1 **Single-cell RNA sequencing reveals stromal evolution into LRRC15[+]**
2 **myofibroblasts as a determinant of patient response to cancer immunotherapy**
3
4
5 Claudia X. Dominguez[1*], Sören Müller[2*], Shilpa Keerthivasan[1], Hartmut Koeppen[3],
6 Jeffrey Hung[3], Sarah Gierke[4], Beatrice Breart[1], Oded Foreman[3], Travis W. Bainbridge[5],
7 Alessandra Castiglioni[1,] Yasin Senbabaoglu[2], Zora Modrusan[6], Yuxin Liang[6], Melissa R.
8 Junttila[7], Christiaan Klijn[2], Richard Bourgon[2] & Shannon J. Turley[1,#]
9
10
11 Departments of [1]Cancer Immunology, [2]Bioinformatics and Computational Biology,
12 [3]Pathology, [4]Center for Advanced Light Microscopy, [5]Protein Chemistry,
13 [6]Microchemistry, Proteomics & Lipidomics, and [7]Translational Oncology, Genentech,
14 South San Francisco, California

15

16 [*]Equal contribution [#]Corresponding author

17

18 Corresponding Author: Shannon J. Turley, Genentech, 1 DNA Way, South San
19 Francisco, CA 94080, USA. Phone: (650)-225-2790 E-mail: turley.shannon@gene.com

20

21 **Disclosure of Potential Conflicts of interest**

22 All authors are employees and all but C.X.D. are stockholders of Genentech/Roche.
23 Otherwise authors declare no competing interests.

24

25 **RUNNING TITLE**

26 Characterization of TGFB-activated LRRC15+ CAFs in PDAC

27

ABSTRACT

With only a fraction of patients responding to cancer immunotherapy, a better understanding of the entire tumor microenvironment is needed. Using single-cell transcriptomics we chart the fibroblastic landscape during pancreatic ductal adenocarcinoma (PDAC) progression in animal models. We identify a population of carcinoma-associated fibroblasts (CAFs) programmed by transforming growth factor beta and expressing the leucine-rich repeat containing 15 (LRRC15) protein. These LRRC15$^+$ CAFs surround tumor islets and are absent from normal pancreatic tissue. The presence of LRRC15$^+$ CAFs in human patients was confirmed in >80,000 single-cells from 22 PDAC patients as well as immunohistochemistry on samples from 70 patients. Furthermore, immunotherapy clinical trials comprising over 600 patients across 6 cancer types revealed elevated levels of the LRRC15$^+$ CAF signature correlated with poor response to anti-PD-L1 therapy. This work has important implications for targeting non-immune elements of the tumor microenvironment to boost responses of cancer patients to immune checkpoint blockade therapy.

SIGNIFICANCE

This study describes the single-cell landscape of cancer-associated fibroblasts in pancreatic cancer during in vivo tumor evolution. A TGFB-driven, LRRC15$^+$ CAF lineage is associated with poor outcome in immunotherapy trial data comprising multiple solid tumor entities and represents a target for combinatorial therapy.

2

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) remains a devastating disease, with a 5-year survival rate of 7%(1). One of the hallmarks of this aggressive cancer is a dramatic desmoplasia driven by carcinoma-associated fibroblasts (CAFs). CAFs not only deposit the extracellular matrix (ECM) that characterizes desmoplasia, but also produce factors that promote tumor growth. Subsequently, CAFs have been targeted in efforts to improve PDAC outcomes, with conflicting results(2–6). The discrepancy in outcomes might be explained by CAF heterogeneity, with different fibroblast populations having separate, perhaps even opposing functions(7,8). Smooth muscle actin (SMA) and fibroblast activating protein (FAP) have been described as showing heterogenous expression on CAF populations(8,9) and SMA high CAFs have further been identified as a tumor adjacent transforming growth factor beta (TGFB) driven population with different inflammatory properties from SMA low CAFs.

Intriguingly, despite the conflicting results of targeting CAFs as a single therapy, modulating CAFs in combination with immunotherapies improved outcomes in several preclinical models(2,4,10). As these studies model cancers that show resistance to immunotherapies alone, they suggest that elucidating CAF functions may provide the understanding needed to design more efficacious immunotherapeutic approaches and address the unmet clinical need in devastating cancers like PDAC. The full scope of CAF functions in the context of cancer immunotherapy remain to be determined, but will necessarily be influenced by the fibroblast state at the tissue-tumor interface.

We sought to provide an unbiased assessment of fibroblast heterogeneity in normal as well as PDAC tissues by using a combination of bulk and single-cell RNA-seq of stromal cells. Normal tissues, non-malignant adjacent, early and advanced tumors from genetically engineered mouse models (GEMMs) were utilized in this study. We hypothesized that the changing microenvironment during tumor progression impacts the phenotype of resident tissue fibroblasts resulting in their development into multiple CAF subsets. Our analyses revealed that pre-existing fibroblast heterogeneity in normal tissue dictated the developmental trajectories of murine CAFs. These data enabled identification of the transcriptional profiles of individual CAF populations, and revealed a TGFB programmed CAF, identifiable by expression of leucine rich repeat containing 15

3

83   (LRRC15), that became the dominant fibroblast in advanced tumors. Combining publicly

84   available human sequencing data with newly acquired immunohistochemistry of 70

85   PDAC patients, we confirmed the identification of these LRRC15[+] CAFs in human

86   patients. The LRRC15[+] CAF signature was used to evaluate their impact on anti-PDL1

87   immunotherapy response in large patient cohorts and revealed that high expression of

88   the LRRC15[+] CAF signature was associated with poor response to anti-PD-L1 therapy

89   in immune excluded tumors.

90

91   **RESULTS**

92   ***PDPN[+] cells are the dominant fibroblast population in normal and PDAC murine***

93   ***pancreas***

94   To characterize the stromal compartment in PDAC we began by optimizing

95   digestion conditions for stromal cell phenotyping from murine pancreas, starting with

96   protocols to isolate the dominant known stromal cell in the pancreas, the stellate cell.

97   Standard stellate cell pronase-based digestion(11) was observed to cleave many

98   surface markers whereas our novel digestion method preserved podoplanin (PDPN)

99   and platelet endothelial cell adhesion molecule-1(PECAM-1/CD31) expression (Fig. 1a).

100 To model PDAC we used the *Pdx1[cre/+];LSL-Kras[G12D/+];p16/p19[flox/flox]* (KPP) mice which

101 form aggressive tumors within 12 weeks(12). While tumors from these mice often show

102 several different carcinoma types including sarcomatoid, acinar, and mucinous

103 subtypes, we observe up to 88% of a given cohort develop substantial regions of PDAC

104 as has been previously reported(13,14) (Suppl. Fig. 1a). Flow cytometry of dissociated

105 pancreases from the KPP mice and normal mice from the same albino B6

106 background(B6(Cg)-Tyrc-2J/J) revealed 3 major populations of stromal cells with similar

107 composition between the two states (Fig.1b and Suppl. Fig. 1b). CD31[+] stromal cells

108 were predominantly PDPN- blood endothelial cells with very few lymphatic endothelial

109 cells(15). The remaining CD31[-] cells were largely PDPN[+] with fibroblast and stellate cell

110 characteristics (Fig. 1b, Suppl. Fig. 1c-f). Immunofluorescence microscopy confirmed

111 the presence of PDPN[+] cells around structures in the normal pancreas including acinar

112 clusters, ducts, and islets as well as a single cell layer of mesothelial cells encapsulating

113 the pancreas (Fig. 1c [left] and Suppl. Fig. 1g). The KPP mice exhibited increased

4

114   PDPN expression, most dramatically bordering tumor islets, but also in some areas

115   distal to tumors (Fig. 1c [middle and right] and Suppl. Fig. 1h). To better characterize

116   changes in the non-endothelial stroma with tumor progression, we harvested tissues

117   from normal mice, mice with early tumors (<5mm), and mice with advanced disease

118   (tumors>5mm). Pancreatic stromal cells were sorted on CD31⁻PDPN⁺PDGFRa⁺ and

119   CD31⁻PDPN⁻ (DN) cells for RNA sequencing of these two populations. This strategy

120   was used to exclude mesothelial cells, which are also PDPN⁺ but negative for PDGFRA

121   (16) (Suppl. Fig. 1i). The transcriptional profiles confirmed that PDPN⁺ stromal cells are

122   enriched for fibroblast signature genes and the DN population for pericyte signature

123   genes(17) (Fig. 1d). Several CAF associated genes were enriched in the PDPN⁺ PDAC

124   population, although in mice *Fap*, *Sma (Acta2)*, *Fsp1*, and *Pdgfrb*, often described as

125   CAF marker genes, were detected to some degree in both stromal populations in

126   normal and PDAC pancreas. Particularly, we find that *Acta2* is highest in normal

127   pericytes and *Fap* is equally high in normal fibrobroblasts (Fig. 1e). While the pericyte

128   enriched population also showed changes between normal and tumor tissues, we

129   focused on the PDPN⁺ populations as they represent the major CAF constituent.

130

131   ***Single-cell RNA-seq identifies several populations of PDPN⁺ cells***

132   While PDPN⁺ stromal cells constituted the majority of CAFs, they expressed

133   individual CAF markers at variable levels between replicates (e.g. *Il6* levels ranged from

134   60 to 240 Reads Per Kilobase of Transcript per Million mapped reads [rPKM]), and

135   furthermore they appeared to simultaneously express markers reported to separate

136   CAF subsets *i.e. Acta2* and *Il6* (Fig. 1e). This implied a significant heterogeneity within

137   the PDPN⁺ stroma. To resolve this heterogeneity, we performed single-cell RNA-

138   sequencing (scRNA-seq) of viable PDPN⁺ stromal cells from the pancreas of KPP and

139   normal mice. To better capture changes that occur with tumor progression we divided

140   the KPP samples into tumor-adjacent tissue, as well as small (1-4mm), and large (5-10

141   mm) tumor samples for scRNA-seq (Fig. 2a). Five animals were pooled per condition in

142   each of two biological replicates and scRNA-seq was performed (Fig. 2a). After quality

143   control and batch correction (Suppl. Fig. 2a, described in Methods) we obtained 13,454

144   high-quality cells for downstream analysis (replicate 1: n=3,315; replicate 2: n=10,139).

145 Graph-based clustering of cells after dimensionality reduction with t-Distributed
146 Stochastic Neighbor Embedding (t-SNE) (Fig. 2b) or Uniform Manifold Approximation
147 and Projection (UMAP, Suppl. Fig. 2b), identified 12 robust groups of cells (Suppl. Table
148 1).

149 Endothelial, myeloid, and acinar cell clusters represent contaminating cells as
150 anti-CD31, -CD45, and -EPCAM, respectively, were used to gate out those populations
151 in our flow protocol prior to sequencing (3% of all cells, Fig. 2a-c and Suppl. Fig. 2c).
152 Within the remaining 97% of cells, 83.5% of cells were identified as fibroblasts, 11.5%
153 were classified as tumor cells undergoing EMT, and 5.1% were mesothelial cells. All
154 clusters were represented in both replicates (Fig. 2b-d). Clusters 5 and 7 had lost
155 *Epcam* expression but retained higher expression of several keratins and genes
156 associated with an epithelial origin. This suggested they might be EMT tumor
157 populations (Suppl. Fig. 2d). To confirm this assignment, we identified *Alcam* as a gene
158 uniquely expressed by these clusters (Fig. 2c). Flow cytometry confirmed ALCAM
159 protein was expressed by a subset of cells found only in some large tumors. Isolation
160 and sequencing of ALCAM$^+$ cells revealed expression of the KRASG12 allele with 98%
161 variant allele frequency (Suppl. Fig. 2e), confirming their identity as tumor cells. Cluster
162 6 was identified as mesothelial cells based on previous work from Buechler, et al. who
163 transcriptionally profiled these cells and their transcriptional differences to fibroblasts
164 using bulk RNAseq, as well as Xie et al. who identified their signature genes with
165 scRNA-seq (16,18). Genes identified by both studies as mesothelial cell markers were
166 strongly enriched in cluster 6, conversely 18 of the 20 most enriched genes in cluster 6
167 were also upregulated in mesothelial cells compared to fibroblasts in the Buechler et al.
168 dataset (Fig. 2c and 2e, Suppl. Table 2). We primarily observed mesothelial cells in
169 normal, and normal adjacent tissues (Fig. 2d).

170 Clusters 0-4, 8 and 9 were identified as fibroblasts by their expression of
171 signature fibroblast genes (Suppl. Fig. 2d). Two clusters of normal tissue fibroblasts
172 (ntFib) derived from normal mice (c3 and c4), as well as five clusters of CAFs were
173 identified (Fig. 2b and Suppl. Fig. 2b; c0, c1, c2, c8, and c9). C0 and c1 were most
174 abundant in tissue adjacent to tumors (~88% of CAFs; Fig. 2f). Meanwhile, the
175 frequency of cells from c8 and especially c2 increased with tumor progression and

6

176   dominated in late stage tumors (>70% of all CAFs) (Fig. 2f). Given the disappearance of

177   ntFib with tumor progression but proximity of normal fibroblast and CAF clusters in tSNE

178   and UMAP space, we hypothesized that heterogeneity at baseline might play a role in

179   subsequent CAF development.

