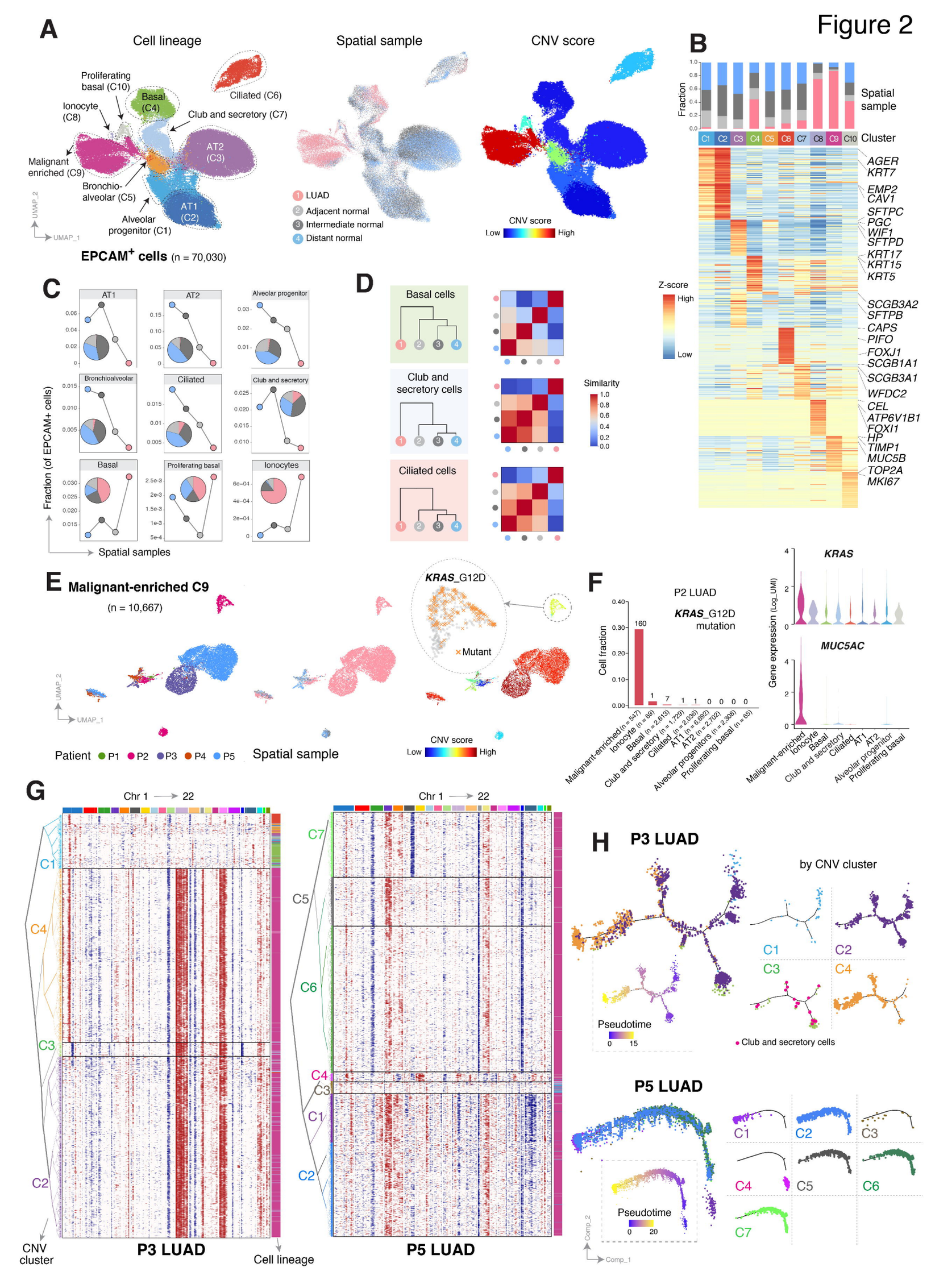
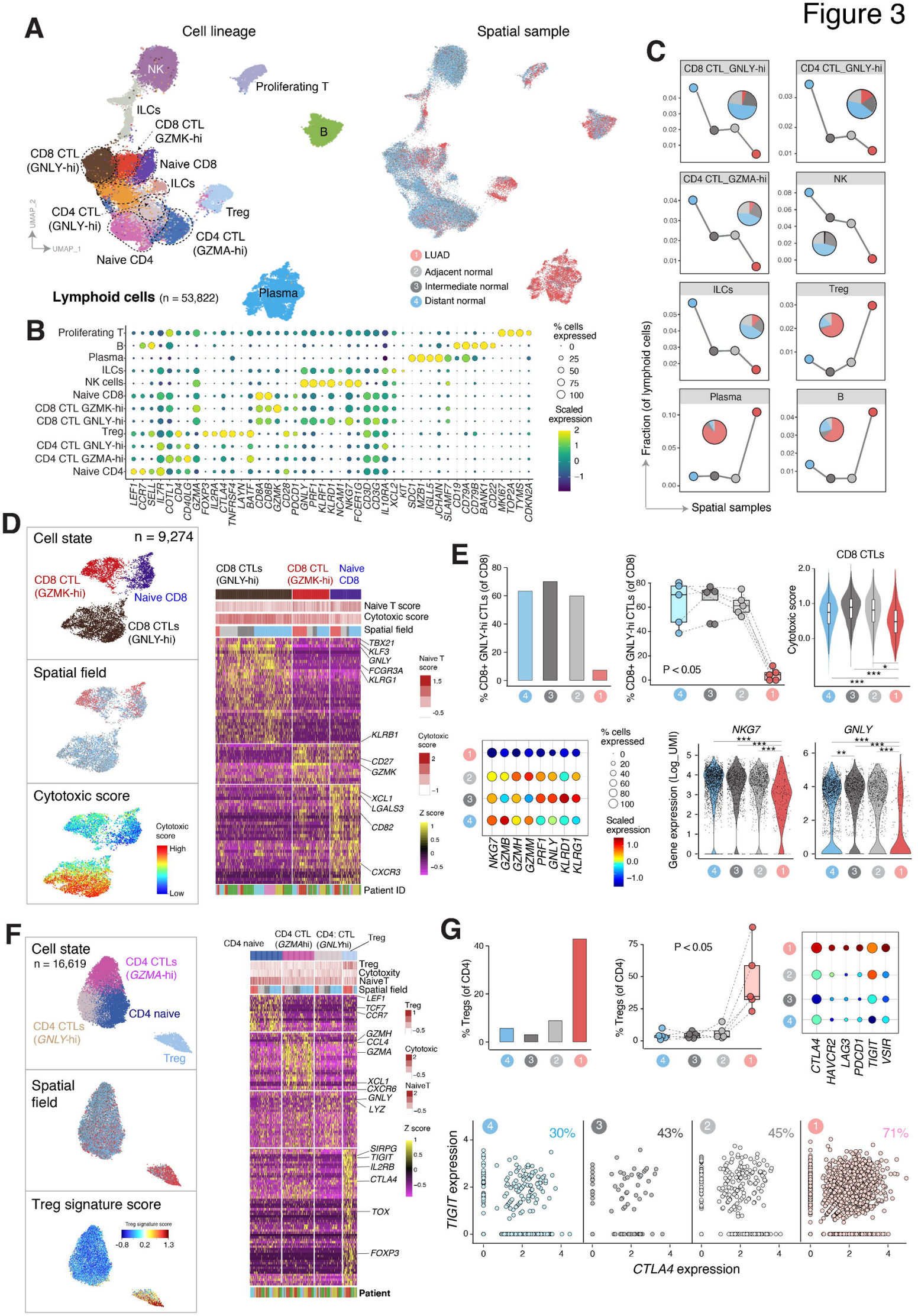


Figure 2
A
Cell lineage
Spatial sample
CNV score
Proliferating basal (C10)
Basal (C4)
Ciliated (C6)
Ionocyte (C8)
Club and secretory (C7)
AT2 (C3)
Malignant enriched (C9)
Bronchio-alveolar (C5)
AT1 (C2)
Alveolar progenitor (C1)
UMAP_2
UMAP_1
EPCAM+ cells (n = 70,030)
1 LUAD
2 Adjacent normal
3 Intermediate normal
4 Distant normal
CNV score
Low
High
B
Fraction
Spatial sample
Cluster
C1 C2 C3 C4 C5 C6 C7 C8 C9 C10
Z-score
High
Low
AGER
KRT7
EMP2
CAV1
SFTPC
PGC
WIF1
SFTPD
KRT17
KRT15
KRT5
SCGB3A2
SFTPB
CAPS
PIFO
FOXJ1
SCGB1A1
SCGB3A1
WFDC2
CEL
ATP6V1B1
FOXI1
HP
TIMP1
MUC5B
TOP2A
MKI67
C
AT1
AT2
Alveolar progenitor
Bronchioalveolar
Ciliated
Club and secretory
Basal
Proliferating basal
Ionocytes
Fraction (of EPCAM+ cells)
Spatial samples
D
Basal cells
Club and secretory cells
Ciliated cells
Similarity
E
Malignant-enriched C9
(n = 10,667)
KRAS_G12D
Mutant
