

Serveur Académique Lausannois SERVAL [serval.unil.ch](http://serval.unil.ch)

## Author Manuscript

Faculty of Biology and Medicine Publication

This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Published in final edited form as:

**Title:** Non-genetic evolution drives lung adenocarcinoma spatial heterogeneity and progression.

**Authors:** Tavernari D, Battistello E, Dheilily E, Petruzzella AS, Mina M, Sordet-Dessimoz J, Peters S, Krueger T, Gfeller D, Riggi N, Oricchio E, Letovanec I, Ciriello G

**Journal:** Cancer discovery

**Year:** 2021 Feb 9

**DOI:** 10.1158/2159-8290.CD-20-1274

In the absence of a copyright statement, users should assume that standard copyright protection applies, unless the article contains an explicit statement to the contrary. In case of doubt, contact the journal publisher to verify the copyright status of an article.

1 **Non-genetic evolution drives lung adenocarcinoma spatial heterogeneity and progression**

2 Daniele Tavernari<sup>1,2,3</sup>, Elena Battistello<sup>1,2,3,4,#</sup>, Elie Dheilly<sup>1,4,+</sup>, Aaron S. Petruzzella<sup>1,4</sup>, Marco  
3 Mina<sup>1,2,3,§</sup>, Jessica Sordet-Dessimoz<sup>4</sup>, Solange Peters<sup>1,5</sup>, Thorsten Krueger<sup>6</sup>, David Gfeller<sup>1,7</sup>, Nicola  
4 Riggi<sup>1,8</sup>, Elisa Oricchio<sup>1,4</sup>, Igor Letovanec<sup>1,8,9,\*</sup>, Giovanni Ciriello<sup>1,2,3,\*</sup>

5

6 <sup>1</sup>Swiss Cancer Center Leman, Lausanne, Switzerland

7 <sup>2</sup>Department of Computational Biology, University of Lausanne (UNIL), Lausanne, Switzerland

8 <sup>3</sup>Swiss Institute of Bioinformatics, Lausanne, Switzerland

9 <sup>4</sup>Swiss Institute for Experimental Cancer Research (ISREC), EPFL, Lausanne, Switzerland

10 <sup>5</sup>Department of Oncology, University Hospital of Lausanne (CHUV), Lausanne, Switzerland

11 <sup>6</sup>Division of Thoracic Surgery, University Hospital of Lausanne (CHUV), Lausanne, Switzerland

12 <sup>7</sup>Department of Oncology, Ludwig Institute for Cancer Research, University of Lausanne (UNIL), Lausanne,  
13 Switzerland

14 <sup>8</sup>Institute of Pathology, University Hospital of Lausanne (CHUV), Lausanne, Switzerland

15 <sup>9</sup>Department of Pathology, Central Institute, Hôpital du Valais, Sion, Switzerland

16 \* Co-corresponding authors: [Igor.Letovanec@hopitalvs.ch](mailto:Igor.Letovanec@hopitalvs.ch), [giovanni.ciriello@unil.ch](mailto:giovanni.ciriello@unil.ch)

17 # Present address: Department of Pathology, New York University (NYU), USA

18 + Present address: Ichnos Sciences SA, La Chaux-de-Fonds, Switzerland

19 § Present address: SOPHiA GENETICS, Saint-Sulpice, Switzerland

20

21 **Running title:**

22 Non-genetic evolution of lung adenocarcinoma heterogeneity

23

24 **Keywords:**

25 Lung adenocarcinoma, tumor heterogeneity, non-genetic evolution, histologic patterns, tumor  
26 microenvironment.

27

28 **Corresponding authors contact details:**

29 Giovanni Ciriello, Génopode, 1015 Lausanne, Switzerland, +41216925450,

30 [Giovanni.Ciriello@unil.ch](mailto:Giovanni.Ciriello@unil.ch)

31 Igor Letovanec, Avenue Grand-Champsec 80, 1951 Sion, Switzerland, +41276034774,

32 [Igor.Letovanec@hopitalvs.ch](mailto:Igor.Letovanec@hopitalvs.ch)

33

34 **Conflict of Interest:** The authors declare no potential conflicts of interests.

35

36 **Abstract**

37 Cancer evolution determines molecular and morphological intra-tumor heterogeneity and  
38 challenges the design of effective treatments. In lung adenocarcinoma, disease progression and  
39 prognosis are associated with the appearance of morphologically diverse tumor regions, termed  
40 histologic patterns. However, the link between molecular and histological features remains elusive.  
41 Here, we generated multi-omics and spatially resolved molecular profiles of histologic patterns  
42 from primary lung adenocarcinoma, which we integrated with molecular data from >2,000 patients.  
43 The transition from indolent to aggressive patterns was not driven by genetic alterations but by  
44 epigenetic and transcriptional reprogramming reshaping cancer cell identity. A signature  
45 quantifying this transition was an independent predictor of patient prognosis in multiple human  
46 cohorts. Within individual tumors, highly multiplexed protein spatial profiling revealed co-existence  
47 of immune desert, inflamed, and excluded regions, which matched histologic pattern composition.  
48 Our results provide a detailed molecular map of lung adenocarcinoma intra-tumor spatial  
49 heterogeneity, tracing non-genetic routes of cancer evolution.

50

51 **Statement of significance**

52 Lung adenocarcinomas are classified based on histologic pattern prevalence. However, individual  
53 tumors exhibit multiple patterns with unknown molecular features. We characterized non-genetic  
54 mechanisms underlying intra-tumor patterns and molecular markers predicting patient prognosis.  
55 Intra-tumor patterns determined diverse immune microenvironments warranting their study in the  
56 context of current immunotherapies.

57

## 58 Introduction

59 Cancer cells evolve by acquiring novel alterations and adapting to changing conditions. Genetic,  
60 epigenetic, and transcriptional changes determine extensive heterogeneity among patients and  
61 within individual tumors, influencing disease prognosis and therapeutic options. Lung  
62 adenocarcinoma (LUAD) is the most common subtype of lung cancer and it encompasses  
63 molecularly and phenotypically diverse diseases (1,2), either associated or unrelated to tobacco  
64 exposure (3,4). LUAD genetic diversity has been documented both across patients (1,5), where it  
65 can determine treatment choices (6–10), and within individual tumors (11,12), where it sustains  
66 disease evolution and treatment resistance (13,14). LUAD inter- and intra-patient molecular  
67 diversity is however not exclusively genetic. Transcriptional and epigenetic heterogeneity has been  
68 reported both among and within patients (15–19). Moreover, molecular diversity can translate in  
69 diverse tumor microenvironments, for example in association with variable tumor mutational  
70 burden (20–23) or presence of specific oncogenic alterations (24).

71 In the clinic, histopathological analyses have revealed heterogeneous tumor tissue morphologies  
72 referred to as histologic patterns. The most frequent patterns are classified as lepidic, papillary,  
73 acinar, and solid (**Fig. 1a**) and 80% of LUAD tumors concurrently exhibit at least 2 of these patterns.  
74 According to the latest WHO Classification (2), histopathological LUAD classification is based on  
75 pattern prevalence, which is a major prognostic indicator (25,26). Indeed, tumors with a prevalent  
76 lepidic pattern are typically considered less aggressive and associated with the early phases of the  
77 disease, whereas solid-prevalent tumors are indicative of poor prognosis. These prognostic  
78 associations define a potential progression of patterns from lepidic to papillary, acinar, and at last  
79 solid (**Fig. 1a**). Whether this progression can also be observed at the molecular level is however  
80 unknown. Indeed, reconciling molecular and histological heterogeneity is hampered by the

81 difficulty of assaying these features in the same tumor material. Recent digital pathology and  
82 spatial genomics technologies coupled with advanced computational approaches provided new  
83 strategies to address this challenge. For example, the combination of multi-region molecular  
84 profiles and histological data was recently used to draw a link between LUAD mutational  
85 heterogeneity and microenvironment composition, with relevant implications for the adoption of  
86 current immunotherapies in this disease (21,22). However, in spite of their prognostic relevance,  
87 comprehensive molecular profiles of morphologically diverse regions in the same patient are  
88 missing. Indeed, the molecular features of LUAD histologic patterns and the molding of their tumor  
89 microenvironment are largely unknown.

90 Here, we performed histopathology-guided multi-region sampling from primary human LUAD to  
91 dissect tumor regions corresponding to unique histologic patterns. Multi-omics and spatially  
92 resolved molecular profiles of these regions allowed us to define tumor intrinsic and extrinsic  
93 processes that determined LUAD pattern progression. We validated and evaluated the prognostic  
94 significance of our results in more than 2,000 LUAD samples from independent patient cohorts.  
95 Importantly, none of these processes could be traced back to specific genetic variants, but rather to  
96 epigenetic and transcriptional reprogramming. Overall, we identified oncogenic processes and  
97 spatial features that support non-genetic evolution as a driver of LUAD heterogeneity and  
98 progression.

99

## 100 **Results**

### 101 **Molecular inter-patient heterogeneity of histologic patterns**

102 We first examined molecular features of 206 LUAD samples from The Cancer Genome Atlas (TCGA)  
103 cohort (5), which had been annotated based on their most prevalent pattern: lepidic (n=8),  
104 papillary (n=47), acinar (n=86), and solid (n=65) (**Supplementary Table 1**). Samples with the same  
105 most prevalent pattern had a similar representation of tumor stages (**Supplementary Fig. S1a**) but  
106 solid-prevalent tumors exhibited significantly higher tumor mutation burden (TMB - **Fig. 1b**),  
107 consistent with recent observations in an independent cohort (27), and number of copy number  
108 alterations at coding genes (**Supplementary Fig. S1b**). A notable exception to this trend was a  
109 lepidic-annotated tumor sample (TCGA-44-7670). However, after reviewing the virtual slides  
110 provided for this dataset, we found highly different histologic pattern in the tumor region  
111 submitted for molecular profiling (01A-TS1) and the one submitted for pathology review (01Z-DX),  
112 suggesting that intra-tumor heterogeneity could explain this inconsistency (**Fig. 1c**). Predicted neo-  
113 antigens increased proportionally with the TMB (**Supplementary Fig. S1c**), potentially predicting  
114 diverse immunogenicity among the histologic patterns. No recurrent genetic lesion (mutation, copy  
115 number alteration, or gene fusion) was found enriched in a specific pattern, except for a few  
116 *PIK3CA* mutations mostly occurring in lepidic samples (3 out of 8 patients, adj. p-value = 0.0004) and a  
117 trend for a higher fraction of *TP53* mutations in solid samples (adj. p-value = 0.096) (**Supplementary**  
118 **Fig. S1d** and **Supplementary Table 2**). The association between solid-prevalent tumors and high  
119 TMB and *TP53* mutations was confirmed in an independent dataset of LUAD patients of East Asian  
120 ancestry (28) (EAS - **Supplementary Fig. S1e,f**) and in a recently analyzed clinical cohort (27).  
121 Conversely, no association was found in these cohorts for *PIK3CA* mutations. Additional analyses to  
122 test candidate weak drivers (29) or alterations converging on the same pathway (30) did not return  
123 significant hits that could be confirmed across datasets (**Supplementary Table 3**). Overall, our  
124 results suggest limited associations between histologic patterns and LUAD genetic features.

125 In contrast, TCGA samples with different prevalent patterns exhibited highly diverse transcriptional  
126 and epigenetic profiles (**Supplementary Table 3**), with at least two-fold more differentially  
127 expressed genes or methylated probes than expected by chance (**Fig. 1d,e**). Interestingly, the most  
128 differentially expressed genes (n = 1,337, adj. p-value < 0.001) and methylated DNA loci (n = 1,753,  
129 adj. p-value < 0.001) among the 4 histologic subtypes did not highlight features unique of each  
130 group but rather progressive changes from lepidic- to solid-prevalent samples. To quantify this  
131 trend, we computed gene expression and DNA methylation fold-changes between each pair of  
132 histologic subtypes, always comparing a more aggressive to a less aggressive subtype  
133 (**Supplementary Fig. S1g**). In this way, progressive changes from lepidic to papillary, acinar and  
134 solid cases would result in all fold-changes having the same sign: all positive for increasing  
135 expression/methylation or all negative for decreasing expression/methylation with pattern  
136 progression (e.g., see the top differentially expressed genes *RAP1GAP* and *ANLN*, **Fig 1f**). Indeed,  
137 concordant positive or negative gene expression (**Fig 1g** - top) or DNA methylation (**Fig. 1g** -  
138 bottom) fold-changes were observed in the majority of the cases, suggesting that histologic  
139 patterns do not represent four independent molecular phenotypes, but rather a transition from  
140 lepidic to solid, driven by epigenetic and transcriptional reprogramming.

141 Differentially expressed genes and methylated gene promoters were enriched for similar functional  
142 categories (**Supplementary Table 4**). Indeed, genes over-expressed in lepidic compared to solid  
143 samples were enriched for cell differentiation, development, and morphogenesis terms, whereas  
144 genes over-expressed in solid compared to lepidic cases were highly enriched for cell proliferation  
145 and markers of immune infiltration (**Fig. 1h**). Transcriptional differences were confirmed in the EAS  
146 dataset (**Supplementary Fig. S1h**) and in an additional LUAD cohort (31) (**Supplementary Fig. S1i**).  
147 Similarly, promoter probes that increased DNA methylation with pattern progression were enriched  
148 for genes involved in cell differentiation and morphogenesis, whereas probes that lost methylation

149 with pattern progression were enriched for immune cell markers (**Fig. 1i**), further suggesting that  
150 aggressive patterns are associated with changes in the tumor microenvironment. To corroborate  
151 this finding, we estimated the presence of distinct non-tumor cell populations from transcriptional  
152 data (32). Lepidic samples were enriched for lung alveolar and epithelial markers, supporting a  
153 similar cell identity between lepidic cancer cells and normal lung tissue, whereas both lymphoid  
154 and myeloid immune cell types were invariably enriched in acinar- and solid-prevalent samples, in  
155 both the TCGA (**Fig. 1j**) and EAS (**Supplementary Fig. S1j**) cohorts. Overall, these results indicated  
156 that LUAD pattern progression is associated with a progressive reprogramming of both tumor cells  
157 and their microenvironment. However, molecular profiles analyzed so far were generated from  
158 single tumor samples annotated by predominant pattern; hence, it remained unclear whether  
159 similar features and plasticity could be observed within individual tumors.