180

181   ***In mouse two separate fibroblast lineages co-evolve during tumor progression***

182   ***driven by TGFB and IL1***

183   UMAP dimensionality reduction of ntFib alone confirmed two major $Pdpn^+Pdgfra^+$

184   cell populations (Fig. 3a), and we identified their transcriptional profiles (Suppl. Fig. 3a

185   and Suppl. Table 3). The c3 population expressed ECM genes associated with elastin

186   fibrils and ECM attachment (i.e. *Emilin2, Mfap5, Fbn1*) while the c4 population was

187   characterized by high expression of ECM proteins that suggested a predominant role in

188   structural support through the production and maintenance of collagen networks and

189   basement membranes (*i.e. Col4a1, Col6a6, Plc*). Consequently, c4 exhibited a

190   significantly higher overall expression of collagens compared to c3 (Wilcoxon's rank-

191   sum test <0.001, Suppl. Fig. 3b). A similar trend was observed with respect to immune

192   regulation, with c4 showing enrichment of several immune chemo-attractants (including

193   *Ccl11, Cxcl14 and Cxcl16*), while c3 showed enrichment of different immunoregulatory

194   genes (*Thbd, CD55, Il33, Dpp4*, and *Ackr3* [*Cxcr7*]). Their expression of non-

195   overlapping genes indicates complementary activities of c3 and c4. Flow cytometry for

196   dipeptidyl peptidase 4 (DPP4) and LY6C, markers for c3, and endoglin (ENG), a c4

197   marker (Suppl. Table 3), confirmed that both these populations can be phenotypically

198   identified (Fig. 3b).

199   To assess the transcriptional changes during tumor progression, we first

200   performed an unbiased principal component analysis (PCA) comprising all CAF and

201   ntFib cells. PC1, the component explaining the strongest variance in the dataset, clearly

202   separated c4 ntFib and c1 and c2 CAFs from c3 ntFib and c0 and c8 CAFs (Fig. 3c,

203   top). Genes driving this unbiased separation were the same found in the supervised

204   differential expression test between the two ntFibs c3 and c4 (Fig. 3c, bottom). This

205   analysis strongly suggests a lineage relationship between CAFs and pre-existing

206   fibroblasts in the tissue. To investigate this further, we calculated a score for each CAF

7

207 cell based on the normal fibroblast ontogeny signature genes (Suppl. Fig. 3c) which

208 enabled tracing of the CAF populations back to their non-malignant ancestor (Fig. 3d,

209 Suppl. Fig. 3c). C3 and c4 ntFibs have separate differentiation trajectories during tumor

210 progression with c4 giving rise to c1 CAFs, which predominantly gives rise to c2 CAFs,

211 meanwhile, c3 ntFibs give rise to c0 CAFs, which then predominantly progress into c8

212 CAFs (Fig. 3d, right, Fig. 3e). We find c9, strongly characterized by high expression of

213 proliferation markers (*Mki67*, *Top2a*), splits into two clusters in UMAP space (Suppl. Fig.

214 2b), one aligning with EMT tumor cells, the other one aligned with c2 CAFs. The

215 proliferating, CAF-proximal cells also exhibit a higher c4 ntFIB score, explaining the

216 observed expansion of descendants of this lineage with tumor progression (Suppl. Fig.

217 3d). We thus conclude that non-tumor cells from c9 are mostly a proliferating subset of

218 c2 CAF.

219   The trajectories for the two separate fibroblast populations were confirmed with

220 pseudo time analysis for each of the lineages(19) (Fig. 3f). Comparing the expression of

221 ECM genes and selected immune regulatory genes across all of the CAF clusters

222 revealed a sharp transcriptional shift in the programming of c2 and c8 (Fig. 3g). In the

223 transition from c4 ntFib there is a significant loss of basement membrane components

224 (i.e. Type IV and Type VI collagens) with a drastic increase in levels of several fibrillar

225 collagens in c2 (Fig. 3g), indicating an increase and reorganization of fibrillar collagen

226 deposition(20,21). CAFs originating from c3 also increase expression of ECM genes,

227 particularly fibrillar collagens (Fig. 3g, top), but the most dramatic changes observed

228 with tumor progression are in chemokine and cytokines expression (Fig. 3g, bottom)

229 such as the upregulation of *Cxcl9/10*, *Cxcl1, and Ccl2* which likely recruit myeloid

230 populations through Cxcr2 and Ccr2 as well as pleiotropic cytokines such as *Il6*(22–24),

231 particularly in late stage fibroblasts (Fig. 3g). Interestingly, while there are differences in

232 expression levels between c2 and c8 CAFs we also see increased expression of genes

233 encoding factors that are known to support tumor cell survival and metastasis, such as

234 *Timp1, Vegf, Il11, Lif, and Pdgf*(25–29) that are increased in both lineages (Fig. 3g).

235 Moreover, with tumor progression both lineages acquire potential immune regulatory

236 gene expression such as increased *Mif* and *Timp1* which can drive infiltration of

237 myeloid-derived suppressor cells and regulatory T cells (Tregs) in PDAC(30,31). Thus,

8

238   while each lineage does appear to have specialized functions reflected by differences in

239   their gene signatures they also share some common programming.

240   To further investigate the signals driving the differences between the two

241   lineages, we queried the promoters of their signature genes (Suppl. Table 1) for

242   transcription factor binding sites. This analysis revealed a strong enrichment of NFKB

243   binding sites in the promoters of c8-specific genes while c2-specific genes showed

244   Smad3 binding site enrichment (Fig. 3h). Pathway enrichment analysis supported these

245   predictions, suggesting signaling through IL1 and TNFa as a driver of the c8

246   transcriptional signature and TGFB-driven activation of c2 (Fig. 3h). Furthermore, we

247   observed a strong enrichment of a TGFB fibroblast gene signature(32) in c2 cells further

248   validating TGFB as a key driver of the c2 phenotype (Fig. 3i, left). Interestingly, their

249   transcriptional signatures suggest these populations may promote their own

250   programming; while c8 cells express Il1a and their chemotactic profile suggests

251   paracrine interactions with myeloid cells, that can also be a primary source of Il1 and

252   TNF (Fig. 3i, right)(33–35), c2 shows expression of *TGFB1* and *TGFB3* (Fig. 3i).

253   To confirm the validity of our fibroblast evolution model, we compared our

254   expression signatures to the previously published fibroblast-enriched data from KPC

255   mice(36). UMAP clustering identified a group of *Pdpn*$^+$ *Pdgfra*$^+$ cells, that could be

256   dissected into three different sub-clusters (Suppl. Fig. 3e): one cluster of *Ly6a/c1+* cells,

257   previously described as "iCAFs", one cluster of *Col15a1+* cells previously described as

258   "myCAFs", and one cluster with high levels of *Cd74*, *H2-Ab1* and *Saa3,* previously

259   described as "apCAFs". When we compared the average expression levels of our two

260   normal fibroblast lineage programs to these clusters, we found that myCAFs clearly

261   clustered with our c4 fibroblasts and iCAFs with c3 fibroblasts, confirming that the

262   lineage hierarchy is similarly present in the KPC model (Fig. 3j). Notably, apCAFs

263   clustered with c6 mesothelial cells from normal pancreas (Markers: Suppl. Table 3).

264   Accordingly, the Il1 c8 CAFs exhibited most similar expression profiles to iCAFs, the

265   TGFB c2 CAFs clustered with myCAFs, and the c6 mesothelial cells with apCAFs

266   (Suppl. Fig. 3f).

267

268   ***Mouse models of PDAC identify LRRC15 as a marker of TGFB driven c2 CAFs***

9

269   We were particularly interested in further characterizing c2 as it increased with

270   tumor progression, dominating the CAF compartment in late stage tumors (Fig. 2f), and

271   because TGFB-associated stroma is correlated with poor prognosis(32,37). Therefore,

272   we sought to identify markers that distinguish TGFB driven c2 CAFs from the other

273   fibroblast stromal subsets in PDAC. Bulk RNA-seq data from early and late stage tumor

274   PDPN[+] CAFs identified leucine rich repeat containing 15 (*Lrrc15)* to be one of the most

275   differentially expressed genes between CAFs and ntFIBs (Fig. 4a). *Lrrc15* encodes a

276   transmembrane domain containing molecule expressed in the stroma of several human

277   tumors and upregulated by TGFB(38). Cross-referencing genes enriched in the TGFB

278   driven c2 to an atlas of proteins experimentally identified to be on the cell surface(39)

279   further validated *Lrrc15* to be a strongly enriched c2 gene encoding a surface protein

280   (Suppl. Fig. 4a).

281   The presence of LRRC15[+] PDPN[+] cells with fibroblast morphology in tumor

282   bearing pancreases was confirmed by immunofluorescence microscopy; LRRC15[+] cells

283   were usually found in nests throughout the tumor-bearing pancreas surrounding tumor

284   cells in KPP GEMMs (Fig. 4b). We further employed subcutaneous models of PDAC,

285   using a cell line derived from KPP GEMMs (KPP14388). Characterization of flank

286   injections of 100k tumor cells showed tumors with similar stromal composition to the

287   KPP mice (Suppl. Fig. 4b). Immunofluorescence microscopy revealed abundant

288   LRRC15[+] PDPN[+] cells in the subcutaneous tumors derived from KPP PDAC line (Fig.

289   4c). Flow cytometry analysis showed that LRRC15 marked a significant portion of

290   PDPN[+] stromal cells, and was largely absent from other cell populations in the TME

291   (Fig. 4d). Notably, LRRC15, unlike many other CAF markers, is also absent in lymph

292   node stroma as well as normal pancreas (Suppl. Fig. 4c).

293   Due to their proximity to tumor islets we decided to assess whether LRRC15[+]

294   CAFs can directly enhance tumor growth. We generated a KPP line expressing

295   Diptheria Toxin Receptor (DTR), that allowed us to remove residual tumor cells and

296   culture isolated CAFs with the addition of DT. 2000 KPP-mApple tumor cells were

297   grown alone or in coculture with LRRC15[+] CAFs compared to LRRC15[-] LY6C[+] CAFs or

298   c3 and c4 ntFIBs and assessed for their ability to promote spheroid growth in 3D

299   culture. Tumor spheroids cultured with any fibroblast population grew larger than those

10

300   in media alone. This demonstrated that all the fibroblasts tested can directly enhance

301   tumor growth (Fig. 4e). Although we cannot rule out that the spheroids themselves

302   reprogrammed the fibroblasts as has been previously reported(40), it suggests that the

303   specific *in situ* positioning of LRRC15$^+$ CAFs next to tumor islets might be one of the

304   keys to their role in a protumor niche rather than a unique ability to promote tumor

305   growth. This experiment also represents a single functional test; given the unique

306   transcriptional differences between the fibroblasts populations we might expect

307   functional differences in other areas, such as immune regulation.

308

309   ***LRRC15$^+$ CAFs are present in human PDAC samples***

310   To translate our findings from mouse models into human cancer, we re-analyzed

311   data from a recently published study of single-cell RNA-seq of human PDAC patients by

312   *Peng, J. et al.* (41). After quality control and filtering we retained 84,276 cells from 22

313   patients for downstream analysis. Clustering in dimensionality-reduced space revealed

314   12 clusters of 11 main cell types (Fig. 5a, Suppl. Table 4). All clusters were comprised

315   of cells from more than 8 patients (Fig. 5b). To confirm our tumor-cell assignment and

316   by extension ensure our fibroblast assignment did not include tumor cells that had

317   undergone EMT, we collected three independent lines of evidence. First, we confirmed

318   only cells in cluster 0 were positive for mRNA with the KRAS G12 missense mutation

319   (Fig. 5a). Second, re-analysis of a separate publicly available microdissection study(42)

320   showed that only markers for tumor cells (cluster 0) are enriched in bulk RNA-seq of

321   microdissected tumor samples compared to samples from microdissected stroma or

322   non-malignant control pancreas (Fig. 5c). Third, we identified large-scale copy number

323   variants in cells from cluster 0 but not cells from the CAF cluster (Suppl. Fig. 5a).

324   After confirming the non-tumor origin of the population identified as CAF cells, we

325   specifically focused on this cluster. Sub-clustering of the 8,931 fibroblasts revealed 3

326   distinct subsets (Suppl. Table 5): 52% of cells expressed high levels of TGFB c2 CAF

327   markers *TAGLN* and *LRRC15*, 3% were strongly enriched in IL1 c8 CAF markers *HAS1*

328   and *CCL2* (Fig. 5d), and 44% of cells expressed high levels of *C7* and *CFD. LRRC15*

329   was also highly expressed in the bulk RNA-seq from microdissected stroma compared

330   to tumor or normal control samples, suggesting that this population is prominent in

11

331   PDAC stroma (Fig. 5e). Conversely, we found *HAS1* lowly expressed across tumor,

332   stroma, and non-malignant samples and enriched in only a small fraction of

333   microdissected stroma samples, likely representing only a minor population in PDAC

334   stroma. These trends were confirmed when the average expression of signature genes

335   for each of the human fibroblast single-cell clusters was compared within the

336   microdissected bulk-seq samples (Fig. 5f). To confirm protein expression of LRRC15 in

337   human CAFs we performed dual Immunohistochemistry on 70 PDAC patients (Suppl.