Patient P1 P2 P3 P4 P5
Spatial sample
CNV score
Low
High
F
P2 LUAD
KRAS_G12D mutation
Cell fraction
Malignant-enriched (n = 547)
Ionocyte (n = 69)
Basal (n = 2,613)
Club and secretory (n = 1,729)
Ciliated (n = 2,036)
AT1 (n = 6,892)
AT2 (n = 2,702)
Alveolar progenitors (n = 2,398)
Proliferating basal (n = 65)
KRAS
MUC5AC
Gene expression (Log_UMI)
G
Chr 1 → 22
C1
C4
C3
C2
C7
C5
C6
CNV cluster
P3 LUAD
Cell lineage
P5 LUAD
H
P3 LUAD
by CNV cluster
Pseudotime
Club and secretory cells
P5 LUAD
Comp_2
Comp_1

Figure 3


A
Cell lineage
Spatial sample
NK
Proliferating T
ILCs
CD8 CTL GZMK-hi
B
CD8 CTL (GNLY-hi)
Naive CD8
ILCs
Treg
CD4 CTL (GNLY-hi)
CD4 CTL (GZMA-hi)
Naive CD4
Plasma
UMAP_2
UMAP_1
Lymphoid cells (n = 53,822)
1 LUAD
2 Adjacent normal
3 Intermediate normal
4 Distant normal
B
Proliferating T
B
Plasma
ILCs
NK cells
Naive CD8
CD8 CTL GZMK-hi
CD8 CTL GNLY-hi
Treg
CD4 CTL GNLY-hi
CD4 CTL GZMA-hi
Naive CD4