160

### 161 **Molecular intra-tumor heterogeneity of histologic patterns**

162 To determine the molecular features of histologic pattern progression within individual tumors, we  
163 selected a cohort of 10 early stage LUAD primary patient samples that exhibited each at least two  
164 distinct patterns (CHUV cohort, **Supplementary Table 5**) and performed *histopathology-guided*  
165 *multi-region sampling*. For each patient, we reviewed and dissected tumor regions from formalin-  
166 fixed paraffin-embedded (FFPE) tissue slides such that each region was composed by a unique  
167 pattern (**Fig. 2a**). In total, we collected 29 tumor regions and 10 normal tissue samples. These  
168 samples were processed by whole-exome sequencing, RNA-sequencing, and DNA methylation EPIC  
169 array (see **Methods**). LUAD driver mutations were predominantly clonal, i.e. observed in all regions,  
170 and not associated with a specific pattern (**Fig. 2b**). In most cases, we confirmed a trend for higher  
171 TMB in more advanced patterns (**Supplementary Fig. S2a**). After accounting for patient-specific

172 features, differentially expressed genes and methylated probes clustered together samples  
173 annotated for the same pattern (**Fig. 2c** and **Supplementary Fig. S2b, Supplementary Table 6**).  
174 Transcriptional differences among patterns in our cohort were consistent with those observed in  
175 the TCGA and EAS cohorts (**Supplementary Fig. S2c,d** and **Supplementary Table 7**). Indeed, genes  
176 over-expressed in lepidic samples were enriched for tissue development and morphogenesis (**Fig 2c**  
177 – blue cluster), while solid but especially acinar samples exhibited over-expression of immune  
178 infiltration markers, in particular of B-cells (**Fig 2c** – orange cluster). Genes over-expressed in solid  
179 samples were more specifically enriched for markers of cell proliferation and over-expression of  
180 matrix metalloproteinase (MMP) genes (**Fig 2c** – red cluster). Both lepidic- and solid-associated  
181 genes were enriched for extra-cellular matrix (ECM) components and regulators (**Fig. 2c**), albeit  
182 exerting opposite functions (33). Indeed, ECM genes up-regulated in solid samples were mostly  
183 enriched for ECM degradation (e.g. MMP genes) and collagen proteins (e.g. *COL1A1* and *COL1A2*),  
184 whose activation is known to alter cell adhesion and promote invasion (34). Vice versa, ECM genes  
185 over-expressed in lepidic samples included several proteins mediating cell adhesion (35), (e.g.  
186 *TNXB*, *FBLN5*, and *MFAP4*), and putative tumor suppressors (e.g. *DLC1* (36) and *FOXF1* (37)).  
187 Importantly, expression of these genes was associated with pattern progression within individual  
188 tumors (**Fig. 2d**). Similarly, immune infiltration predicted from gene expression (32) increased from  
189 lepidic to solid pattern (**Fig. 2e**) within 8 out of 10 patients, and intra-tumor patterns showed a  
190 different enrichment for markers of normal lung tissue (enriched in lepidic) and immune cell  
191 markers (enriched in acinar and solid) (**Supplementary Fig. S2e**). Altogether, transcriptional and  
192 epigenetic differences observed among patients classified by predominant pattern paralleled  
193 expression and methylation changes observed within individual tumors. Importantly, these  
194 differences pointed at both tumor *intrinsic* (differentiation, migration, proliferation) and *extrinsic*  
195 (immune infiltration) processes as key determinants of LUAD histologic patterns.

196

### 197 **Cancer cell plasticity underlies pattern progression**

198 To explore tumor intrinsic features of pattern progression, independent of the extent of immune  
199 infiltration, we analyzed single cell RNA-seq data for three LUAD samples (38). Differential  
200 expression analysis between tumor and non-tumor cells allowed us to extract 2,410 genes that  
201 were highly expressed only in tumor cells (*cancer-specific genes*, **Fig. 3a**, see **Methods**). First, we  
202 selected cancer-specific genes that were significantly differentially expressed between lepidic and  
203 solid tumor regions in our cohort (adj p-value < 0.1 and absolute fold-change > 2) to determine  
204 lepidic (n = 36) and solid (n = 21) cancer cell markers (**Fig. 3b**). These genes confirmed the  
205 enrichment for cell proliferation (solid) and differentiation (lepidic) terms (**Fig. 3c** and  
206 **Supplementary Table 8**). Next, using these genes as cancer cell markers of lepidic and solid  
207 patterns, we derived a transcriptional score for each single cancer cell to quantify their lepidic-like  
208 or solid-like transcriptional state (see **Methods**). Single cell transcriptional scores from these  
209 patient samples showed a transition of states consistent with plastic reprogramming (**Fig. 3d**): cells  
210 from sample S1 exhibited predominantly lepidic features, sample S2 instead harbored tumor cells  
211 that lost lepidic markers and exhibited variable expression of solid markers, lastly sample S3  
212 comprised cells spanning the whole transition from lepidic to solid (**Fig. 3d**). To explore the origin of  
213 these transcriptional changes, we algorithmically predicted which master transcriptional regulators  
214 (TRs) were most likely to modulate differentially expressed genes between lepidic and solid  
215 samples (39). Results in the TCGA and our cohorts were extremely concordant (**Fig. 3e**) and  
216 identified, among solid master TRs, cell cycle regulators such as E2F transcription factors,  
217 minichromosome maintenance (MCM) complex components, which regulate DNA replication and  
218 elongation, and the Forkhead Box M1 (*FOXM1*) transcription factor, which is a key regulator of cell

219 proliferation and over-expressed in several cancer types (40). Among lepidic master TRs, we found  
220 genes associated with tumor suppressive functions, such as the circadian repressor *CRY2*, which  
221 degrades the *MYC* oncogene (41), and the zinc-finger transcription factor *ZBTB4* (42), as well as  
222 transcription factors involved in cell differentiation and development, such as *CASZ1* (43,44), and  
223 the YAP repressor *WWC1* (45,46). In both the TCGA and our cohorts, lepidic master TRs and lepidic  
224 cancer cell markers exhibited on average higher promoter DNA methylation in solid samples  
225 (**Supplementary Fig. 3a**), suggesting that downregulation of lepidic TRs and markers is at least in  
226 part driven by epigenetic silencing. Interestingly, data from a high-throughput CRISPR knock-out  
227 screening (47) revealed that, in lung adenocarcinoma cell lines, loss of TRs enriched in the solid  
228 pattern was largely deleterious and many, though not all, were classified as essential genes, due to  
229 their role on cell proliferation (**Fig. 3f**). Conversely, in the same cells, knock-out of TRs enriched in  
230 the lepidic pattern led to moderate effects on cell viability and sometimes even improved cell  
231 fitness (**Fig. 3f**), consistent with a putative tumor suppressive function.

232 Next, we combined cancer-specific lepidic and solid markers to generate a unique mRNA signature  
233 and quantify lepidic-to-solid transition (L2S signature). L2S signature scores in TCGA and CHUV  
234 samples were consistent with patient and intra-tumor classifications and pattern progression  
235 (**Supplementary Fig. 3b,c**) and, indeed, normal lung tissues had the lowest scores, followed by  
236 lepidic, papillary, acinar, and finally solid samples, which on average had the highest scores.  
237 Interestingly, L2S scores correctly predicted the pattern of the misannotated TCGA sample (TCGA-  
238 44-7670 – **Fig 1c**) and, unlike the classification based on predominant pattern, it stratified TCGA  
239 samples in classes with significantly different prognosis (**Fig. 3g** and **Supplementary Fig. S3d**). This  
240 signature gave us the possibility of estimating pattern progression and assess its prognostic value in  
241 a much larger ensemble of LUAD tumors, where transcriptional profiles were available, but  
242 histopathology annotations were not. In total, we analyzed and scored >2,000 LUAD human

243 samples, from 10 patient cohorts (5,23,28,31,48–51). Multi-variate Cox regression confirmed that  
244 tumor stage and L2S scores were orthogonal and independent prognostic factors in all except one  
245 of the tested cohorts (i.e. cohorts comprising more than 100 patients) (**Fig. 3h**, **Supplementary Fig.**  
246 **S3e**, and **Supplementary Table 1**). Furthermore, across all cohorts, L2S scores were strongly  
247 associated with the predicted activity of lepidic and solid TRs (**Supplementary Fig. S3f**) and  
248 microenvironment composition (**Fig. 3i**). Intriguingly, the highest correlation between L2S scores  
249 and immune cell markers was with markers of T-cell exhaustion, suggesting that mechanisms of  
250 immune evasion occur in tumor samples with solid pattern features.

251

### 252 **The tumor microenvironment of LUAD histologic patterns**

253 The reproducible association between our L2S signature and immune infiltration across  
254 independent LUAD patient cohorts (**Fig. 3i**) prompted us to investigate the spatial composition of  
255 the tumor immune microenvironment in correspondence of different patterns. First, we analyzed  
256 FFPE tumor tissue slides from our patient cohort and from 3 additional patients with solid patterns  
257 by multi-color immunofluorescence to detect proliferating cells (Ki-67+), B-cells (CD20+), CD4+ and  
258 CD8+ T-cells, and macrophages (CD68+). We distinguished LUAD patterns and tumor cells by  
259 Hematoxylin and Eosin (H&E) staining (**Fig. 4a**) and TTF1 staining (**Fig. 4b**), respectively, and  
260 quantified fluorescent signal intensities (**Fig. 4c**) by designing a spatial grid quantification approach  
261 (GridQuant) that averaged fluorescence signals within pixels of variable size (**Fig. 4d**, see **Methods**).  
262 These analyses revealed striking differences in the extent and geographical organization of immune  
263 cell infiltration across LUAD patterns. Solid regions exhibited significantly stronger Ki-67 intensity  
264 than the other patterns, whereas immune cell markers increased intensity with pattern progression  
265 but were highest in acinar regions (**Fig. 4e**). Interestingly, in several tumors we observed the

266 formation of tertiary lymphoid structures (TLS) (**Fig. 4f**), sometimes characterized by a Ki-67  
267 positive core of proliferating B-cells (**Fig. 4f** - right) resembling germinal centers. TLS formation has  
268 been associated with improved prognosis and response to immunotherapies (52,53), hence we  
269 assessed their distribution across patterns in our samples. We automatically identified all TLS in our  
270 slides and found that these were absent in normal lung tissue and lepidic cancer regions but  
271 prevalently observed within acinar regions and, less frequently, in papillary and solid regions (**Fig.**  
272 **4g**). Altogether, these results suggested that immune infiltration increased with pattern  
273 progression but was maximal in acinar and not in solid patterns.

274 Next, we investigated the spatial organization of the tumor microenvironment in different patterns,  
275 by assessing co-localization of tumor and non-tumor cells. TTF1+ and TTF1- signals were positively  
276 correlated in normal lung and lepidic regions, likely due to the presence of cell-depleted lung  
277 alveolar structures, lacked correlation in papillary and acinar, but were highly anti-correlated in  
278 solid regions (**Fig. 4h**), consistent with low intermixing of cancer and non-cancer cells. Similarly, co-  
279 localization of immune cell markers and Ki-67, which here could be used to mark tumor cells  
280 (**Supplementary Fig. S4a**), was lowest at solid regions independently of the pixel size  
281 (**Supplementary Fig. S4b**). Consistent with these trends, we noticed that lymphoid cells and  
282 macrophages localize at the boundary of solid regions within individual tumor slides (**Fig. 4i**). To  
283 quantify these observations, we used GridQuant to extract average signal intensities at different  
284 distances from the periphery of each solid tumor region towards its core (**Fig. 4j**). In all cases, the  
285 density of immune cells was higher at the periphery than at the core of the tumor region (**Fig. 4k,l**)  
286 indicating that the spatial distribution of immune cells in solid patterns was consistent with an  
287 immune excluded phenotype.

288 To corroborate this evidence and explore in more detail the molecular profiles and immune  
289 microenvironment of the core and periphery of lung adenocarcinoma, we performed digital spatial  
290 profiling (DSP - Nanostring GeoMX) in 5 tissue slides from 5 patients. Briefly, in each slide, we  
291 selected and analyzed with a panel of 58 antibodies 12 regions of interest (n = 60 ROIs in total),  
292 located either at the core or periphery of different histologic patterns (**Supplementary Table 9** and  
293 **Supplementary Fig. S5a**). ROI localization was not associated with immune infiltration, measured  
294 by either ratio of CD45 positive cells or protein expression, except for solid ROIs (**Fig. 5a**). Indeed,  
295 out of six solid ROIs from Patient 8 (**Fig. 5b**), two were localized at the core of the tumor (R9 and  
296 R8) and had lowest levels of immune infiltration, and four were selected at the tumor periphery  
297 (R6, R7, R10, R12) and all exhibited high immune infiltration (**Fig. 5a**). Solid core ROIs expressed  
298 high levels of cancer cell specific markers (PanCK and EpCAM), Ki-67, and the interleukin 7 receptor  
299 (IL7R or CD127) (**Fig. 5c**). Although IL7R can be expressed by both tumor and immune cells, the low  
300 content of immune cells in these ROIs suggested that IL7R was here expressed by cancer cells.  
301 Importantly, cancer cell expression of IL7R has been associated with poor prognosis in non-small  
302 cell lung cancer (54). Conversely, periphery ROIs exhibited high expression of immune cell markers,  
303 including the immunosuppressive regulatory T-cell (T-regs) marker Tim3 and the immune  
304 checkpoint VISTA (**Fig. 5c**). The highly different extent of immune infiltration at the core and  
305 periphery of the solid tumor region challenged the possibility of comparing features specific of  
306 either cancer cells or immune cells. To overcome this challenge, we profiled 36 additional ROIs  
307 from 3 solid tumor regions, and, in each ROI, we separately analyzed immune cells (CD45+) and  
308 cancer cells (PanCK+) (**Fig 5d, Supplementary Fig. S5b**). By selectively retaining the signal coming  
309 from either one or the other cell population (see an example of the masking strategy on **Fig 5d**), we  
310 first compared cancer cells at the core and periphery of solid tumor regions. Here, we found that  
311 cancer cells at the periphery actually exhibited significantly higher levels of the proliferation marker

312 Ki-67 than cancer cells at the core of the tumor, consistent with an invasive margin, as well as Pan-  
313 AKT and p53 (**Fig 5e** and **Supplementary Table 9**). Next, although immune infiltration was low at  
314 the core of solid ROIs, specific comparison of CD45+ cells at the core and periphery showed that  
315 immune cells infiltrated at the core of solid regions were significantly enriched for markers of  
316 immunosuppressive effector T-regs, such as FOXP3, CD25 and Tim3, and immune checkpoints, such  
317 as CTLA4, VISTA, and ICOS (**Fig 5f**, **Supplementary Fig. S5c**). Overall, whereas all solid tumors  
318 exhibited features of immune exclusion, the residual immune infiltration was consistent with an  
319 immunosuppressive microenvironment associated in particular with the presence of effector T-  
320 regs.

321

## 322 **Discussion**

323 Cancer heterogeneity across and within patients is apparent at both the molecular and histological  
324 levels. However, how and whether genomic features determine cell morphology and spatial  
325 organization is largely unexplored. Here, we reconciled molecular and histological heterogeneity in  
326 lung adenocarcinoma by introducing an approach based on histopathology-guided multi-region  
327 sampling. Our results showed evidence of non-genetic mechanisms of tumor evolution as  
328 determinants of histological heterogeneity and disease progression. Indeed, progression from  
329 lepidic to solid histology was associated with plastic reprogramming of differentiation cell markers,  
330 increased cell proliferation, and a transition from an immune *desert* (lepidic), to an *inflamed*  
331 (papillary and especially acinar) and eventually *excluded* and *suppressive* (solid) microenvironment  
332 (**Fig. 6**). Importantly, the transition of both cancer cell intrinsic features and microenvironment  
333 composition was evident within individual tumors and matched intra-tumor pattern heterogeneity.

334 The concomitant evidence of plastic reprogramming of cancer cells and changing  
335 microenvironment prompts questions on the origin of such changes. Does tumor cell de-  
336 differentiation activate specific immune shaping and responses? Or, are dynamic changes of the  
337 tumor immune microenvironment triggering cancer cell plasticity? To address these questions,  
338 models of lung adenocarcinoma that mimic transcriptional, epigenetic, and morphological features  
339 of the human disease are required. Mouse models of non-small cell lung cancer recapitulate some  
340 of the histologic patterns observed in the human disease (55,56), but the molecular features of  
341 these patterns remain to be investigated. Interestingly, recent evidence has shown that LUAD  
342 progression in a genetically engineered mouse model (GEMM) is accompanied by plastic  
343 reprogramming driving cell dedifferentiation (57,58). A detailed comparison of the cell state  
344 transitions that we observed in our human cohort with those in the LUAD GEMM will be important  
345 to investigate the possibility of genetically and therapeutically manipulate specific transcriptional  
346 regulator driving LUAD progression. Moreover, cross talks between the tumor microenvironment  
347 and changes in cancer cell epigenetic features have been observed in the context of neoantigen  
348 presentation, cytokine production, and epigenetic regulation of PD-L1 expression (59,60). However,  
349 it is challenging to establish the relative timing and causative interactions between cancer cell  
350 reprogramming and immune surveillance. Towards this goal, detailed single-cell spatial  
351 characterization of tumor molecular profiles from primary patient samples or longitudinal analysis  
352 of tumor spatial features could inform and complement functional assays in experimental models.

