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

3

4 Ansam Sinjab [1]*, Guangchun Han [2]*, Kieko Hara [1], Warapen Treekitkarnmongkol [1],

5 Patrick Brennan [3], Minghao Dang [2], Dapeng Hao [2], Ruiping Wang [2], Enyu Dai [2], Hitoshi

6 Dejima [1], Jiexin Zhang [4], Elena Bogatenkova [3], Beatriz Sanchez-Espiridion [1], Kyle

7 Chang [5], Danielle R. Little [6], Samer Bazzi [7], Linh Tran [8], Kostyantyn Krysan [8], Carmen

8 Behrens [9], Dzifa Duose [1], Edwin R. Parra [1], Maria Gabriela Raso [1], Luisa M. Solis [1],

9 Junya Fukuoka [10], Jianjun Zhang [9], Boris Sepesi [10], Tina Cascone [9], Lauren Byers [9],

10 Don L. Gibbons [9], Jichao Chen [6], Seyed Javad Moghaddam [6], Edwin J. Ostrin [11], Daniel

11 G. Rosen [12], Junya Fujimoto [1], John V. Heymach [9], Paul Scheet [1,2,5], Steven Dubinett [8],

12 Ignacio I. Wistuba [1], Christopher S. Stevenson [13], Avrum E. Spira [13,14], Linghua Wang

13 [2]#, Humam Kadara [1]#

14

15 Departments of Translational Molecular Pathology [1], Genomic Medicine [2], Pathology [3],

16 Bioinformatics and Computer Biology [4], Epidemiology [5], Pulmonary Medicine [6],

17 Thoracic, Head and Neck Medical Oncology [9], Cardiovascular and Thoracic Surgery [10]

18 and General Internal Medicine [11], The University of Texas MD Anderson Cancer Center,

19 Houston, TX, USA. Department of Medicine [7], University of Balamand, Koura, Lebanon.

20 Department of Medicine [8], The University of California Los Angeles, Los Angeles, CA,

21 USA. Department of Pathology [10], Nagasaki University Graduate School of Biomedical

22 Sciences, Nagasaki, Japan. Department of Pathology and Immunology [12], Baylor

23 College of Medicine, Houston, TX, USA. Lung Cancer Initiative at Johnson and Johnson

24 [13], Boston, MA, USA. Section of Computational Biomedicine [14], Boston University,

25 Boston, MA, USA.

26

27 *: equally contributing co-first authors

28 #: equally contributing senior co-corresponding authors

29

Running title: Spatial single-cell atlas of lung adenocarcinoma

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

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.

50 **ABSTRACT**

51 Little is known of the geospatial architecture of individual cell populations in lung

52 adenocarcinoma (LUAD) evolution. Here, we perform single-cell RNA sequencing of

53 186,916 cells from five early-stage LUADs and fourteen multi-region normal lung tissues

54 of defined spatial proximities from the tumors. We show that cellular lineages, states,

55 and transcriptomic features geospatially evolve across normal regions to the LUADs.

56 LUADs exhibit pronounced intratumor cell heterogeneity within single sites and

57 transcriptional lineage-plasticity programs driven by *KRAS* mutations. T regulatory cell

58 phenotypes are increased in normal tissues with closer proximity to LUAD, in contrast to

59 diminished signatures and fractions of cytotoxic CD8+ T cells, antigen-presenting

60 macrophages and inflammatory dendritic cells. Further, the LUAD ecosystem harbors

61 gain of ligand-receptor based interactions involving increased expression of CD24

62 antigen on epithelial cells and SIGLEC10 on myeloid subsets. These data provide a

63 spatial atlas of LUAD evolution, and a resource for identification of targets for treatment.

64 **Statement of significance:** The geospatial ecosystem of the peripheral lung and early-

65 stage LUAD is not known. Our multi-region single-cell sequencing analyses unravel cell

66 populations, states, and phenotypes in the spatial and ecological evolution LUAD from

67 the lung that comprise high-potential targets for early interception.

68

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 and following results from the National Lung Screening Trial, 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, which in turn heavily rests on an improved understanding of the molecular and cellular processes underlying early LUAD development.

Previous studies have 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 have 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 spatial 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. Recent reports have employed single-cell sequencing technologies to chart the immune microenvironment of advanced lung cancers as well as to probe mechanisms underlying lung tumor metastasis and response 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, has remained 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

4

100 evolution. We performed deep scRNA-seq analysis of 19 spatial regions, including

101 enriched epithelial populations, from five early-stage LUADs and 14 multi-region

102 normal-appearing lung tissues with differential and defined spatial proximities from the

103 tumors. Our study unravels tumor evolutionary trajectories as well as geospatial

104 evolution in cell populations, their expression signatures, and ligand-receptor based

105 interactions that portray how early-stage LUAD develops from the lung ecosystem.

106

107

108 **RESULTS**

109

110 **Single-cell spatial landscape of early-stage LUAD**

111 To begin to chart a comprehensive single-cell atlas of early-stage LUAD and the

112 peripheral lung, we performed scRNA-seq on all cells from a freshly resected early-

113 stage LUAD (P1) as well as matched tumor-adjacent and relatively more distant normal

114 lung tissues (**Fig. 1A)**. The spatial locations of multi-region normal tissues were

115 carefully defined with respect to the tumor edge (Supplementary Methods), such that

116 the studied samples span a spatial continuum and, thus, enable interrogation of

117 geospatial relationships among early-stage LUAD and the peripheral lung tissues,

118 hereby defined as LUAD "field". Unsupervised clustering of 15,132 QC-passed cells

119 revealed cell clusters representing 5 major cellular lineages, namely epithelial,

120 endothelial, myeloid, lymphoid, and stromal cell subsets (**Fig. 1B, Supplementary Fig.**

121 **S1**). Epithelial (*EPCAM*+) cell fractions were 3.7%, 5.4%, and 3.5% for tumor, tumor-

122 adjacent and -distant normal samples, respectively, at an average of 4.2% and in line

123 with previous studies interrogating different subtypes and/or stages of lung cancer

124 ((8,10); **Fig. 1B and C**).

125

126 To increase the throughput and to better capture patterns of cellular heterogeneity

127 based on distance from LUADs, in particular within the epithelial lineage, we performed

128 separate scRNA-seq analysis on epithelial (*EPCAM*+) and non-epithelial (*EPCAM*-)

129 single cells enriched from early-stage LUADs of four additional patients (P2-P5,

130 Methods), each with three matching normal lung tissues of defined spatial proximities to

5

131   LUADs: tumor-adjacent, -intermediate and -distant (total of 19 samples and 35 scRNA-

132   seq libraries from P1-P5; **Fig. 1A**). A total of 186,916 cells were retained for subsequent

133   analyses with a median of 1,844 genes detected per cell (**Supplementary Fig. S1A-**

134   **S1C; Supplementary Table S1**). By single-cell sequencing two separate fractions

135   (*EPCAM*-positive and *EPCAM*-negative), we were able to profile samples markedly

136   enriched with epithelial cells (37.6%, n = 70,030 epithelial cells) in comparison to the

137   unbiased approach (4.2%, non-*EPCAM*-enriched) in P1 (**Fig. 1D**). Cells were uniformly

138   derived from all spatial samples and their lineage cluster fractions spatially distributed

139   along the continuum of LUAD-to-distant normal tissues (**Fig. 1E**). We next computed

140   the Euclidean distance metric and from this inferred hierarchical relationships among

141   major cell lineages across the spatial fields (**Fig. 1F,** see **Methods**). Overall, cells of the

142   LUAD tumor samples were transcriptomically distinct from those of the normal samples

143   across major cell lineages. Of note, we found that cells from adjacent normal samples in

144   3 out 5 patients (P2, P4 and P5) were clustered more closely with those of LUAD tumor

145   samples (i.e. lymphoid and myeloid lineages) than those of intermediate and distant

146   normal tissues, thereby reflecting possible transcriptomic gradients (**Fig. 1F**).

147

148   We then further classified lymphoid and myeloid lineages and determined major cell

149   types, such as T cells, B cells, natural killer (NK) cells, macrophages, or dendritic cells

150   (DC) based on expression of canonical marker genes (**Fig. 1G; Supplementary Fig.**

151   **S1E, Methods**). Analysis of spatial cell composition revealed distinct topological

152   gradients including greatly increased fractions of B cells and decreased abundance

153   levels of NK cells with greater proximity to the tumors (**Fig. 1H; Supplementary Table**

154   **S2**), which were evident and consistent across patients (**Fig. 1I**). These observations

155   highlight geospatial transcriptomic heterogeneity in single-cell tumor microenvironment

156   landscape of early-stage LUADs.

157

158   **Spatial diversity and intratumoral heterogeneity of lung epithelial lineage**

159   We next interrogated spatial epithelial features of the LUADs and multi-region normal

160   tissues. The 70,030 epithelial cells formed 10 distinct clusters representing different

161   airway lineages including alveolar type I (AT1; C2, *AGER*+), AT2 (C3; *SFTPC*+), basal

162   (C4; *KRT15+*), bronchioalveolar (C5; *SFTPC+/SCGB1A1+*), ciliated (C6; *PIFO+*) and

163   club/secretory (C7; *BPIFB1+*) cells (**Figs. 2A and B; Supplementary Fig. S2A;**

164   **Supplementary Table S3**). We also identified, as distinct clusters, the recently-

165   described and rare ionocytes (C8; *FOXI1+/CFTR+*; (11)), understudied bipotent alveolar

166   progenitors (C1; (12)), as well as unique cell states such as proliferating basal cells

167   (C10; *TOP2A+*). In addition, we noted a cluster of malignant-enriched (C9) cells of

168   mixed lineage (8) mostly from the LUADs (**Fig. 2A-C**). Cells from the five LUADs were

169   predominantly, yet not exclusively, comprised of C9 cells (**Fig. 2B; Supplementary Fig.**

170   **S2B; Supplementary Table S4**). Interestingly, few cells from the normal tissues were

171   found in the C9 cluster (**Fig. 2A-C**). Therefore, to distinguish *bona fide* malignant cells

172   from non-malignant subsets, we employed a strategy that infers copy number variations

173   (CNVs) from scRNA-seq data in every epithelial cell (13) (Methods). We found overall

174   high CNV (inferCNV) scores in cells from C9 (**Fig. 2A**, right), thereby supporting the

175   overall malignant assignment of this cluster. We also noted pronounced steady-state

176   enrichment or depletion of epithelial subsets with closer spatial proximity to the tumors

177   (**Fig. 2C and D**). Relative to cells from tumor-intermediate or -distant normal sites, cells

178   from tumor-adjacent normal tissues were, overall, more transcriptomically similar

179   (clustered closely) to those from the LUADs (**Fig. 2E)**, signifying spatial epithelial

180   patterns that are possibly tumor-pertinent.

181

182   Alveolar differentiation hierarchies have been shown to partake in lung tumor

183   development *in vivo* (12,14,15). In our cohort, alveolar cells with definitive lineage

184   features (e.g. AT1, AT2, and alveolar progenitors) were marked depleted in LUAD

185   tissues (**Fig. 2C and D)**, which prompted us to dissect potential alveolar differentiation

186   trajectories. Pseudotemporal ordering of alveolar cells revealed a developmental

187   hierarchy that was initiated by AT2 cells and that followed a main trajectory of

188   differentiation into AT1 cells **(Supplementary Fig. S2C and S2D)** in close agreement

189   with previous studies in mice (12,14,15). The differentiation trajectory included

190   increased NOTCH signaling score (**Supplementary Fig. S2E and S2F)** in line with

191   previously reported role of NOTCH in AT2-to-AT1 differentiation and alveolar repair

192   (16).

7

193

194   To further investigate malignant programs, we performed subclustering of malignant-

195   enriched C9 cells (n = 10,667 cells) while projecting inferCNV scores, which separated

196   likely malignant cells from subsets of normal tissues such as those evident in P2 (**Fig.**

197   **2F; Supplementary Fig. S2G; Supplementary Table S5**). We noted overall low

198   inferCNV scores in malignant-enriched cells derived from the LUAD of P2 **(Fig. 2F)**.

199   Interestingly, among all P2 epithelial subsets, 29% of malignant-enriched cells (160 of

200   547) harbored the *KRAS* G12D mutation (**Fig. 2F and G; Supplementary Fig. S2H**).

201   These cells exhibited distinctively high expression of *MUC5AC* (**Fig. 2G**) and tumor

202   markers (e.g., *CEACAM5*, **Supplementary Fig. S2I; Supplementary Table S6**). They

203   also displayed elevated levels of *LCN2* and reduced expression of *NKX2-1* when

204   compared to *KRAS* wild type cells from the same LUAD (**Supplementary Fig. S2I;**

205   **Supplementary Table S6**), suggestive of mucinous differentiation (17,18) and in line

206   with the histological (mucinous) pattern of this tumor (**Supplementary Table S1**). These

207   findings underscore spatial heterogeneity dynamics comprising driver mutation, lineage

208   plasticity, and stable chromosomal alterations that are unique to the cellular ecosystem

209   of *KRAS*-mutant LUAD.

210

211   Unlike P2, C9 cells in P3 and P5 were almost exclusively derived from the LUAD

212   tissues **(Fig. 2F)**. In P3 LUAD, we identified large-scale chromosomal alterations

213   (**Supplementary Fig. S2G; Supplementary Table S5)**, based on which unsupervised

214   clustering analysis revealed 4 clusters with differential CNV profiles. Among them, three

215   clusters (C2, C3, and C4) exhibited pronounced CNVs that were indicative of malignant

216   cell features (**Fig. 2H**, left**)**. Notably, we found an additional CNV event (i.e. gain of 1p)

217   unique to cells of cluster C4 but not C2 or C3, possibly signifying a late event in the

218   evolutionary trajectory of P3 LUAD. When inferCNV clusters were projected on the

219   pseudotime trajectory plot, we observed a branched differentiation trajectory that mainly

220   started with cells of C2 and C3 and comprised few "normal cells" with club and

221   secretory lineage, and that later branched into cells of inferCNV cluster C4 **(Fig. 2I**, top**)**

222   -- suggesting that P3 LUAD perhaps originated from club/secretory cells and that indeed

223   C4 evolved from C2 and C3. P5 LUAD comprised 6 distinct inferCNV clusters that are

224   indicative of high degree of genotypic heterogeneity within the same tumor region **(Fig.**

225   **2H**, right**)**. Among these clusters, cells of C1 had more intensive CNVs, while cells of C4

226   clustered the closest to those of C3 which comprised mostly non-malignant cells **(Fig.**

227   **2H**, right**)**. Pseudotime trajectory analysis revealed a C4-to-C1 evolutionary path, which

228   in contrast to P3 was unidirectional and unbranched, suggesting that C4 and C1 in P5

229   signified malignant cells from early and late developmental states, respectively **(Fig. 2I**,

230   bottom**)**. Together, our single-cell interrogation of a large number of epithelial cells from

231   multi-region tissues identified diverse epithelial identities, developmental malignant

232   trajectories, as well as high resolution intratumor cell heterogeneity.