LEF1 CCR7 SELL IL7R CCTL1 CD4 CD40LG GZMA FOXP3 IL2RA CTLA4 TNFRSF4 LAYN BATF CD8A CD8B GZMK CD28 PDCD1 GNLY PRF1 KLRF1 KLRD1 NCAM1 NKG7 FCER1G CD3D CD3G IL10RA XCL2 KIT SDC1 MZB1 IGLL5 JCHAIN SLAMF7 CD19 CD79A CD79B BANK1 CD22 MKI67 TOP2A TYMS CDKN2A
% cells expressed
0 25 50 75 100
Scaled expression
2 1 0 −1
C
CD8 CTL_GNLY-hi
CD4 CTL_GNLY-hi
CD4 CTL_GZMA-hi
NK
ILCs
Treg
Plasma
B
Fraction (of lymphoid cells)
Spatial samples
D
Cell state
n = 9,274
CD8 CTL (GZMK-hi)
Naive CD8
CD8 CTLs (GNLY-hi)
Spatial field
Cytotoxic score
High
Low
CD8 CTLs (GNLY-hi)
CD8 CTL (GZMK-hi)
Naive CD8
Naive T score
Cytotoxic score
Spatial field
TBX21
KLF3
GNLY
FCGR3A
KLRG1
KLRB1
CD27
GZMK
XCL1
LGALS3
CD82
CXCR3
Naive T score
Cytotoxic score
Z score
Patient ID
E
% CD8+ GNLY-hi CTLs (of CD8)
P < 0.05
CD8 CTLs
Cytotoxic score
% cells expressed
0 20 40 60 80 100
Scaled expression
1.0 0.0 −1.0
NKG7 GZMB GZMH GZMM PRF1 GNLY KLRD1 KLRG1
NKG7
GNLY
Gene expression (Log_UMI)
F
Cell state
n = 16,619
CD4 CTLs (GZMA-hi)