353 Intriguingly, even by selecting intra-tumor regions characterized by a unique pattern, we did not  
354 find evidence of markers discriminating the four patterns into separate and independent classes.  
355 Instead, our results suggested a transition between two extreme states, lepidic and solid, with  
356 papillary and acinar as possible intermediate states. To quantify this transition, we proposed a  
357 transcriptional signature (called L2S signature) derived from the comparison of pure lepidic and

358 pure solid tumor regions. Importantly, L2S scores were independent predictors of patient's overall  
359 survival and immune infiltration in multiple independent LUAD cohorts and highlighted sample  
360 misannotations due to intra-tumor heterogeneity. With the expanding use of molecular profiling  
361 technologies for diagnostic purposes, signatures like the one we identified could provide a  
362 complement to histopathology. In particular, intra-tumor pattern heterogeneity is prevalent in  
363 early stage LUAD. With the recent success and possible increased adoption of screenings to detect  
364 the disease (61,62), cases diagnosed at an early stage are expected to augment. It will be critical to  
365 discriminate those more likely to progress or relapse after treatment, surgery and/or radiotherapy,  
366 and better select those needing adjuvant treatment and type of therapy.

367 In advanced/metastatic stage adenocarcinoma, immunotherapy as a single agent or in combination  
368 with other drugs is now the treatment of choice in an important portion of cases (63). Furthermore,  
369 immune checkpoint inhibitors (ICI) in the neo-adjuvant setting are currently of great interest in  
370 non-small cell lung cancer, as shown by recent results with the PD-1 inhibitor nivolumab, which led  
371 to major pathological response in 45% of the patient (64). Additional clinical trials on early-stage  
372 tumors are ongoing and results are expected in 2021 (65,66). As new data will become available, it  
373 will be interesting to test the association between lung adenocarcinoma histologic pattern  
374 composition or signature and response to ICI in both adjuvant and neo-adjuvant setting.

375

376 The emergence of histological heterogeneity with disease progression is not a feature exclusively  
377 observed in lung adenocarcinoma. Evidence of morphological changes have been reported in  
378 several tumor types such as breast cancer (67) and hepatocellular carcinoma (68). In these tumor  
379 types, integrating histopathology-guided multi-region sampling with molecular profiling could  
380 prove to be an effective strategy to study the evolution of the disease. In particular, recently

381 developed spatial genomics technologies need to be coupled with computational approaches able  
382 to exploit spatial information to provide novel insight on interactions between tumor and non-  
383 tumor cells and on how such interactions shape cancer cell identity. Overall, reconciling molecular  
384 and phenotypic heterogeneity is a critical first step to understand and integrate genetic and non-  
385 genetic mechanisms of cancer evolution.

386

387

388

## 389 **Methods**

### 390 Statistical analyses

391 Details of all statistical analyses and tests performed and referred to in the main text and methods  
392 sections are outlined in Supplementary Table 2. Standard statistical tests (Chi-Square, Fisher,  
393 Kruskal-Wallis, t, Wilcoxon Rank-Sum, correlation coefficients) were performed using the  
394 appropriate functions from R 'stats' package. Dunn and Tukey post-hoc tests were performed with  
395 R packages 'FSA' (v 0.8.22) (<https://github.com/droglenc/FSA>) and 'multcomp' (v1.4-13) (69)  
396 respectively. Multiple hypotheses corrections were made using Benjamini-Hochberg procedure. All  
397 analyses were performed implementing custom scripts in bash and R (v3.4 and 3.5) languages.

### 398 TCGA dataset

399 Molecular and clinical data for The Cancer Genome Atlas lung adenocarcinoma cohort (TCGA-LUAD)  
400 were downloaded from the Genomic Data Commons (GDC) (70) and GDAC FireHose  
401 (<https://gdac.broadinstitute.org/>) repositories. The dataset included somatic point mutations  
402 (whole exome sequencing, MAF file version: mc3 v0.2.8 (71)), copy number changes (Illumina SNP6

403 array, segmentation files and gene-level copy number generated with GISTIC (72)), gene expression  
404 profiled by RNA sequencing (Illumina HiSeq, HTSeq raw counts and FPKM-normalized values), DNA  
405 methylation data (Illumina Infinium HM450k array, beta values), H&E stained images and clinical  
406 data. Neoantigen counts were retrieved from [https://gdc.cancer.gov/about-](https://gdc.cancer.gov/about-data/publications/panimmune)  
407 [data/publications/panimmune](https://gdc.cancer.gov/about-data/publications/panimmune) (73) (metric: 'numberOfBindingExpressedPMHC'). Only primary and  
408 normal samples were considered; in case multiple samples for the same patient were available, the  
409 sample with the latest plate number was retained as recommended by the GDAC guidelines. All  
410 data was generated and processed by The Cancer Genome Atlas research network (74). Consistent  
411 predominant histologic pattern annotation for a subset of tumors (N=206) was obtained by merging  
412 clinical tables from GDC, GDAC and Supplementary Table 1 from TCGA-LUAD 2014 manuscript (74).  
413 Patients ambiguously annotated to different patterns in different clinical tables were excluded from  
414 this pattern-annotated subset.

#### 415 CHUV dataset

416 We retrieved from the database of the Pathology Institute of the Lausanne University Hospital  
417 resected stage I and II lung adenocarcinoma. 13 adenocarcinomas were selected (**Supplementary**  
418 **Table 5**) fulfilling following criteria: (1) presence of at least 2 of the main histological patterns of  
419 lung adenocarcinoma (lepidic, acinar, papillary and solid); (2) sufficient material available with  
420 appropriate surface and delineation of each pattern allowing microscopic dissection and material  
421 harvesting for molecular and imaging analyses. For every patient, a FFPE block of normal lung tissue  
422 distant from the tumor was selected. 2 to 5 tumor regions/FFPE blocks were also selected for every  
423 patient each with a specific dissectible pattern. For 3 patients (1, 3 and 8) 2 non-adjacent regions of  
424 the same pattern were also selected for intra-tumor comparisons. 5 to 20 micrometer thick  
425 unstained slides were obtained from every selected block. Slides were dried and deparaffinized

426 using Xylol and Ethanol (100% and 70%) and stained with Toluidine Blue 0.024% isopropanol 30%.  
427 Macroscopic and microscopic dissection were performed and unwanted tissue (non-pattern specific  
428 and non-tumor tissue) was removed from the slides. 4 micrometer thick slides were also taken  
429 before, after as well as in between the dissected slides and stained with Hematoxylin Eosin to  
430 improve morphological control. Slides taken at the same time were subsequently used for  
431 molecular and imaging analyses. Local Ethical Committee approval was obtained to perform all  
432 mentioned analyses, under authorization N. 2017-00334.

433

434

#### 435 Other datasets

436 All additional datasets were generated and processed as described in the corresponding  
437 publications and are summarized in **Supplementary Table 1**. Chen/EAS (28) dataset including  
438 somatic point mutations, copy number changes, gene expression profiled by RNA sequencing, H&E  
439 stained images and clinical data was downloaded from OncoSG (75,76). Ding (31) gene expression  
440 dataset was downloaded from the UCSC Xena browser (77). TRACERx (12) raw RNA sequencing  
441 data for each tumor region was downloaded from the European Genome-Phenome Archive (ID:  
442 EGAS00001003458) and processed as described in the paragraph 'RNA sequencing and data  
443 processing'; lymph node metastases were excluded; access was provided by the authors upon  
444 request. Micke (48), Yokota (49), Beg (50), Shedden (51) and Pintilie (78) gene expression datasets  
445 were downloaded from Gene Expression Omnibus (accession numbers, respectively: GSE37745,  
446 GSE31210, GSE72094, GSE68465, GSE50081).

#### 447 Whole exome sequencing and data processing

448 Whole exome sequencing was performed on 27 tumor regions and 9 adjacent normal lung  
449 specimens by Genewiz with a protocol optimized for FFPE samples. Briefly, genomic DNA samples  
450 were fragmented into 200-500 base pairs fragments; libraries were prepared using the Agilent  
451 SureSelect Exome library preparation kit and sequenced on Illumina HiSeq in High Output mode  
452 with a 2x150 bases paired end sequencing configuration. On average, total number of reads per  
453 sample was around 98 million, with a mean quality score of 38.42 and 92% of bases with quality  
454  $\geq 30$ . Sequencing reads were checked for quality using FastQC v0.11.7, trimmed with TrimGalore  
455 v0.4.5 to remove Illumina universal adapter contamination (parameters: -q 15 -phred33 --illumina -  
456 -length 20 --paired --retain\_unpaired -r1 24 -r2 24) and aligned to human genome (hg38 build,  
457 downloaded from GATK (79) resource bundle) using bwa-mem (80) v0.7.17. Duplicates were  
458 removed with Picard tool MarkDuplicates (v2.18.4). Reads were processed with GATK v4.0.3.0 Best  
459 Practices workflow (81) using the tools AddOrReplaceReadGroups, BaseRecalibrator, ApplyBQSR  
460 and AnalyzeCovariates (see <https://gatk.broadinstitute.org/hc/en-us/articles/360035535912> for  
461 details) and tagging known variant sites with the VCF files dbsnp\_146.hg38.vcf,  
462 Mills\_and\_1000G\_gold\_standard.indels.hg38.vcf, af-only-gnomad.hg38.vcf,  
463 Homo\_sapiens\_assembly38.known\_indels.vcf (downloaded from GATK resource bundle). The BAM  
464 files thus processed were indexed with samtools v1.6 and used for somatic variant calling.

#### 465 Variant calling and filtering

466 Somatic point mutations and short insertions-deletions for each tumor region were called with  
467 GATK v4.0.3.0 tool Mutect2 using the matched normal lung sample from the same patient (the only  
468 exceptions were tumor regions of patient 8, for which variants were called in tumor-only mode)  
469 and a panel of normal samples. Additional parameters used were '--af-of-alleles-not-in-resource  
470 "0.0000025"' and '--disable-read-filter MateOnSameContigOrNoMappedMateReadFilter'. Variants

471 were filtered using FilterMutectCalls, CollectSequencingArtifactMetrics and FilterByOrientationBias  
472 (--artifact-modes "G/T" and "C/T"); the latter was designed specifically to filter out transitions likely  
473 resulting from FFPE-related deamination of cytosines. Resulting VCF files were converted into MAF  
474 files using vcf2maf (<https://github.com/mskcc/vcf2maf>) v1.6.16 with default parameters. Variants  
475 were then tagged with OncoKB (82) for oncogenicity and retained only if they satisfied all of the  
476 following conditions: (1) GnomAD population frequency < 0.01, (2) not being tagged by Mutect2  
477 filters panel\_of\_normals, artifact\_in\_normal, germline\_risk, str\_contraction, multiallelic and  
478 clustered\_events, (3) variant allele frequency in the tumor sample is at least twice greater than the  
479 one in the normal sample, (4) tumor depth is greater than 6 and Mutect2 filter value is 'PASS' or,  
480 alternatively, the same variant is shared by more than one tumor region of the same patient. The  
481 rationale behind these choices consists in not filtering out variants whose evidence is supported by  
482 independent regions (point 4), provided that they are not germline (points 2 and 3). Known  
483 oncogenic variants were manually verified and recovered using IGV genome browser (83) and  
484 inspecting aligned RNA reads.

#### 485 RNA sequencing and data processing

486 RNA sequencing was performed on 29 tumor regions by Genewiz with a protocol optimized for  
487 FFPE samples, which involved ribosomal RNA depletion and 2x150 bases paired end sequencing  
488 with Illumina HiSeq. On average, total number of reads per sample was around 65 million, with a  
489 mean quality score of 38 and 91% of bases with quality  $\geq 30$ . Sequencing reads were checked for  
490 quality using FastQC v0.11.7, trimmed with TrimGalore v0.4.5 to remove Illumina universal adapter  
491 contamination (parameters: -q 15 -phred33 --illumina --length 20 --paired --retain\_unpaired -r1 24  
492 -r2 24) and processed with RSEM (84) v1.3.0, performing the following steps: (1) alignment with  
493 STAR (85) v2.5.4b to human genome (hg38) using GTF annotation file

494 'gencode.v27.primary\_assembly.annotation.gtf' (downloaded from GATK resource bundle) and  
495 default RSEM parameters; (2) removal of reads mapping to tRNA and rRNA regions (retrieved using  
496 UCSC table browser); (3) estimation of isoform-level and gene-level expression as Transcripts Per  
497 Million (TPM) and RSEM expected counts using rsem-calculate-expression. A batch effect detected  
498 with Principal Component Analysis was corrected in the TPM expression matrix using the ComBat  
499 function implemented in the R package sva (86) v3.30.1. Conversions between Ensembl IDs and  
500 gene symbols were performed using BioMart (87).

#### 501 DNA methylation array and data processing

502 Pattern-specific tumor samples were extracted as described in the section 'CHUV dataset' and  
503 collected in deparafinization solution from the GeneRead™ DNA FFPE kit (cat# 180134, Quiagen);  
504 DNA was extracted according to the manufacturer's instructions. DNA concentration was assessed  
505 using the Qubit High Sensitivity Assay, and DNA quality was monitored using the Infinium HD FFPE  
506 QC Assay (cat# WG-321-1001, Illumina), according to the manufacturer's instructions. Samples for  
507 which at least 1 µg was available and with good overall quality (Delta Cq from the Infinium HD FFPE  
508 QC Assay lower than 5) were selected for bisulfite conversion. Bisulfite conversion was performed  
509 on 1 µg of DNA using the EZ DNA Methylation™ Kit (cat# D5001, Zymo Research), using the  
510 alternative protocol for CT conversion (optimized for the Illumina Infinium Methylation Assay). FFPE  
511 restoration was then performed using Infinium HD FFPE DNA restore kit (cat# WG-321-1002,  
512 Illumina) and samples were processed using the Infinium MethylationEPIC 850k Kit (iGE3, University  
513 of Geneva). Raw signal intensities were processed, quantile-normalized and converted into beta  
514 values using the R package minfi (88) v1.28.4. Probes were annotated using the hg38 EPIC manifest  
515 file generated by (89) (version of September 2018) and filtered according to the corresponding  
516 masking column. Probes mapping to sex chromosomes and having a detection p-value above 0.01

517 were also removed, thus obtaining a final set of 730257 probes. A batch effect detected with  
518 Principal Component Analysis was corrected using the ComBat function implemented in the R  
519 package sva (86) v3.30.1. Probes were annotated to FANTOM5 (90) gene promoters, and promoter  
520 methylation for each gene was computed as the average beta value of all probes mapping to the  
521 gene main promoter (p1 or p).

#### 522 Copy number alterations and fusions calling

523 Gene-level copy number alterations were called from DNA methylation data using the R package  
524 conumee v1.16.0 and GISTIC (72) v83 with default parameters. Gene fusions were called from  
525 adapter-trimmed RNA-seq FASTQ files using STAR-Fusion (91) v2.6.1d with default parameters.