233

234   **Lymphoid reprogramming towards a protumor microenvironment**

235   We further characterized lymphoid spatial dynamics **(Fig. 1H and I)**, including unique

236   cellular states across all lymphoid clusters (n = 53,882 cells) in the LUAD field samples

237   (see Methods; **Supplementary Fig. S3A**). This clustering analysis unveiled 10

238   transcriptomically distinct lymphoid cell types/states **(Fig. 3A and 3B, Supplementary**

239   **Table S7)** that were, overall, spatially modulated by tumor proximity **(Fig. 3C;**

240   **Supplementary Fig. S3B)**. Relative to normal tissues, LUADs were heavily enriched

241   with plasma cells *(SDC1+/MZB1+)*, B cells (*CD19+/CD22+*), and regulatory T cells

242   (Treg; *FOXP3*+) **(Fig. 3A-C)**. With increasing tumor proximity, we noted a gradual

243   decrease in NK cells (*GNLY*+), innate lymphoid cells (ILCs), both *GZMA*-hi and *GNLY*-

244   hi CD4+ cytotoxic T lymphocytes (CTL; *CD40LG+, BATF+*), and *GNLY*-hi CD8+ CTLs,

245   all of which were, overall, depleted in the LUADs **(Fig. 3A-C)**.

246

247   We further performed subclustering analysis of CD8+ T cells, which revealed 3

248   subpopulations, naïve CD8+ T cells, *GZMK*-hi CD8+ CTLs, and *GNLY*-hi CD8+ CTLs,

249   with differential expression of cell state signatures (**Fig. 3D; Supplementary Tables S8**

250   **and S10).** We also computed the naïve and cytotoxic T cell scores using curated gene

251   signatures of naïve and cytotoxic T cells (see Methods), respectively. Consistently, the

252   naïve CD8+ T cells showed high naïve T cell scores but low cytotoxic T cell scores and

253   were composed of cells across all samples in the LUAD fields. In contrast, the *GNLY*-hi

254   CD8+ CTLs exhibited high expression levels of cytotoxicity genes (*TBX21, KLF3,*

255 *FCGR3A, KLRG1*, *KLRB1*) and high cytotoxic activity but were specifically depleted in

256 LUAD samples of all patients **(Fig. 3C-E).** In line with this, we also observed a

257 significant spatial pattern of reduced cytotoxic activity in P3 and P4 (**Supplementary**

258 **Fig. S3D).** Overall, CD8+ CTLs showed significant and spatially-modulated reduction in

259 cytotoxic signature score (depleted in tumors of all patients) and decreased expression

260 of major cytotoxicity genes, including reduced *NKG7* and *GNLY* expression levels **(Fig.**

261 **3E)**.

262

263 Spatial analysis of CD4+ T cell states (**Supplementary Tables S9 and S10**) showed

264 that LUAD tissues were specifically enriched with *FOXP3*+ Tregs **(Fig. 3F and 3G;**

265 **Supplementary Table S11)** and the Treg signature scores were significantly and

266 spatially increased with closer proximity to all LUADs **(Supplementary Fig. S3E)**. The

267 Tregs cells also expressed high levels of pro-tumor immune checkpoints including

268 *TIGIT, CTLA4, LAG3,* or *PDCD1* (**Fig. 3G**). Of note, the fraction of Tregs co-expressing

269 both *CTLA-4* and *TIGIT* immune checkpoints was progressively higher along the

270 spectrum of distant normal sites to more adjacent (to the tumor) regions up to the

271 LUADs **(Fig. 3G**, bottom**)**. In contrast, we noted a significant reduction of cytotoxic

272 CD4+ CTLs characterized by high expression of *GZMA*, *GNLY*, or both *GZMA* and

273 *GZMH*, with increasing proximity to all LUADs **(Supplementary Fig. S3F-H)**.

274

275 We further examined the spatial enrichment of LUADs with plasma cells and B cells

276 **(Fig. 3A-C; Supplementary Table S11).** Expression profiling of plasma cells revealed

277 spatial changes in isotype-switching, such as increased *IGHA1/2* and decreased

278 *IGHG1/3* with increasing proximity to P3 and P5 LUADs **(Supplementary Fig. S3I-K;**

279 **Supplementary Table S12)**. Based on differential gene expression profiles, we

280 identified 3 distinct subsets of B cell states **(Supplementary Table S13)**, including a

281 LUAD-enriched subcluster (C0) with high expression levels of *RAC2*+ and *ACTG*+

282 **(Supplementary Fig. S3L-N)**, that are known to play key roles in synapse formation in

283 B cells (19). Importantly, when analyzed in an external cohort of matched normal lung,

284 preneoplasia and LUAD (20), the B cell signature (C0) was progressively increased

285 across atypical adenomatous hyperplasias (AAH), the preneoplastic precursors of

286 LUAD, and invasive LUAD lesions compared to matched normal lung tissues

287 **(Supplementary Fig. S3O).** Taken together, these analyses propose spatial properties

288 in lymphoid cell states that underlie pro-tumor immune remodeling in early-stage LUAD.

289

290 **Depletion of antigen presenting macrophages and inflammatory dendritic cells**

291 Spatial LUAD field patterns in myeloid cells **(Fig. 1H and 1I)** prompted us to further

292 investigate myeloid subsets and cellular states in the early-stage LUADs and their

293 normal multi-region tissues **(Supplementary Fig. S4A)**. In total, 45,803 myeloid cells

294 were clustered into 13 distinct subsets: classical monocytes (*S100A8+, S100A9+*), non-

295 classical monocytes *(CDKN1C+)*, mast cells *(MS4A2+)*, neutrophils (*IL1A+*), M2-like

296 macrophages C1 (*TREM2+)*, M2-like macrophages C5 (*CD163+*), alveolar

297 macrophages (*MARCO+)*, classical dendritic cells 1 (cDC1; *CLEC9A+*), cDC2

298 (*CLEC10A+*), plasmacytoid DC (pDC; *PLD4+*), other DCs (*CCL22+)* and proliferating

299 myeloid cells *(TOP2A+)* **(Fig. 4A and 4B; Supplementary Tables S14 and S15)**.

300 Mapped to the field samples, M2-like macrophages C5, monocytes (classical and non-

301 classical), and mast cells were gradually depleted with increasing tumor proximity,

302 whereas M2-like macrophages C1, proliferating myeloid subsets and cDC2 cells were

303 steadily enriched in the tumors **(Fig. 4A and 4C; Supplementary Fig. S4B and S4C).**

304

305 We next performed subclustering analysis of monocytes and macrophages (n = 27,664

306 cells) which revealed five distinct subclusters, and confirmed the unique enrichment of

307 M2-like macrophages C1 in the LUAD tissues **(Fig. 4D; Supplementary Tables S14)**.

308 Further, we characterized monocyte and macrophage subsets by antigen presentation

309 capability (i.e. antigen presentation signature score), and found that the M2-like

310 macrophages C1 showed diminished antigen presentation scores compared to cells of

311 M2-like macrophages C5 which were mainly enriched in normal samples **(Fig. 4D;**

312 **Supplementary Table S16)**. In addition, we interrogated the spatial expression of

313 antigen presentation-related genes in this cluster. We found markedly reduced

314 expression levels of antigen presentation signature genes, including MHC class I and

315 MHC class II members with increasing spatial proximity to the LUADs **(Fig. 4E and 4F)**.

316 Furthermore, the spatially-driven pattern of macrophage antigen presentation depletion

11

317   was evident across M2-like macrophages combined from both clusters C1 and C5 and

318   was statistically significant in 4 of the 5 LUAD patient fields **(Fig. 4G)**.

319

320   To characterize spatial DC dynamics in early-stage LUAD with respect to cellular states,

321   we first examined gene and signature score differences between different subsets of

322   DCs (n = 8,694). Spatial field patterns were evident in cDC2 and pDC subsets **(Fig. 4H)**.

323   We observed differential expression of an inflammatory gene signature between the

324   three cDC2 subclusters with C1, cells of which exhibited the lowest inflammatory

325   scores, heavily enriched in the LUADs **(Fig. 4I and 4J)**. In addition, unsupervised

326   subclustering analysis of cDC2 cells using a different approach (SC3, see Methods)

327   identified three distinct cDC2 subclusters, characterized by differential expression of the

328   chronic inflammatory gene signature (highly enriched in C2) and MHC class II genes

329   (enriched in C0/1) **(Supplementary Fig. S4D; Supplementary Table S17)** that were

330   previously shown to discriminate inflammatory from non-inflammatory cDCs (21).

331   Reduced expression of pro-inflammatory genes and increased expression of anti-

332   inflammatory genes was evident in all cDC2 cells along the continuum of normal-to-

333   LUAD space **(Fig. 4K).** cDC2 subcluster with the highest inflammatory score (C2) was

334   markedly under-represented in the LUADs and was further characterized by a spatially

335   diluted inflammatory signature score with increasing tumor proximity **(Fig. 4L and 4M**).

336   Notably, the inflammatory signature score significantly and progressively decreased

337   along the continuum from normal lung tissues, to matched premalignant AAHs and

338   invasive LUADs (**Fig. 4N**) in sharp contrast to non-inflammatory DC expression

339   components **(Supplementary Fig. S4E).** We also studied pDC subsets and found

340   spatial enrichment of *FOS, FOSB,* and *JUN* with increasing proximity to the tumors

341   **(Supplementary Fig. S4F and S4G)**. Altogether, these data describe spatial immune

342   remodeling dynamics comprising gradual loss of antigen presentation in macrophage

343   subclusters and of inflammatory phenotypes in subsets of DCs in early-LUAD

344   microenvironment.

345

346   We also identified multiple stromal populations in our dataset **(Supplementary Fig.**

347   **S5A-S5C)**, including tumor enrichment of vascular and airway smooth muscle cells,

12

348   adventitial fibroblasts C5, and endothelial cell (EC) venule clusters **(Supplementary**

349   **Fig. S5D and S5E)**. We pinpointed significantly differentially expressed gene sets in the

350   EC venule subpopulation and altered stromal signatures **(Supplementary Fig. S5F)**

351   that support the observed immune-related changes in the LUAD field, and that are in

352   line with previous observations in a cohort of lung cancers of multiple subtypes (10).

353   These comprised tumor-specific activation of extracellular matrix reorganization,

354   syndecan-2 pathway, and neutrophil degranulation, as well as decreased JAK-STAT

355   signaling and reduced antigen-processing cross-presentation **(Supplementary Fig.**

356   **S5G and S5H)**.

357

358   **Ligand-receptor based cell-cell communication networks**

359   Crosstalk between tumor cells and elements in the tumor microenvironment, such as

360   immune cells, has been implicated in the progression of solid tumors largely in part by

361   mediating immune-evasive and pro-tumor phenotypes (22). To profile cell-cell

362   communication networks in early-stage LUAD space, we utilized iTALK (23) to leverage

363   signals from our scRNA-seq dataset and visualize ligand-receptor (L-R)-mediated

364   intercellular cross-talk that is spatially modulated within the tumor microenvironment

365   **(Fig. 5A; Supplementary Table S18)**. Computational analysis and annotation were

366   carried out using the iTALK's built-in database focusing on immune checkpoint-receptor

367   pairs (n = 55) and cytokine-receptor pairs (n = 327) **(Fig. 5A)**. Overall, we found

368   reduced overlap of L-R interactions between the tumor and distance normal tissues

369   than that between the tumor and more proximal (adjacent, intermediate) regions (**Fig.**

370   **5B**).

371

372   By comparative analysis, we identified altered cellular interactions that were significantly

373   and differentially increased or decreased in LUADs *versus* their respective spatial

374   normal tissues (**Fig. 5C; Supplementary Fig. S6A)**. Specifically, we noted increased

375   interactions between immune checkpoint proteins *CD24* or *LGALS9 (Galectin-9)* on

376   tumor epithelial cells, and *SIGLEC10* on macrophages or *HAVCR2 (TIM-3)* on DCs,

377   respectively, and which were shared across multiple patients (**Fig. 5C and 5D**;

378   **Supplementary Fig. S6A and S6B; Supplementary Table S19)**. These interactions

13

379   were differentially enriched in tumors *versus* normal tissues at different distances from

380   the tumor **(Fig. 5C; Supplementary Fig. S6A)**. Expression levels of the checkpoints

381   *CD24* and *LGALS9 (Galectin-9)* were increased in epithelial cells from LUAD tissues in

382   general, and in malignant epithelial cells in particular, and accompanied by increased

383   expression of their cognate predicted ligands *SIGLEC10* and *HAVCR2 (TIM-3)*,

384   respectively, on myeloid cells and with increasing tumor proximity (**Fig. 5D;**

385   **Supplementary Fig. S6B).** These interactions suggest a spatial pattern of modulation

386   in cell-cell communication that is mediated by epithelial inhibitory checkpoint signaling to

387   myeloid cells. Targeted analysis of *CD24* expression in an external cohort using the

388   Nanostring platform revealed progressively and markedly increased expression of the

389   antigen across normal lung tissues, AAHs, and LUADs **(Fig. 5F)**. *CD24* positively

390   correlated with expression of the epithelial marker *EPCAM* as well as with levels of pro-

391   tumor and immune suppressive features (*TIGIT, CTLA4, FOXP3, CCL19*)**,** in contrast to

392   negatively correlating with anti-tumor immune markers (*GZMB, GZMH, PRF1*) **(Fig. 5F;**

393   **Supplementary Fig. S6G).** These pro-tumor and epithelial patterns of *CD24* were

394   further validated in LUADs (n = 51) from The Cancer Genome Atlas (TCGA) cohort (24)

395   compared to matched adjacent normal lung tissues **(Fig. 5G and 5H; Supplementary**

396   **Fig. S6H)**, thereby suggesting a potential role for epithelial *CD24* signaling in pro-tumor

397   immune reprogramming in LUAD evolution.

398

399   We also identified cytokine-receptor pairs that were differentially modulated in the LUAD

400   space, such as communication between tumor epithelial cells with high expression of

401   the cytokine *CX3CL1*, and DCs or macrophages expressing increased levels of its

402   cognate receptor *CX3CR1* **(Supplementary Fig. S6C, S6D, S6E, S6F;**

403   **Supplementary Table S18)**. *CX3CR1* was increasingly expressed on macrophages

404   and DCs but decreased on CD8 T cells of tumor samples **(Fig. 5F)**, in line with previous

405   reports implicating macrophage *CX3CR1* with pro-tumor features (25). These findings

406   signify a potential spatial enrichment of pro-tumorigenic and tumor-mediated cytokine

407   signaling to macrophages infiltrating the LUAD space.