CD4 naive
CD4 CTLs (GNLY-hi)
Treg
Spatial field
Treg signature score
Treg signature score
-0.8 0.2 1.3
CD4 naive
CD4 CTL (GZMAhi)
CD4: CTL (GNLYhi)
Treg
Treg
Cytotoxity
NaiveT
Spatial field
LEF1
TCF7
CCR7
GZMH
CCL4
GZMA
XCL1
CXCR6
GNLY
LYZ
SIRPG
TIGIT
IL2RB
CTLA4
TOX
FOXP3
Treg
Cytotoxic
NaiveT
Z score
Patient
G
% Tregs (of CD4)
P < 0.05
CTLA4 HAVCR2 LAG3 PDCD1 TIGIT VSIR
30%
43%
45%
71%
TIGIT expression
CTLA4 expression

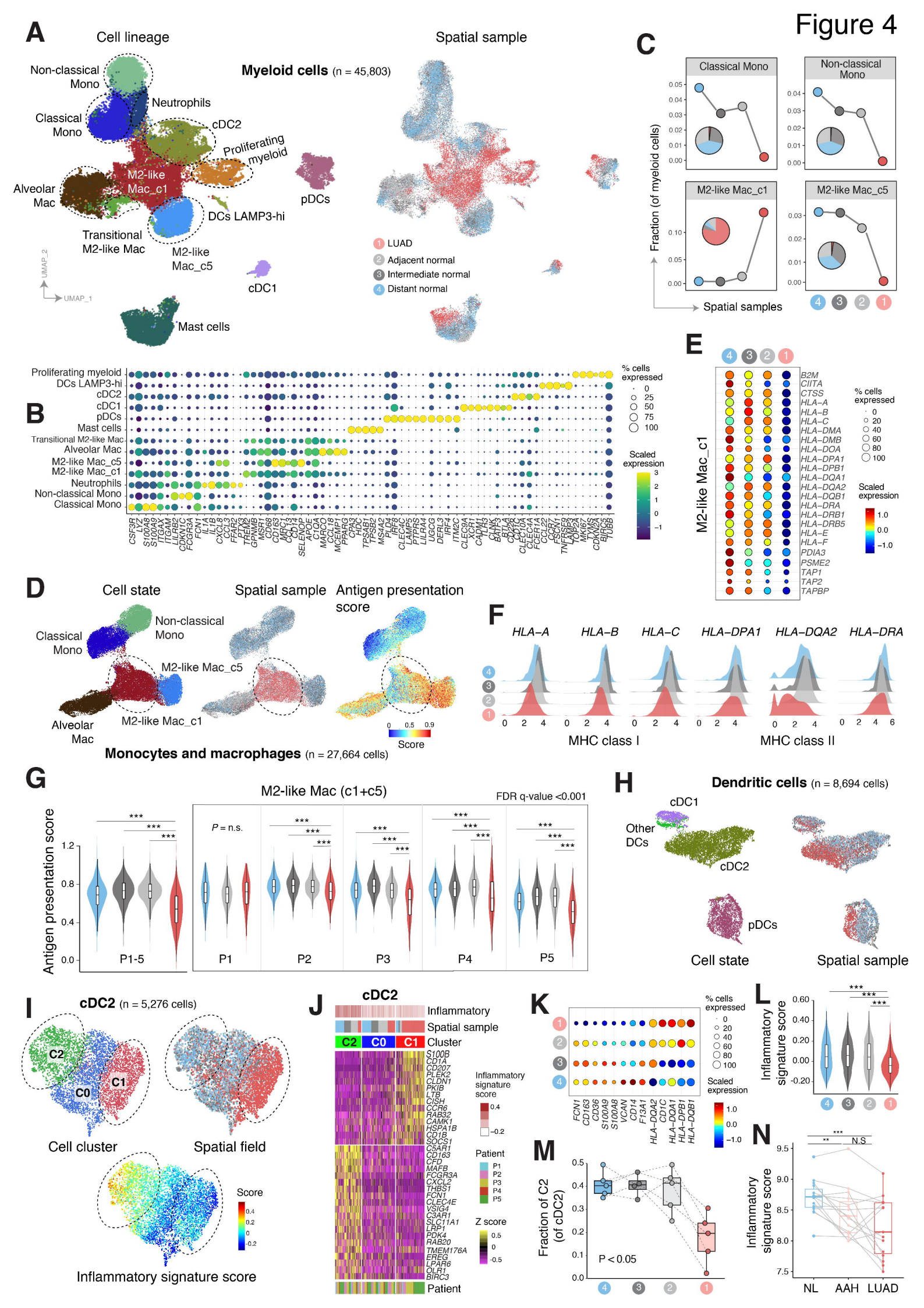


Figure 4
A
Cell lineage
Spatial sample
Myeloid cells (n = 45,803)
Non-classical Mono
Neutrophils
Classical Mono
cDC2
Proliferating myeloid
M2-like Mac_c1
Alveolar Mac
pDCs
DCs LAMP3-hi
Transitional M2-like Mac
M2-like Mac_c5
cDC1
Mast cells
UMAP_2
UMAP_1
1 LUAD
2 Adjacent normal
3 Intermediate normal
4 Distant normal
C
Classical Mono
Non-classical Mono
M2-like Mac_c1
M2-like Mac_c5
Fraction (of myeloid cells)
Spatial samples
B
Proliferating myeloid
DCs LAMP3-hi
cDC2