526

527

#### 528 Genetic alteration differences among patterns

529 Fraction of genome altered for each sample was computed as the fraction of genes having a gene-  
530 level GISTIC value equal to 2 or -2. Differences among patterns in terms of tumor mutational  
531 burden and fraction of genome altered were tested with Kruskal-Wallis test followed by post-hoc  
532 Dunn test. The potential confounding role of purity in the association between mutational burden  
533 and patterns was assessed by performing a main-effects ANCOVA with log-scaled number of  
534 mutations as dependent variable, purity (CPE values from (92)) as continuous covariate and  
535 histologic pattern as categorical independent variable. In TCGA pattern-annotated dataset, two lists  
536 of driver genetic alterations were inspected for differences among patterns: 1) a published binary  
537 genomic alteration matrix (93) which included point mutations, copy number changes and gene  
538 fusions (results shown in figures); 2) a list of drivers, which included 'weak drivers' (as described in

539 (29)) obtained by running FunSeq2 (94) algorithm as web service and retaining coding and non-  
540 coding variants with score  $\geq 1.5$ . Driver alterations occurring in at least 3 samples were then  
541 tested for differences among the 4 patterns with a Chi-Square test. Residuals were inspected in  
542 order to determine which pattern was enriched for a given genetic event. P-values were adjusted  
543 for multiple hypotheses with Benjamini-Hochberg procedure. FunSeq2 was applied also to EAS  
544 dataset and the same downstream analysis was performed. For pathway-level analysis, driver  
545 alterations called by FunSeq2 were used and the following steps were performed: 1) relevant  
546 pathways in cancer and their gene components annotated as 'oncogenes', 'tumor suppressor  
547 genes' or 'unknown' were retrieved from (30); 2) a pathway was called 'altered' in TCGA pattern-  
548 annotated dataset if at least one gene was altered in FunSeq2 calls or the gene harbored GISTIC-  
549 based deep deletion (in case of 'tumor suppressor genes' or 'unknown') or amplification (in case of  
550 'oncogenes' or 'unknown'); 3) Pathway alterations were then tested for differences among the 4  
551 patterns with a Chi-Square test, residuals were inspected in order to determine which pattern was  
552 enriched for a given altered pathway and p-values were adjusted for multiple hypotheses with  
553 Benjamini-Hochberg procedure.

#### 554 Differential expression analyses

555 Differential expression analyses among patterns were performed on RNA-seq read counts (HTSeq  
556 counts for TCGA and RSEM expected counts for Chen and CHUV datasets) using R packages limma  
557 v3.38.3 (95) and edgeR (96) v3.24.3 with a standard published pipeline (97). Only genes expressed  
558 (counts per million  $> 1$ ) in at least 3 (CHUV) or 50% (TCGA and Chen) of samples of any pattern were  
559 tested. P-values were adjusted using Benjamini-Hochberg FDR-controlling procedure. Pairwise  
560 pattern comparisons were performed with limma function decideTests. In the CHUV dataset, the  
561 patient corresponding to each tumor region was inserted as covariate in the limma model. For

562 purely graphical purposes, in the heatmaps the patient-specific batch effect was removed with  
563 limma function 'removeBatchEffect'. For TCGA, a null model was constructed by randomly  
564 permuting the assignment of patterns to samples and re-performing differential expression analysis  
565 on 100 random permutations. Differential expression analysis on Ding dataset was performed on  
566 microarray gene expression profiles between samples with  $\geq 50\%$  (solid-like) and  $<50\%$  (lepidic-  
567 like) of solid pattern prevalence.

#### 568 Differential methylation analyses

569 Differential methylation analyses were performed on M-values, which were derived from beta  
570 values, on all probes and with the same pipeline as for expression data. Pairwise pattern  
571 comparisons were performed with limma function decideTests. In the CHUV dataset the patient  
572 corresponding to each tumor region was inserted as covariate in the limma model. For purely  
573 graphical purposes, in the heatmaps the patient-specific batch effect was removed with limma  
574 function 'removeBatchEffect'. For TCGA, a null model was constructed by randomly permuting the  
575 assignment of patterns to samples and re-performing differential methylation analysis on 100  
576 random permutations.

#### 577 Gene ontology analyses

578 Gene ontology analyses were performed on differentially expressed genes and genes targeted by  
579 differentially methylated promoters using MSigDB (98) and Gene Ontology gene sets (99)  
580 (Biological Process and Molecular Function categories), using an FDR cutoff of 0.01 and retrieving a  
581 maximum of 100 categories.

#### 582 Extraction of lung adenocarcinoma tumor and non-tumor markers

583 Genes expressed preferentially in lung adenocarcinoma tumor cells were extracted from a  
584 previously published lung cancer single cell RNA-seq dataset (38), which included 3 lung  
585 adenocarcinoma patients (2 from the 'discovery cohort' and 1 from the 'validation cohort'), in the  
586 following way: (1) For the two 'discovery cohort' patients, genes x cells expression matrix and cell  
587 IDs for each cell type including cancer cells were downloaded from ArrayExpress (ID: E-MTAB-6149)  
588 and SCoPe repositories, and the expression matrix was filtered such that only single cells coming  
589 from tumor regions belonging to the two lung adenocarcinoma patients were retained; (2) For the  
590 tumor regions of the lung adenocarcinoma patient in the 'validation cohort' only raw FASTQ files  
591 were available (ArrayExpress ID: E-MTAB-6653), and thus they were aligned filtered and processed  
592 as described in the Lambrecht et al. study (38) using Cell Ranger v3.0.2 and Seurat v3.0.0; clusters  
593 were detected with Seurat function 'FindClusters' (resolution = 0.5) and 12 out of 16 of them could  
594 be assigned to the main immune/stromal cell types using the expression of the markers reported in  
595 Figure S1 of (38); of the remaining 4 clusters, 2 showed increased expression of several keratin  
596 genes and they were thus assigned to cancer cells (3) for each of the three patients separately,  
597 markers for cells classified as 'cancer' were extracted using the function 'FindMarkers' (with  
598 min.pct = 0) of Seurat (100); (4) markers were further filtered to have adjusted p-value < 0.05 and  
599 average  $\log_2(\text{fold-change}) > 0.25$ ; (5) in order to better accommodate for inter-patient  
600 heterogeneity in tumor cell expression, the union of the three lists of patient-specific cancer  
601 markers was extracted as final list. This procedure yielded a list of 2410 cancer-specific genes.  
602 Markers of lung alveolar and epithelial cells were extracted with FindMarkers applied to the lung  
603 adenocarcinoma samples of the 'discovery cohort' dataset using more stringent thresholds  
604 (adjusted p-value < 0.0001 and average  $\log_2(\text{fold-change}) > 10$ ) in order to increase specificity.

605 Quantification of immune cell infiltration

606 Methylation-based immune cell infiltration fractions for TCGA LUAD samples were downloaded  
607 from a previous study (92). Bulk RNA-seq deconvolution was performed with consensusTME (32)  
608 implemented in the corresponding R package (v 0.0.1.9000, parameters: cancer = 'LUAD',  
609 statMethod = 'ssgsea'). Three cell types were added to the available ones (T cells exhausted, lung  
610 epithelial cells and lung alveolar cells) and their score in each sample was quantified in the same  
611 way as done by consensusTME, namely by computing single sample gene set enrichment analysis  
612 (ssgsea, implemented in GSVA (101) R package v1.30.0 with ssgsea.norm = T) using markers of each  
613 of these cell types. Markers for T cells exhaustion used were PD-1, PD-L1, LAG3, TIGIT, PD-L2, B7-H3  
614 (CD276), HAVCR2 (TIM-3), CD244, CTLA4, CD160 (102) and ([https://www.rndsystems.com/product-](https://www.rndsystems.com/product-highlights/adoptive-cell-transfer-monitor-t-cell-exhaustion)  
615 [highlights/adoptive-cell-transfer-monitor-t-cell-exhaustion](https://www.rndsystems.com/product-highlights/adoptive-cell-transfer-monitor-t-cell-exhaustion)); markers of lung epithelial and alveolar  
616 cells were extracted as described in the paragraph 'Extraction of lung adenocarcinoma tumor and  
617 non-tumor markers'.

#### 618 Construction and scoring of Lepidic-to-Solid signature

619 A differential expression analysis restricted to the 2410 LUAD cancer-specific genes was performed  
620 between lepidic and solid samples of the CHUV dataset. Differentially expressed genes were  
621 extracted by filtering for adjusted p-value < 0.1 and absolute  $\log_2(\text{fold-change}) > 1$ , thus obtaining  
622 36 cancer-specific lepidic markers and 21 cancer-specific solid markers. For bulk RNA-seq datasets,  
623 sample-wise enrichment scores for these markers were computed using the singscore (103) R  
624 package (v1.0.0). Singscore outputs a unified score for the complete signature ('TotalScore') as well  
625 as scores for the upregulated (lepidic) and downregulated (solid) genes separately. The sign of  
626 scores relative to the lepidic markers set were changed for graphical reasons (such that lepidic-like  
627 samples would harbor higher lepidic markers signature scores). For the single cell RNA-seq dataset,  
628 a different strategy was adopted in order to account for the high dropout rate of single cell profiles:

629 1) cancer cells from the 3 lung adenocarcinoma patients extracted as reported in paragraph  
630 'Extraction of lung adenocarcinoma tumor and non-tumor markers' were further filtered to retain  
631 only cells with more than 2000 genes expressed; 2) a differential expression analysis was  
632 performed between the 10 most lepidic-like TCGA patients (i.e. harboring the lowest 'TotalScore'  
633 computed as described above) and the 10 most solid-like TCGA patients (with highest 'TotalScore'),  
634 restricting the set of genes tested to the list of 2410 cancer-related genes; 3) Differentially  
635 expressed genes were again extracted by filtering for adjusted p-value < 0.1 and absolute  $\log_2$ (fold-  
636 change) > 1, in this case obtaining 279 cancer-specific lepidic markers and 215 cancer-specific solid  
637 markers, due to the higher statistical power achieved with increased sample size (most of the  
638 lepidic and solid markers obtained from the CHUV dataset as reported above were among these  
639 augmented lepidic and solid marker lists, and none was found in the wrong list); 4) this augmented  
640 signature was used to score single cells obtained at point (1) for lepidic-like or solid-like features  
641 using AUCell (104) (with default parameters), a tool specifically designed for scRNA-seq datasets.

642

643

#### 644 Purity correction of DNA methylation profiles

645 Tumor purity was estimated from DNA methylation profiles in the CHUV dataset using the Lump  
646 algorithm (92). Briefly, we retrieved DNA methylation of purified leukocytes profiled with Illumina  
647 850k EPIC array (105) and probes with a beta value below 0.1 in all leukocyte samples were  
648 intersected with probes having a beta value above 0.85 in the top 7 TCGA purest samples as  
649 estimated by Lump, which yielded a set of 24 probes methylated in tumor cells and unmethylated  
650 in immune cells. Finally, purity in heterogeneous samples was computed as mean beta value of  
651 these 24 probes. Purity estimates were very correlated with ConsensusTME's immune scores

652 (Spearman's  $r = 0.87$ ) and different thresholds yielded consistent results. Next, in order to estimate  
653 tumor cells specific methylation, beta values observed for a mixture of tumor and immune cells  
654 were modeled as a linear combination of tumor and immune cells beta values, i.e. for probe  $i$ :  
655  $\beta_{\text{observed}}[i] = p\beta_{\text{tumor}}[i] + (1-p)\beta_{\text{immune}}[i]$ , where  $p$  is the purity computed as described.  $\beta_{\text{tumor}}$  was  
656 estimated in the following way: 1) Variance of beta values across all leukocyte samples was  
657 computed for each probe, and probes showing a variance greater than 0.01 were discarded; 2)  
658  $\beta_{\text{immune}}$  for the remaining probes was computed as the mean beta value in leukocyte samples; 3)  
659  $\beta_{\text{tumor}}$  was thus computed with the equation above and probes whose purity-corrected values were  
660 not within  $[0,1]$  were discarded. The rationale behind step 1 was to select probes for which a  
661 reliable estimate of leukocyte methylation could be obtained. The conservative choices at steps 1  
662 and 3 resulted in a decrease of the number of probes for which tumor cells-specific methylation  
663 could be computed (571349 for the CHUV and 91228 for the TCGA datasets), with the advantage,  
664 however, of a higher reliability for the purity correction. Differential methylation analyses (as  
665 described before) were performed on purity-corrected methylation profiles to assess methylation  
666 differences shown in Fig. S3a.

667

#### 668 Master transcriptional regulator activity analysis

669 Virtual Inference of Protein-activity by Enriched Regulon analysis (39) (implemented in the 'viper' R  
670 package, v1.16.0) was used to estimate transcriptional regulator (TR) activity, following guidelines  
671 contained in the user manual. Briefly, 'msviper' function takes as input (1) a co-expression  
672 regulatory network estimated with ARACNe methodology (106), for which a lung adenocarcinoma-  
673 specific regulatory network was downloaded as a Bioconductor package (DOI:  
674 10.18129/B9.bioc.aracne.networks), (2) a gene expression signature generated with viper function

675 'rowTtest' between two biological conditions and (3) a null model obtained with viper function  
676 'ttestNull' and 1000 random permutations. The output is a list of TRs driving the biological  
677 conditions of interest and their corresponding enrichment p-value, FDR, and normalized  
678 enrichment score. VIPER analyses were thus performed on lepidic and solid samples of CHUV and  
679 TCGA datasets separately, restricting the gene expression matrix in input to contain only cancer-  
680 specific genes (see 'Extraction of lung adenocarcinoma tumor and non-tumor markers' section).  
681 Lastly, TR activity was estimated in each single sample of all datasets using the 'viper' function  
682 again restricted to cancer-specific genes.

### 683 Survival analyses

684 Survival analyses were performed using the R package survival v 2.44-1.1. Cox regression models  
685 included sex, age and stage (numeric) as covariates. Signature scores were recalibrated such that  
686 hazard ratios represented the effect of a 10% increase of the signature value from its theoretical  
687 minimum (-1) to its theoretical maximum (+1).

688

689

### 690 Analyses of genetic dependency screenings

691 CRISPR-based gene dependency scores and gene expression of cell lines were downloaded from the  
692 DepMap portal (<https://depmap.org/portal/download/>), CRISPR/Avana: 'Achilles\_gene\_effect.csv',  
693 expression: 'CCLE\_expression.csv') along with cell line annotations. Only cell lines with  
694 lineage\_sub\_subtype = NSCLC\_adenocarcinoma were investigated. Essential genes were also  
695 retrieved from the DepMap portal (union of 'common\_essentials.csv' and

696 'Achilles\_common\_essentials.csv'). Transcriptional regulators (TRs) that were differentially active in  
697 lepidic vs solid in CHUV and TCGA cohorts were tested for dependency.

#### 698 H&E, TTF1 and multicolor immunofluorescence staining

699 Hematoxylin and Eosin (H&E) staining was performed with standard protocol to retrieve the  
700 histologic patterns on all samples that were used for molecular or imaging assays. TTF1 staining was  
701 performed using mouse anti-TTF1 antibody, clone 8G7G3/1 (Invitrogen 18-0221) on Roche-Ventana  
702 Benchmark Ultra using the following protocol: (1) retrieval in Ventana Cell Conditioning 1 solution  
703 at 95°C for 64 min; (2) incubation for 32 min at 37°C, with the 8G7G3/1 being diluted 1/15; (3)  
704 application of Ventana ultraView Universal DAB Detection Kit followed by Ventana Hematoxylin as  
705 a nuclear counterstain. The multicolor immunofluorescence assay was performed using the  
706 Ventana Discovery ULTRA automate (Roche Diagnostics, Rotkreuz, Switzerland). All steps were  
707 performed automatically with Ventana solutions except if mentioned. Dewaxed and rehydrated  
708 paraffin sections were pretreated with heat using the CC1 solution for 40 minutes at 95°C. Primary  
709 antibodies were applied and revealed sequentially either with a rabbit Immpress HRP (Ready to  
710 use, Vector laboratories Laboratories) or a mouse Immpress HRP (Ready to use, Vector laboratories  
711 Laboratories) followed by incubation with a fluorescent tyramide. A heat denaturation step was  
712 performed after every revelation. The primary antibodies sequence was: rabbit anti CD8 (clone:  
713 Sp57, fluorophore: R6G), rabbit anti CD4 (clone: Sp35, fluorophore: DCC), mouse anti CD68 (clone:  
714 KP-1, fluorophore: R610), rabbit anti Ki67 (clone: 30-9, fluorophore: Cy5) and mouse anti CD20  
715 (clone: L26, fluorophore: FAM). All sections (H&E, TTF1 and multicolor immunofluorescence) were  
716 mounted with FluoromountG (Bioconcept) and scanned at 20x magnification using an Olympus  
717 VS120 whole slide scanner equipped with specific filters.