408

409

DISCUSSION

Previously, spatial molecular changes have been documented in the local niche of LUAD including loss-of-heterozygosity in 3p and 9p, point mutations and tumor suppressor methylation (26,27). Earlier work also underscored transcriptome profiles, somatic driver variants, as well as genome-wide allelic imbalance that are shared between lung cancer and adjacent normal-appearing airway cells but that are absent in distant normal cells, thereby pointing to putative drivers of lung oncogenesis (4,5). As such, the interrogation of spatial dynamics in LUADs has been a topic of interest, particularly owing to its utility as a surrogate for studying difficult-to-obtain longitudinal patient (including precursor) samples. However, much of these studies have focused on understanding, by bulk profiling methods, spatiotemporal profiles in an aggregate of cells, thereby inadvertently obscuring the individual contributions of epithelial and tumor microenvironmental cues to the pathogenesis of lung cancer. Our knowledge of the spatial architecture of individual cell populations in the evolution of this malignancy remains poorly understood. By single-cell interrogation of multi-region samples, we here characterized spatial and ecological maps comprising many epithelial and non-epithelial subsets and underlying emergence of early-stage LUAD from its local niche.

Multi-region or spatial analyses have been employed to interrogate intratumor heterogeneity (ITH) in solid tumors including LUADs, in order to understand evolutionary trajectories and therapy response (28). Cohorts based on multi-region sampling and microdissection strategies have increased the resolution for detecting subtle, rare and subclonal events by bulk sequencing approaches. Our analyses here showed that ITH is evident at the tumor epithelial cell and intra-site level, i.e. within the same tumor region or tissue. We applied an integrative approach to dissect ITH of malignant cells and characterized cell clusters with differential transcriptomic profiles, evolution trajectories, CNV burdens, and/or driver mutations. We also found "normal" cells in the LUAD tissues themselves that are close in the inferred trajectory paths to specific malignant cell subsets (e.g. club/secretory cells for the mucinous LUAD in P2) and perhaps represent the cells-of-origin for tumor cells. It is noteworthy that we found

15

441 in the normal-appearing samples cells with features of malignant-enriched subsets and

442 demonstrating heterogeneous CNV profiles. Whether these cells comprise early LUAD

443 precursors, mutagenic clones that do not progress to malignancy (e.g. negatively

444 selected by the immune microenvironment), or putative molecular field cancerization

445 remains to be investigated. It is important to mention that LUAD exhibits remarkable

446 inter-patient heterogeneity in histological differentiation patterns, driver alterations and

447 suggested tumor cells-of-origin (29). We could not characterize the full spectrum in the

448 heterogeneity of these features (e.g., programs in tumor cells-of-origin of different

449 lineages) given the limited number of patients profiled. Nonetheless, our in-depth

450 analysis of a relatively large number of epithelial cells unveiled different characteristics

451 (airway lineage trajectories, malignant cell programs, potential tumor cells-of-origin, and

452 cellular ITH) of the epithelial architecture of early-stage LUAD development that could

453 be extended in the future to a larger and more diverse array of LUADs.

454

455 We pinpointed G12D mutations in *KRAS*, the most frequently mutated oncogene in

456 LUAD, to a unique subset of malignant cells within P2 tumor. These cells had overall

457 low copy number alterations, perhaps reminiscent of findings in LUADs driven by strong

458 driver genes *in vivo* (30). These cells comprised increased expression of genes

459 associated with *KRAS*-mutant cancer such as *LCN2* which is a marker of inflammation

460 (17,31) and elevated levels of genes associated with hyper-mutation such as *CDA* (32).

461 *KRAS*-mutant cells, compared to other malignant-enriched cells from P2, exhibited

462 reduced expression of the lineage-specific oncogene *NKX2-1* (33). Earlier studies

463 demonstrated that increased LCN2 and reduced NKX2-1 is observed in mucinous

464 *KRAS*-mutant LUADs or those undergoing gastric differentiation (17,18), which is in line

465 with the histological variant of P2 LUAD. It is noteworthy that *NKX2-1* exhibits context-

466 dependent roles in LUAD development. While Nkx2-1 was shown to enhance *Egfr*-

467 driven lung tumorigenesis (34), Winslow et al demonstrated that the transcription factor

468 suppresses *Kras*-mutant LUAD by controlling tumor differentiation as well as limiting

469 progression and metastatic potential (35). This is in line with our finding of reduced

470 expression of airway lineage-specific genes (e.g. *SCGB3A1*, *SFTPB*) suggestive of loss

471 of differentiation in the *KRAS*-mutant cells compared to other malignant-enriched cell

472 subclusters from the same LUAD tissue. Our observations are also in accordance with

473 the acquisition of a mixed-lineage phenotype by malignant lung epithelial cells recently

474 reported by Laughney and colleagues (8). Our findings also allude to the possibility that

475 tumor cell lineage plasticity may occur at the early stage of *KRAS*-mutant LUAD

476 carcinogenesis – a supposition that can be better explored with sequencing of a larger

477 and more diverse repertoire of *KRAS*-mutant cells.

478

479 Earlier studies have shown that immunosuppressive T regulatory cells are crucial for

480 immune evasion in lung cancer (36,37). By single-cell analysis of diverse lymphoid

481 subsets, we found Tregs co-expressing both *TIGIT* and *CTLA-4* immune checkpoints

482 and that were progressively enriched with increasing geospatial proximity to the LUADs

483 – suggesting a value in combinatorial targeting of multiple checkpoints for

484 immunotherapy of early-stage LUAD. We pinpointed B cell signatures implicated in actin

485 remodeling and immunological synapse formation that are spatially enriched in the

486 LUADs and progressively increased along the course of normal to preneoplasia and

487 invasive LUAD. These data suggest important yet unexplored roles for B cell

488 phenotypes in immune evolution of LUAD. Our spatial single-cell analysis also pointed

489 to mechanisms by which the myeloid immune microenvironment permits LUAD

490 pathogenesis. Our interrogation of macrophage subsets revealed downregulation/loss

491 of antigen presentation in tumor-specific M2-like macrophages, whereby expression

492 levels of MHC genes as well as genes involved in peptide transport and loading (*TAP1,*

493 *TAP2, TAPBP;* (38)) were markedly reduced. Our study also unveiled relatively

494 understudied immune subsets in LUAD. We found DC subclusters and programs that

495 were recently reported in other sites (21), including a non-inflammatory cDC signature

496 with higher expression of MHC class II genes and that is enriched in LUAD relative to

497 matched preneoplastic or normal tissues, thereby possibly informing of immune cues

498 that could be harnessed to manipulate the immunogenicity of tumors. We also identified

499 an inflammatory tumor-depleted cDC cluster with increased expression of *CD163*,

500 *CD36*, *CD14*, *S100A9*, and *S100A8*. Interestingly, dendritic CD36 is a scavenger for the

501 acquisition and presentation of cell surface antigens and subsequent sculpting of the

502 Treg repertoire (39), yet the precise consequences of its stark absence in a LUAD-

17

503 specific cDC subset and the effect(s) on the tumor immune microenvironment warrant

504 further investigation. Nevertheless, these findings, along with our observation of

505 decreased expression of the inflammatory DC signature across normal lung to

506 premalignant AAH up to LUAD, suggest that unique cDC subsets play critical roles in

507 LUAD pathogenesis and, thus, could be potential targets for immune-based

508 interception.

509

510 Ligand-receptor mediated interactions are known to be the main forms of interactions

511 responsible for cellular signaling mechanisms and cell-to-cell communication, which is

512 especially important during carcinogenesis and immune response (22). A better

513 understanding of cell-to-cell communication networks, particularly ligand-receptor

514 mediated interactions, may help elucidate tumor-immune co-evolution and immune

515 reprogramming, and thus, help identify potential therapeutic targets. However, this has

516 been hampered by the lack of bioinformatics tools for efficient data analysis and

517 visualization. In this study, we applied the iTALK tool (23) developed by our group and

518 performed a deep analysis of cellular interaction networks. We identified significant

519 immune checkpoint- (e.g. *CD24*, *Galectin-9*, or *TIM-3)* and cytokine- (e.g. *CX3CL1)*

520 receptor interactions whose enrichment or depletion in the LUAD space signified a

521 highly pro-tumorigenic milieu. These findings are in accordance with the CD24–Siglec-

522 10 interaction and subsequent "do not eat me" signal recently highlighted in breast

523 cancer (40). Additionally, our findings of prominently increased expression of *CD24* in

524 hyperplastic precursors (AAH) and LUADs compared to normal lung, and its association

525 with pro-tumor and immunosuppressive phenotypes in this pathologic continuum,

526 propose immune functions for CD24 in lung cancer that extend beyond its prognostic

527 role (41) and that can be harnessed for early intervention. Overall, our interactome

528 analyses implicate novel or understudied immune subsets as culprits in LUAD

529 pathogenesis and provide a rationale for expanding the armamentarium of immune-

530 based cancer therapies.

531

532 In summary, our results provide a spatial atlas of early-stage LUAD and its nearby and

533 distant lung ecosystem. This atlas comprises high cellular heterogeneity as well as

18

534   spatial dynamics in cell populations, cell states and their transcriptomic features that

535   underlie evolution of the LUAD from the peripheral lung ecosystem. Our extensive

536   transcriptomic dataset of lung epithelial and immune cells, other populations such as

537   stromal and endothelial subsets, as well as of tumor-pertinent cell-cell interactions

538   constitutes a valuable resource to functionally interrogate LUAD trajectories at high

539   resolution and generate strategies for its early treatment. Also, our study's multi-region

540   sampling design in conjunction with single-cell analysis could help address specific

541   questions in early malignant and immune biology of other solid tumors.

542

543

544   **METHODS**

545   *Additional description of methods can be found in the Supplementary Data file.*

546

547   **Multi-region sampling of surgically resected LUADs and spatial normal lung**

548   **tissue specimens**

549   Patients undergoing surgical resection for primary early-stage lung adenocarcinoma (I-

550   IIIA) and who had not received any preoperative therapy were carefully selected for

551   derivation of multi-region samples for single cell analysis (**Supplementary Table S1**).

552   All patients were evaluated at the University of Texas MD Anderson Cancer Center and

553   had provided informed consents under approved institutional review board protocols.

554   Immediately following surgery, resected tissues were processed on ice by an

555   experienced pathologist assistant (PB). One side of the specimen was documented and

556   measured, followed by identification of tumor margin, mapping defined collection sites in

557   one direction along the length of the specimen and spanning the entire lobe, and

558   making incisions into the surface of the lobe for specimen collection according to the

559   following criteria. Based on the placement of the tumor within the specimen, tumor-

560   adjacent and -distant normal parenchyma specimen were assigned at 0.5 cm from the

561   tumor edge and from the periphery of the overall specimen/lobe (furthest distance from

562   the tumor), respectively. An additional tumor-intermediate normal tissue was selected

563   for P2-5 that ranged between 3-5 cm from the edge of the tumor. Sample collection was

564   initiated at normal lung tissues that are farthest from the tumor moving inwards towards

565   the tumor to minimize cross-contamination during collection, with each sample being

566   immediately placed in ice-cold RPMI medium supplemented with 2% fetal bovine

567   serum, and transported to the processing lab on ice.

568

569   Single cell derivation and scRNA-seq processing

570   Tumor and spatial normal parenchyma tissues (n = 19 samples) were immediately

571   minced and enzymatically digested, as previously described ((42), Supplementary

572   Data). Following red blood cell removal, cells were filtered and counted. Cells were

573   sorted (by FACS) for viable singlets cells (and also EPCAM+/- fractions from P2-P5).

574   Sorted fractions were then loaded on individual channels of 10X Chromium Microfluidic

575   chips, targeting up to 10,000 cells per fraction. Gene expression libraries were

576   generated according to manufacturer's instructions using Chromium Next GEM Single

577   Cell 5' Gel Bead Kit v1.1 (1000169, 10X Genomics), and sequenced on Illumina

578   NovaSeq 6000 platform (Supplementary Data).

579

580   scRNA-seq data analysis

581   Single-cell analyses were performed using available computational framework. Raw

582   scRNA-seq data were pre-processed, demultiplexed, aligned to human GRCh38

583   reference and feature-barcodes generated using CellRanger (10X Genomics, version

584   3.0.2). Details of quality control including quality check, data filtering, identification and

585   removal of cellular debris, doublets and multiplets, batch effect evaluation and

586   correction are found in Supplementary Methods. Following quality filtering, a total of

587   186,916 cells were retained for downstream analysis. Raw unique molecular identifier

588   (UMI) counts were log normalized and used for principal component analysis using

589   Seurat (43). The output was then processed by Harmony (44) for batch effect correction,

590   followed by unsupervised clustering analysis using Seurat (43). Uniform Manifold

591   Approximation and Projection (UMAP) clustering (45) was used for visualization.

592   *EPCAM+* cells were partitioned into major airway lineage clusters, followed by

593   subclustering within each compartment/lineage to identify subpopulations. In addition,

594   we applied single-cell consensus clustering (SC3) approach (46) for unsupervised

595 clustering analysis. SC3 was run with default parameters and independent of cell

596 lineage annotation. Differentially expressed genes (DEGs) for each cell cluster were

597 identified using the *FindAllMarkers* function in Seurat R package. We applied an

598 integrative approach to define the cell type and determine its functional state of each

599 cluster, based on the enrichment of canonical marker genes, top-ranked DEGs in each

600 cell cluster, and the global cluster distribution, as previously described (47). For

601 transcriptional signature analysis, single-sample GSVA (ssGSVA) was applied to the

602 scRNA-seq data and pathway scores were calculated for each cell using gsva function

603 in GSVA software package (48).