338   Table 6) for LRRC15 and CD8. We found 100% of patients showed LRRC15 staining in

339   non-normal areas of pancreas and LRRC15 appeared fibrillar and was largely excluded

340   from tumor islets. Mostly, it was found surrounding them, additionally it was frequently

341   seen in proximity to CD8 T cells in the area (Fig.5g and Suppl. Fig.5b). Flow cytometry

342   on 4 patient samples further confirmed LRRC15 was largely restricted to the EPCAM$^-$

343   CD45$^-$ stromal gate and marked the majority of CAFs (Fig. 5h). Altogether, we find that

344   LRRC15$^+$ TGFB cluster 0 (hC0) CAFs are the most prominent fibroblast population in

345   multiple human PDAC data sets confirming our findings from the mouse model.

346   While human fibroblast clusters 0 and 2 (hC0 and hC2) exhibited overlapping

347   genes of mouse TGFB c2 and IL1 c8 CAFs in a cross-species comparison, respectively

348   (Fig. 5i), cluster 1 (hC1) did not obviously match the early CAF populations observed in

349   mouse. Although HC1 was characterized by high levels of mouse c4 relative to c3

350   genes (Suppl. Fig. 5c), individual cells did not show a clear phenotype of one or the

351   other population. To test if, in contrast to mice, human pancreatic fibroblasts are a

352   homogeneous population, we performed *in silico* isolation of single fibroblast cells from

353   11 non-malignant pancreatic tissues, published as part of Peng, J. *et al.*(41).

354   Dimensionality reduction with UMAP and clustering in reduced space revealed a

355   population of 1,407 fibroblasts (*DCN*, *LUM,* Suppl. Fig. 5d, Suppl. Table 7). Sub-

356   clustering of these cells identified two clusters, of which the minor one (<5% of cells)

357   exhibited high levels of *EPCAM* and other epithelial markers and likely represents an

358   artifact of our *in silico* isolation (Suppl. Fig.5e). The other cluster was characterized by

359   strong expression of *C7* and *CFD.* The strong similarity of markers of these non-

360   malignant fibroblasts to hC1 suggested that hC1 have not undergone extensive

361   transcriptional changes relative to non-malignant fibroblasts. This was confirmed by

12

362   comparing the average expression profile of these two cell types (Pearson correlation:

363   0.97), where only a few genes changed expression (Fig. 5j). Based on this analysis we

364   were able to identify genes enriched in non-malignant fibroblasts and hC1 CAFs, but not

365   hC0 TGFB CAFs or hC2 IL1 CAFs, as well as genes that are induced in all three CAF

366   populations compared to non-malignant fibroblasts (Suppl. Fig. 5f). Together, the results

367   suggested that hC1 CAFs are early CAFs (eCAFs) and the predecessor to both hC2 IL1

368   and hC0 TGFB CAFs. To test this hypothesis, we performed principal component

369   analysis of all fibroblasts including those from non-malignant pancreas. Strikingly, PC1

370   separated cells in the order of non-malignant fibroblasts, hC1 eCAFs, hC2 Il1 CAFs and

371   hC0 TGFB CAFs (Fig. 5k). The minimum spanning tree fit to this dimensionality reduced

372   data supported that starting from non-malignant fibroblasts cells undergo a

373   transformation into C1 eCAFs, upregulating type 1 collagen, SPARC, and other

374   extracellular matrix proteins. From this intermediate state cells either become hC2 IL1

375   CAFs or hC0 TGFB CAFs. Interestingly, both the hC1 eCAFs and the hC0 TGFB CAFs

376   make up almost all fibroblasts in the single-cell dataset under investigation. We did not

377   see evidence of any cells with a mesothelial or apCAF signature in these data, however

378   we find that all human CAFs expressed *CD74* and *HLA-DRA* (Suppl. Fig. 5g), which is

379   consistent with data found in Elyada, et al(36). We have summarized our findings in

380   both human and mouse in a model (Fig. 5l).

381

382   ***A LRRC15[+] CAF signature can be found across several human cancer indications***

383   As it had been previously reported that stromal LRRC15 expression could be

384   observed in several tumor types(38) we performed a pan-cancer analysis across tumors

385   in The Cancer Genome Atlas (TCGA, n=9,736) and compared these data to matched

386   non-malignant tissues from the GTEx database (n=8,587). We found that *LRRC15*

387   expression was consistently low/absent across normal tissues, but upregulated in a

388   variety of tumors including but not limited to pancreatic, breast, and head and neck

389   cancers (Fig. 6a). To verify that the *LRRC15* signal was derived from TGFB-activated

390   CAFs in tumor types other than PDAC, we first identified a more robust expression

391   signature of TGFB CAFs from our human PDAC scRNA-seq analysis. We focused on

392   genes significantly enriched in TGFB CAFs compared to all other fibroblast populations

13

393 that showed no/low expression by any other cell type in the full dataset (Fig. 6b). The

394 gene set was strongly enriched in microdissected PDAC bulk stroma vs tumor samples

395 (Fig. 6c), suggesting that their combined signal allows conclusions about the

396 presence/absence of TGFB fibroblasts in bulk RNA-seq data.

397   To next confirm the presence of this population in other tumor types, we

398 exemplarily re-analyzed single-cell RNA-seq data from 18 head and neck squamous

399 cell carcinoma (HNSC) patients(43). Dimensionality reduction with UMAP of 3,363 cells

400 from the TME free of somatic mutations confirmed the cell-type annotations provided by

401 the authors (Fig. 6d, top). Clustering of the mesenchymal cells revealed 5 sub-clusters,

402 of which two were pericytes (25% of all cells, *PDGFRB*, *MCAM*, *RGS5, ACTA2*), two

403 were fibroblasts (15% of all cells, *LUM*, *DCN*) and one was myoblastic-like (2% of all

404 cells). Confirming the results from the pancreatic cancer single-cell data, expression of

405 our TGFB CAF marker gene set was almost entirely restricted to one of the two

406 fibroblast clusters (Fig. 6d, bottom right; cluster 2, 60% of all fibroblasts). This

407 underscores the presence of LRRC15[+] TGFB CAFs also in head and neck cancer.

408 Furthermore, the majority of genes showed low/no expression by tumor cells (Suppl.

409 Fig. 6a), indicating that the main signal of these genes in bulk RNA-seq data is, as we

410 have shown in pancreatic cancer, likely primarily derived from TGFB CAFs and less so

411 from EMT tumor cells.

412   Based on these results we used an 11 gene signature (*MMP11, COL11A1,*

413 *C1QTNF3, CTHRC1, COL12A1, COL10A1, COL5A2, THBS2, AEBP1, LRRC15,*

414 *ITGA11*) to infer the presence of TGFB CAFs across different cancer types from bulk

415 RNA-seq TCGA data. In this pan-cancer analysis comprising 31 different cancer types

416 from TCGA we found a positive correlation between the average expression and the

417 average gene-wise correlation of our core signature across cancer types (Fig. 6e). The

418 positive correlation indicates that in cancer types where the TGFB CAF is present (high

419 average expression), there is a true signal in the bulk coming from this population that

420 leads to a high gene-wise correlation. In cancer types lacking TGFB CAFs (low average

421 expression) there is just a "noise" signal and the genes are uncorrelated. The analysis

422 points, besides PAAD, to TGFB CAFs playing a strong role also in breast cancer

423 (BRCA), lung cancer (LUSC, LUAD), ovarian cancer (OV), colon cancer (COAD), renal

14

424   cancer (READ), esophageal cancer (ESCA), Stomach Adenocarcinoma (STAD),

425   bladder cancer (BLCA, [Suppl. Fig. 6b]) as well as head and neck cancer (HNSC). In

426   summary these data suggest that LRRC15$^+$ CAFs are a prominent population across

427   multiple human cancer types that emerges from a LRRC15$^-$ fibroblast population.

428

429   **A LRRC15$^+$ CAF signature predicts poor clinical response to checkpoint blockade**

430   Having shown that *Lrrc15*$^+$ CAFs are present in human cancers, we next sought

431   to test the clinical impact of this population. Due to the known roles of TGFB in

432   modulating immunotherapy(32,44) we evaluated the clinical significance of the newly

433   identified LRRC15$^+$ CAF in response to cancer immunotherapy. Multiple reports have

434   shown that the molecular makeup of bladder cancer is similar to pancreatic cancer with

435   shared subtypes(45,46). Further, we have shown in our previous analysis that LRRC15$^+$

436   CAFs are a frequent population also in bladder cancer (Fig. 6e). We found that the

437   markers for LRRC15$^+$ CAFs identified in PDAC were also significantly co-expressed in

438   RNA-seq data from our recent bladder cancer immunotherapy trial (32) (Suppl. Fig. 7a).

439   Moreover, the signature was associated with worse outcome for patients receiving anti-

440   PD-L1 (Atezolizumab) therapy (p=0.03, HR= 1.4, Suppl. Fig. 7b). This effect is

441   explained by the increased expression of the signature in patients that fail to respond to

442   anti-PD-L1 therapy exclusively in immune excluded tumors, but not in tumors with

443   inflamed or desert immune phenotype (Fig. 7a). Consequently, we observe a significant

444   association of the LRRC15$^+$ CAF signature with worse outcome specifically in patients

445   with immune excluded tumors (p<0.001, HR=2.3, Fig. 7b). Our result represents an

446   improvement over the fibroblast TGFB response signature obtained through *in vitro*

447   activation of fibroblasts with TGFB in regards to their cell-type specificity (Fig. 7c) and

448   predictive power (Fig. 7d). Immunoscoring of PDAC patients revealed the majority

449   showed an immune excluded phenotype, suggesting these findings might also apply to

450   PDAC patients (Suppl. Fig 7c). Importantly, the LRRC15$^+$ CAF signature was also

451   predictive of response to Atezolizumab in a second trial comprising multiple other

452   cancer types, such as renal cell carcinoma, head and neck cancer, and non-small cell

453   lung cancer (Fig. 7e, Suppl. Fig. 7d, p=0.01, HR=2.01). This effect (HR>1.5) was

454   apparent across several individual cancer indications, despite their small individual

455   sample sizes (Suppl. Fig 7e). It remains unclear what the nature of LRRC15[+] CAF

456   immunosuppression might be, but these data provide a strong basis to further elucidate

457   the functions of these cells. The correlation between LRRC15[+] CAFs and poor outcome

458   in immunotherapy treatment suggests that multiple tumor indications may benefit from

459   LRRC15 CAF reprogramming combined with immunotherapy.

460

461   **DISCUSSION**

462   Here, we took advantage of our novel digestion method to profile PDPN[+] CAFs

463   *ex vivo* using scRNA-seq. Our approach identified two separate populations of normal

464   tissue fibroblasts(ntFib) in mouse pancreas. Their expression signatures suggest

465   disparate functions, with one more primed to provide structural support and another

466   appearing more immunoregulatory. These two separate lineages evolve separately into

467   IL1- and a TGFB- driven CAFs in the context of PDAC. Further work on localization and

468   lineage tracking will allow us to distinguish whether there is a physical

469   compartmentalization or niche that results in differential exposure to IL1 and TGFB or

470   whether the ntFibs fundamentally have differing potential to respond to IL1 and TGFB

471   that results in the two separate CAF trajectories we observe.

472   Our findings from murine late-stage tumors support previous observations

473   identifying IL1 driven "iCAF" and TGFB driven "myCAFs"(40,47). Furthermore, we

474   provide new insights into how resting fibroblast heterogeneity pre-determines the fate of

475   stromal cells in the TME. It also seems from our cross comparison that "iCAFs" include

476   both our early CAF1 and IL1 CAF while "myCAFs" include both our early CAF2 and

477   TGFB CAF. We do observe a difference with our identification of a population previously

478   designated as an "apCAF"(36), that we identify a mesothelial cell population. This

479   discrepancy might be due to mesothelial cells acquiring some fibroblasts genes in the

480   KPC system as mesothelial to mesenchymal transition has been described in some

481   tissues(48). In fact, in the KPC system a relationship between the "apCAF" and

482   "myCAF" is apparent in UMAP space and the "apCAFs" express *Pdgfra* suggesting

483   changes from the normal tissue state. However, the dominant mesothelial genes driving

484   the clustering of that population, including the antigen presenting genes *CD74* and *H2-*

485   *Ab1*, were present in normal pancreas. We also did not observe these cells in the

16

486   tumor, which may be a consequence of our dissection method where the majority of the
487   mesothelium would be included in the adjacent normal tissue or it could reflect a
488   difference between the KPC and KPP models.

489   Comparison to several different human patient cohorts revealed both similarities
490   and differences with the mouse which we have modelled (Fig. 5I). While the
491   conservation of IL1 and TGFB CAFs was quite obvious, we were struck by several
492   differences between the mouse models and human patient data. First, we do not
493   observe baseline heterogeneity in the human non-malignant tissue fibroblasts. Rather
494   human fibroblasts from non-malignant tissue show a transcriptional profile that
495   combines the mouse ntFIB signatures. Subsequently, non-malignant human fibroblasts
496   transition to a single early CAF which then gives rise to either a TGFB or IL1
497   programmed CAF. However, given the human tissues we analyzed were not truly
498   normal we cannot rule out that non-malignant fibroblasts had already undergone
499   changes that masked baseline heterogeneity. We also find that *CD74* and *HLA-DRA*
500   are expressed by all human CAF populations, revealing a potentially important
501   functional difference between human and mouse CAFs. We do not observe a specific
502   population with a similar transcriptional signature to "apCAF" or mesothelial cells
503   suggesting this population is absent or very rare in the TME of PDAC.