#### 718 Spatially resolved cell quantification framework (GridQuant)

719 H&E, TTF1 and multicolor immunofluorescence stained images were visualized using QuPath (107)  
720 v0.2.0 software. Histologic patterns were assessed and drawn with QuPath on H&E images and  
721 annotations were then transferred with minor adjustments to the nearby adjacent TTF1-stained  
722 and immunofluorescence-stained images. Regions that were too small were discarded. Next,  
723 quantifications and downstream statistical analyses were performed by developing a gridding  
724 framework named GridQuant, which involved the following steps for each image: (1) automated  
725 cell detections on QuPath with available algorithms; (2) acquisition of the coordinates of histologic  
726 pattern boundaries drawn on the image; (3) setup of a pixel grid spanning the entire image, with  
727 tunable grid spacing (pixel size); (4) for each cell type, summarization of cell detections at the pixel  
728 level as counts of the number of cells whose centroid fell within each pixel of the grid, thus  
729 obtaining one matrix for each cell type and for each pixel size considered; (5) various downstream  
730 spatially-resolved statistical analyses across cell types and histologic patterns (described in the  
731 following paragraphs). For TTF1-stained images, cell detection algorithm used on QuPath was  
732 'positive cell detection' and cell types investigated were TTF1-positive and TTF1-negative. For  
733 multicolor immunofluorescence-stained images, 'Watershed cell detection' algorithm was used to  
734 detect macrophages (CD63-positive cells), CD4 T cells (CD4-positive cells), CD8 T cells (CD8-positive  
735 cells), B cells (CD20-positive cells) and proliferating cells (Ki67-positive cells). In order to  
736 accommodate for potential variability in signal intensities among cell types and slides, the  
737 parameters of these cell detection algorithms were tuned with visual inspection in each slide and  
738 for each cell type. Steps 1 and 2 of GridQuant were implemented in Groovy programming language,  
739 steps 3, 4 and 5 were implemented in R v3.5.

740 GridQuant - cell density across patterns

741 Cell densities for each pixel were computed as the pixel counts divided by the pixel area.  
742 Distributions of densities across patterns were derived aggregating all pixels annotated to each  
743 pattern across all slides.

#### 744 GridQuant - cell types colocalizations across patterns

745 For each pattern-annotated region, the colocalization between cell types X and Y was computed as  
746 the Spearman's correlation coefficient between X and Y densities across all pixels falling within the  
747 region boundary. Pixels having total cell densities (summing densities of all available cell types)  
748 below the 5<sup>th</sup> percentile or below 200 N/mm<sup>2</sup> were discarded as likely representing empty regions  
749 corresponding to alveoli.

#### 750 GridQuant - solid pattern boundary analysis

751 Regions annotated to the solid pattern were partitioned into three internal sub-regions according  
752 to their distance from the boundary delimitating the solid-annotated region and other patterns,  
753 including the normal lung (boundaries between the solid-annotated region and uncertain  
754 transitioning patterns or the external cut delimiting the end of the slide were removed). The first  
755 sub-region was constituted by the pixels displaced at a distance between 0 and 0.5 mm, the second  
756 at a distance between 0.5 and 1 mm and the third was composed of all the remaining internal  
757 pixels (core). Moreover, an external region with pixels at a distance between 0 and 0.5 mm outside  
758 the boundary was also considered (**Fig. 4j**). The mean and distribution of densities of each cell type  
759 was then computed across all pixels falling within each of these 4 regions.

#### 760 In-situ detection and quantification of Tertiary Lymphoid Structures (TLS-finder)

761 Tertiary Lymphoid Structures (TLSs) were modeled as clusters of B cells detected with  
762 immunofluorescence staining. An automatic detection pipeline (named TLS-finder) was developed

763 to quantify their presence across histologic patterns. TLS-finder involved the following steps: (1)  
764 GridQuant framework was used to generate high resolution (pixel size = 20  $\mu m$ ) matrices of B cell  
765 counts; (2) each matrix was imported in Fiji (108) image analysis software as text image; (3) B cell  
766 count matrices were binarized using Fiji 'Convert to mask' function; (4) TLSs were modeled as  
767 connected components of the binarized B cell matrices, which were detected and labelled using Fiji  
768 plugin 'Find connected regions', with a minimum number of pixels to call a connected component  
769 set to 25 (corresponding to a minimum TLS size of 10000  $\mu m^2$ ); (5) Labelled TLSs were then  
770 processed with R scripts for downstream statistical analyses. The density of TLSs across patterns  
771 was computed as the number of distinct TLSs divided by the area of each pattern-annotated region  
772 (Fig. 4g).

### 773 Digital Spatial Profiling

774 Two runs of highly-multiplexed and spatially-resolved proteomic profiling of tumor and immune  
775 cells were performed with the GeoMx Digital Spatial Profiler (NanoString) as previously described  
776 (52) on 5 (batch1) and 3 (batch2) FFPE tissue sections. In batch1, immunofluorescence assays were  
777 performed using antibodies against CD3, CD20, CD45 and DAPI, and the multicolor images were  
778 used to guide the selection of 12 regions of interest (ROIs) for each slide; for each ROI, digital  
779 counts from barcodes corresponding to protein probes (52 immune and tumor-related proteins,  
780 **Supplementary Table 9**) were obtained using nCounter (NanoString). In each ROI, automated cell  
781 detection was performed on the immunofluorescence images to count nucleated cells (DAPI-  
782 positive) and, among them, CD45-positive cells; Immune ratio was computed as the ratio between  
783 the number of CD45-positive cells and DAPI-positive cells. In batch2, Pan-cytokeratin (PanCK),  
784 CD45, CD3 and DAPI antibodies were used; from each ROI, two areas of interest (AOIs) were  
785 extracted with image segmentation, one containing pixels positive for PanCK (tumor compartment)

786 and one for CD45 (immune compartment); digital counts were then obtained for each AOI  
787 separately (73 immune and tumor-related proteins, **Supplementary Table 9**); in PanCK-positive and  
788 CD45-positive AOIs only tumor-related and immune-related proteins, respectively, were tested in  
789 downstream analyses. Levels of 3 housekeeping proteins (GAPDH, histone H3, S6) and 3 negative  
790 controls (Ms IgG1, Ms IgG2a, Rb IgG) were also measured. Digital counts for each protein were  
791 normalized with internal spike-in controls (ERCC) and signal-to-noise ratio (SNR), i.e. the ratio  
792 between the ERCC-normalized counts of the protein and the geometric mean of the negative  
793 controls assayed in the ROI/AOI considered. In all downstream analyses, only proteins having  
794 SNR>2 in at least 3 ROIs/AOIs were tested. ROIs/AOIs were annotated according to their histologic  
795 pattern and location with respect to the pattern boundary (center/core or periphery) using  
796 adjacent H&E-stained tissue sections. Some ROIs were taken also from Tertiary Lymphoid  
797 Structures (TLS) but were not used for the analyses presented in this study. Differences between  
798 locations in solid pattern regions were tested with T-test (batch1, one solid pattern region and 6  
799 ROIs) and with two-way ANOVA controlling for sample and testing separately AOIs from immune  
800 and tumor compartments (batch2, three samples with one solid pattern region in each of them, 26  
801 total AOIs for each tissue compartment).

#### 802 Data and code availability

803 All datasets analyzed and the corresponding accession numbers are reported in **Supplementary**  
804 **Table 1**. Data generated in this study has been deposited in two Zenodo repositories, one  
805 containing raw images for H&E, TTF1 and immunofluorescence staining and Digital Spatial Profiling  
806 data (DOI: 10.5281/zenodo.3941450) and one containing processed molecular data (somatic  
807 mutations, gene expression tables and DNA methylation beta values, DOI:  
808 10.5281/zenodo.4443496). Source code for GridQuant and TLS-finder pipelines is available in two

809 public GitHub repositories (<https://github.com/CSOgroup/GridQuant> and

810 <https://github.com/CSOgroup/TLS-finder>).

811

## 812 **Acknowledgements**

813 We would like to thank Leila Akkari and Luca Magnani for critical reading of the manuscript and  
814 insightful discussions. This work was supported by the Gabriella Giorgi-Cavaglieri foundation (G.  
815 Ciriello), University of Lausanne Interdisciplinary Research Grant (G. Ciriello, I. Letovanec, N. Riggi),  
816 the Swiss National Science Foundation (SNSF, grant 310030\_169519) (G. Ciriello, D. Tavernari), and  
817 Emma Muschamp foundation (D. Tavernari).

818

## 819 **References**

- 820 1. Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, et al. Distinct patterns of  
821 somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat*  
822 *Genet.* 2016;48:607–16.
- 823 2. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015  
824 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and  
825 Radiologic Advances Since the 2004 Classification. *Journal of Thoracic Oncology.*  
826 2015;10:1243–60.
- 827 3. Subramanian J, Govindan R. Lung Cancer in Never Smokers: A Review. *JCO. American Society*  
828 *of Clinical Oncology;* 2007;25:561–70.
- 829 4. Samet JM, Avila-Tang E, Boffetta P, Hannan LM, Olivo-Marston S, Thun MJ, et al. Lung cancer  
830 in never smokers: clinical epidemiology and environmental risk factors. *Clin Cancer Res.*  
831 2009;15:5626–45.
- 832 5. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung  
833 adenocarcinoma. *Nature.* 2014;511:543–50.
- 834 6. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus  
835 standard chemotherapy as first-line treatment for European patients with advanced EGFR  
836 mutation-positive non-small-cell lung cancer (EORTC): a multicentre, open-label,  
837 randomised phase 3 trial. *Lancet Oncol.* 2012;13:239–46.
- 838 7. Solomon BJ, Mok T, Kim D-W, Wu Y-L, Nakagawa K, Mekhail T, et al. First-Line Crizotinib versus  
839 Chemotherapy in ALK-Positive Lung Cancer. *New England Journal of Medicine.*  
840 *Massachusetts Medical Society;* 2014;371:2167–77.

- 841 8. Peters S, Camidge DR, Shaw AT, Gadgeel S, Ahn JS, Kim D-W, et al. Alectinib versus Crizotinib in  
842 Untreated ALK-Positive Non–Small-Cell Lung Cancer. *New England Journal of Medicine*.  
843 Massachusetts Medical Society; 2017;377:829–38.
- 844 9. Shaw AT, Ou S-HI, Bang Y-J, Camidge DR, Solomon BJ, Salgia R, et al. Crizotinib in ROS1-  
845 Rearranged Non–Small-Cell Lung Cancer. *New England Journal of Medicine*. Massachusetts  
846 Medical Society; 2014;371:1963–71.
- 847 10. Planchard D, Besse B, Groen HJM, Souquet P-J, Quoix E, Baik CS, et al. Dabrafenib plus  
848 trametinib in patients with previously treated BRAF(V600E)-mutant metastatic non-small cell  
849 lung cancer: an open-label, multicentre phase 2 trial. *Lancet Oncol*. 2016;17:984–93.
- 850 11. Bruin EC de, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal  
851 diversity in genomic instability processes defines lung cancer evolution. *Science*.  
852 2014;346:251–6.
- 853 12. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al.  
854 Tracking the Evolution of Non–Small-Cell Lung Cancer. *New England Journal of Medicine*.  
855 2017;376:2109–21.
- 856 13. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev*  
857 *Clin Oncol*. 2018;15:81–94.
- 858 14. Turke AB, Zejnullahu K, Wu Y-L, Song Y, Dias-Santagata D, Lifshits E, et al. Pre-existence and  
859 clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell*. 2010;17:77–88.
- 860 15. Hayes DN, Monti S, Parmigiani G, Gilks CB, Naoki K, Bhattacharjee A, et al. Gene Expression  
861 Profiling Reveals Reproducible Human Lung Adenocarcinoma Subtypes in Multiple  
862 Independent Patient Cohorts. *Journal of Clinical Oncology*. 2006;24:5079–90.
- 863 16. Dietz S, Lifshitz A, Kazdal D, Harms A, Endris V, Winter H, et al. Global DNA methylation reflects  
864 spatial heterogeneity and molecular evolution of lung adenocarcinomas. *International*  
865 *Journal of Cancer*. 2019;144:1061–72.
- 866 17. Sharma A, Merritt E, Hu X, Cruz A, Jiang C, Sarkodie H, et al. Non-Genetic Intra-Tumor  
867 Heterogeneity Is a Major Predictor of Phenotypic Heterogeneity and Ongoing Evolutionary  
868 Dynamics in Lung Tumors. *Cell Reports*. 2019;29:2164-2174.e5.
- 869 18. Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW, et al. Single-cell RNA sequencing demonstrates  
870 the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nature*  
871 *Communications*. Nature Publishing Group; 2020;11:2285.
- 872 19. Hua X, Zhao W, Pesatori AC, Consonni D, Caporaso NE, Zhang T, et al. Genetic and epigenetic  
873 intratumor heterogeneity impacts prognosis of lung adenocarcinoma. *Nature*  
874 *Communications*. Nature Publishing Group; 2020;11:2459.
- 875 20. Ghorani E, Reading JL, Henry JY, Massy MR de, Rosenthal R, Turati V, et al. The T cell  
876 differentiation landscape is shaped by tumour mutations in lung cancer. *Nature Cancer*.  
877 Nature Publishing Group; 2020;1:546–61.

- 878 21. AbdulJabbar K, Raza SEA, Rosenthal R, Jamal-Hanjani M, Veeriah S, Akarca A, et al. Geospatial  
879 immune variability illuminates differential evolution of lung adenocarcinoma. *Nature*  
880 *Medicine*. Nature Publishing Group; 2020;1–9.
- 881 22. Joshi K, de Massy MR, Ismail M, Reading JL, Uddin I, Woolston A, et al. Spatial heterogeneity of  
882 the T cell receptor repertoire reflects the mutational landscape in lung cancer. *Nature*  
883 *Medicine*. Nature Publishing Group; 2019;25:1549–59.
- 884 23. Rosenthal R, Cadieux EL, Salgado R, Bakir MA, Moore DA, Hiley CT, et al. Neoantigen-directed  
885 immune escape in lung cancer evolution. *Nature*. Nature Publishing Group; 2019;567:479–  
886 85.
- 887 24. Klemm F, Maas RR, Bowman RL, Kornete M, Soukup K, Nassiri S, et al. Interrogation of the  
888 Microenvironmental Landscape in Brain Tumors Reveals Disease-Specific Alterations of  
889 Immune Cells. *Cell*. Elsevier; 2020;181:1643-1660.e17.
- 890 25. Travis WD, Brambilla E, Geisinger KR. Histological grading in lung cancer: one system for all or  
891 separate systems for each histological type? *European Respiratory Journal*. European  
892 Respiratory Society; 2016;47:720–3.
- 893 26. Mäkinen JM, Laitakari K, Johnson S, Mäkitaro R, Bloigu R, Pääkkö P, et al. Histological features  
894 of malignancy correlate with growth patterns and patient outcome in lung adenocarcinoma.  
895 *Histopathology*. 2017;71:425–36.
- 896 27. Caso R, Sanchez-Vega F, Tan KS, Mastrogiacomo B, Zhou J, Jones GD, et al. The Underlying  
897 Tumor Genomics of Predominant Histologic Subtypes in Lung Adenocarcinoma. *Journal of*  
898 *Thoracic Oncology*. 2020;15:1844–56.
- 899 28. Chen J, Yang H, Teo ASM, Amer LB, Sherbaf FG, Tan CQ, et al. Genomic landscape of lung  
900 adenocarcinoma in East Asians. *Nature Genetics*. Nature Publishing Group; 2020;52:177–86.
- 901 29. Kumar S, Warrell J, Li S, McGillivray PD, Meyerson W, Salichos L, et al. Passenger Mutations in  
902 More Than 2,500 Cancer Genomes: Overall Molecular Functional Impact and Consequences.  
903 *Cell*. 2020;180:915-927.e16.
- 904 30. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling  
905 Pathways in The Cancer Genome Atlas. *Cell*. 2018;173:321-337.e10.
- 906 31. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, et al. Somatic mutations  
907 affect key pathways in lung adenocarcinoma. *Nature*. 2008;455:1069–75.
- 908 32. Jiménez-Sánchez A, Cast O, Miller ML. Comprehensive Benchmarking and Integration of  
909 Tumor Microenvironment Cell Estimation Methods. *Cancer Res*. 2019;79:6238–46.
- 910 33. Nallanthighal S, Heiserman JP, Cheon D-J. The Role of the Extracellular Matrix in Cancer  
911 Stemness. *Front Cell Dev Biol*. 2019;7:86
- 912 34. Lu P, Takai K, Weaver VM, Werb Z. Extracellular matrix degradation and remodeling in  
913 development and disease. *Cold Spring Harb Perspect Biol*. 2011;3.