cDC1
pDCs
Mast cells
Transitional M2-like Mac
Alveolar Mac
M2-like Mac_c5
M2-like Mac_c1
Neutrophils
Non-classical Mono
Classical Mono
% cells expressed
Scaled expression
E
M2-like Mac_c1
B2M
CIITA
CTSS
HLA–A
HLA–B
HLA–C
HLA–DMA
HLA–DMB
HLA–DOA
HLA–DPA1
HLA–DPB1
HLA–DQA1
HLA–DQA2
HLA–DQB1
HLA–DRA
HLA–DRB1
HLA–DRB5
HLA–E
HLA–F
PDIA3
PSME2
TAP1
TAP2
TAPBP
% cells expressed
Scaled expression
D
Cell state
Spatial sample
Antigen presentation score
Classical Mono
Non-classical Mono
M2-like Mac_c5
Alveolar Mac
M2-like Mac_c1
Score
Monocytes and macrophages (n = 27,664 cells)
F
HLA–A
HLA–B
HLA–C
HLA–DPA1
HLA–DQA2
HLA–DRA
MHC class I
MHC class II
G
M2-like Mac (c1+c5)
FDR q-value <0.001
Antigen presentation score
P = n.s.
P1-5
P1
P2
P3
P4
P5
H
Dendritic cells (n = 8,694 cells)
cDC1
Other DCs
cDC2
pDCs
Cell state
Spatial sample
I
cDC2 (n = 5,276 cells)
C2
C0
C1
Cell cluster
Spatial field
Score
Inflammatory signature score
J
cDC2
Inflammatory