604 To study hierarchical relationships among cell types identified in this study, pairwise

605 Spearman correlations were calculated from average expression level (*Seurat* function

606 *AverageExpression*) of each cell type, based on which Euclidean distances between

607 cell types were calculated. Unsupervised hierarchical cluster analysis was performed

608 using the R function *hclust*, and the dendrogram was drawn using R package

609 dendextend. Monocle 2 (version 2.10.1) (49) was applied to construct single-cell

610 trajectories. The tool inferCNV (https://github.com/broadinstitute/inferCNV) was applied

611 to infer large-scale CNVs from scRNA-seq data using NK cells from the same dataset

612 as a control. Malignant cells were distinguished from non-malignant subsets based on

613 information integrated from multiple sources including cluster distribution of the cells,

614 marker genes expression, inferred large-scale CNVs, and presence of the *KRAS*

615 mutation. For single-cell somatic *KRAS* mutation analysis, the reads were extracted

616 from the original BAM files using cell-specific barcodes and genomic coordinates of

617 *KRAS* hot spot mutations and then subjected to quality filtering and duplicates removal,

618 followed by mutation identification and annotation. Extracted alignments were manually

619 evaluated using IGV (50). To identify significant ligand-receptor pairs among major cell

620 lineages (CD4 T cells, CD8 T cells, B cells, macrophages, dendritic cells, endothelial

621 cells, epithelial cells and fibroblasts), the top 30% of most highly expressed genes were

622 included in the analysis. Significant cellular interactions were identified using iTALK as

623 previously described (23). For ligand-receptor annotation, the iTALK built-in ligand-

624 receptor database was used.

21

625

Statistical analysis

627 In addition to the bioinformatics approaches described above for scRNA-seq data

628 analysis, all other statistical analyses were performed using R package v3.6.0. Pseudo-

629 bulk gene expression values for defined cell clusters were calculated by taking mean

630 expression of each gene across all cells in a specific cluster. Pearson's correlation

631 analysis was used to identify genes significantly correlated with *CD24* expression. All

632 statistical significance testing was two-sided, and results were considered statistically

633 significant at p - value < 0.05. The Benjamini-Hochberg method was applied to control

634 the false discovery rate (FDR) in multiple comparisons and to calculate adjusted p -

635 value (q-values).

636

Data Availability

638 All sequencing data generated in this study will be deposited in the Gene Expression

639 Omnibus.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA: a cancer journal for clinicians **2018**;68(1):7-30 doi 10.3322/caac.21442.
2. Goldstraw P, Ball D, Jett JR, Le Chevalier T, Lim E, Nicholson AG*, et al.* Non-small-cell lung cancer. Lancet (London, England) **2011**;378(9804):1727-40 doi 10.1016/s0140-6736(10)62101-0.
3. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM*, et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. The New England journal of medicine **2011**;365(5):395-409 doi 10.1056/NEJMoa1102873.
4. Abbosh C, Venkatesan S, Janes SM, Fitzgerald RC, Swanton C. Evolutionary dynamics in pre-invasive neoplasia. Current opinion in systems biology **2017**;2:1-8 doi 10.1016/j.coisb.2017.02.009.
5. Saab S, Zalzale H, Rahal Z, Khalifeh Y, Sinjab A, Kadara H. Insights Into Lung Cancer Immune-Based Biology, Prevention, and Treatment. Front Immunol **2020**;11:159 doi 10.3389/fimmu.2020.00159.
6. Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL*, et al.* Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. Cell **2020** doi https://doi.org/10.1016/j.cell.2020.07.017.
7. Lavin Y, Kobayashi S, Leader A, Amir ED, Elefant N, Bigenwald C*, et al.* Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. Cell **2017**;169(4):750-65.e17 doi 10.1016/j.cell.2017.04.014.
8. Laughney AM, Hu J, Campbell NR, Bakhoum SF, Setty M, Lavallée V-P*, et al.* Regenerative lineages and immune-mediated pruning in lung cancer metastasis. Nature Medicine **2020**;26(2):259-69 doi 10.1038/s41591-019-0750-6.
9. Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW*, et al.* Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. Nature Communications **2020**;11(1):2285 doi 10.1038/s41467-020-16164-1.
10. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O*, et al.* Phenotype molding of stromal cells in the lung tumor microenvironment. Nature Medicine **2018**;24(8):1277-89 doi 10.1038/s41591-018-0096-5.
11. Plasschaert LW, Žilionis R, Choo-Wing R, Savova V, Knehr J, Roma G*, et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature **2018**;560(7718):377-81 doi 10.1038/s41586-018-0394-6.
12. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH*, et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature **2014**;509(7500):371-5 doi 10.1038/nature13173.
13. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K*, et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature **2016**;539(7628):309-13 doi 10.1038/nature20123.
14. Little DR, Gerner-Mauro KN, Flodby P, Crandall ED, Borok Z, Akiyama H*, et al.* Transcriptional control of lung alveolar type 1 cell development and maintenance

by NK homeobox 2-1. Proc Natl Acad Sci U S A **2019**;116(41):20545-55 doi 10.1073/pnas.1906663116.

15. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. Nature **2014**;507(7491):190-4 doi 10.1038/nature12930.

16. Finn J, Sottoriva K, Pajcini KV, Kitajewski JK, Chen C, Zhang W*, et al.* Dlk1-Mediated Temporal Regulation of Notch Signaling Is Required for Differentiation of Alveolar Type II to Type I Cells during Repair. Cell Rep **2019**;26(11):2942-54.e5 doi 10.1016/j.celrep.2019.02.046.

17. Treekitkarnmongkol W, Hassane M, Sinjab A, Chang K, Hara K, Rahal Z*, et al.* Augmented Lipocalin-2 is Associated with COPD and Counteracts Lung Adenocarcinoma Development. Am J Respir Crit Care Med **2020** doi 10.1164/rccm.202004-1079OC.

18. Skoulidis F, Byers LA, Diao L, Papadimitrakopoulou VA, Tong P, Izzo J*, et al.* Co-occurring genomic alterations define major subsets of KRAS-mutant lung adenocarcinoma with distinct biology, immune profiles, and therapeutic vulnerabilities. Cancer Discov **2015**;5(8):860-77 doi 10.1158/2159-8290.CD-14-1236.

19. Arana E, Vehlow A, Harwood NE, Vigorito E, Henderson R, Turner M*, et al.* Activation of the small GTPase Rac2 via the B cell receptor regulates B cell adhesion and immunological-synapse formation. Immunity **2008**;28(1):88-99 doi 10.1016/j.immuni.2007.12.003.

20. Sivakumar S, Lucas FAS, McDowell TL, Lang W, Xu L, Fujimoto J*, et al.* Genomic Landscape of Atypical Adenomatous Hyperplasia Reveals Divergent Modes to Lung Adenocarcinoma. Cancer research **2017**;77(22):6119-30 doi 10.1158/0008-5472.Can-17-1605.

21. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J*, et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science (New York, NY) **2017**;356(6335) doi 10.1126/science.aah4573.

22. Maman S, Witz IP. A history of exploring cancer in context. Nat Rev Cancer **2018**;18(6):359-76 doi 10.1038/s41568-018-0006-7.

23. Wang Y, Wang R, Zhang S, Song S, Jiang C, Han G*, et al.* iTALK: an R Package to Characterize and Illustrate Intercellular Communication. bioRxiv **2019**:507871 doi 10.1101/507871.

24. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature **2014**;511(7511):543-50 doi 10.1038/nature13385.

25. Schmall A, Al-Tamari HM, Herold S, Kampschulte M, Weigert A, Wietelmann A*, et al.* Macrophage and cancer cell cross-talk via CCR2 and CX3CR1 is a fundamental mechanism driving lung cancer. Am J Respir Crit Care Med **2015**;191(4):437-47 doi 10.1164/rccm.201406-1137OC.

26. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA*, et al.* Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res **2002**;62(8):2370-7.

27. Powell CA, Klares S, O'Connor G, Brody JS. Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. Clin Cancer Res **1999**;5(8):2025-34.

28. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J*, et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science **2014**;346(6206):256-9 doi 10.1126/science.1256930.

29. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong K-K. Non-small-cell lung cancers: a heterogeneous set of diseases. Nature reviews Cancer **2014**;14(8):535-46 doi 10.1038/nrc3775.

30. McFadden DG, Politi K, Bhutkar A, Chen FK, Song X, Pirun M*, et al.* Mutational landscape of EGFR-, MYC-, and Kras-driven genetically engineered mouse models of lung adenocarcinoma. Proceedings of the National Academy of Sciences of the United States of America **2016**;113(42):E6409-E17 doi 10.1073/pnas.1613601113.

31. Moschen AR, Gerner RR, Wang J, Klepsch V, Adolph TE, Reider SJ*, et al.* Lipocalin 2 Protects from Inflammation and Tumorigenesis Associated with Gut Microbiota Alterations. Cell Host Microbe **2016**;19(4):455-69 doi 10.1016/j.chom.2016.03.007.

32. Zauri M, Berridge G, Thézénas ML, Pugh KM, Goldin R, Kessler BM*, et al.* CDA directs metabolism of epigenetic nucleosides revealing a therapeutic window in cancer. Nature **2015**;524(7563):114-8 doi 10.1038/nature14948.

33. Snyder EL, Watanabe H, Magendantz M, Hoersch S, Chen TA, Wang DG*, et al.* Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. Mol Cell **2013**;50(2):185-99 doi 10.1016/j.molcel.2013.02.018.

34. Yamaguchi T, Hosono Y, Yanagisawa K, Takahashi T. NKX2-1/TTF-1: an enigmatic oncogene that functions as a double-edged sword for cancer cell survival and progression. Cancer Cell **2013**;23(6):718-23 doi 10.1016/j.ccr.2013.04.002.

35. Winslow MM, Dayton TL, Verhaak RG, Kim-Kiselak C, Snyder EL, Feldser DM*, et al.* Suppression of lung adenocarcinoma progression by Nkx2-1. Nature **2011**;473(7345):101-4 doi 10.1038/nature09881.

36. Quezada SA, Peggs KS, Simpson TR, Allison JP. Shifting the equilibrium in cancer immunoediting: from tumor tolerance to eradication. Immunol Rev **2011**;241(1):104-18 doi 10.1111/j.1600-065X.2011.01007.x.

37. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q*, et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med **2018**;24(7):978-85 doi 10.1038/s41591-018-0045-3.

38. Seliger B, Maeurer MJ, Ferrone S. TAP off--tumors on. Immunol Today **1997**;18(6):292-9 doi 10.1016/s0167-5699(97)01052-9.

39. Perry JSA, Russler-Germain EV, Zhou YW, Purtha W, Cooper ML, Choi J*, et al.* Transfer of Cell-Surface Antigens by Scavenger Receptor CD36 Promotes Thymic Regulatory T Cell Receptor Repertoire Development and Allo-tolerance. Immunity **2018**;48(5):923-36.e4 doi 10.1016/j.immuni.2018.04.007.

40. Barkal AA, Brewer RE, Markovic M, Kowarsky M, Barkal SA, Zaro BW*, et al.* CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. Nature **2019**;572(7769):392-6 doi 10.1038/s41586-019-1456-0.

41. Kristiansen G, Schlüns K, Yongwei Y, Denkert C, Dietel M, Petersen I. CD24 is an independent prognostic marker of survival in nonsmall cell lung cancer patients. British Journal of Cancer **2003**;88(2):231-6 doi 10.1038/sj.bjc.6600702.

42. Slyper M, Porter CBM, Ashenberg O, Waldman J, Drokhlyansky E, Wakiro I*, et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. Nature Medicine **2020**;26(5):792-802 doi 10.1038/s41591-020-0844-1.

43. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol **2015**;33(5):495-502 doi 10.1038/nbt.3192.

44. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K*, et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. Nature Methods **2019**;16(12):1289-96 doi 10.1038/s41592-019-0619-0.

45. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG*, et al.* Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol **2018** doi 10.1038/nbt.4314.

46. Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T*, et al.* SC3: consensus clustering of single-cell RNA-seq data. Nature Methods **2017**;14(5):483-6 doi 10.1038/nmeth.4236.

47. Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, de Boer CG, Jenkins RW*, et al.* Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. Cell **2018**;175(4):998-1013.e20 doi 10.1016/j.cell.2018.10.038.

48. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics **2013**;14(1):7 doi 10.1186/1471-2105-14-7.

49. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ*, et al.* The single-cell transcriptional landscape of mammalian organogenesis. Nature **2019**;566(7745):496-502 doi 10.1038/s41586-019-0969-x.

50. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. Cancer Res **2017**;77(21):e31-e4 doi 10.1158/0008-5472.Can-17-0337.

FIGURE LEGENDS

Figure 1. Dissecting early-stage LUAD and the peripheral lung ecosystem using single-cell RNA sequencing. A, Workflow showing multi-region sampling strategy of five LUADs and fourteen spatially defined normal lung tissues for analysis by scRNA-seq. Dis, distant normal; Int, intermediate normal; Adj, adjacent normal; LUAD, tumor tissue. **B**, Uniform manifold approximation and projection (UMAP) embedding of cells from tumor, adjacent normal and distant normal samples of patient one (P1). Cells are colored by their inferred cell types. **C**, Cell composition in absolute cell numbers (stacked bar plots) and relative fractions (pie charts) in each spatial field sample derived from P1. **D**, UMAP view of cells from all five patients, including EPCAM+ and EPCAM-pre-enriched cells from P2-P5. Colors represent assigned major cell types. Cycling, cycling cells. **E**, UMAP view of cell types by spatial fields. Colors represent assigned cell types as in panel D. Lollipop plots indicate, for each major cell type, the relative fraction of cells from each spatial field. Spatial fields are indicated by numbers according to panel A. **F**, Dendrograms showing hierarchical relationships of cells among the spatial fields based on the computed Euclidean distance using transcriptomic features. Dendrograms are shown for five major cell types (from left to right), for all patients together (top) and by patient (bottom). **G**, Same UMAP as in panel D, with further subclustering of lymphoid and myeloid cells. Colors correspond to the cell type annotation in panel H for EPCAM- cells. **H-I**, An area plot showing changes in the relative fractions among the EPCAM- subsets across spatial fields for all patients together and by patient (pie charts, panel I). Stacked bar plots in panel I show absolute cell numbers of the fractions by patient and spatial field**.**

Figure 2. Epithelial lineage diversity and intratumoral heterogeneity in the spatial ecosystem of early-stage LUAD. A, UMAP visualization of all EPCAM+ cells from P1-P5 colored (from left to right) by their assigned cell types, spatial fields, and inferred copy number variation (inferCNV) scores. **B**, Heatmap of major lineage marker genes for EPCAM+ cell clusters (C1-10 as shown in panel A, left), with corresponding bar blots

27

836   outlining relative contribution by spatial field. **C**, Area plot showing changes in EPCAM+

837   subset fractions across spatial fields. **D**, Representative line plots displaying changes in

838   abundance of non-malignant epithelial cell clusters across spatial fields (identified by

839   the colors of the circles, in the same order as in panel C), with embedded pie charts

840   showing the relative contribution of each spatial field. **E**, Hierarchical relationships of 3

841   representative subsets of epithelial cells (from top to bottom) among the spatial fields

842   based on the computed Euclidean distance using transcriptomic features (left), and

843   corresponding heatmaps quantifying similarity levels among spatial fields (right).