- 914 35. Liot S, Aubert A, Hervieu V, Kholti NE, Schalkwijk J, Verrier B, et al. Loss of Tenascin-X  
915 expression during tumor progression: A new pan-cancer marker. *Matrix Biology Plus*. 2020;6–  
916 7:100021.
- 917 36. Yuan B-Z, Jefferson AM, Baldwin KT, Thorgeirsson SS, Popescu NC, Reynolds SH. DLC-1  
918 operates as a tumor suppressor gene in human non-small cell lung carcinomas. *Oncogene*.  
919 Nature Publishing Group; 2004;23:1405–11.
- 920 37. Tamura M, Sasaki Y, Koyama R, Takeda K, Idogawa M, Tokino T. Forkhead transcription factor  
921 FOXF1 is a novel target gene of the p53 family and regulates cancer cell migration and  
922 invasiveness. *Oncogene*. Nature Publishing Group; 2014;33:4837–46.
- 923 38. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, et al. Phenotype molding of  
924 stromal cells in the lung tumor microenvironment. *Nature Medicine*. 2018;24:1277.
- 925 39. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization  
926 of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet*.  
927 2016;48:838–47.
- 928 40. Myatt SS, Lam EW-F. The emerging roles of forkhead box (Fox) proteins in cancer. *Nature*  
929 *Reviews Cancer*. Nature Publishing Group; 2007;7:847–59.
- 930 41. Huber A-L, Papp SJ, Chan AB, Henriksson E, Jordan SD, Kriebs A, et al. CRY2 and FBXL3  
931 Cooperatively Degrade c-MYC. *Molecular Cell*. 2016;64:774–89.
- 932 42. Roussel-Gervais A, Naciri I, Kirsh O, Kasprzyk L, Velasco G, Grillo G, et al. Loss of the Methyl-  
933 CpG-Binding Protein ZBTB4 Alters Mitotic Checkpoint, Increases Aneuploidy, and Promotes  
934 Tumorigenesis. *Cancer Research*. 2017;77:62–73.
- 935 43. Liu Z, Yang X, Li Z, McMahon C, Sizer C, Barenboim-Stapleton L, et al. CASZ1, a candidate  
936 tumor-suppressor gene, suppresses neuroblastoma tumor growth through reprogramming  
937 gene expression. *Cell Death & Differentiation*. Nature Publishing Group; 2011;18:1174–83.
- 938 44. Charpentier MS, Christine KS, Amin NM, Dorr KM, Kushner EJ, Bautch VL, et al. CASZ1  
939 promotes vascular assembly and morphogenesis through the direct regulation of an  
940 EGFL7/RhoA-mediated pathway. *Dev Cell*. 2013;25:132–43.
- 941 45. Knight JF, Sung VYC, Kuzmin E, Couzens AL, Verteuil DA de, Ratcliffe CDH, et al. KIBRA (WWC1)  
942 Is a Metastasis Suppressor Gene Affected by Chromosome 5q Loss in Triple-Negative Breast  
943 Cancer. *Cell Reports*. 2018;22:3191–205.
- 944 46. Wang Z, Katsaros D, Biglia N, Shen Y, Fu Y, Tiirikainen M, et al. Low expression of WWC1, a  
945 tumor suppressor gene, is associated with aggressive breast cancer and poor survival  
946 outcome. *FEBS Open Bio*. 2019;9:1270–80.
- 947 47. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational  
948 correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in  
949 cancer cells. *Nat Genet*. 2017;49:1779–84.

- 950 48. Mezheyeuski A, Bergsland CH, Backman M, Djureinovic D, Sjöblom T, Bruun J, et al.  
951 Multispectral imaging for quantitative and compartment-specific immune infiltrates reveals  
952 distinct immune profiles that classify lung cancer patients. *The Journal of Pathology*.  
953 2018;244:421–31.
- 954 49. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes  
955 upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer*  
956 *Res*. 2012;72:100–11.
- 957 50. Schabath MB, Welsh EA, Fulp WJ, Chen L, Teer JK, Thompson ZJ, et al. Differential association  
958 of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and  
959 immune surveillance in lung adenocarcinoma. *Oncogene*. 2016;35:3209–16.
- 960 51. Shedden K, Taylor JMG, Enkemann SA, Tsao M-S, Yeatman TJ, Gerald WL, et al. Gene  
961 expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation  
962 study. *Nature Medicine*. Nature Publishing Group; 2008;14:822–7.
- 963 52. Helmink BA, Reddy SM, Gao J, Zhang S, Basar R, Thakur R, et al. B cells and tertiary lymphoid  
964 structures promote immunotherapy response. *Nature*. Nature Publishing Group;  
965 2020;577:549–55.
- 966 53. Sautès-Fridman C, Petitprez F, Calderaro J, Fridman WH. Tertiary lymphoid structures in the  
967 era of cancer immunotherapy. *Nature Reviews Cancer*. Nature Publishing Group;  
968 2019;19:307–25.
- 969 54. Suzuki K, Kachala SS, Kadota K, Shen R, Mo Q, Beer DG, et al. Prognostic Immune Markers in  
970 Non-Small Cell Lung Cancer. *Clinical Cancer Research*. 2011;17:5247–56.
- 971 55. Nikitin AY, Alcaraz A, Anver MR, Bronson RT, Cardiff RD, Dixon D, et al. Classification of  
972 Proliferative Pulmonary Lesions of the Mouse: Recommendations of the Mouse Models of  
973 Human Cancers Consortium. *Cancer Res*. American Association for Cancer Research;  
974 2004;64:2307–16.
- 975 56. Meuwissen R, Berns A. Mouse models for human lung cancer. *Genes Dev*. 2005;19:643–64.
- 976 57. Marjanovic ND, Hofree M, Chan JE, Canner D, Wu K, Trakala M, et al. Emergence of a High-  
977 Plasticity Cell State during Lung Cancer Evolution. *Cancer Cell*. 2020;38:229-246.e13.
- 978 58. LaFave LM, Kartha VK, Ma S, Meli K, Del Priore I, Lareau C, et al. Epigenomic State Transitions  
979 Characterize Tumor Progression in Mouse Lung Adenocarcinoma. *Cancer Cell*. 2020;38:212-  
980 228.e13.
- 981 59. Cao J, Yan Q. Cancer Epigenetics, Tumor Immunity, and Immunotherapy. *Trends in Cancer*.  
982 Elsevier; 2020;6:580–92.
- 983 60. Villanueva L, Álvarez-Errico D, Esteller M. The Contribution of Epigenetics to Cancer  
984 Immunotherapy. *Trends in Immunology*. 2020;41:676–91.

- 985 61. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al.  
986 Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *New*  
987 *England Journal of Medicine*. Massachusetts Medical Society; 2020;382:503–13.
- 988 62. Kauczor H-U, Baird A-M, Blum TG, Bonomo L, Bostantzoglou C, Burghuber O, et al. ESR/ERS  
989 statement paper on lung cancer screening. *Eur Radiol*. 2020;30:3277–94.
- 990 63. Blons H, Garinet S, Laurent-Puig P, Oudart J-B. Molecular markers and prediction of response  
991 to immunotherapy in non-small cell lung cancer, an update. *J Thorac Dis*. 2019;11:S25–36.
- 992 64. Forde PM, Chaft JE, Smith KN, Anagnostou V, Cottrell TR, Hellmann MD, et al. Neoadjuvant PD-  
993 1 Blockade in Resectable Lung Cancer. *N Engl J Med*. 2018;378:1976–86.
- 994 65. Owen D, Chaft JE. Immunotherapy in surgically resectable non-small cell lung cancer. *J Thorac*  
995 *Dis*. 2018;10:S404–11.
- 996 66. Vansteenkiste J, Doooms C. MS14.01 Basic Science Rationale of IO in Early Stage NSCLC? *Journal*  
997 *of Thoracic Oncology*. 2018;13:S268–9.
- 998 67. Patani N, Barbashina V, Lambros MBK, Gauthier A, Mansour M, Mackay A, et al. Direct  
999 evidence for concurrent morphological and genetic heterogeneity in an invasive ductal  
1000 carcinoma of triple-negative phenotype. *J Clin Pathol*. 2011;64:822–8.
- 1001 68. Friemel J, Rechsteiner M, Frick L, Böhm F, Struckmann K, Egger M, et al. Intratumor  
1002 Heterogeneity in Hepatocellular Carcinoma. *Clin Cancer Res*. 2015;21:1951–61.
- 1003 69. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. *Biom J*.  
1004 2008;50:346–63.
- 1005 70. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a Shared  
1006 Vision for Cancer Genomic Data. *N Engl J Med*. 2016;375:1109–12.
- 1007 71. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandath C, Stewart C, et al. Scalable Open  
1008 Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines.  
1009 *Cell Systems*. 2018;6:271-281.e7.
- 1010 72. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhir R, Getz G. GISTIC2.0 facilitates  
1011 sensitive and confident localization of the targets of focal somatic copy-number alteration in  
1012 human cancers. *Genome Biology*. 2011;12:R41.
- 1013 73. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The Immune  
1014 Landscape of Cancer. *Immunity*. 2018;48:812-830.e14.
- 1015 74. Collisson EA, Campbell JD, Brooks AN, Berger AH, Lee W, Chmielecki J, et al. Comprehensive  
1016 molecular profiling of lung adenocarcinoma. *Nature*. Nature Publishing Group;  
1017 2014;511:543–50.
- 1018 75. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics  
1019 Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer*  
1020 *Discov*. American Association for Cancer Research; 2012;2:401–4.

- 1021 76. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of  
1022 complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6:pl1.
- 1023 77. Goldman MJ, Craft B, Hastie M, Repčička K, McDade F, Kamath A, et al. Visualizing and  
1024 interpreting cancer genomics data via the Xena platform. *Nature Biotechnology*. Nature  
1025 Publishing Group; 2020;38:675–8.
- 1026 78. Der SD, Sykes J, Pintilie M, Zhu C-Q, Strumpf D, Liu N, et al. Validation of a histology-  
1027 independent prognostic gene signature for early-stage, non-small-cell lung cancer including  
1028 stage IA patients. *J Thorac Oncol*. 2014;9:59–64.
- 1029 79. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome  
1030 Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing  
1031 data. *Genome Res*. 2010;20:1297–303.
- 1032 80. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.  
1033 arXiv:1303.3997. 2013.
- 1034 81. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for  
1035 variation discovery and genotyping using next-generation DNA sequencing data. *Nature  
1036 Genetics*. Nature Publishing Group; 2011;43:491–8.
- 1037 82. Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: A Precision  
1038 Oncology Knowledge Base. *JCO Precision Oncology*. 2017;1–16.
- 1039 83. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative  
1040 genomics viewer. *Nature Biotechnology*. Nature Publishing Group; 2011;29:24–6.
- 1041 84. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a  
1042 reference genome. *BMC Bioinformatics*. 2011;12:323.
- 1043 85. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal  
1044 RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
- 1045 86. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch  
1046 effects and other unwanted variation in high-throughput experiments. *Bioinformatics*.  
1047 2012;28:882–3.
- 1048 87. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, et al. The BioMart community  
1049 portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res*.  
1050 Oxford Academic; 2015;43:W589–98.
- 1051 88. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a  
1052 flexible and comprehensive Bioconductor package for the analysis of Infinium DNA  
1053 methylation microarrays. *Bioinformatics*. Oxford Academic; 2014;30:1363–9.
- 1054 89. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of  
1055 Infinium DNA methylation BeadChip probes. *Nucleic Acids Res*. Oxford Academic;  
1056 2017;45:e22–e22.

- 1057 90. Forrest ARR, Kawaji H, Rehli M, Kenneth Baillie J, de Hoon MJL, Haberle V, et al. A promoter-  
1058 level mammalian expression atlas. *Nature*. Nature Publishing Group; 2014;507:462–70.
- 1059 91. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript  
1060 detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome*  
1061 *Biology*. 2019;20:213.
- 1062 92. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun*.  
1063 2015;6:8971.
- 1064 93. Mina M, Raynaud F, Tavernari D, Battistello E, Sungalee S, Saghafinia S, et al. Conditional  
1065 selection of genomic alterations dictates cancer evolution and Oncogenic dependencies.  
1066 *Cancer Cell*. 2017;32:155–68.
- 1067 94. Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, et al. FunSeq2: a framework for prioritizing  
1068 noncoding regulatory variants in cancer. *Genome Biology*. 2014;15:480.
- 1069 95. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression  
1070 analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
- 1071 96. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential  
1072 expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
- 1073 97. Law CW, Alhamdoosh M, Su S, Dong X, Tian L, Smyth GK, et al. RNA-seq analysis is easy as 1-2-  
1074 3 with limma, Glimma and edgeR. *F1000Res*. 2018;5:1408.
- 1075 98. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular  
1076 signatures database (MSigDB) 3.0. *Bioinformatics*. Oxford Academic; 2011;27:1739–40.
- 1077 99. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for  
1078 the unification of biology. *Nat Genet*. 2000;25:25–9.
- 1079 100. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive  
1080 Integration of Single-Cell Data. *Cell*. 2019;177:1888-1902.e21.
- 1081 101. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-  
1082 Seq data. *BMC Bioinformatics*. 2013;14:7.
- 1083 102. Danaher P, Warren S, Lu R, Samayoa J, Sullivan A, Pekker I, et al. Pan-cancer adaptive immune  
1084 resistance as defined by the Tumor Inflammation Signature (TIS): results from The Cancer  
1085 Genome Atlas (TCGA). *J Immunother Cancer*. 2018;6:63.
- 1086 103. Foroutan M, Bhuvu DD, Lyu R, Horan K, Cursons J, Davis MJ. Single sample scoring of molecular  
1087 phenotypes. *BMC Bioinformatics*. 2018;19:404.
- 1088 104. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al.  
1089 SCENIC: Single-cell regulatory network inference and clustering. *Nat Methods*.  
1090 2017;14:1083–6.

- 1091 105. Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library  
1092 for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina  
1093 HumanMethylationEPIC BeadArray. *Genome Biology*. 2018;19:64.
- 1094 106. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, et al. ARACNE: An  
1095 Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular  
1096 Context. *BMC Bioinformatics*. 2006;7:S7.
- 1097 107. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath:  
1098 Open source software for digital pathology image analysis. *Scientific Reports*. Nature  
1099 Publishing Group; 2017;7:16878.
- 1100 108. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-  
1101 source platform for biological-image analysis. *Nature Methods*. Nature Publishing Group;  
1102 2012;9:676–82.
- 1103
- 1104
- 1105

1106 **Figure Legends**

1107 **Figure 1: Inter-patient heterogeneity among LUAD histologic patterns**

1108 **a)** H&E staining of lung adenocarcinoma histologic patterns (from left to right): lepidic, papillary,  
1109 acinar, and solid.