Spatial sample
Cluster
C2
C0
C1
Inflammatory signature score
Patient
P1
P2
P3
P4
P5
Z score
Patient
K
% cells expressed
Scaled expression
L
Inflammatory signature score
M
Fraction of C2 (of cDC2)
P < 0.05
N
Inflammatory signature score
NL
AAH
LUAD
N.S

Figure 5

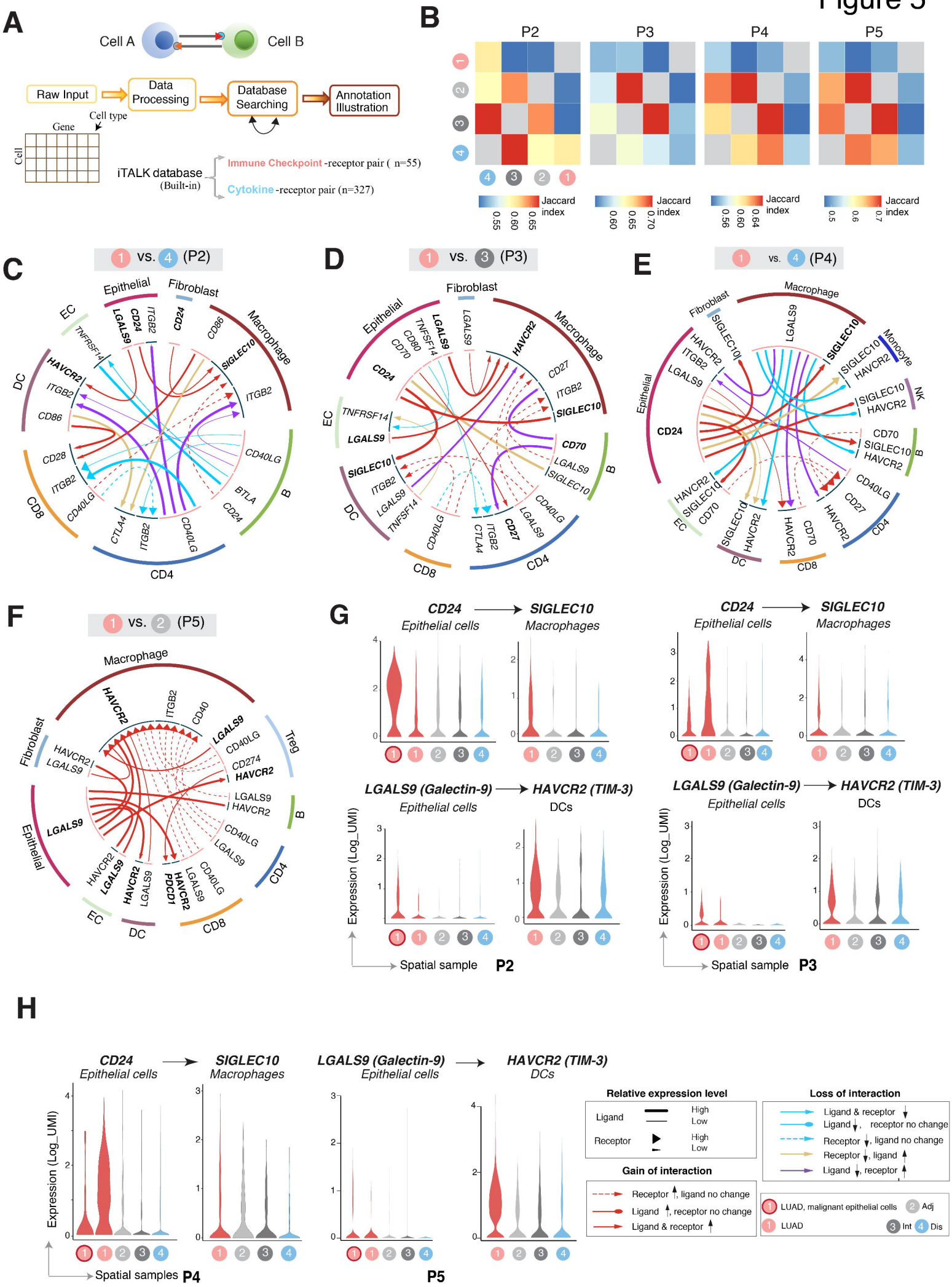
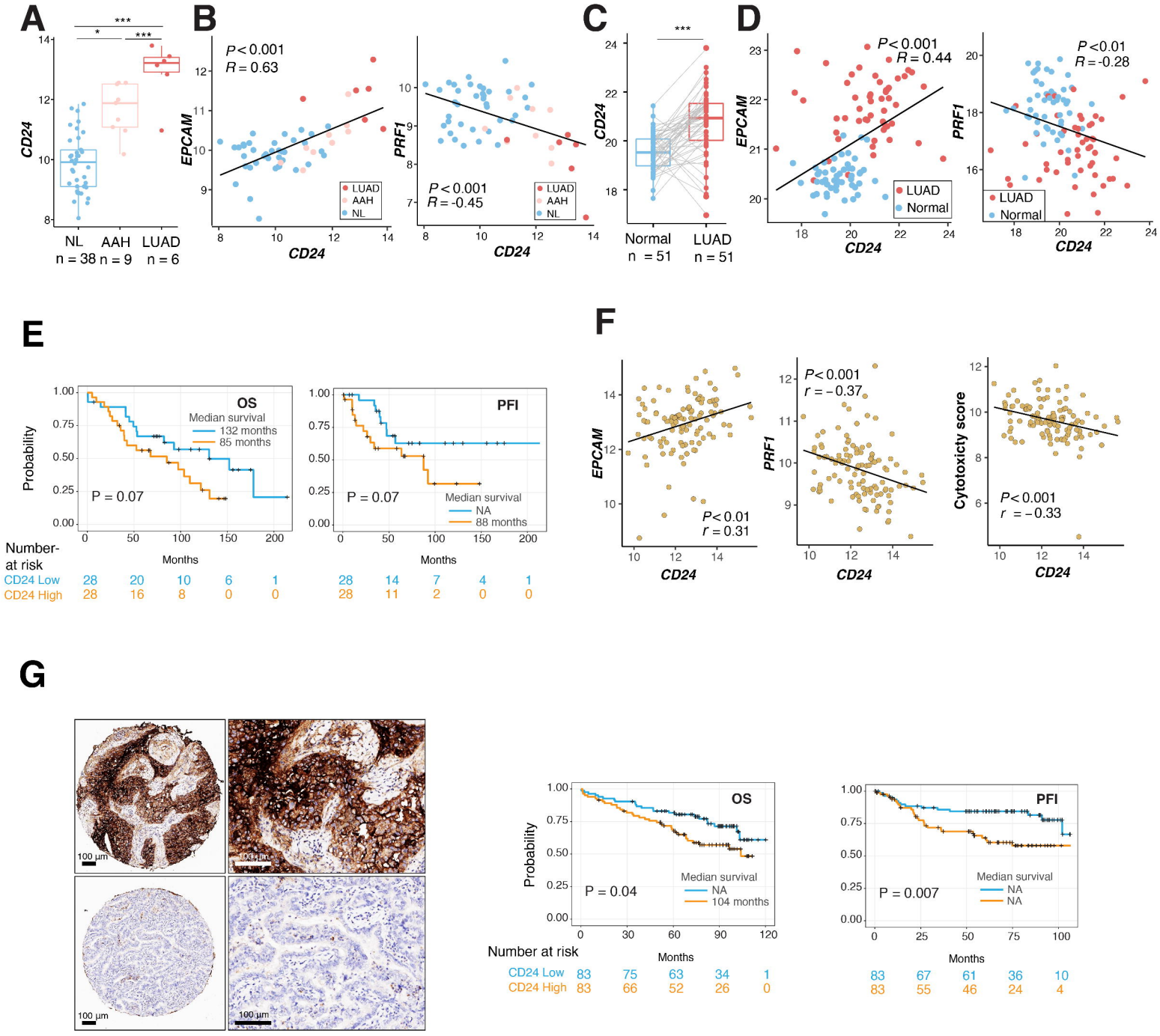

A
Cell A
Cell B
Raw Input
Data Processing
Database Searching
Annotation Illustration
Cell type
Gene
Cell
iTALK database (Built-in)
Immune Checkpoint -receptor pair (n=55)
Cytokine -receptor pair (n=327)
B
P2
P3
P4
P5
Jaccard index
C
1 vs. 4 (P2)
D
1 vs. 3 (P3)
E
1 vs. 4 (P4)
F
1 vs. 2 (P5)
G
CD24 → SIGLEC10
Epithelial cells
Macrophages
LGALS9 (Galectin-9) → HAVCR2 (TIM-3)
DCs
Expression (Log_UMI)
Spatial sample
H
Spatial samples
Relative expression level
Ligand
Receptor
High
Low
Gain of interaction
Receptor ↑, ligand no change
Ligand ↑, receptor no change
Ligand & receptor ↑
Loss of interaction
Ligand & receptor ↓