844   Similarity score is defined as one minus the Euclidean distance. **F**, UMAP plots of cells

845   in the malignant-enriched cluster C9 (panel A), colored by their corresponding patient

846   origin (left), spatial field (middle), and inferCNV score (right). The zoom in view of the

847   right panel shows *KRAS* G12D mutant cells in P2. **G**, Fraction of cells carrying *KRAS*

848   G12D mutation (left bar plot), with numbers indicating the absolute cell numbers, as well

849   as expression levels of *KRAS* (violin plot, top right) and *MUC5AC* (violin plot, bottom

850   right), within cells of each epithelial lineage cluster of P2. **H**, Unsupervised clustering of

851   CNV profiles inferred from scRNA-seq data from patient P3 (left) and P5 (right) tumor

852   samples and demonstrating intratumoral heterogeneity in CNV profiles. Chromosomal

853   amplifications (red) and deletions (blue) are inferred for all 22 chromosomes (color bars

854   on the top). Each row represents a single cell, with corresponding cell type annotated

855   on the right (same as in panel A). **I**, Potential developmental trajectories for EPCAM+

856   cells from P3 (top) and P5 (bottom) inferred by Monocle 3 analysis. Cells on the tree are

857   colored by pseudotime (dotted boxes) and inferCNV clusters.

858

859   **Figure 3. Spatial reprogramming of lymphoid subsets towards protumor**

860   **phenotypes in early-stage LUAD. A**, UMAP visualization of lymphoid cell subsets

861   from P1-P5 colored by cell lineage (left) and spatial field (right). CTL, cytotoxic T

862   lymphocyte; Treg, T regulatory cell; ILC, innate lymphoid cell; NK, natural killer cell. **B**,

863   Bubble plot showing the expression of lineage markers. Both the fraction of cells

864   expressing (indicated by the size of the circle) as well as their scaled expression levels

865   (indicated by the color of the circle) are shown. **C**, Changes in the abundance of

866   lymphoid cell lineages and cellular states across the LUADs and spatial normal

867   samples. Embedded pie charts show the contribution of each spatial sample to the

868   indicated cell subtype/state. **D**, UMAP plots of CD8+ T lymphocytes colored by cell

869   states (top), spatial field (middle), and cytotoxic score (bottom). The heatmap on the

870   right shows normalized expression of marker genes for defined CD8 T cell subsets.

871   Each column represents a cell. Top annotation tracks indicate (from top to bottom) cell

872   states, naïve T cell scores and cytotoxic scores calculated using curated gene

873   signatures, and the corresponding spatial field of each cell. **E**, Depletion of CD8+

874   GNLY-hi CTLs in the tumor microenvironment of LUADs. Bar plot (top left) and boxplot

875   (top middle) showing percentage of CD8+ GNLY-hi CTLs among total CD8+ cells from

876   all patients across the spatial fields. Each circle in the boxplot represents a patient

877   sample. *P* - value was calculated using Kruskal-Wallis test. Cytotoxicity signature score

878   (violin plot, top right) of CD8+ CTLs across spatial fields (*, *P* < 0.05; **, *P* < 0.01; ***, *P*

879   < 0.001). *P* - values were calculated using Wilcoxon rank sum test. The percentage of

880   CD8+ CTLs expressing cytotoxic signature genes (indicated by the size of the circle)

881   and their scaled expression levels (indicated by the color of the circle) across the

882   LUADs and spatial normal lung samples (bubble plot, bottom left). Expression levels of

883   *NKG7* and *GNLY* in CD8+ CTLs across the spatial samples (violin plots, bottom right).

884   **F**, UMAP plots of CD4+ T lymphocytes colored by cell states (top), spatial field (middle),

885   and Treg signature score (bottom). The heatmap on the right shows normalized

886   expression of marker genes for CD4+ T cells grouped by defined subcluster. Each

887   column represents a cell. Top annotation tracks indicate (from top to bottom) cell states,

888   Treg signature score, cytotoxic scores, and naïve T cell score calculated using curated

889   gene signatures, and the corresponding spatial field of each cell. **G**, Enrichment of

890   CD4+ T regulatory cells (Treg) in the tumor microenvironment of LUADs. Bar plot (top

891   left) and boxplot (top middle) showing percentage of CD4 Tregs among total CD4+ cells

892   from all patients across the spatial fields. Each circle in the boxplot represents a patient

893   sample. *P* - value was calculated using Kruskal-Wallis test. Percentage of CD4+ Tregs

894   expressing inhibitory immune checkpoint genes (indicated by the size of the circle) and

895   their scaled expression levels (indicated by the color of the circle, color assignment

896   same as panel E) across the spatial samples (bubble plot, top right). Frequency of

897   CD4+ Treg cells co-expressing *CTLA4* and *TIGIT* immune checkpoints across the

898   spatial samples (scatter plots, bottom). The fractions of *CTLA4+TIGIT+* Tregs are

899   labeled on each plot.

900

901   **Figure 4. Reduced signatures of antigen presentation and inflammatory dendritic**

902   **cells in the microenvironment of early-stage LUAD. A**, UMAP visualization of

903   myeloid cell lineages colored by cell type/state (left) and the spatial fields (right). Mac,

904   macrophages; Mono, monocytes; DC, dendritic cell; cDC, classical dendritic cell; pDC,

905   plasmacytoid dendritic cell. **B**, Bubble plot showing the percentage of myeloid cells

906   expressing lineage specific marker genes (indicated by the size of the circle) as well as

907   their scaled expression levels (indicated by the color of the circle). **C**, Changes in the

908   abundance of myeloid cell subsets across the LUADs and spatial normal lung samples.

909   Embedded pie charts show the contribution of each spatial sample to the indicated cell

910   subtype/state. **D**, UMAP plot of monocyte and macrophage subpopulations, color coded

911   by cell type/state (left), spatial field (middle), and antigen presentation score (right). **E**,

912   The percentage of M2-like macrophages cluster 1 expressing antigen presentation

913   genes (indicated by the size of the circle) and their scaled expression levels (indicated

914   by the color of the circle) across the spatial samples (bubble plot). **F**, Ridge plots

915   showing the expression levels of MHC class I and MHC class II genes in M2-like

916   macrophages cluster 1, and across LUADs and spatial normal lung samples. **G**, Violin

917   plots showing the antigen presentation score in M2-like macrophages (clusters 1 and 5)

918   across LUADs and spatial normal lung samples for all patients together (left) and within

919   patients (right) (***, $P < 0.001$). P – values were calculated by Wilcoxon rank sum test.

920   **H**, UMAP plots of dendritic cells, color coded by cell state (left) and spatial field (right). **I**,

921   UMAP plots showing unsupervised subclustering of cDC2 cells colored by cluster ID

922   (top left), spatial field (top right) and the computed inflammatory signature score

923   (bottom). **J**, Heatmap showing normalized expression of marker genes of cDC2 cell

924   subsets. The top annotation tracks indicate (from top to bottom) the inflammatory

925   signature scores, spatial field tissue of origins, and cDC2 cell clusters. **K**, Bubble plot

926   showing the percentage of cDC2 cells expressing inflammatory and non-inflammatory

927   signature genes (indicated by the size of the circle) as well as their scaled expression

928   levels (indicated by the color of the circle) in the LUADs and spatial normal lung tissues.

929   **L**, **M**, Depletion of inflammatory cDC2 cluster C2 cells in the tumor microenvironment of

930   LUADs. Bar plot (**L**) and boxplot (**M**) showing fraction of cDC2 C2 cells among total

931   cDC2 cells and across the LUADs and spatial normal lung tissues. Individual circles in

932   (**M**) correspond to patient samples. *P* - value was calculated using Kruskal-Wallis test.

933   **N**, Violin plot showing the inflammatory signature scores in cDC2 C2 cells across

934   LUADs and spatial normal lung samples. **O,** boxplot showing the inflammatory signature

935   score in normal lung (NL), in premalignant atypical adenomatous hyperplasia (AAH) and

936   in LUAD from an independent cohort (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; N.S, $P >$

937   0.05 of the Wilcoxon rank sum test).

938

939   **Figure 5. Enriched ligand-receptor cell-cell communication networks between**

940   **LUADs and their immune microenvironment**. **A**, Computational analysis workflow of

941   cell-cell communication using iTALK to identify, from a database of curated ligand-

942   receptor (L-R) pairs, the highly expressed immune checkpoint- and cytokine-receptor

943   pairs, that are significantly and differentially altered (i.e. interactions lost or gained)

944   between LUADs and spatial normal lung tissues. **B**, Heatmaps showing the overlap

945   (quantified by Jaccard index) of predicted ligand-receptor based interactions among

946   individual LUADs and their corresponding spatially distributed normal lung tissues. **C**,

947   Representative circos plots showing details of immune checkpoint-mediated L-R pairs

948   compared between each of the LUADs of patients 2, 3, and 5, and selected matching

949   spatial normal lung samples. **D**, Violin plots showing expression of the ligand and

950   receptor genes (selected from panel C) involving immune checkpoints and showing

951   spatial gain-of-interaction patterns as highlighted in panel C. **E,** Boxplot showing *CD24*

952   expression levels in an independent cohort of normal lung tissues (NL), premalignant

953   atypical adenomatous hyperplasias (AAH) and LUADs assessed using the Nanostring

954   immune Counter panel (see Supplementary Methods) (*, $P < 0.05$; **, $P < 0.01$; ***, $P <$

955   0.001; N.S, $P > 0.05$ of the Wilcoxon rank sum test). **F**, Scatterplots using Pearson

956   correlation coefficients between levels of *CD24* with *EPCAM* and *PRF1* in the NL, AAH,

957   and LUAD samples. **G**, Boxplot depicting *CD24* expression levels in LUADs and

958   matched normal lung tissues (NL) from the TCGA LUAD cohort (*, $P < 0.05$; **, $P <$

959   0.01; ***, $P < 0.001$; N.S, $P > 0.05$ of the Wilcoxon rank sum test). **H**, Scatterplots

960   showing correlation of expression using Pearson's correlation coefficients between

961   *CD24* with *EPCAM* and *PRF1*.

962   **Supplementary Figure Legends**

963

964   **Supplementary Figure S1. Quality control metrics and expression of cell lineage**

965   **markers across the spatial LUAD scRNA-seq dataset. A,** Statistical summary of cells

966   passing quality control (QC) and showing cell number (left), fraction of mitochondrial

967   genes (middle), and the number of detected genes (right) per sample. Dis, distant

968   normal; Int, intermediate normal; Adj, adjacent normal; LUAD, tumor tissue. **B-C,** UMAP

969   plots showing cells colored by patient ID (B) and sample batch (C). **D-E,** Bubble plots

970   showing the percentage of cells expressing lineage markers (indicated by the size of the

971   circle) as well as their scaled expression levels (indicated by the color of the circle)

972   across all cells (D; related to main Fig. 1D and E) or selected cell types (E; related to

973   main Fig. 1 G and H). NK; natural killer cell, DC; dendritic cell, EC; endothelial cell.

974

975   **Supplementary Figure S2. Trajectory analysis of alveolar cells and copy number**

976   **inference of malignant-enriched cells. A,** UMAP view showing *EPCAM+* cells colored

977   by sample batch. **B,** Bar plot showing absolute numbers of cells for each lung epithelial

978   cell lineage. Dis, distant normal; Int, intermediate normal; Adj, adjacent normal; LUAD,

979   tumor tissue; AT1, alveolar type 1; AT2, alveolar type 2. **C,** Potential developmental

980   trajectory for alveolar cells inferred by pseudotime analysis. Cells were ordered by

981   pseudotime (dotted box) and colored by alveolar cell state. **D,** Bubble plots showing the

982   percentage (indicated by the size of the circle) of cells expressing markers of alveolar

983   states shown in trajectory analysis from panel C as well as their scaled expression

984   levels (indicated by the color of the circle). **E,** Pseudotime trajectory in C showing cells

985   colored by Notch signaling signature score. **F,** Plots showing Notch signaling signature

986   score among alveolar cell states in trajectory analysis from panel E. **G,** Heatmap

987   showing inferCNV score of cells in the malignant-enriched cluster. **H,** UMAP plots

988 showing reclustering of malignant-enriched cells and highlighting those from patient 2 in

989 red (P2; top). Bottom UMAP shows reclustering of P2 malignant-enriched cells and

990 identifies (with red asterisk) those harboring *KRAS*-G12D mutation. **I,** Heatmap showing

991 highly expressed genes among sub-clusters (subcluster of cells enriched with *KRAS*

992 G12D mutations compared to subclusters lacking the mutation) identified in the bottom

993 UMAP of panel H.

994

995 **Supplementary Figure S3. Reprogramming of lymphoid cell subsets towards an**

996 **immune suppressive tumor microenvironment in early-stage LUAD. A,** UMAP of

997 lymphoid cells colored by sample batch. **B,** Bar plots showing the absolute numbers of

998 each lymphoid cell subset in each spatial sample. Dis, distant normal; Int, intermediate

999 normal; Adj, adjacent normal; LUAD, tumor tissue; CTL, cytotoxic T lymphocyte; Treg, T

1000 regulatory cell; ILC, innate lymphoid cell; NK, natural killer cell. **C,** Changes in the

1001 abundance of specific lymphoid cellular lineages and states across the LUADs and

1002 spatial normal samples. Embedded pie charts show the contribution of each spatial

1003 sample to the indicated cell subtype/state. **D,** Cytotoxicity signature score in CD8+

1004 *GNLY*_hi CTLs in patient three (P3, left) and in patient 4 (P4, right) across spatial fields

1005 (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ of the Wilcoxon rank sum test). **E,** Treg

1006 signature score in T regulatory cells across spatial fields in all patients (left), P3 (middle)

1007 and in patient 5 (P5, right) (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ of the Wilcoxon rank

1008 sum test). **F,** Depletion of CD4+ CTLs in the tumor microenvironment of LUAD. Boxplot

1009 showing percentage of CD4+ CTLs *GZMA*-hi among total CD4+ cells from all patients

1010 across the spatial fields. Each circle represents a patient sample. *P* - value was

1011 calculated using Kruskal-Wallis test. **G,** Violin plots showing cytotoxic signature score in

1012 CD4+ CTL *GZMA*-hi cells from all patients (left) and P5 (right; *, $P < 0.05$; **, $P < 0.01$;

1013 ***, $P < 0.001$ of the Wilcoxon rank sum test). **H,** Frequency of CD4+ CTL *GZMA*-hi

1014 cells co-expressing *GZMA* and *GZMH* across the spatial samples. The fractions of

1015 *GZMA+GZMH+ CD4+* CTLs are labeled on each plot. **I-K,** Expression profiling of

1016 different isotypes of plasma cells across the spatial samples. **I,** Heatmap showing

1017 isotype genes expression in plasma cells. **J,** Bar plots showing plasma cell isotype

1018 composition across spatial fields. **K,** Dot plots showing the fractional change of IGHA1/2

1019 and IGHG3 across all patients and in each patient (left to right). **L-N,** Reprogramming of

1020 B cells in early-stage LUAD. **L,** Heatmap showing unsupervised clustering of B cells

1021 sub-populations. **M,** UMAP showing re-clustering of B cells into sub-populations. **N,**

1022 Ridge plots (left) showing *RAC2* and *ACTG* expression levels among B cell sub-

1023 clusters, and scatter plots (right) showing the frequency of B cells co-expressing *RAC2+*

1024 and *ACTG+* across spatial fields. The fractions of *RAC2+ACTG+* B cells are labeled on

1025 each plot. **O,** Boxplot showing the B cell C0 signature score in normal lung tissues (NL),

1026 premalignant atypical adenomatous hyperplasias (AAH) and LUADs in an independent

1027 validation cohort. (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; N.S, $P > 0.05$). P - values

1028 were calculated using Wilcoxon rank sum test.