1110 **b)** Total number of somatic coding mutations (Y-axis) in TCGA samples (colored points) stratified by  
1111 histologic pattern classification (X-axis). The outlier lepidic-annotated TCGA sample is highlighted by  
1112 its patient ID. P-values are computed by Wilcoxon two-tailed test.

1113 **c)** Representative H&E images of two tumor tissue slides for the TCGA-44-7670 sample: an image  
1114 taken from slide corresponding to the tumor sample used for histopathology review (left) and an  
1115 image corresponding to the tumor sample used for molecular analyses (right). Complete images  
1116 can be accessed at: <https://cancer.digitalslidearchive.org/> .

1117 **d-e)** Number of **(d)** significantly differentially expressed genes and **(e)** significantly differentially  
1118 methylated probes identified based on different FDR thresholds (X-axis) by comparing patients  
1119 grouped by the real prevalent histologic pattern (black) or after randomizing the histologic pattern  
1120 labels (gray). Error bars correspond to one standard deviation upon 100 label permutations.

1121 **f)** mRNA expression of two top differentially expressed genes (Y-axis) in TCGA samples stratified by  
1122 prevalent pattern (X-axis). Colored dots on the right indicate the median expression value of each  
1123 group and arrows represent the direction of the fold-change (FC): upward (downward) arrows  
1124 indicate that the more aggressive pattern have lower (higher) median expression than the less  
1125 aggressive pattern. Pairwise FCs always compare the more aggressive to the less aggressive  
1126 pattern, hence upward arrows correspond to negative FCs and downward arrows to positive FCs.

1127 **g)** Pie chart distributions of the sign of pairwise FCs computed for all differentially expressed genes  
1128 (top) and differentially methylated probes (bottom).

1129 **h)** Significantly enriched gene sets among genes over-expressed in lepidic-prevalent samples (blue  
1130 bars) and in solid-prevalent samples (red and yellow bars).

1131 **i)** Significantly enriched gene sets among promoter probes with lower DNA methylation in lepidic-  
1132 prevalent samples than solid-prevalent samples (blue bars) or with lower DNA methylation in solid-  
1133 prevalent samples than in lepidic-prevalent samples (yellow bars).

1134 **j)** Mean mRNA expression scores for multiple cell types (rows) within each pattern subtype  
1135 (columns). Values are normalized by rows (Z-scores) to show relative differences among patterns.

1136

1137

1138 **Figure 2: Intra-tumor heterogeneity among LUAD histologic patterns**

1139 **a)** Schematic representation of histopathology-guided multi-region sampling: FFPE slides were  
1140 reviewed for pattern identification, tumor regions corresponding to a unique pattern were  
1141 dissected and molecularly profiled (left). We have collected 29 tumor regions (+10 adjacent normal  
1142 tissue) from 10 primary lung adenocarcinoma samples (right).

1143 **b)** Occurrence of recurrent LUAD genetic mutations in molecularly profiled regions. Regions from  
1144 the same patient are grouped together; patients are numbered (top) and histologic patterns are  
1145 color coded (annotation bar).

1146 **c)** Heatmap representation of differentially expressed genes among LUAD histologic patterns (rows,  
1147 adjusted p-value < 0.001). Samples (columns) are identified by patient number followed by a letter  
1148 corresponding to individual tumor regions. Histologic patterns are color coded. Cellular processes  
1149 associated to significantly enriched gene sets are annotated on the right.

1150 **d)** mRNA expression differences among histologic patterns for a selected panel of extra-cellular  
1151 matrix components and/or regulators (over-expressed in solid regions at the top, over-expressed in  
1152 lepidic regions at the bottom). Expression values within each patient were normalized to the mean  
1153 of the corresponding lepidic regions. Samples corresponding to same patient are connected by a  
1154 dashed line and color coded based on concordance between intra-tumor differences and pattern  
1155 progression.

1156 **e)** Immune score predicted within each tumor region for each patient.

1157

1158 **Figure 3: Tumor intrinsic features of LUAD histologic patterns**

1159 **a)** Schematic representation of single cell RNA-seq analysis of 3 tumor samples from 3 patients. For  
1160 each sample, differential expression analysis between tumor (red) and non-tumor (gray) cells led to  
1161 identify 2,410 genes preferentially expressed in cancer cells.

1162 **b)** Volcano plot showing mRNA expression fold-changes of cancer-specific genes between lepidic  
1163 and solid tumor regions ( $\log_2$  – Y-axis) and corresponding p-values ( $-\log_{10}$  – X-axis). Significant  
1164 genes are color coded (red: over-expressed in solid regions; blue: over-expressed in lepidic regions)

1165 **c)** Significantly enriched gene sets among genes over-expressed in lepidic regions (blue bars) and in  
1166 solid regions (red bars).

1167 **d)** Scatterplot of single tumor cells from 3 patients scored by lepidic-like single cell signature (Y-axis)  
1168 and solid-like single cell signature (X-axis). Single cells are color coded by combined signature scores  
1169 (main scatterplot) and separately shown for each patient sample (top right insets).

1170 **e)** Transcriptional regulator (TR) activity scores obtained with the VIPER algorithm upon comparing  
1171 lepidic and solid annotated regions in the CHUV (X-axis) and TCGA (Y-axis) cohorts. Significant TRs  
1172 are color coded (red: solid associated, blue: lepidic associated) and the top scoring are labeled.

1173 **f)** Gene dependency scores obtained from the AVANA CRISPR screening dataset. Negative (positive)  
1174 values indicate fitness decrease (increase) upon gene knock-out. Values for lepidic (blue) and solid  
1175 (red) transcriptional regulators (TR) are the mean obtained upon gene KO in lung adenocarcinoma  
1176 cell lines.

1177 **g)** Overall survival difference (Kaplan-Meier curve) between TCGA samples within the top (red) and  
1178 bottom (blue) quartiles of L2S scores. P-value was computed by log-rank test.

1179 **h)** Hazard ratios associated with increasing values of the L2S signature score (10% increase) in 7  
1180 independent LUAD datasets comprising >100 patients each (# column). P-values were computed by  
1181 multi-variate Cox regression. The size of the dots is proportional to the number of sample. 95%  
1182 confidence intervals are reported as horizontal lines.

1183 **i)** Correlation values between gene-set mRNA expression scores for multiple cell types and L2S  
1184 scores in 10 independent datasets.

1185

1186

1187 **Figure 4: Spatial immune profiles of LUAD histologic patterns**

1188 **a-b)** H&E (a) and TTF1 (b) staining of LUAD tissue sample (Patient 1). Lepidic (blue) and acinar  
1189 (orange) patterns are contoured.

1190 **c)** Multi-color immunofluorescence staining for a LUAD tissue sample (Patient 1). Images show  
1191 separately fluorescence staining for Ki-67 (top left), CD8 (top right), CD20 (bottom left), and CD4  
1192 (bottom right).

1193 **d)** Schematic representation of GridQuant: each image is binned into a grid with bins/pixels of  
1194 variable sizes. In this study we tested pixel sizes varying from 10 to 500  $\mu\text{m}$  (left). Fluorescence  
1195 intensity is then averaged for each bin. An example for CD8 fluorescence is shown (right).

1196 **e)** Boxplot distribution of cell densities (number of cells per  $\text{mm}^2$ ,  $\text{N}/\text{mm}^2$ ) for cells that were  
1197 positive for each of the tested antibodies (X-axis). For each antibody, cell density values are  
1198 computed for each pixel and values obtained for pixel from regions with a different histologic  
1199 pattern are compared. Pixel size = 200  $\mu\text{m}$ .

1200 **f)** Multicolor IF staining of tertiary lymphoid structures (TLS).

1201 **g)** Quantification of TLS density across different regions corresponding to a unique histologic  
1202 pattern (colored points).

1203 **h)** Correlation between the number of TTF1+ and TTF1- cells within each pixel of a given region  
1204 corresponding to a unique histologic pattern (colored points). Pixel size = 200  $\mu\text{m}$ .

1205 **i)** H&E (left) and multicolor IF (center) staining of LUAD tissue sample (Patient 8) and zoom-in of the  
1206 IF staining of the solid pattern (right). Histologic patterns are contoured and color coded.

1207 **j)** Schematic representation of spatial quantification based on distance from the tumor boundary  
1208 (red line). Contoured regions define discrete subsets of pixels within a certain interval of distances  
1209 from the tumor boundary.

1210 **k)** Spatial quantification based on distance from the tumor boundary for the solid region of Patient  
1211 8 (8B). Signal intensities were averaged among pixels within a certain interval of distances from the  
1212 tumor boundary (gray line) and at the tumor core, defined as >1mm inside the boundary. Pixel size  
1213 = 100  $\mu\text{m}$ .

1214 **l)** Spatial quantification based on distance from the tumor boundary for all solid regions. Pixel size =  
1215 100  $\mu\text{m}$ .

1216

1217

1218 **Figure 5: Digital spatial profiling (DSP) of LUAD histologic patterns**

1219 **a)** Immuno-score of all tumor regions of interest (ROIs) for 5 patients defined as fraction of CD45  
1220 positive cells by immunofluorescence analysis (top barplot, bars are color coded based on the  
1221 pattern of the corresponding region). ROIs are annotated by tumor core or periphery localization  
1222 (black and white circles, respectively) and by DSP protein expression of the top correlated protein  
1223 with CD45+ immuno-score (bottom heatmap). DSP values are normalized by signal-to-noise ratio  
1224 (SNR) and by z-score for each patient.

1225 **b)** Top: H&E staining of LUAD tissue sample (Patient 8). Bottom: Schematic diagram of Patient 8  
1226 normal tissue regions and tumor histologic patterns (color coded) and selected tumor regions of  
1227 interest analyzed by DSP.

1228 **c)** Differentially expressed proteins between ROIs at the core of the solid region (R8, R9) and at the  
1229 periphery of the solid region (R6, R7, R10, R12). Values are normalized by SNR and by z-score for  
1230 each patient.

1231 **d)** Schematic representation of DSP analysis with ROI masks: ROIs at the core or periphery of the  
1232 tumor (left) were first analyzed by IF for PanCK (green), CD3 (light blue), DNA (dark blue), and CD45  
1233 (red); next, CD45 and PanCK fluorescence was used to build two masks (white: selected, black:  
1234 unselected) to selectively analyze either PanCK+ cells or CD45+ cells.

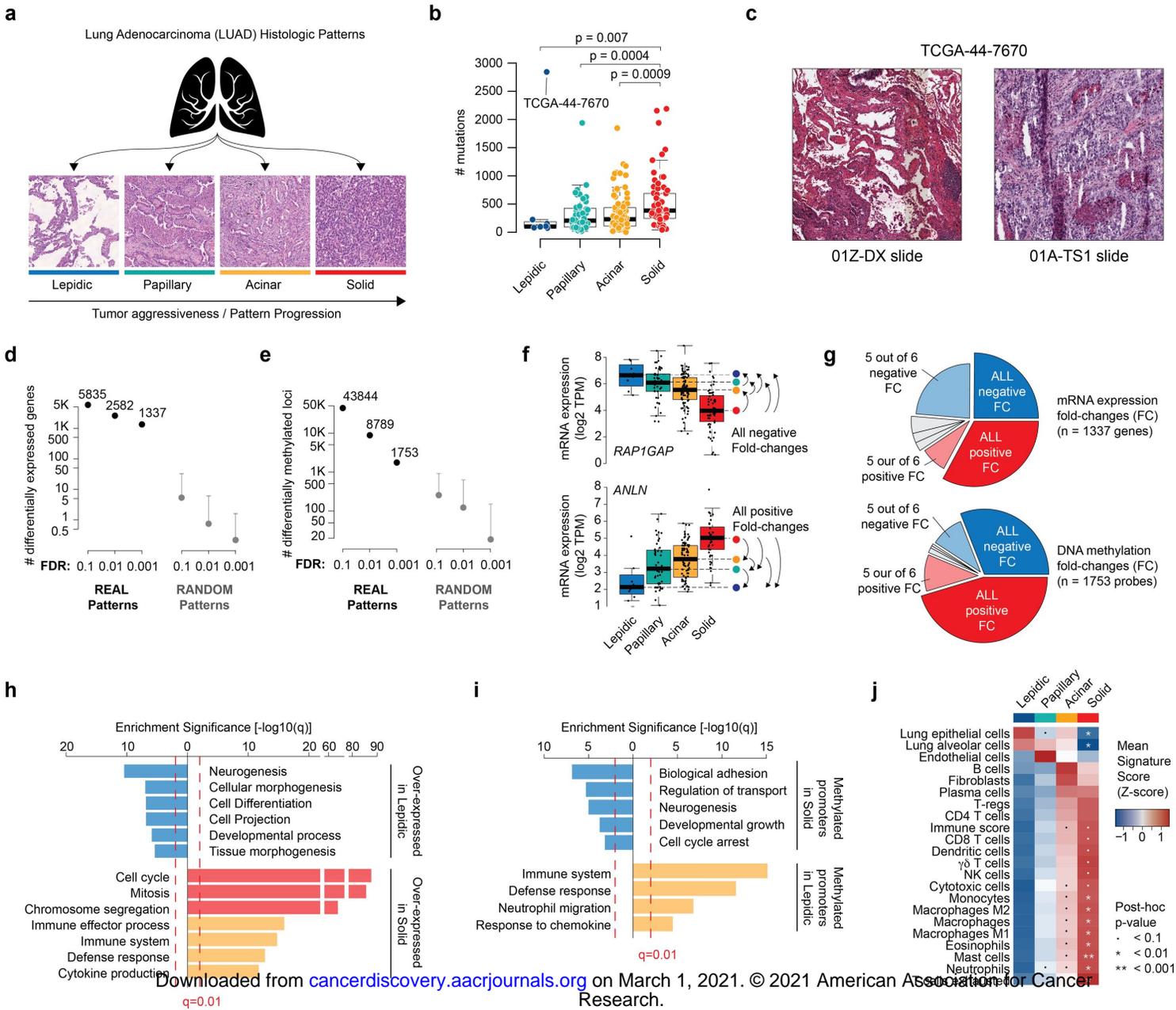
1235 **e-f)** Differentially expressed proteins between core (black circles) and periphery (white circle) ROIs  
1236 exclusively comprising **(e)** PanCK+ cells or **(f)** CD45+ cells from solid tumor region of 3 patients  
1237 (patient of origin is annotated on the left). Values are normalized by SNR and by z-score for each  
1238 patient.