Ligand ↓, receptor no change
Receptor ↓, ligand no change
Receptor ↓, ligand ↑
Ligand ↓, receptor ↑
LUAD, malignant epithelial cells
LUAD
Adj
Int
Dis

Figure 6


A
CD24
NL
AAH
LUAD
n = 38
n = 9
n = 6
B
P < 0.001
R = 0.63
EPCAM
CD24
LUAD
AAH
NL
P < 0.001
R = -0.45
PRF1
C
CD24
Normal
LUAD
n = 51
n = 51
D
P < 0.001
R = 0.44
EPCAM
LUAD
Normal
P < 0.01
R = -0.28
PRF1
E
OS
Median survival
132 months
85 months
P = 0.07
Probability
Months
PFI
P = 0.07
Median survival
NA
88 months
Number-at risk
CD24 Low 28 20 10 6 1 28 14 7 4 1
CD24 High 28 16 8 0 0 28 11 2 0 0
F
EPCAM
P < 0.01
r = 0.31
PRF1
P < 0.001
r = −0.37
Cytotoxicity score
P < 0.001
r = −0.33
CD24
G
100 μm
OS
Median survival
NA
104 months
P = 0.04
Probability
Months
Number at risk
CD24 Low 83 75 63 34 1
CD24 High 83 66 52 26 0
PFI
Median survival
NA
NA
P = 0.007
Months
83 67 61 36 10
83 55 46 24 4
H
MDA-F471 sgCt
MDA-F471 sgCd24a
Tumor volume (mm³)
Days
control IgG
anti-CD24
N.S.