1029

1030 **Supplementary Figure S4. Reprogramming of myeloid cells in the tumor**

1031 **microenvironment of early-stage LUAD. A,** UMAP view showing myeloid cells

1032 colored by sample batch. **B,** Stacked bar plots showing absolute numbers of each

1033 lymphoid cell subset across the spatial samples. Dis, distant normal; Int, intermediate

1034 normal; Adj, adjacent normal; LUAD, tumor tissue; mono; monocytes, mac;

1035 macrophages, cDC; classical dendritic cells, pDC; plasmacytoid dendritic cell. **C,**

1036 Changes in the abundance of specific lymphoid cellular lineages and states across the

1037 LUADs and spatial normal samples. Embedded pie charts show the contribution of each

1038 spatial sample to the indicated cell subtype/state. **D,** Heatmap showing the expression

1039 of inflammatory/non-inflammatory signature genes in cDC2 cells. **E,** Boxplot showing

1040 the non-inflammatory *versus* inflammatory signature score in normal control (NL), in

1041 normal lung tissues (NL), premalignant atypical adenomatous hyperplasias (AAH) and

1042 LUADs in an independent validation cohort. (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$;

1043 N.S, $P > 0.05$). P - values were calculated using Wilcoxon rank sum test. **F,** Heatmap

1044 showing DEGs between LUAD and normal samples in pDCs. **G,** Ridge plots showing

1045 the expression level changes of *FOS*, *JUN* and *FOSB* in pDCs and across spatial

1046 samples.

1047

1048 **Supplementary Figure S5. Composition and gene expression changes in stromal**

1049 **and endothelial sub-populations in early-stage LUAD as well as adjacent and**

1050 **more distant normal lung tissues. A,** UMAP view of stromal and endothelial cells

1051 colored by sample batch. **B**, UMAP view of stromal and endothelial sub-populations. **C**,

1052 Bubble plot showing the percentage of stromal and endothelial cells expressing lineage

1053 markers (indicated by the size of the circle) as well as their scaled expression levels

1054 (indicated by the color of the circle). **D,** Bar plot showing the absolute number of cells

1055 from each stromal and endothelial subset. Dis, distant normal; Int, intermediate normal;

1056 Adj, adjacent normal; LUAD, tumor tissue; EC, endothelial cells. **E**, Changes in the

1057 abundance of stromal and endothelial cell lineages across the LUADs and spatial

1058 normal samples. Embedded pie charts show the contribution of each spatial sample to

1059 the indicated stromal and endothelial subtype. **F**, Heatmap showing DEGs between

1060 LUAD and normal samples in EC venule sub-populations. **G**, Bar plot showing

1061 significantly enriched pathways of up/down regulated DEGs in panel F. **H,** Ridge plots

1062 showing the expression level changes of *HLA-DPB1*, *IL33*, and *IGFBP7* in EC venule

1063 sub-populations and across spatial samples.

1064

1065 **Supplementary Figure S6. Enriched ligand-receptor cell-cell communication**

1066 **networks between LUAD and the immune microenvironment. A-B,** Circos plot

1067 showing additional details of the immune checkpoint-mediated L-R pairs compared

1068 between each of the LUADs of patients 5 and select matching spatial normal lung

1069 samples. And **B**) Violin plot showing expression of epithelial *LGALS9* (ligand) and

1070 dendritic cells *HAVCR2* (receptor) and showing spatial gain-of-interaction patterns as

1071 highlighted in the circos plot of patient 5 (P5) in panel A. **C,** Circos plot showing details

1072 of chemokine-mediated L-R pairs comparison between the LUAD and tumor-

1073 intermediate normal tissue of patient 3 (P3). **D,** Violin plots showing spatial expression

1074 of the chemokine and receptor genes in P3 and highlighted in panel C. **E,** Circos plots

1075 showing details of chemokine-mediated L-R pairs comparison between the LUAD and

1076 tumor-intermediate normal tissue of patient 2 (P2, panel E left) and patient 5 (P5, panel

1077 E right). **F**, Violin plots showing spatial expression of the chemokine and receptor genes

1078 in P5 and highlighted in panel E right. **G-H,** Scatterplots showing the correlation of

1079 expression between *CD24* and other immune-regulated genes in GSE10251 (G) and in