504   We identify the TGFB-driven cell population as the most prevalent CAF in late-
505   stage tumors, and show that surface expression of LRRC15 enables experimental
506   isolation and manipulation of these CAFs both in mouse models and human patient
507   samples. Furthermore, we find that the LRRC15[+] CAF signature correlates with poor
508   response to checkpoint blockade in several different human tumors. These cells have
509   myofibroblastic properties and a dominant ECM gene signature. We find they constitute
510   the majority of CAFs in PDAC patients, which are dominantly of an immune excluded
511   phenotype. This suggests an immunoregulatory role for these cells. It will be valuable to
512   further explore whether early CAFs can be prevented from adopting the pro-tumorigenic
513   fate of the LRRC15[+] CAF or whether the LRRC15[+] CAF phenotype can be reverted to
514   improve immunotherapy efficacy.

515   While we chose to focus on the LRRC15[+] CAFs, due to their prevalence in
516   PDAC, IL1 CAFs have a transcriptional program clearly suggesting immune regulation

17

517 of the TME. Inhibition of Jak signaling in PDAC has shown both a reduction in IL1 CAFs

518 and reduced tumor burden(47), although it is difficult to distinguish direct effects of

519 these inhibitors on the tumor cells(28,49–51) from the effects of iL1 CAF loss in the

520 TME. It is also important to note that while both of these CAFs have many

521 transcriptional differences they also both express genes associated with myofibroblast

522 characteristics and both express various immune regulatory and even inflammatory

523 mediators. Thus, we describe them by their most conserved characteristics, their major

524 transcriptional programming; IL1 CAF and TGFB CAF. For the TGFB CAF we have

525 identified LRRC15 expression as a good proxy across several cancer indications.

526   We chose to focus on fibroblasts to generate a robust data set that would be a

527 good representation of the heterogeneity of a somewhat rare population. Our work also

528 identifies other populations with fibroblast properties: PDPN$^-$ CD31$^-$ cells which were

529 enriched for pericytes, but adopted expression of some CAF genes, as well as two

530 populations of PDPN$^+$ cells, which we classified as tumors cells undergoing EMT. There

531 are also various stromal and non-stromal cells that we chose to leave out of the focus

532 this study. These cells are all part of the TME and future research into their functions is

533 sure to yield a more complete understanding of their interactions and contributions to

534 tumor progression and response to therapy.

535

536

537

18

538   **METHODS**

539   *Mice*

540   WT B57BL/6 mice (colony 000664) and albino WT B6(Cg)-Tyrc-2J/J mice (colony

541   000058), mice were purchased from Jackson. We licensed KrasLSL.G12D from Tyler

542   Jacks (Massachusetts Institute of Technology, Boston), p16/p19fl/fl from Anton Berns

543   (NKI, Amsterdam), and Pdx1.Cre from Andy Lowy (University of Ohio). Pdx1cre/+;LSL-

544   KrasG12D/+;p16/p19flox/flox (KPP) mice were generated as previously described(12).

545   Age- and sex-matched mice were used for experiments. The mice were housed at

546   Genentech in standard rodent micro-isolator cages and were acclimated to study

547   conditions for at least 3 days before tumor cell implantation. Animals were 6-12 weeks

548   old.

549   KPP GEMM mice were euthanized at median ages of 9 wk. This age reflects the

550   disease state with high penetrance adenocarcinoma and at which the GEMMs with

551   moderate and large tumors are observed.

552   All animals were monitored according to the guidelines from the Institutional Animal

553   Care and Use Committee (IACUC) at Genentech, Inc.

554

555   *Cell lines*

556   The KPP14388 murine pancreatic adenocarcinoma cell line was generated by the

557   Junttila group at Genentech from the KPP GEMMs. Transgenic lines where created as

558   follows: mApple MSCV retrovirus was transfected into KPP14388 and a single clone

559   was grown out; KPP-DTR was obtained by transducing KPP-14388 with DTR-efp

560   lentivirus and single-cell clone was selected. Cancer cells were cultured in High

561   Glucose Dulbecco's Modified Eagle Medium (DMEM) medium plus 2 mM L-glutamine

562   with 10% fetal bovine serum (FBS; HyClone, Waltham, MA). All cell lines were tested

563   for mycoplasma by qPCR. For all injected tumors cells where used within the first three

564   passages.

565

566   *Tissue digestion, cell isolation and flow cytometry of murine tissues*

567   To isolate pancreases, first the omentum was removed then the pancreas collected with

568   careful exclusion of draining lymph nodes. The pancreas was then minced. For stellate

19

569   cell enrichment tissue was digested as previously described(11). Enzymatic digestion

570   was used with 0.02% Pronase (Roche cat 10165921001), 0.05% Collagenase P

571   (Roche, cat. 11249002001), and 0.1% DNAse (Roche, cat. 10104159001) in Gey's

572   balanced salt solution (Sigma cat G9779) for 50 min. Digested tissue was then filtered

573   through a 100-μm nylon mesh. Cells were centrifuged at 1300rpm for 5min, washed and

574   then resuspended in GBSS containing 0.3% bovine serum albumin (Sigma Aldrich cat

575   A2153). The cell suspension was centrifuged, decanted and resuspended into 8ml

576   28.7% (wt/vol) solution of Nycodenz (Sigma, no longer available. Also used 16.7%

577   Optiprep[Sigma-Aldrich D1556]) overlayed with 6ml .3% BSA GBSS, then centrifuged at

578   1400g with no break for 20 min. The cells of interest separated into a fuzzy band just

579   above the interface of the Nycodenz cushion and the GBSS. This band was harvested,

580   and the cells were washed and resuspended in MACs buffer.

581   For our new digest, which was modified from Fletcher, et al.(52), 20 ug/ml anti-

582   trypsin (Sigma Aldrich cat 10109886001) was used in first round of digest incubations

583   which were prepared as follows. Pancreases were enzymatically digested using

584   800ug/ml Dispase), 400ug/ml collagenase P, and 100ug/ml DNaseI at 37°C. Fractions

585   were collected into Macs buffer and digest media was refreshed 2 more times after

586   15min, 10min, and 5 min incubations. At this point either: 1) for comparison to stellate

587   cell enrichment the cell suspension was centrifuged at 1300rpm for 5min, decanted,

588   subjected to a density gradient as described above. 2) For subsequent digests without

589   gradient enrichment, samples underwent RBC lysis and were spun for 4 min at 50g to

590   pellet debris, supernatant was collected, spun down and resuspended in media and

591   cells were counted using a Vi-CELL XR (Beckman Coulter, Brea, CA).

592   Pancreatic tumors were similarly treated with the addition of 2U/ml hyaluronidase

593   (Worthington cat LS002592) and 20U/ml purified collagenase (Worthington LS005273)

594   and 100ug/ml CollagenaseP. In PDAC experiments control normal pancreases were

595   digested using same enzymatic cocktail as tumors. Tumors also often required 1-2

596   additional digest incubations to break down tissue.

597   Subcutaneous tumors were collected with care to avoid draining lymph node and

598   epidermis. Subcutaneous tumors were weighed and enzymatically digested using the

20

599 same enzymatic mix as the previously described normal pancreas samples, without the

600 trypsin inhibitor addition.

601   Cells were labeled with mAbs purchased from eBioscience, BioLegend, or BD

602 Biosciences at 1:200 for 20-30 min, unless otherwise noted. Prior to cell surface

603 staining with the following fluorescently labeled antibodies, cells were blocked with Fc

604 block (2.4G2; 1:500). Surface staining for experiments was performed using antibodies

605 described in Suppl. Table 8 for 25min at 4c, washed 2.5 times with MACs buffer, then

606 either fixed (Biolegend cat 420801) or resuspended in 7AAD (1:50; BD cat 559925) and

607 Calcein Blue(1:1000; Invitrogen, cat. C1429) for cytometry analysis. For intracellular

608 staining, cells were surface stained as above, washed, and then fixed and

609 permeabilized using the FoxP3 ICS kit (eBiosciences cat 00-5523-00), per the

610 manufacturer's directions. Cells were then incubated with antibodies described in Suppl.

611 Table 8 for 1 hour in perm buffer. Data were acquired on a Fortessa, Symphony or

612 LSRII (BD Biosciences) and analyzed using FlowJo (Tree Star).

613

614 ***Aldefluor assay***

615 Cells were isolated as above. Once a single-cell suspension was obtained, cells were

616 plated and resuspended in aldefluor assay buffer, as part of the aldefluor kit

617 (STEMCELL, cat. 01700) in FACS tubes (500uL) or 96 well U bottom plates (200uL)

618 with Fc block. Then, 10uL (FACS tubes) or 4uL (96 well plate) of assay buffer was

619 aliquoted into quench tubes. 2.5uL (FACS tubes) or 1uL (96 well plates) of DEAB buffer

620 was aliquoted into quench tubes. 5uL (FACS tubes) or 2uL (96 well plate) aldefluor

621 reagent was then added to cells as rapidly as possible. Cells were then mixed and ½ of

622 the volume was added to quench tubes with DEAB. Samples were incubated at 37C for

623 15-20 minutes, then spun down and surfaced stained as above on ice for 20-30 minutes

624 before FACS analysis.

625

626 ***Anti-LRRC15 antibodies for flow cytometry and murine imaging***

627 Gene synthesis and cloning was performed following reverse translation and codon

628 optimization for Chinese hamster ovary (CHO) cells of the amino acid sequences

629 encoding the variable heavy (VH) and variable light (VL) domains of huM25 (flow) and

21

630   huAD208.4.1 (imaging) anti-LRRC15 clones, as published via patent 10195209(53). For

631   huM25, expression constructs of human-mouse chimeric antibodies were generated by

632   subcloning the VL and VH sequences into mammalian expression vectors containing

633   mouse kappa light chain and mouse IgG2a heavy chain frameworks, respectively. For

634   huAD208.4.1, expression constructs of human antibodies were generated by subcloning

635   the VL and VH sequences into mammalian expression vectors containing human kappa

636   light chain and human IgG1 heavy chain frameworks, respectively. Both antibodies

637   were produced by transiently-transfected CHO cells and purified using standard

638   antibody purification methods and were confirmed to be human/mouse LRRC15 cross-

639   reactive by surface plasmon resonance (data not shown).

640

641   ***Immunofluorescence and image analysis of mouse tissues***

642   Mouse pancreas, PDAC tumor tissue, and subcutaneous KPP and KPR tumors where

643   fixed overnight in 1% PFA, embedded in optimal cutting temperature medium (Sakura

644   Finetek) and frozen for storage at -80C. 5-12 microns thick sections were

645   cryosectioned, immune-stained, and imaged with confocal microscope Leica TCS SP8.

646   Images were processed with Fiji software (ImageJ v2.0.0-rc-69/1.52i). For staining

647   slides were fixed for 5min in 4%PFA, blocked and permeabilized in stain buffer(2% BSA

648   5%Goat Serum in PBS) with .3%triton for 1 hour. Primary antibodies were added for 1hr

649   RT to overnight at 4C, secondaries were added for 30m-2hours at RT slides were

650   mounted in hardening media (Dako S3023). Details of the antibodies used can be found

651   in Suppl. Table 8.

652

653   ***Tumor implantation***

654   Cells between passage 1-2 were trypsinized, filtered, counted, and resuspended in 50%

655   PBS and 50% Matrigel (Corning cat 356231) at a concentration of $1x10^6$ cells/mL for

656   injection into mice. The mice were housed at Genentech in standard rodent micro-

657   isolator cages and were acclimated to study conditions for at least 3 days before tumor

658   cell implantation. Animals were 6-10 weeks old. Only animals that appeared to be

659   healthy and free of obvious abnormalities were used for studies. Mice were inoculated

660   in the right flank with $1x10^5$ cancer cells in 100 μL of PBS:Matrigel (1:1). 16-24 days

22

661   after tumor injection, mice were euthanized and tumors collected for either IF or flow
662   cytometry analysis.

663

Spheroid cultures

665   Donor B6 mice were injected with KPP-DTR cancer cells as described above. Tumors
666   were collected and digested. Single-cell suspensions were spun down at 1300RPM for
667   3min, decanted, and resuspended in fibroblast media (10% FBS αMEM, supplemented
668   1x L-Glutamine, 1x Pen/Strep, and 1x HEPES (all from Gibco) and 10% batch tested
669   low IgG Fcs (Gemini)). Bead depletion was performed with anti-cd45-bio, anti-cd24-bio,
670   and anti-cd31-bio in conjunction with the Easy-sep Biotin selection kit(STEMCELL cat
671   17655)). Cells were then sorted for LRRC15+ CAFs and LRRC15- Ly6C+ CAFs, as
672   controls normal pancreas was sorted for ENG+LY6c- and ENG-LY6C+ fibroblasts as
673   described in flow cytometry panels. Sorted cells were cultured in fibroblast media, 25ng
674   of Diptheria Toxin (DT, Enzo cat BML-G135-0001) was added to kill Tumor cells, this
675   was done in 100mm tissue culture treated dishes with 20ml media. Plates were rinsed
676   the next day with PBS and media with DT was replaced. Cells were cultured for 7-10
677   days week, with one passage at confluence, to expand fibroblasts. Cultured fibroblast
678   cells were trypsinized counted, then combined with KPP-mApple cancer cells.