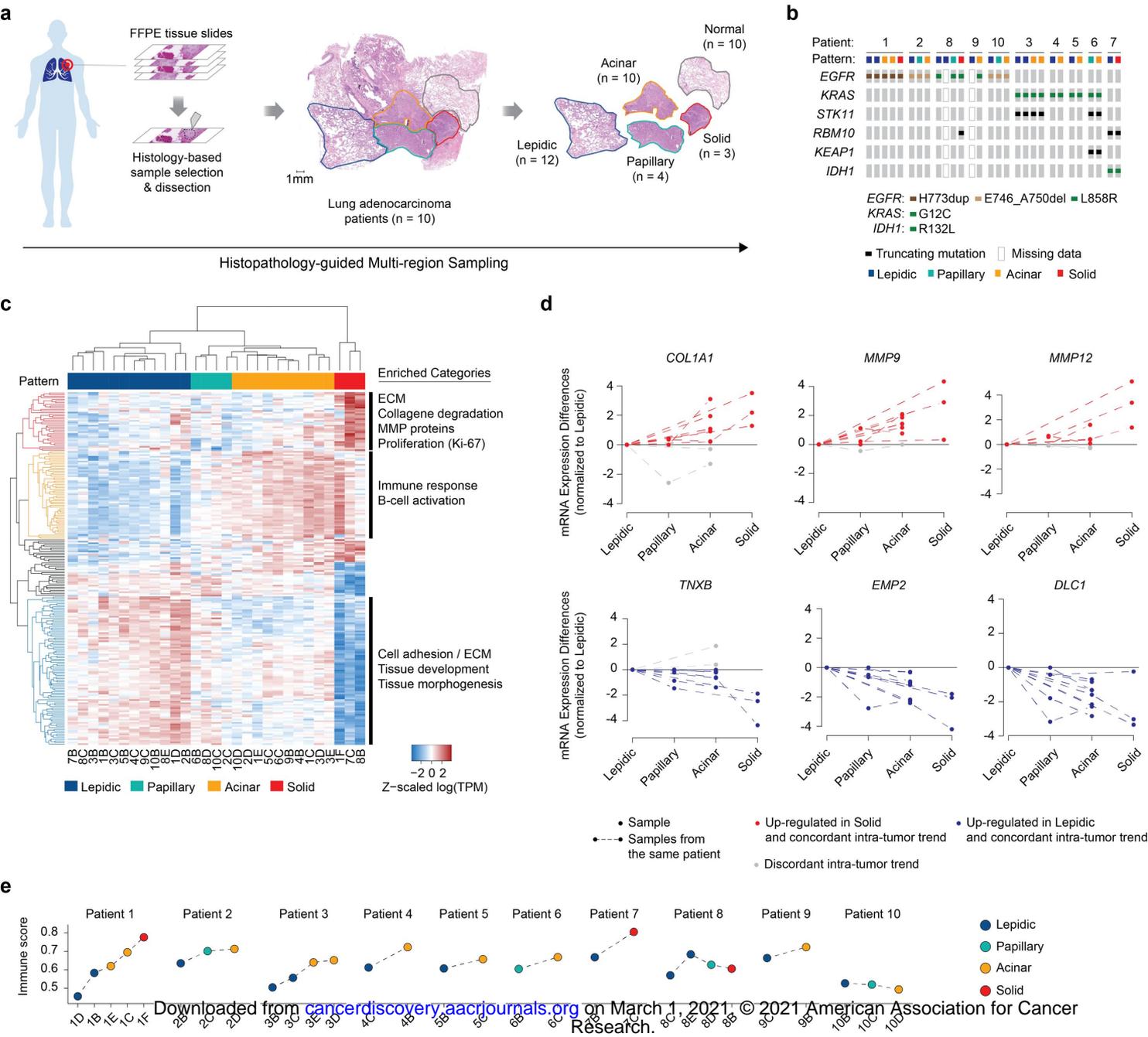
1239

1240 **Figure 6**

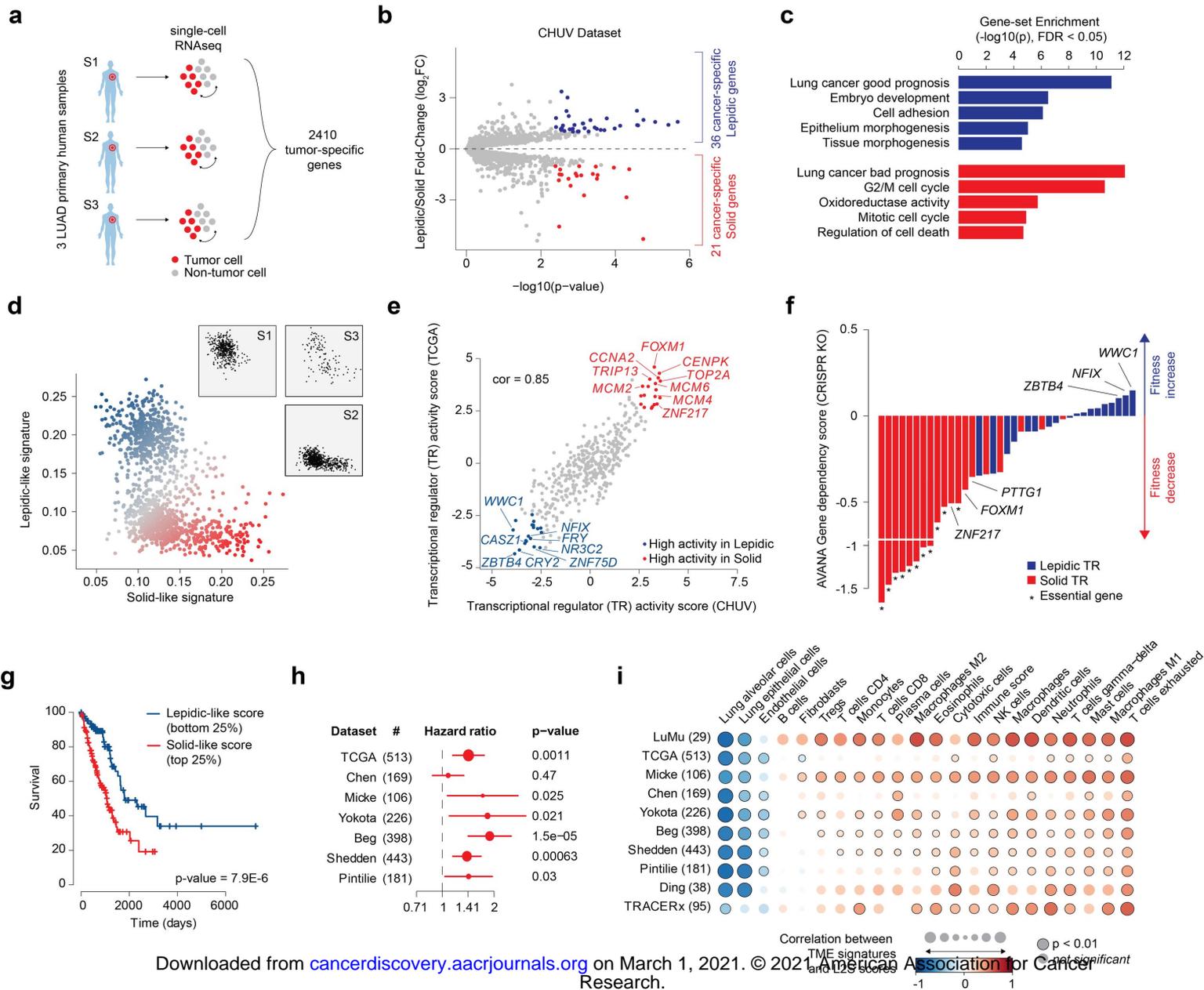
1241 Schematic representation of cancer cell (top) and microenvironment (bottom) evolution in the  
1242 progression from lepidic to papillary, acinar, and at last solid patterns.

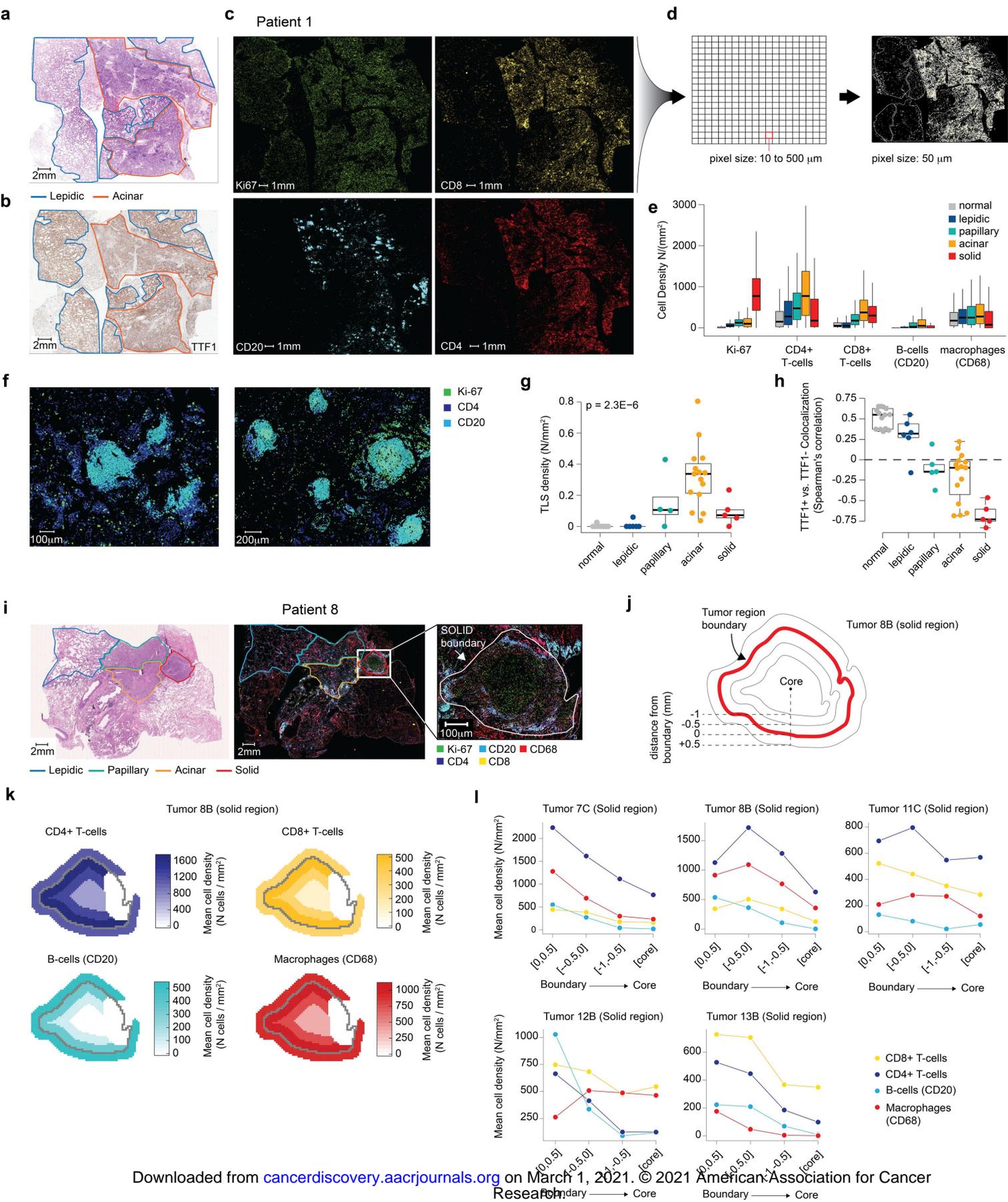
**Figure 1**

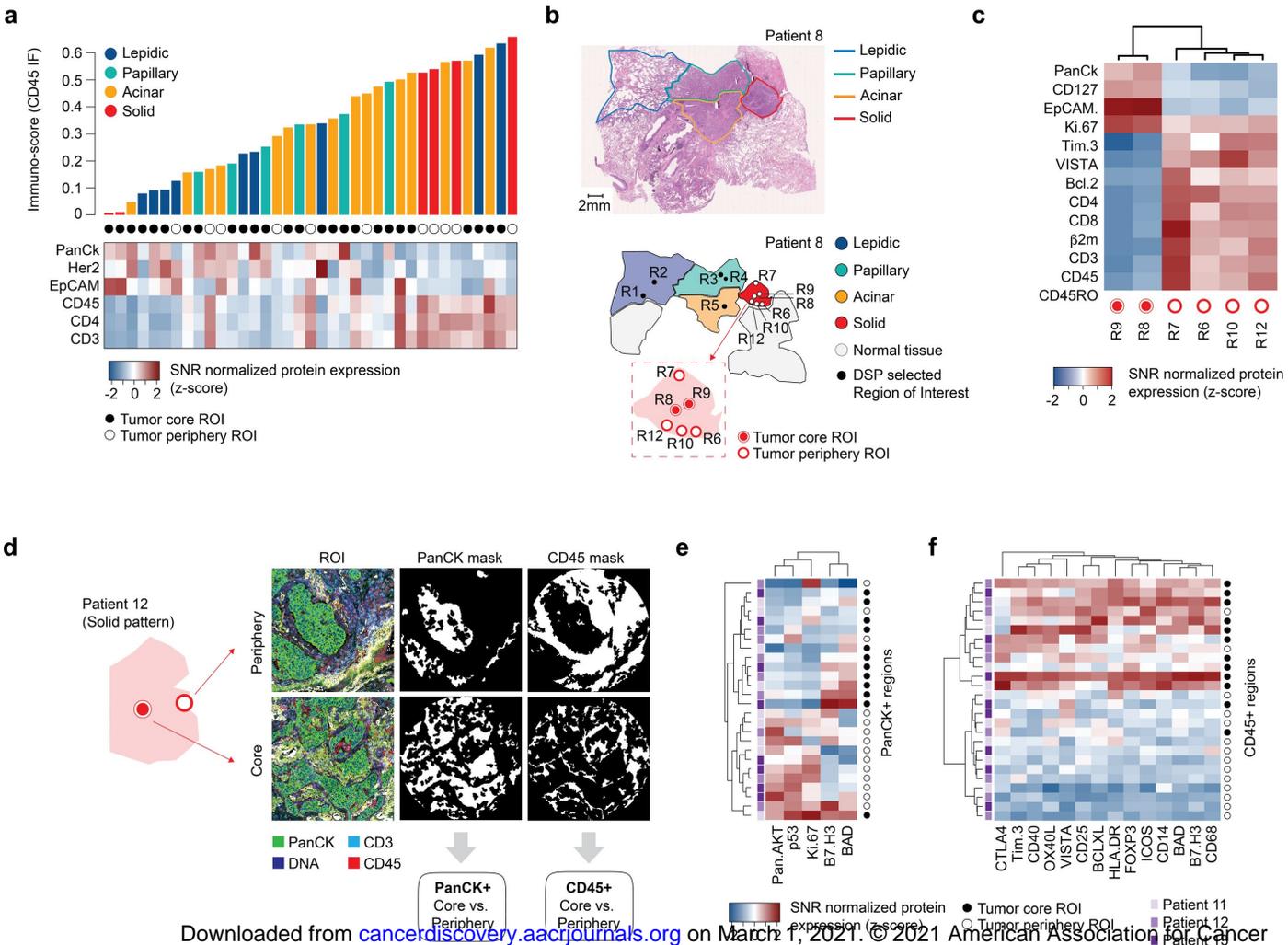


**Figure 2**

**Figure 3**





**Figure 5**

# Figure 6

Cell Differentiation

*ZBTB4*  
*WWC1*  
Alveolar/Epithelial  
markers

Migration/Proliferation

*FOXM1*  
*MMP*  
ECM degradation  
Poorly differentiated

Lepidic Papillary Acinar Solid

COLD

HOT

EXCLUDED

Tertiary  
Lymphoid  
Structures

Immune  
infiltration  
Immune  
evasion  
Effector T-regs

Lepidic Papillary Acinar Solid

Downloaded from [cancerdiscovery.aacrjco.org](http://cancerdiscovery.aacrjco.org)

# CANCER DISCOVERY

## Non-genetic evolution drives lung adenocarcinoma spatial heterogeneity and progression

Daniele Tavernari, Elena Battistello, Elie Dheilily, et al.

*Cancer Discov* Published OnlineFirst February 9, 2021.

<b>Updated version</b>	Access the most recent version of this article at: doi: <a href="https://doi.org/10.1158/2159-8290.CD-20-1274">10.1158/2159-8290.CD-20-1274</a>
<b>Supplementary Material</b>	Access the most recent supplemental material at: <a href="http://cancerdiscovery.aacrjournals.org/content/suppl/2021/01/27/2159-8290.CD-20-1274.DC1">http://cancerdiscovery.aacrjournals.org/content/suppl/2021/01/27/2159-8290.CD-20-1274.DC1</a>
<b>Author Manuscript</b>	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

<b>E-mail alerts</b>	<a href="#">Sign up to receive free email-alerts</a> related to this article or journal.
<b>Reprints and Subscriptions</b>	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a> .
<b>Permissions</b>	To request permission to re-use all or part of this article, use this link <a href="http://cancerdiscovery.aacrjournals.org/content/early/2021/02/01/2159-8290.CD-20-1274">http://cancerdiscovery.aacrjournals.org/content/early/2021/02/01/2159-8290.CD-20-1274</a> . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.