35

1080   TCGA (H) LUAD cohorts. Pearson correlation test was used for p-values show in panel
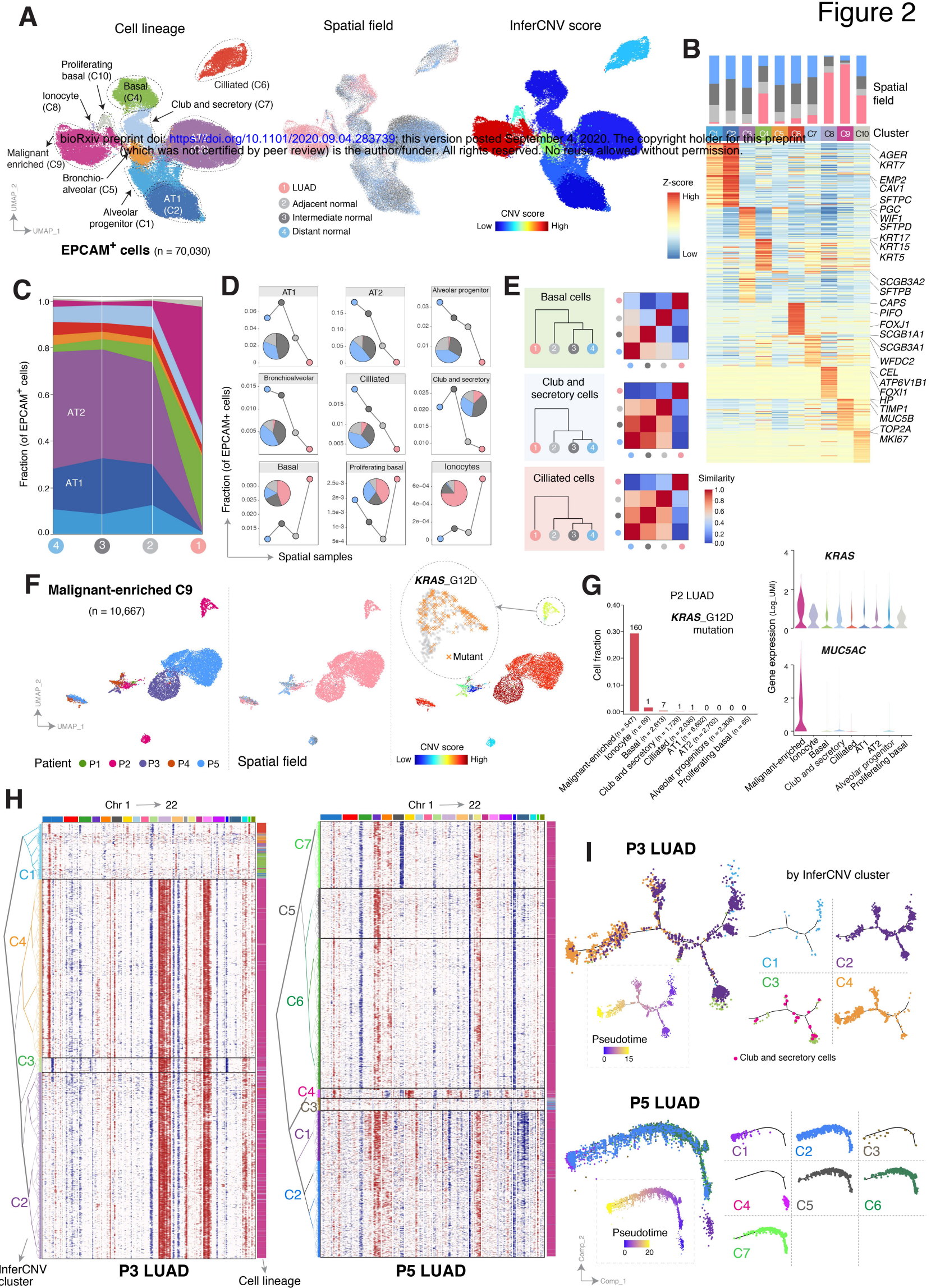
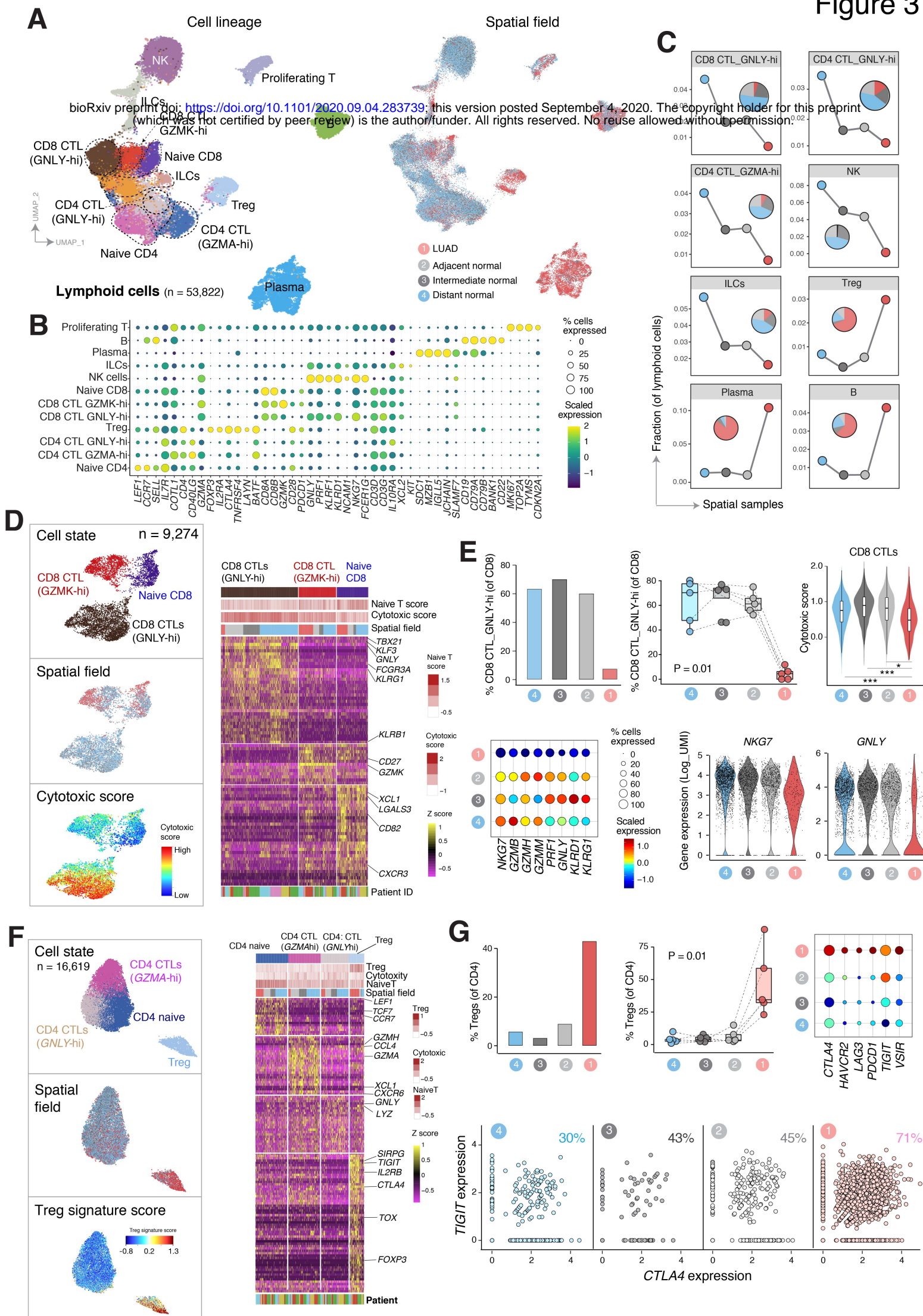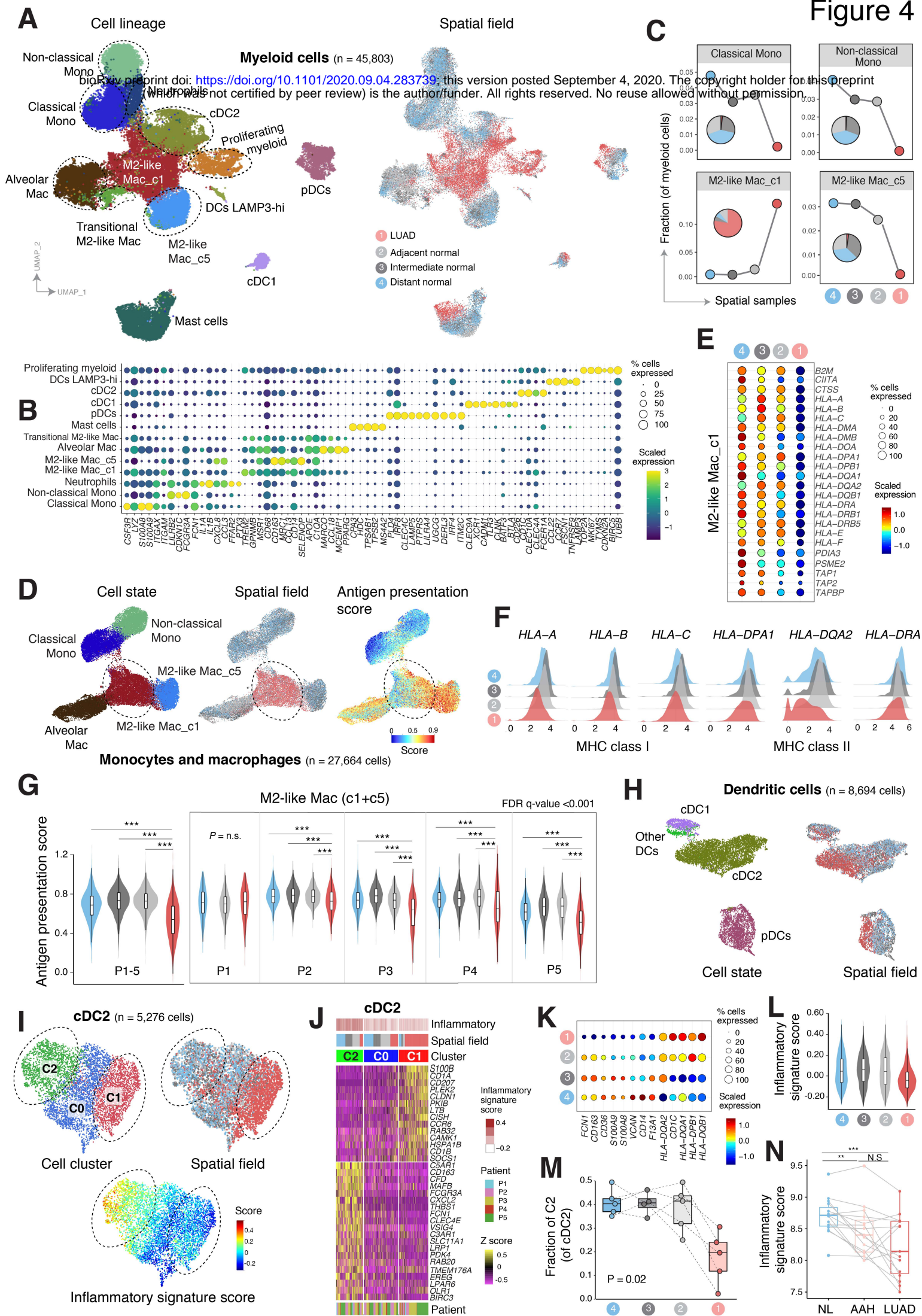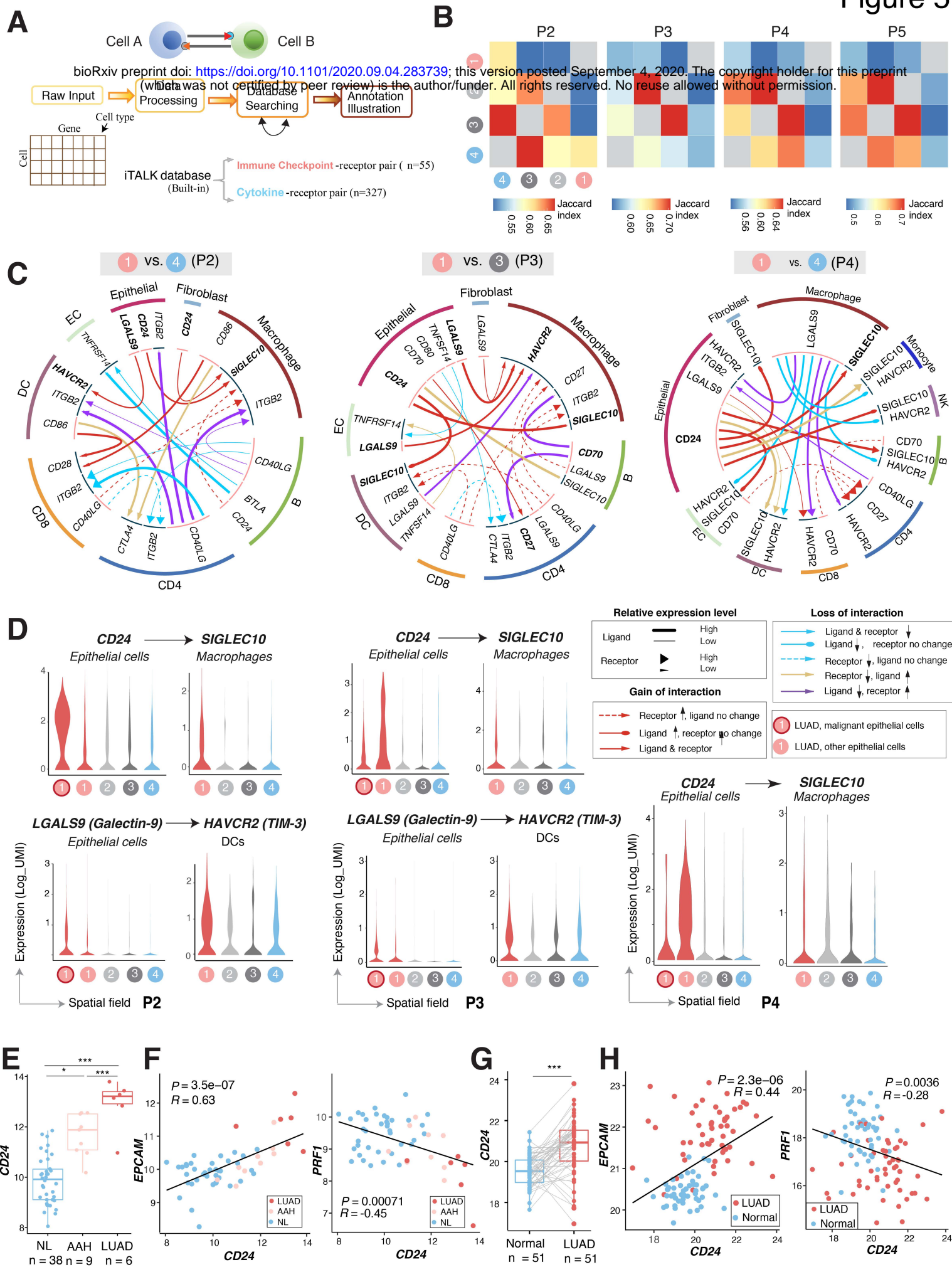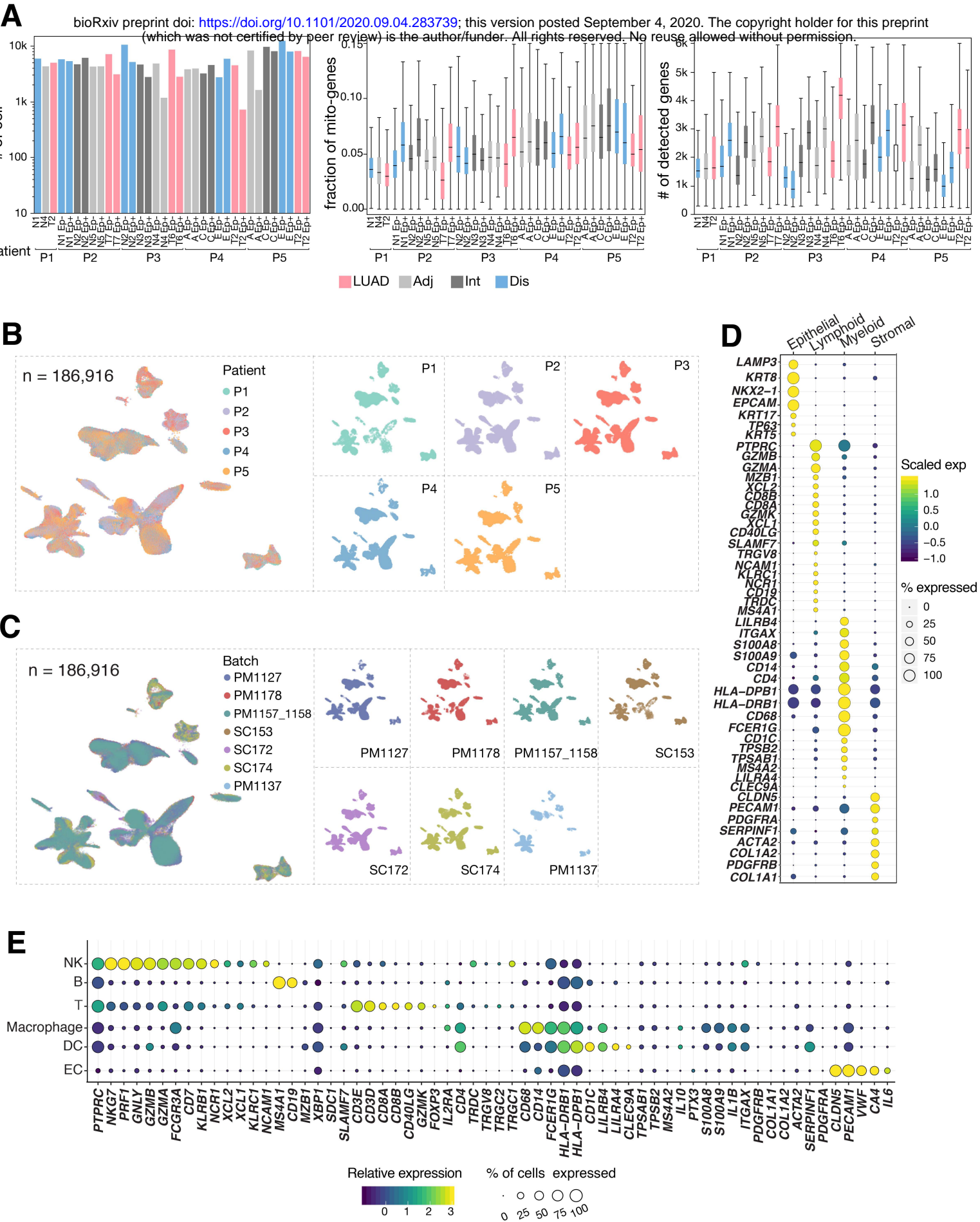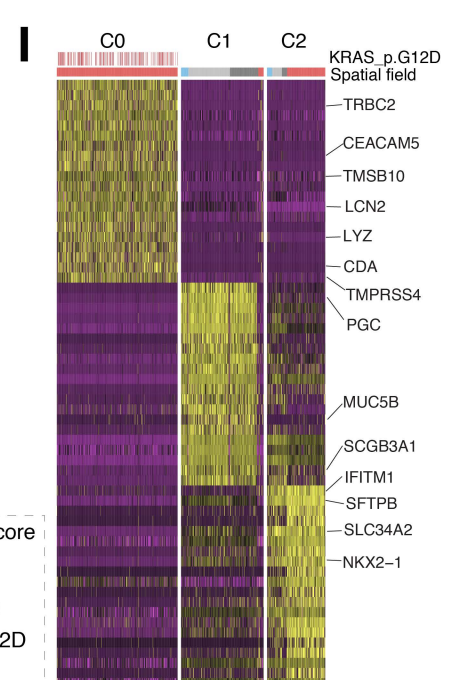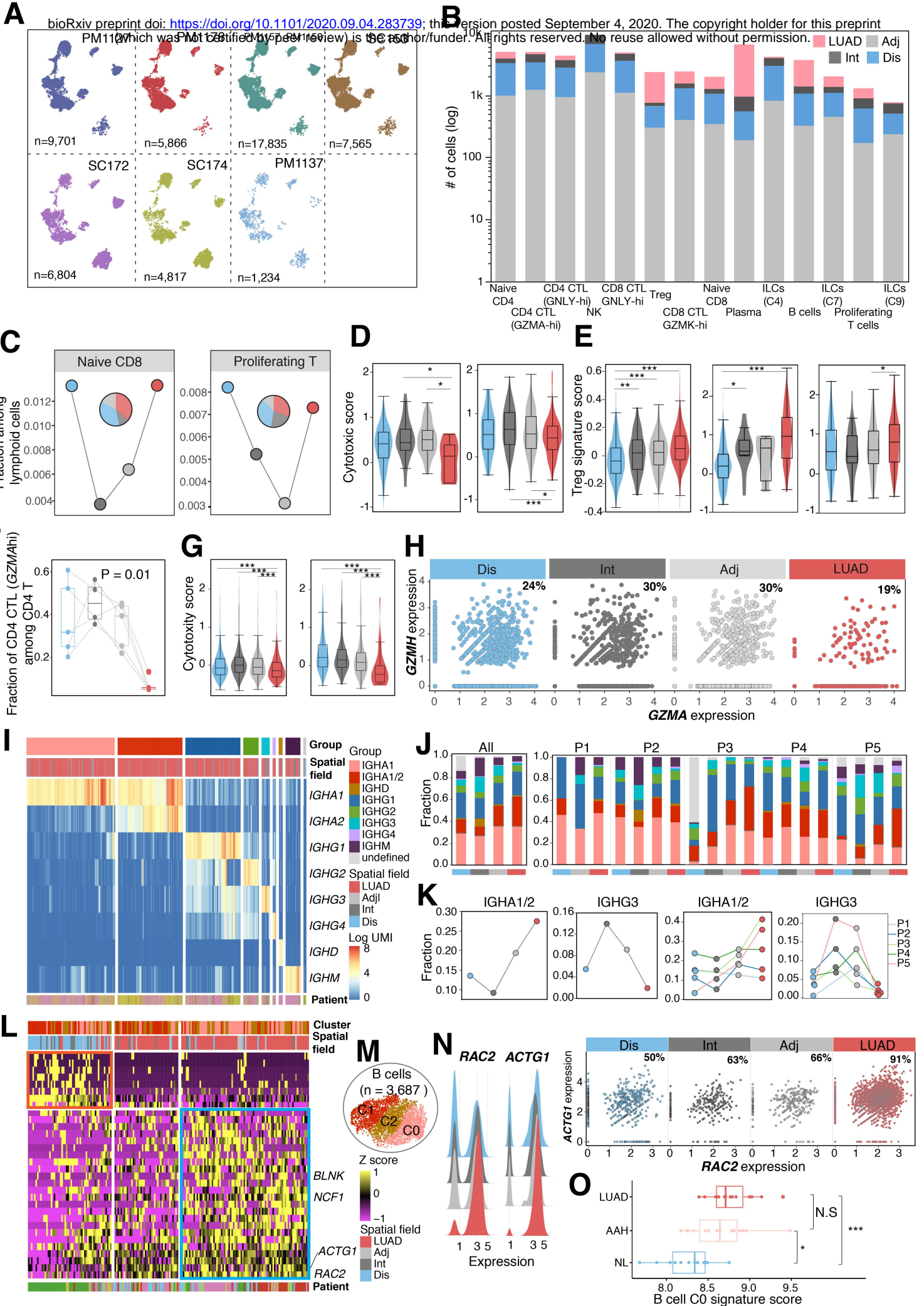1081   G and panel H.
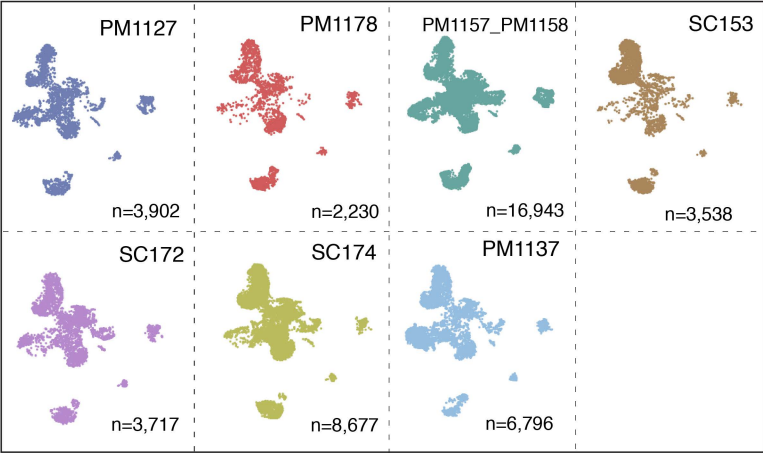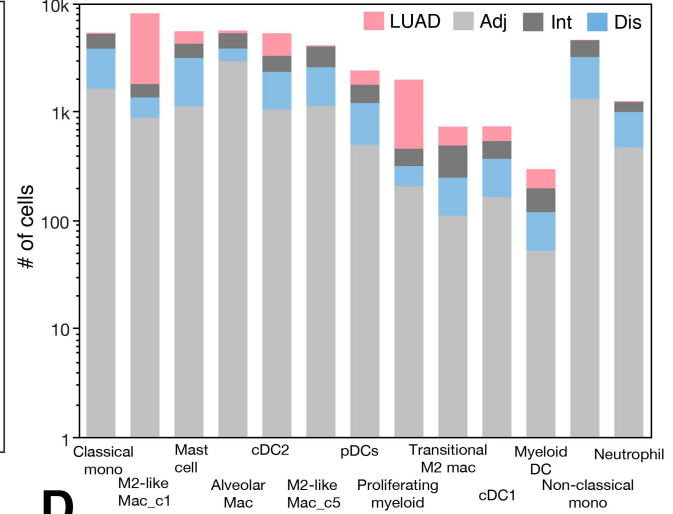
Figure 1
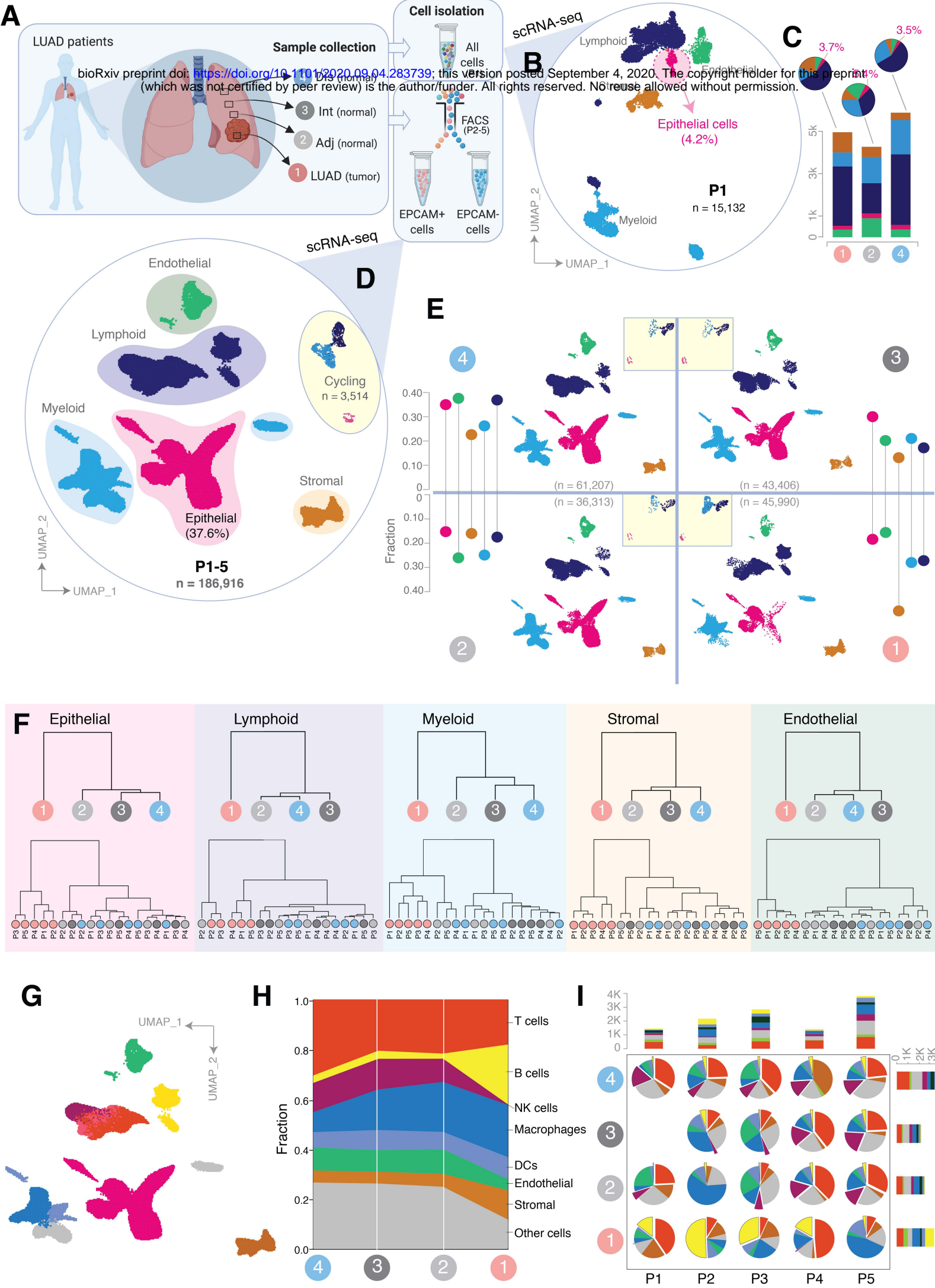
Figure 2