679   For spheroids, 2000 KPP-mApple cancer cells were seeded in 100ul matrigel
680   (Corning cat 356231) with or without different fibroblasts populations(15k). Before
681   plating cells, 100ul of Matrigel was spread on each well of a 24-well glass-bottom
682   plate(Mattek) and allowed to polymerize, to keep cells from directly colonizing glass. 2
683   mls of fibroblast media was added to each well. The plate was imaged 2-4 times a week
684   with a 4x Plan Fluor objective (NA: 0.13, Nikon) on a Nikon Ti-E inverted microscope
685   equipped with a Neo scMOS camera (Andor, Oxford Instruments), a linear encoded
686   automated stage (Applied Scientific Instrumentation), 37C/5% CO2 environmental
687   chamber (Okolab), all run by NIS Elements software (Nikon). Image sets in TRITC and
688   brightfield of the Matrigel bubble were stitched and focused into one image projection
689   with an extended depth of focus module (EDF, Nikon). The resulting TRITC EDF image
690   was analyzed in Matlab (vR2018a, Mathworks) to measure total mApple spheroid area.

691

23

Tissue digestion, cell isolation and flow cytometry of human tissues

Human PDAC tumor samples were obtained and digested using a previous published protocol(9). PDAC samples were fragmented into small pieces (around $1mm^3$) and digested in CO2-independent medium (Gibco, cat18045-054) supplemented with 5% fetal bovine serum (FBS, PAA, catA11-151), 2 mg/ml collagenase I (Sigma-Aldrich, catC0130), 2 mg/ml hyaluronidase (Sigma-Aldrich, catH3506) and 25 mg/ml DNase I (Roche, cat11284932001) for 45 min at 37C with shaking (180-200rpm). After tissue digestion, cells were filtered using a cell strainer (40 mm, Fisher Scientific, cat223635447) and resuspended in PBS+ solution supplemented with 2 mM EDTA and 1% Human serum (Sigma P2918) to a final concentration at approximately 5x105 cells in 50 ul. Tissue single cell suspension was stained with antibody cocktail described in Suppl.Table 8.

Immunohistochemistry and image analysis of human tissues

Tissues were fixed in 10% neutral Buffered formalin for 24 hours, then dehydrated and paraffin embedded, and sectioned into 4um slices. Slides were de-paraffinized and antigen retrieved in CC1 buffer (TRIS-EDTA pH 8.1) at 95C for 64min. Sequential staining with elution step after first antibody detection was completed. First antibody anti-LRRC15 Abcam ab150376 used at 2.5ug/ml. Elution was done with CC2 buffer (Citrate-Acetate based with SDS 0.3% (pH 6.0)) Time: 8 min @ 100C. Second antibody was anti-cd8 Abcam ab101500 used at 1:200 and the isotype control for both was Naïve rabbit monoclonal CST cat 3900S. The detection system used was OmniMap-Rbt-HRP with DAB for CD8 and OmniMap-Rbt-HRP with Discovery purple for LRRC15.

Patient slides were immunoscored by pathologist with expertise in the field, H.K. Bright field images were acquired by a Hamamatsu Nanozoomer automated slide-scanning platform at a final magnification of 200x. The images were analyzed with the 2019a version of the Matlab software package (MathWorks). LRRC15[+] fibroblasts and CD8 cell nuclei were segmented by intensity thresholding and simple morphological filtering of the image.

Generation of bulk-sorted RNA sequencing

24

723 Single-cell suspensions were isolated as described above. For sorting described in

724 Figure1; 4 individual animals/group were sorted from the following: healthy albino B6

725 pancreas, KPP pancreas bearing tumors <4mm, or KPP pancreas with tumors > 10mm.

726 Samples were stained and sorted for EPCAM-, CD45-, TER119-, ITGA6-, CD31-, viable

727 (life Technologies) and then PDPN+ PDGFRa+ or PDPN- populations where sorted

728 directly into trizol, purity was assessed on a small aliquot sorted into PBS as 95% or

729 higher. RNA was isolated according to Universal RNeasy kit (Qiagen). RNA was profiled

730 with the Bioanalyzer Pico RNA Kit (Agilent Technologies). Low-input RNA kit (Clontech)

731 was used to generate cDNA libraries. RNA-seq libraries were multiplexed and

732 sequenced using HiSeq4000 to generated 30 M single end 50 bp reads per library.

733

734 ***Generation of Single-Cell sequencing libraries***

735 For single-cell sequencing. Albino B6 and KPP age and sex matched mice animals

736 were sacrificed, for each of 2 replicates; 5 albino B6 pancreases where used for

737 "normal", 5 KPP animals where used for other samples with tissues being divided into

738 "adjacent" (no masses), "small tumors" (tumors <4mm), and "large tumors" (tumors

739 5mm-10mm). Single-cell suspensions were isolated as described above. Samples were

740 stained and sorted for CD45-, TER119-, CD24a-, CD31-, 7AAD-, Calcein Violet+,

741 PDPN+. Sorted single-cell suspensions were converted to barcoded scRNA-seq

742 libraries by using the Chromium Single Cell 3' Library, Gel Bead & Multiplex Kit and

743 Chip Kit (10x Genomics), loading an estimated 6,000 cells per library and following the

744 manufacturer's instructions. Samples were processed using kits pertaining to either the

745 V2 barcoding chemistry of 10x Genomics. Single samples were processed in a single

746 well of a PCR plate, allowing all cells from a sample to be treated with the same master

747 mix and in the same reaction vessel. For each replicate, all samples (non-malignant and

748 tumor) were processed in parallel in the same thermal cycler. The final libraries were

749 profiled using the Bioanalyzer High Sensitivity DNA Kit (Agilent Technologies) and

750 quantified using the Kapa Library Quantification Kit (Kapa Biosystems). Each single-cell

751 RNA-seq library was sequenced twice in two lanes of HiSeq4000 (Illumina) to obtain

752 single end, 98 bp, ~500M reads per library.

753

Bioinformatic processing of mouse scRNA-seq data

Single-cell RNA-seq data for each replicate were processed with cellranger count (CellRanger 2.1.0 [10x Genomics]) using a custom reference package based on mouse reference genome GRCm38 and GENCODE gene models. Individual count tables were merged using cellranger aggr to reduce batch effects. Subsequent data analysis was carried out in R 3.5.1 and the Seurat package (v 2.3.4). From an initial set of 14,916 cells, counts of transcripts measured as unique molecular identifiers (UMIs) in each cell were normalized and log transformed to log(CPM/100+1) [CPM=UMI counts per million]. Cells with at least 1,200 measured genes per cell were considered for analysis. To remove noise from droplets containing more than one cell, we focused on cells with at most 5,000 measured genes. Dead cells were excluded by retaining cells with less than 3% mitochondrial reads leaving 13,454 cells for final analysis. Genes induced due to dissociation stress of single cells published previously(54) were used to score the dissociation stress in each cell with the *AddModuleScore* function in Seurat (see section Calculation of single cell scores for details). Subsequently, normalized data was scaled to regress out the number of distinct UMIs and the stress signature score.

Prior to dimensionality reduction we performed batch correction with Harmony (55) version 0.0.0.9000 as described in the tutorial at http://htmlpreview.github.io/?https://github.com/immunogenomics/harmony/blob/master/docs/SeuratV2.html We adjusted the cluster membership penalty parameter to theta to 1, in order to put a less strong force on combining cells across replicates. Dimensionality reduction was carried out with the Seurat package(56). Prior to principal component analysis we identified the 1,000 most variable genes (Seurat, *FindVariableGenes* using the mean of logged values and the variance to mean ratio (VMR) in non-logspace) and applied PCA to cells in this gene space. Principal components 1 to 20 were provided as an input for dimensionality reduction via tSNE and UMAP with default parameters in Seurat. Clusters of cells were identified based on a shared-nearest neighbor graph between cells and the smart moving (SLM) algorithm (k=40, resolution = 0.7). Markers for each cluster were identified by reducing the number of candidate genes to those genes which were a) at least log(0.25) fold higher expressed in the cluster under consideration compared to all other clusters and b)

26

785   expressed in at least 10% of cells in the cluster under consideration. For genes passing

786   those criteria significance between cells in the cluster vs all other cells was calculated

787   using Wilcoxon's rank sum test and adjusted with the Benjamini Hochberg method.

788

789   ***Calculation of single-cell scores***

790   Scores for single-cells were calculated as the average relative expression of a gene set

791   of interest, minus the average relative expression of a control gene set to account for

792   technical differences between cells, as described by Tirosh et al. (57) and implemented

793   in the Seurat AddModuleScore function. To obtain the geneset for lineages P3 and P4,

794   we used the 20 most significantly upregulated (adj. p-val. <0.00001, sorted by average

795   logFC) genes in each of the two normal populations to score all cells for these two

796   expression programs. To obtain a Collagen signature, we used all Collagen encoding

797   genes expressed in the full dataset to score each cell.

798

799   ***Pseudotime reconstruction***

800   Single-cell pseudotime trajectories were constructed with Monocle version 2.8.0

801   independent for each of the two fibroblast lineages, as Monocle 2 does not support

802   trajectories with multiple roots. For each trajectory we collected a set of 400 ordering

803   genes that defined CAF progression by testing each gene for differential expression

804   between normal fibroblasts from the respective lineage and fibroblasts from late-stage

805   tumors (adj. p<0.001, sorted by logFC, 200 most up and 200 down regulated genes).

806   Expression profiles were reduced to two dimensions using the DDRTree algorithm

807   included with Monocle 2 via the reduceDimension method and cells ordered along the

808   trajectory using the orderCells method, both with default parameters.

809

810   ***Enrichment analysis***

811   Pathway and Gene Ontology enrichments for cluster-specific genes were calculated

812   using ConsensusPathDB(58) and DAVID(59), respectively. Genes with an adj. p

813   <0.00001, log(0.25) fold higher expressed in the cluster, and expressed in >10% of cells

814   served as an input for analysis. Transcription factor binding site (TFBS) enrichment was

815   calculated using the OPOSSUM web service(60). Pathways with adjusted p-value <0.05

816   were considered significantly enriched for pathway analysis, TF motifs with a z-score

817   >10 were considered enriched. The z-score calculation in OPOSSUM uses the normal

818   approximation to the binomial distribution to compare the rate of occurrence of a TFBS

819   in the target set of genes to the expected rate estimated from a pre-computed

820   background set.

821

822   ***Bioinformatic processing of bulk RNA-seq data***

823   Bulk RNA-seq data from GEMM and subcutaneous mouse models was processed as

824   described previously (32). Briefly, reads were aligned to the mouse reference genome

825   (mm10) using GSNAP version '2013-10-10', allowing a maximum of two mismatches

826   per 75 base sequence (parameters: '-M 2 -n 10 -B 2 -i 1 -N 1 -w 200000 -E 1 --pairmax-

827   rna=200000 --clip-overlap'). To quantify gene expression levels, the number of reads

828   mapped to the exons of each gene was calculated in a strand-specific manner using the

829   functionality provided by the R package GenomicAlignments. For heatmap

830   visualizations, per-gene counts were normalized to Reads Per Kilobase Million (RPKM)

831   within each sample to account for differences in transcript length and sequencing depth.

832   Differentially expressed genes between groups were determined using the R

833   package limma (61) after trimmed mean of M-values (TMM) normalization, which

834   implements an empirical Bayesian approach to estimate gene expression changes

835   using moderated t-tests.

836

837   ***Bioinformatic processing of human PDAC scRNA-seq data***

838   Data from 24 PDAC patients and 11 control pancreas tissues was obtained from the

839   Genome Sequence Archive under project PRJCA001063 in FASTQ format. Single-cell

840   RNA-seq data for each patient was processed with cellranger *count* (Cell Ranger 3.0.2

841   [10x Genomics]) using standard parameters and supplying a custom reference package

842   based on human reference genome GRCh38 and GENCODE gene models. Samples of

843   22 patients and 11 control tissues for which the correct chemistry was detected by Cell

844   Ranger from the sequencing data were used for downstream analysis.

845   Subsequent data analysis was carried out in R 3.5.1 and the Seurat package (v

846   3.0.2). Cells with at least 300 measured genes per cell were considered for analysis. To

28

847 remove noise from droplets containing more than one cell, we focused on cells with at

848 most 6,000 measured genes. Dead cells were excluded by retaining cells with less than

849 15% mitochondrial reads leaving 84,276 cells for final analysis. Subsequently, data was

850 normalized to log(CPM/100+1) and scaled regressing out the number of distinct UMIs

851 and the fraction of mitochondrial reads during scaling.

852 Dimensionality reduction was carried out with the Seurat package. Prior to

853 principal component analysis we identified the 2,000 most variable genes and applied

854 PCA to cells in this gene space. Principal components 1 to 20 were provided as an input

855 for dimensionality reduction via UMAP with default parameters. Clusters of cells were

856 identified based on a shared-nearest neighbor graph between cells and the smart

857 moving (SLM) algorithm (resolution = 0.1). Markers for each cluster were identified by

858 reducing the number of candidate genes to those genes which were a) at least

859 log(0.25)-fold higher expressed in the cluster under consideration compared to all other

860 clusters and b) expressed in at least 10% of cells in the cluster under consideration. For

861 genes passing those criteria, significance between cells in the cluster vs. all other cells

862 was calculated using Wilcoxon's rank sum test and adjusted with the Benjamini

863 Hochberg method(62). Average expression within individual clusters was calculated

864 with the *AverageExpression* function in Seurat and subsequently z-score transformed

865 for each gene. The minimum spanning to infer global lineage structure of CAFs was

866 calculated using Slingshot(19) with default parameters and defining normal fibroblasts

867 (leftmost population of PC1) as starting and TGFB CAFs (rightmost population of PC1)

868 as end point. KRAS G12X mutations in each cell were manually identified from

869 individual reads in BAM alignment files visualized *via* IGV (63) and assigned to a cell via

870 the CB tag in the BAM file. Copy-number alterations were inferred from single-cell RNA-

871 seq data with the CONICS R package(64) using non-malignant acinar cells as reference

872 cells. The default filtering and normalization procedures were followed, as outlined in

873 https://goo.gl/tFYLEh.

874

875 ***Bioinformatic processing of human HNSC scRNA-seq data***

876 Normalized data from 18 HNSC patients(43) was obtained from GEO (GSE103322) as

877 log(CPM/10+1) transformed gene-by-cell count matrix. Annotations of cell types

29

878   (malignant/non-malignant, as well as immune and stromal cell types for non-tumor cells)

879   for each cell were downloaded from the same GEO repository. Data was scaled

880   regressing out the number of distinct UMIs and the different usage of enzymes for

881   scRNA-seq library preparation during scaling. Dimensionality reduction was carried out

882   with the Seurat package. Prior to principal component analysis we identified the 2,000

883   most variable genes and applied PCA to cells in this gene space. Principal components

884   1 to 30 were provided as an input for dimensionality reduction via UMAP with default

885   parameters in Seurat (v3.0.2). Clusters of cells were identified based on a shared-

886   nearest neighbor graph between cells and the smart moving (SLM) algorithm (resolution

887   = 0.4).

888

Bioinformatic processing of mouse KPC scRNA-seq data

889

890   Normalized fibroblast-enriched data from 4 KPC mice(36) was obtained from GEO

891   (GSE129455) as log (number of UMIs in each cell is equal to the median UMI count

892   across the dataset) transformed gene-by-cell count matrix. Ensembl IDs were converted

893   to gene names using the biomart R package(65) Data was scaled regressing out the

894   number of distinct UMIs during scaling. Dimensionality reduction was carried out with

895   the Seurat package. Prior to principal component analysis we identified the 3,000 most

896   variable genes and applied PCA to cells in this gene space. Principal components 1 to

897   20 were provided as an input for dimensionality reduction via UMAP with default

898   parameters in Seurat (v3.0.2). Clusters of cells were identified based on a shared-

899   nearest neighbor graph between cells and the smart moving (SLM) algorithm (resolution

900   = 0.2).

901

Bioinformatic processing of human bulk RNA-seq data

902

903   Comparisons of normalized *LRRC15* levels between tumors from TCGA and their host

904   tissues from GTEx were retrieved from the GEPIA (66) web platform and filtered for

905   tumor types with significant differences between normal and tumor tissue. Raw

906   expression counts per sample of microdissected stroma (n=122) and tumor (n=65)

907   samples(42) were downloaded from GEO (GSE93326). Raw expression counts per

908   sample of normal pancreas RNA-seq (n=247) was downloaded from GTEx (67). Both

30

909 datasets were normalized to Log2(CPM+1) and heatmaps were generated using the

910 pheatmap R package (https://cran.r-project.org/web/packages/pheatmap/) using

911 complete linkage clustering and with Euclidean distance as distance measure.

912 For pan-cancer TCGA data analysis the TCGA Pan-Cancer (PANCAN) batch effects

913 normalized mRNA data was downloaded from the UCSC XenaBrowser

914 (https://xenabrowser.net) providing a gene by samples matrix of log2(norm_value+1)

915 counts table and patient metadata. From the initial set of n=11,060 samples, we only

916 utilized those samples that were annotated as "Primary Tumor" or "Additional - New

917 Primary" in the metadata table resulting in 9,712 samples from 31 different cancer

918 types.

919

920 ***Analysis of Immunotherapy trial data***

921 Whole transcriptome data from patients enrolled in anti PD-L1 (Atezolizumab)

922 immunotherapy trial imvigor 210 (NCT02951767, NCT02108652)(68), were generated

923 as described previously (32). Data from anti PD-L1 (Atezolizumab) immunotherapy trial

924 PCD (NCT01375842) were generated as given in(69). For calculation of LRRC15 CAF

925 scores from expression data, the signature is computed by using the

926 *eigenWeightedMean* method from the MultiGSEA R package

927 (https://github.com/lianos/multiGSEA). Briefly, the expression of each gene in a

928 signature is first z-score transformed. Then, a principal component analysis was

929 performed, weights for the genes are calculated by the percent of which they contribute

930 to the first principal component indicated by eigengene. Last, a weighted average per

931 sample is calculated as the final score. This approach has the advantage of focusing

932 the score for the set on the largest block of well-correlated (or anti-correlated) genes in

933 the set, while downweighting contributions from genes that do not track with other set

934 members.

935 For survival analysis, patients were split into low and high expression groups by

936 median. Kaplan-Meier curves were generated using the RMS R package (

937 http://biostat.mc.vanderbilt.edu/wiki/Main/Rrms).

938

939 ***Data availability***

31

940   The single-cell RNA-seq data from mouse PDAC KPP GEMMs are available from the

941   ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) under accession number E-

942   MTAB-8483.

943   .

944

945

32

Acknowledgements

Author Contributions

S.J.T and C.X.D designed the experiments within this study. S.J.T, C.X.D, and S.M. wrote the manuscript. C.X.D performed the studies described with contribution on animal immunofluorescence from B.B. and in vivo help from A.C. S.M. performed critical bioinformatic analysis of all sequencing data, both generated for this study or publicly available, with contribution from Y.S. and supervision from C.K. and R.B. S.K. and H.K. performed all staining of human samples and H.K performed scoring of human PDAC patient slide samples. J.H. provided digital pathology expertise. Z.M. and Y.L. performed all sequencing generated in this study. M.J. provided parental cell lines for in vivo tumor injection studies as well as samples and expertise on GEMM models. O.F. provided H&E imaging and expertise on mouse models of PDAC. S.G. provided analysis of spheroid growth. T.W.B. generated the anti-LRRC15 antibodies used in these studies.

33

REFERENCES

1. Stark AP, Sacks GD, Rochefort MM, Donahue TR, Reber HA, Tomlinson JS, et al. Long-term survival in patients with pancreatic ductal adenocarcinoma. Surgery. 2016;159:1520–7.

2. Kraman M, Bambrough PJ, Arnold JN, Roberts EW, Magiera L, Jones JO, et al. Suppression of Antitumor Immunity by Stromal Cells Expressing Fibroblast Activation Protein–α. Science. 2010;330:827–30.

3. Lo A, Wang L-CS, Scholler J, Monslow J, Avery D, Newick K, et al. Tumor-Promoting Desmoplasia Is Disrupted by Depleting FAP-Expressing Stromal Cells. Cancer Res. 2015;75:2800–10.

4. Özdemir BC, Pentcheva-Hoang T, Carstens JL, Zheng X, Wu C-C, Simpson TR, et al. Depletion of Carcinoma-Associated Fibroblasts and Fibrosis Induces Immunosuppression and Accelerates Pancreas Cancer with Reduced Survival. Cancer Cell. 2015;28:831–3.

5. Santos AM, Jung J, Aziz N, Kissil JL, Puré E. Targeting fibroblast activation protein inhibits tumor stromagenesis and growth in mice. J Clin Invest. 2009;119:3613–25.

6. Wang L-CS, Lo A, Scholler J, Sun J, Majumdar RS, Kapoor V, et al. Targeting Fibroblast Activation Protein in Tumor Stroma with Chimeric Antigen Receptor T Cells Can Inhibit Tumor Growth and Augment Host Immunity without Severe Toxicity. Cancer Immunol Res. 2014;2:154–66.

7. LeBleu VS, Kalluri R. A peek into cancer-associated fibroblasts: origins, functions and translational impact. Dis Model Mech. 2018;11:dmm029447.

8. Avery D, Govindaraju P, Jacob M, Todd L, Monslow J, Puré E. Extracellular matrix directs phenotypic heterogeneity of activated fibroblasts. Matrix Biol. 2018;67.

9. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, et al. Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. Cancer Cell. 2018;33.

10. Feig C, Jones JO, Kraman M, Wells RJ, Deonarine A, Chan DS, et al. Targeting CXCL12 from FAP-expressing carcinoma-associated fibroblasts synergizes with anti–PD-L1 immunotherapy in pancreatic cancer. Proc National Acad Sci. 2013;110:20212–7.

11. Apte M, Haber P, Applegate T, Norton I, McCaughan G, Korsten M, et al. Periacinar stellate shaped cells in rat pancreas: identification, isolation, and culture. Gut. 1998;43:128–33.

34

1018

1019 12. Aguirre AJ, Bardeesy N, Sinha M, Lopez L, Tuveson DA, Horner J, et al. Activated
1020 Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal
1021 adenocarcinoma. Gene Dev. 2003;17:3112–26.

1022

1023 13. Chung W-J, Daemen A, Cheng JH, Long JE, Cooper JE, Wang B, et al. Kras mutant
1024 genetically engineered mouse models of human cancers are genomically
1025 heterogeneous. Proc National Acad Sci. 2017;114:E10947–55.

1026

1027 14. Hruban RH, Adsay VN, Albores-Saavedra J, Anver MR, Biankin AV, Boivin GP, et
1028 al. Pathology of Genetically Engineered Mouse Models of Pancreatic Exocrine Cancer:
1029 Consensus Report and Recommendations. Cancer Res. 2006;66:95–106.

1030

1031 15. Erkan M, Reiser-Erkan C, Michalski CW, Deucker S, Sauliunaite D, Streit S, et al.
1032 Cancer-Stellate Cell Interactions Perpetuate the Hypoxia-Fibrosis Cycle in Pancreatic
1033 Ductal Adenocarcinoma. Neoplasia. 2009;11:497–508.

1034

1035 16. Buechler MB, Kim K-W, Onufer EJ, Williams JW, Little CC, Dominguez CX, et al. A
1036 Stromal Niche Defined by Expression of the Transcription Factor WT1 Mediates
1037 Programming and Homeostasis of Cavity-Resident Macrophages. Immunity.
1038 2019;51:119-130.e5.

1039

1040 17. Cremasco V, Astarita JL, Grauel AL, Keerthivasan S, MacIsaac KD, Woodruff MC,
1041 et al. FAP delineates heterogeneous and functionally divergent stromal cells in immune-
1042 excluded breast tumors. Cancer Immunol Res. 2018;6:canimm.0098.2018.

1043

1044 18. Xie T, Wang Y, Deng N, Huang G, Taghavifar F, Geng Y, et al. Single-Cell
1045 Deconvolution of Fibroblast Heterogeneity in Mouse Pulmonary Fibrosis. Cell Reports.
1046 2018;22:3625–40.

1047

1048 19. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage
1049 and pseudotime inference for single-cell transcriptomics. Bmc Genomics. 2018;19:477.

1050

1051 20. Kalluri R. Basement membranes: structure, assembly and role in tumour
1052 angiogenesis. Nat Rev Cancer. 2003;3:422–33.

1053

1054 21. Groulx J-F, Gagné D, Benoit YD, Martel D, Basora N, Beaulieu J-F. Collagen VI is a
1055 basement membrane component that regulates epithelial cell–fibronectin interactions.
1056 Matrix Biol. 2011;30:195–206.

1057

1058 22. Steele CW, Karim SA, Leach J, Bailey P, Upstill-Goddard R, Rishi L, et al. CXCR2
1059 Inhibition Profoundly Suppresses Metastases and Augments Immunotherapy in
1060 Pancreatic Ductal Adenocarcinoma. Cancer Cell. 2016;29:832–45.

1061

1062 23. Sano M, Ijichi H, Takahashi R, Miyabayashi K, Fujiwara H, Yamada T, et al.
1063 Blocking CXCLs–CXCR2 axis in tumor–stromal interactions contributes to survival in a

35

mouse model of pancreatic ductal adenocarcinoma through reduced cell invasion/migration and a shift of immune-inflammatory microenvironment. Oncogenesis. 2019;8:8.

24. Long KB, Gladney WL, Tooker GM, Graham K, Fraietta JA, Beatty GL. IFNγ and CCL2 Cooperate to Redirect Tumor-Infiltrating Monocytes to Degrade Fibrosis and Enhance Chemotherapy Efficacy in Pancreatic Carcinoma. Cancer Discov. 2016;6:400–13.

25. Jackson HW, Defamie V, Waterhouse P, Khokha R. TIMPs: versatile extracellular regulators in cancer. Nat Rev Cancer. 2017;17:38.

26. Goel H, Mercurio AM. VEGF targets the tumour cell. Nat Rev Cancer. 2013;13:871–82.

27. Liang M, Ma Q, Ding N, Luo F, Bai Y, Kang F, et al. IL-11 is essential in promoting osteolysis in breast cancer bone metastasis via RANKL-independent activation of osteoclastogenesis. Cell Death Dis. 2019;10:353.

28. Shi Y, Gao W, Lytle NK, Huang P, Yuan X, Dann AM, et al. Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring. Nature. 2019;569:131–5.

29. Pietras K, Sjöblom T, Rubin K, Heldin C-H, Östman A. PDGF receptors as cancer drug targets. Cancer Cell. 2003;3:439–43.

30. D'Costa Z, Jones K, Azad A, van Stiphout R, Lim SY, Gomes AL, et al. Gemcitabine-induced TIMP1 attenuates therapy response and promotes tumor growth and liver metastasis in pancreatic cancer. Cancer Res. 2017;77:canres.2833.2016.

31. Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur B, et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. Nat Cell Biol. 2015;17:816–26.

32. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018;554:544.

33. Ling J, Kang Y, Zhao R, Xia Q, Lee D-F, Chang Z, et al. KrasG12D-Induced IKK2/β/NF-κB Activation by IL-1α and p62 Feedforward Loops Is Required for Development of Pancreatic Ductal Adenocarcinoma. Cancer Cell. 2012;21:105–20.

34. Schmid MC, Avraamides CJ, Foubert P, Shaked Y, Kang S, Kerbel RS, et al. Combined Blockade of Integrin-α4β1 Plus Cytokines SDF-1α or IL-1β Potently Inhibits Tumor Inflammation and Growth. Cancer Res. 2011;71:6965–75.

36

1110 35. Tjomsland V, Bojmar L, Sandström P, Bratthäll C, Messmer D, Spångeus A, et al.
1111 IL-1α Expression in Pancreatic Ductal Adenocarcinoma Affects the Tumor Cell
1112 Migration and Is Regulated by the p38MAPK Signaling Pathway. Plos One.
1113 2013;8:e70874.
1114
1115 36. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, et al. Cross-
1116 Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-
1117 Presenting Cancer-Associated Fibroblasts. Cancer Discov. 2019;9:1102–23.
1118
1119 37. Calon A, Espinet E, Palomo-Ponce S, Tauriello D, Iglesias M, Céspedes M, et al.
1120 Dependency of Colorectal Cancer on a TGF-β-Driven Program in Stromal Cells for
1121 Metastasis Initiation. Cancer Cell. 2012;22:571–84.
1122
1123 38. Purcell JW, Tanlimco SG, Hickson JA, Fox M, Sho M, Durkin L, et al. LRRC15 is a
1124 novel mesenchymal protein and stromal target for antibody-drug conjugates. Cancer
1125 Res. 2018;78:canres.0327.2018.
1126
1127 39. Bausch-Fluck D, Hofmann A, Bock T, Frei AP, Cerciello F, Jacobs A, et al. A Mass
1128 Spectrometric-Derived Cell Surface Protein Atlas. Plos One. 2015;10:e0121314.
1129
1130 40. Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, et al.
1131 Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer.
1132 J Exp Med. 2017;214:jem.20162024.
1133
1134 41. Peng J, Sun B-F, Chen C-Y, Zhou J-Y, Chen Y-S, Chen H, et al. Single-cell RNA-
1135 seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic
1136 ductal adenocarcinoma. Cell Res. 2019;1–14.
1137
1138 42. Maurer C, Holmstrom SR, He J, Laise P, Su T, Ahmed A, et al. Experimental
1139 microdissection enables functional harmonisation of pancreatic cancer subtypes. Gut.
1140 2019;68:1034.
1141
1142 43. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-Cell
1143 Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and
1144 Neck Cancer. Cell. 2017;171:1611-1624.e24.
1145
1146 44. Principe DR, Park A, Dorman MJ, Kumar S, Viswakarma N, Rubin J, et al. TGFβ
1147 Blockade Augments PD-1 Inhibition to Promote T-Cell Mediated Regression of
1148 Pancreatic Cancer. Mol Cancer Ther. 2018;18:molcanther.0850.2018.
1149
1150 45. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza GS, Hoadley KA, et al. Virtual
1151 microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic
1152 ductal adenocarcinoma. Nat Genet. 2015;47:1168.
1153
1154 46. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Yang T-H, et al. The
1155 Immune Landscape of Cancer. Immunity. 2018;48:812-830.e14.

1156

1157 47. Biffi G, Oni TE, Spielman B, Hao Y, Elyada E, Park Y, et al. IL-1-induced JAK/STAT
1158 signaling is antagonized by TGF-beta to shape CAF heterogeneity in pancreatic ductal
1159 adenocarcinoma. Cancer Discov. 2018;CD-18-0710.

1160

1161 48. Koopmans T, Rinkevich Y. Mesothelial to mesenchyme transition as a major
1162 developmental and pathological player in trunk organs and their cavities. Commun
1163 Biology. 2018;1:170.

1164

1165 49. Corcoran RB, Contino G, Deshpande V, Tzatsos A, Conrad C, Benes CH, et al.
1166 STAT3 Plays a Critical Role in KRAS-Induced Pancreatic Tumorigenesis. Cancer Res.
1167 2011;71:5020–9.

1168

1169 50. Shien K, Papadimitrakopoulou VA, Ruder D, Behrens C, Shen L, Kalhor N, et al.
1170 JAK1/STAT3 Activation through a Proinflammatory Cytokine Pathway Leads to
1171 Resistance to Molecularly Targeted Therapy in Non–Small Cell Lung Cancer. Mol
1172 Cancer Ther. 2017;16:2234–45.

1173

1174 51. Grivennikov SI, Karin M. Dangerous liaisons: STAT3 and NF-κB collaboration and
1175 crosstalk in cancer. Cytokine Growth F R. 2010;21:11–9.

1176

1177 52. Fletcher AL, Malhotra D, Acton SE, Lukacs-Kornek V, Bellemare-Pelletier A, Curry
1178 M, et al. Reproducible Isolation of Lymph Node Stromal Cells Reveals Site-Dependent
1179 Differences in Fibroblastic Reticular Cells. Front Immunol. 2011;2:35.

1180

1181 53. Morgan-Lappe S, GISH KC, Hickson JA, Purcell JW. ANTI-huLRRC15 ANTIBODY
1182 DRUG CONJUGATES AND METHODS FOR THEIR USE.

1183

1184 54. van den Brink SC, Sage F, Vértesy Á, Spanjaard B, Peterson-Maduro J, Baron CS,
1185 et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue
1186 subpopulations. Nat Methods. 2017;14:935–6.

1187

1188 55. Korsunsky I, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, et al. Fast,
1189 sensitive, and accurate integration of single cell data with Harmony. Biorxiv.
1190 2018;461954.

1191

1192 56. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al.
1193 Comprehensive Integration of Single-Cell Data. Cell. 2019;177:1888-1902.e21.

1194

1195 57. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, et al. Single-
1196 cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature.
1197 2016;539:309.

1198

1199 58. Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and interpreting genome
1200 data at the network level with ConsensusPathDB. Nat Protoc. 2016;11:1889–907.

1201

1202   59. Huang D, Sherman BT, Tan Q, Collins JR, Alvord GW, Roayaei J, et al. The DAVID
1203   Gene Functional Classification Tool: a novel biological module-centric algorithm to
1204   functionally analyze large gene lists. Genome Biol. 2007;8:R183.
1205
1206   60. Sui SJ, Mortimer JR, Arenillas DJ, Brumm J, Walsh CJ, Kennedy BP, et al.
1207   oPOSSUM: identification of over-represented transcription factor binding sites in co-
1208   expressed genes. Nucleic Acids Res. 2005;33:3154–64.
1209
1210   61. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers
1211   differential expression analyses for RNA-sequencing and microarray studies. Nucleic
1212   Acids Res. 2015;43:e47–e47.
1213
1214   62. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and
1215   Powerful Approach to Multiple Testing. J Royal Statistical Soc Ser B Methodol.
1216   1995;57:289–300.
1217
1218   63. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al.
1219   Integrative genomics viewer. Nat Biotechnol. 2011;29:24.
1220
1221   64. Müller S, Cho A, Liu SJ, Lim DA, Diaz A. CONICS integrates scRNA-seq with DNA
1222   sequencing to map gene expression to tumor sub-clones. Bioinformatics.
1223   2018;34:3217–9.
1224
1225   65. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration
1226   of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc.
1227   2009;4:1184.
1228
1229   66. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and
1230   normal gene expression profiling and interactive analyses. Nucleic Acids Res.
1231   2017;gkx247-.
1232
1233   67. Consortium Gte, Laboratory D, Group S, groups E, Fund N, NIH/NCI, et al. Genetic
1234   effects on gene expression across human tissues. Nature. 2017;550:204.
1235
1236   68. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi
1237   A, et al. Atezolizumab in patients with locally advanced and metastatic urothelial
1238   carcinoma who have progressed following treatment with platinum-based
1239   chemotherapy: a single-arm, multicentre, phase 2 trial. Lancet. 2016;387:1909–20.
1240
1241   69. Kowanetz M, Zou W, Gettinger SN, Koeppen H, Kockx M, Schmid P, et al.
1242   Differential regulation of PD-L1 expression by immune and tumor cells in NSCLC and
1243   the response to treatment with atezolizumab (anti–PD-L1). Proc National Acad Sci.
1244   2018;115:201802166.
1245
1246

1247 **FIGURE LEGENDS**

1248 **Figure 1. Podoplanin (PDPN) expression identifies the majority of tissue**

1249 **fibroblasts in normal and tumor-bearing pancreas. a)** PDPN and CD31 expression

1250 on cells digested according to standard stellate cell pronase protocols (left) or new

1251 digestion protocol (right) and enriched by gradient centrifugation, gated on live CD45$^-$

1252 Epcam$^-$ cells. **b)** PDPN and CD31 expression in normal or tumor-bearing KPP GEMM

1253 pancreases (left) with quantification (right). Blood endothelial cells (BECs) are PDPN-

1254 CD31+ as highlighted in the contour plots (green). **c)** Immunofluorescence imaging of

1255 PDPN staining in normal pancreas (left, arrowheads highlight examples of PDPN$^+$ cells

1256 surrounding acinar clusters) and KPP GEMM in advanced PDAC, either in "distal" tissue

1257 not directly contacting tumor (middle) or "proximal" where tumor cells were visibly

1258 contacting non-tumor tissue (right). Scale bar represents 50um. **d)** Heatmap showing

1259 relative gene expression levels from RNA-seq analysis of PDPN$^+$CD31$^-$PDGFRa$^+$

1260 stroma and the DN populations demonstrating fibroblastic nature of the PDPN$^+$

1261 population. **e)** Expression of selected CAF-associated genes in respective fibroblasts

1262 populations (log2(rPKM+1)). Statistical comparison between all groups performed with

1263 Tukey's test, bars designate pairwise comparisons where p<.05. All dot plots are

1264 representative of flow cytometry data from single-cell dissociated tissues. a) is

1265 representative of 4 independent experiments with 5 animals pooled per condition. b)

1266 Combined data from 5 independent experiments with total n=12 for normal and n=25 for

1267 PDAC GEMM. Statistical test used was Sidak's multiple comparisons test ***p<0.0005;

1268 **** p<0.0001 c) Representative image from normal n=4 for PDAC GEMMs n= 10. d) for

1269 normal samples n=3-4, with 5 animals pooled/single sequenced sample, for tumor n=4

1270 with 1-2 animals pooled/single sequenced sample.

1271

1272 **Figure 2. PDPN expression is a feature of several stromal populations. a)**

1273 Experimental design of the scRNA-seq experiment. **b)** Left: t-Distributed Stochastic

1274 Neighbor Embedding (tSNE) embedding of 13,454 single cells sorted from n=20 mice

1275 across all conditions (normal, adjacent, small and large tumors). Clusters identified

1276 through graph-based clustering are indicated by color. Right: Heatmap showing the

1277 relative average expression of the most strongly enriched genes for each cluster

40

1278   identified by log-fold change of cells within a cluster to all other cells in the dataset. Two

1279   representative genes are highlighted for each cluster. fEMT: full EMT, pEMT: partial

1280   EMT. **c)** tSNE embedding as in b), color indicates normalized expression level

1281   (log(CPM/100+1) of indicated genes. **d)** Fraction of cells in each cluster (z-scored per

1282   row) from each condition (column). Two adjacent rows per cluster visualize the fraction

1283   in each replicate. **e)** Left: Comparison of gene expression from bulk RNA-seq data

1284   between normal mesothelial cells and fibroblasts based on log2 fold-change (x-axis)

1285   and -log10(adj. p-value; limma). Genes enriched in cluster 6 of the scRNA-seq data in

1286   Fig. 2b are highlighted in red, genes upregulated in clusters 3 and 4 are highlighted in

1287   green. Right: Heatmap of the relative average expression of markers for mesothelial

1288   cells identified by both Xie et al. (scRNA-seq) and Buechler et al. (bulk RNA-seq) in

1289   clusters 0,1,2,3,4,6, and 8 from b). **f)** The fraction of fibroblast cells from clusters 0, 2, 8,

1290   and 1 (y-axis) in tumor adjacent tissue, tissue from small, and tissue from large tumors

1291   (x-axis; columns sum to 1).

1292

1293   **Figure 3. Two normal tissue fibroblasts follow two separate differentiation**

1294   **trajectories driven by IL1 and TGFB. a)** Left: UMAP embedding of cells from normal

1295   pancreas. Color and numbers indicate dot density

1296   (grey:low;blue:low/medium/;red:medium/high;yellow:high). Middle: Color indicates

1297   cluster membership. Right: Color indicates marker gene expression. **b)** Representative

1298   flow cytometry plots of fibroblasts gated on live PDPN$^+$ fibroblasts from normal mouse

1299   pancreas stained for DPP4 and ENDOGLIN (ENG) or Ly6c and ENG. **c)** Top: Density

1300   distribution of cells from individual fibroblast clusters (color) along the first principal

1301   component of a PCA analysis. Bottom: PC1 loadings of genes highlighted in Suppl. Fig.