Figure 3

Figure 4

Figure 5

Supplementary figure S1

Supplementary figure S2

Supplementary figure S3

Supplementary figure S4
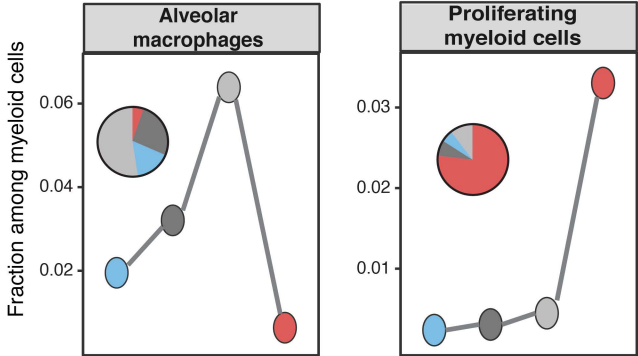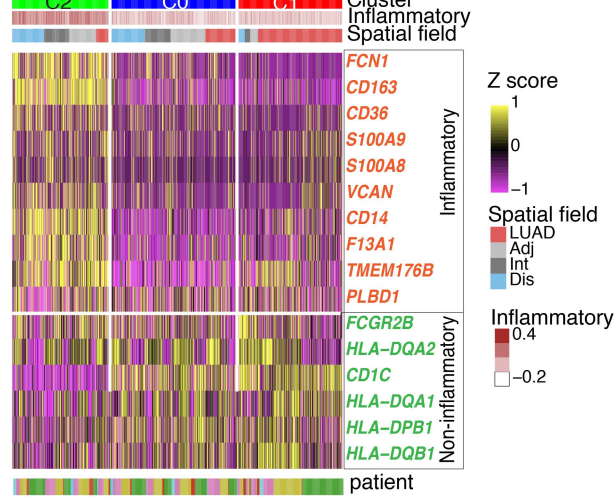
Supplementary figure S5

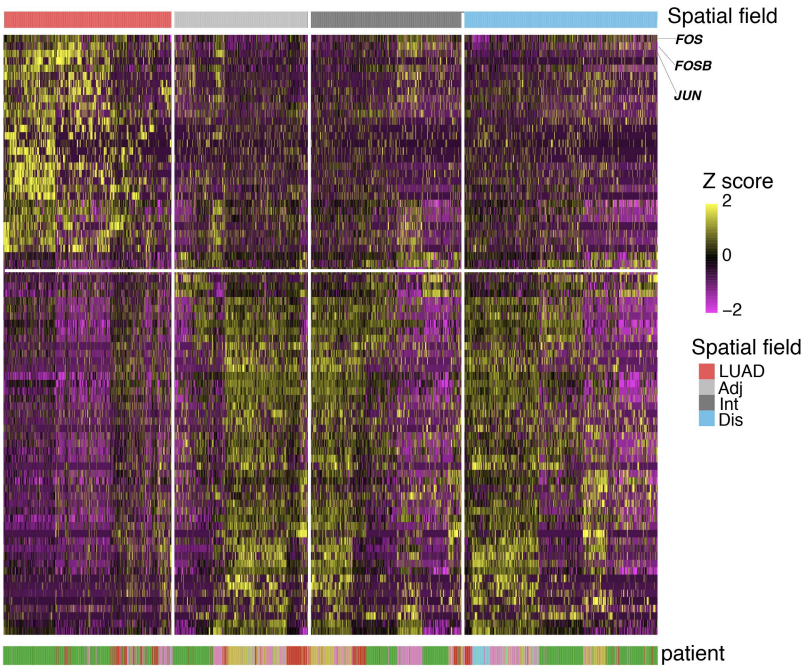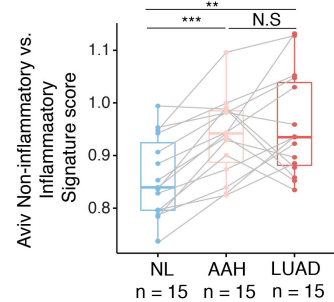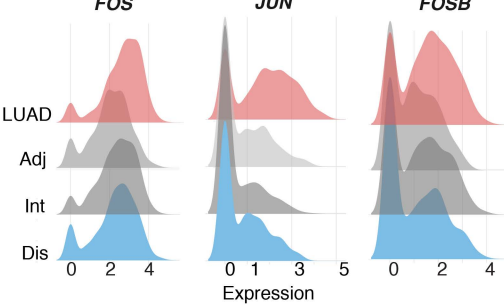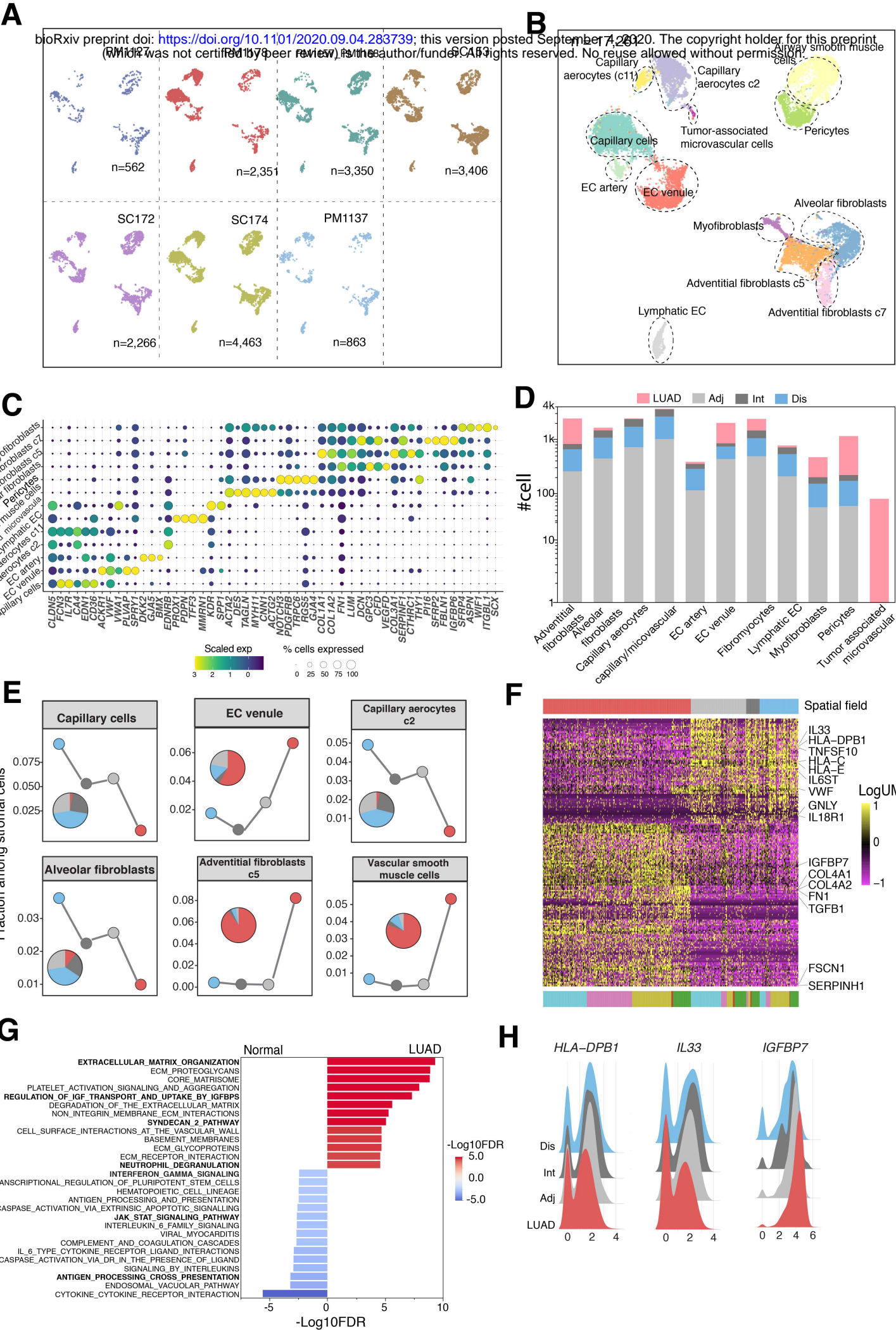# Supplementary figure S6