1302   3a. **d)** Left: tSNE from Fig. 2b) restricted to fibroblasts. Color indicates the score for

1303   expression of marker genes for two populations from normal pancreas shown in a.

1304   Right: Boxplots outline the distribution of scores in each cluster. **e)** Heatmap visualizing

1305   the relative average expression of indicated genes (rows) in fibroblast clusters

1306   (columns). **f)** Top: Monocle2 pseudo-time trajectory of c4 normal fibroblasts and c4-

1307   derived CAFs. Cells are colored by cluster. Bottom: same as top, but for c3 normal

1308   fibroblasts and c3-derived CAFs. **g)** Heatmaps visualizing the relative average

41

1309   expression of extracellular matrix encoding genes (top) and chemokines/cytokines
1310   (bottom) across the 6 main fibroblast clusters (rows). Columns were clustered using
1311   complete linkage clustering and Euclidean distance as distance measure. **h)** Left:
1312   Transcription factor motif enrichment analysis in promoters (+-10kb of TSS) of genes
1313   specific to CAF populations c2 and c8. Right: Pathway enrichment analysis for genes
1314   specific to CAF clusters 8 (top) and 2 (bottom). **i)** Left: Heatmap visualizing the relative
1315   average expression of genes from the F-TBRS signature across fibroblast clusters.
1316   Columns were clustered with complete linkage clustering using Euclidean distance.
1317   Right: Relative average expression of indicated genes (columns) in CAF clusters (rows).
1318   Columns were clustered with complete linkage clustering using Euclidean distance. **j)**
1319   Heatmap comparing the relative average expression of markers genes (rows) of normal
1320   fibroblast clusters 3 and 4, as well as normal mesothelial cells between iCAFs,
1321   myCAFs, apCAFs, c3 normal fibroblasts, c4 normal fibroblasts, and normal mesothelial
1322   cells (rows). Rows and columns were clustered with complete linkage clustering using
1323   Euclidean distance as distance measure.

1324

1325   **Figure 4. TGFB-responsive CAF can be identified by LRRC15 expression. a)** Log2
1326   fold-change of gene expression (dots) between normal fibroblasts and early PDPN[+]
1327   PDAC CAFs (x-axis) and normal fibroblasts and late-stage PDAC CAFs (y-axis). Genes
1328   significant (adj. p-val<0.1 and absolute(log2FC)>1) are indicated in blue, genes
1329   significant in only late stage CAFs in green and only in early CAFs in red. **b)**
1330   Immunofluorescent image of PDPN (green), LRRC15 (white), EPCAM (red), and DAPI
1331   (blue) expression in tumor-bearing KPP GEMM pancreas, yellow arrowheads highlight
1332   examples of LRRC15[+] clusters. **c)** Immunofluorescent image of PDPN (green), LRRC15
1333   (white), EPCAM (red), and DAPI (blue) expression in subcutaneous KPP tumor showing
1334   PDPN[+]LRRC15[+] fibroblasts. **d)** Representative plot showing LRRC15 staining by flow
1335   cytometry in the PDPN[+] fibroblast gate (left), quantification of LRRC15[+] cells by
1336   frequency of population and numbers normalized to weight in the KPPsc model. **e)** Top:
1337   Representative images from single wells (from 24-well plate) of KPP-mApple spheroids
1338   after 12 days of 3D culture. KPP-mApple spheroids were cultured alone or co-cultured
1339   with the designated fibroblast population. Bottom: Quantification of total area of mApple[+]

42

1340   spheroids per whole well over time. These data are combined from 2 independent

1341   experiments, for each experiment n=4 wells/condition. Dunnett's multiple comparison

1342   test, comparing tumor alone against all other conditions, showed a significant difference

1343   ****p<.0001 for all conditions.

1344

1345   **Figure 5. TGFB-responsive LRRC15$^+$ CAFs are the most frequent fibroblast**

1346   **population in human PDAC. a)** Left: UMAP embedding of 84,276 high-quality cells

1347   from 22 PDAC patients. Clusters identified through graph-based clustering are indicated

1348   by color. Cells with identified KRAS single-nucleotide variation identified from scRNA-

1349   seq reads are highlighted in orange. Labels for each cluster were identified by markers

1350   on the right. Right: Heatmap showing the relative average expression of the most

1351   strongly enriched genes for each cluster identified by log-fold change of cells within a

1352   cluster to all other cells in the dataset. Two representative genes are highlighted for

1353   each cluster. **b)** Bar plots representing the number of patients that contributed at least

1354   10 cells to a cluster given in a). **c)** Relative expression of marker genes for clusters 0, 5,

1355   and 6 from a) in bulk RNA-seq samples from microdissected tumor (n=65), stroma (n=

1356   122) and normal pancreas (n=247). **d)** Left: UMAP embedding of 8,931 fibroblast cells.

1357   Clusters identified through graph-based clustering are indicated by color. Right:

1358   Heatmap showing the most strongly enriched genes for each cluster identified by

1359   Wilcoxon's ranks sum test, all p<1e-10. Three representative genes are highlighted for

1360   each cluster. **e)** Distribution of left: *LRRC15* and right: *HAS1* expression in bulk RNA-

1361   seq data from 65 tumor, 122 stroma, and 247 normal samples. **f)** Average expression

1362   (+- standard error of mean) of signature genes from d) in 122 microdissected PDAC

1363   stroma bulk RNA-seq samples. **g)** Representative IHC image from a PDAC patient

1364   sample (1 of 70. More images in Suppl. Fig. 5b). Purple: LRRC15 staining Blue: Nuclear

1365   counterstain Brown: CD8 stain. Dashed line demarcates tumor islet. Arrowheads point

1366   of CD8 T cells. Scale bar represents 200um. **h)** Representative flow cytometry plots of

1367   mesenchymal cells, gated on live EPCAM$^-$, CD45$^-$, CD31$^-$ cells from PDAC tissue

1368   stained for LRRC15 and EPCAM. **i)** Heatmap visualizing the relative average

1369   expression of mouse IL1 CAF markers and mouse TGFB CAF markers and their

1370   respective homologues in human across human and mouse IL1 CAFs, human and

1371   mouse normal fibroblasts, as well as human and mouse TGFB CAFs (rows). Rows and

1372   columns were clustered using complete linkage clustering and Euclidean distance as

1373   distance measure. Representative genes for each of the two main clusters are

1374   highlighted and represented by their human gene symbol. **j)** Top: Scatter plot comparing

1375   the average expression of genes in fibroblast single-cell cluster 5 from normal pancreas

1376   (Suppl. Fig. 5d) to the average expression of CAF cluster 1 from d). Bottom: Heatmap

1377   visualizing the relative average expression of indicated genes (rows) in mouse CAF

1378   clusters from figure 2b). Columns were clustered using complete linkage clustering and

1379   Euclidean distance as distance measure. **k)** Principle component analysis of normal

1380   fibroblasts from Suppl. Figure 5d in purple and human CAFs colored by clusters from d).

1381   Dots and dashed lines represent the cluster-based minimum spanning tree. **l)**

1382   Schematic representation of mouse and human PDAC fibroblast evolution. Rectangles

1383   demarcate different stages of tumor progression; purple is the non-malignant stage*,

1384   orange is the early stage of tumor development, red is the established tumor stage.

1385   Proteins shown have been validated as markers that identify the respective population,

1386   genes that are shown were among the most significantly enriched for that population.

1387   Currently we cannot identify all populations by protein markers. The pie charts show the

1388   frequency of CAF populations found in late tumors for mouse, and overall in patient

1389   PDAC tumor samples based on the scRNA-seq experiments described in Figs. 2 and 5.

1390   *in mice this is the normal tissue baseline, in human the control tissues come from

1391   patients with either duodenal tumors, bile duct tumors, or non-malignant pancreatic

1392   tumors that were scored by a pathologist to have "no visible inflammation"(41)

1393

1394   **Figure 6. Pan-Cancer analysis identifies LRRC15$^+$ CAFs as a frequent population**

1395   **across several human tumor types. a)** Distribution of *LRRC15* expression (Log2

1396   TPM) across indicated cancer types (TCGA, n=4,848) compared to their host tissues

1397   (GTEx, n=2,810). **b)** Top: Expression of the 14 most significantly enriched (Wilcoxon's

1398   rank sum test <1e-285) genes in TGFB CAF cluster 0 compared to cluster 1 as well as

1399   cluster 2 from figure 5d) that are expressed by less than 10 percent of the other cells in

1400   the complete PDAC single-cell dataset. Bottom: Relative average expression in CAF

1401   clusters from Figure 5d). **c)** Relative expression of genes from b) in 122 microdissected
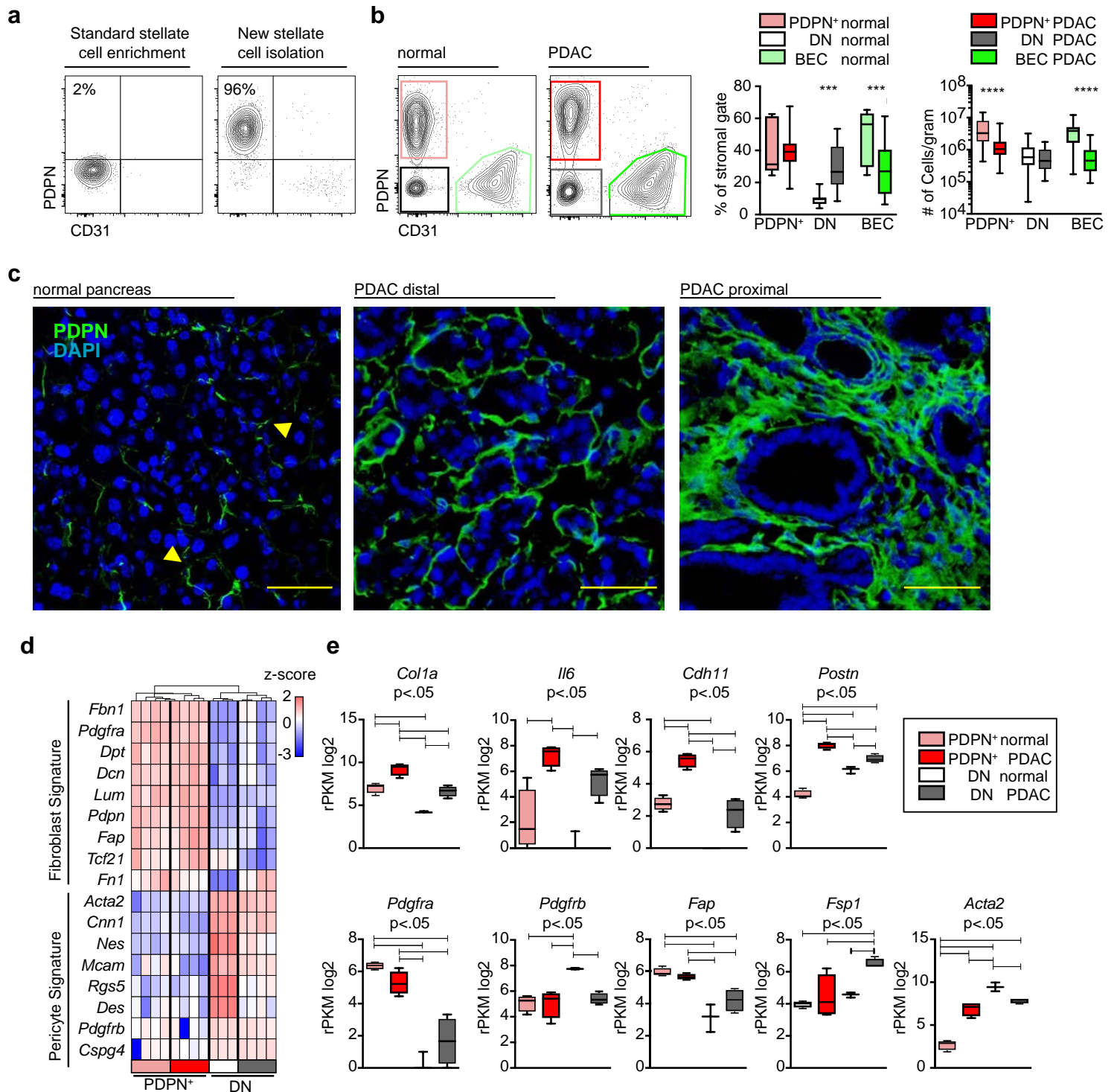
44

1402   stroma and 65 microdissected tumor samples. **d)** Top left: UMAP embedding of 3,363
1403   non-malignant cells from 18 HNSC biopsies. Cell type assignments provided by the
1404   authors are indicated by color. Top right: UMAP reduction as on the left, colored by
1405   expression (Log(CPM/10+1)) of indicated genes. Bottom left: UMAP as on top, clusters
1406   identified through graph-based clustering are indicated by color. Bottom right: Heatmap
1407   visualizing the relative average expression of indicated genes (rows) in clusters given
1408   on the bottom left. **e)** Top: Gene-by gene correlation matrix visualizing the pairwise
1409   Spearman's correlation coefficients in bulk RNA-seq TCGA data from pancreatic cancer
1410   (left, n=178) and uveal melanoma (right, n=80) patients. Bottom: Scatter plot comparing
1411   the average pairwise correlations (x-axis) and average expression (y-axis) of genes
1412   from given on top across 31 cancer indication from TCGA (Total of 9,712 samples from
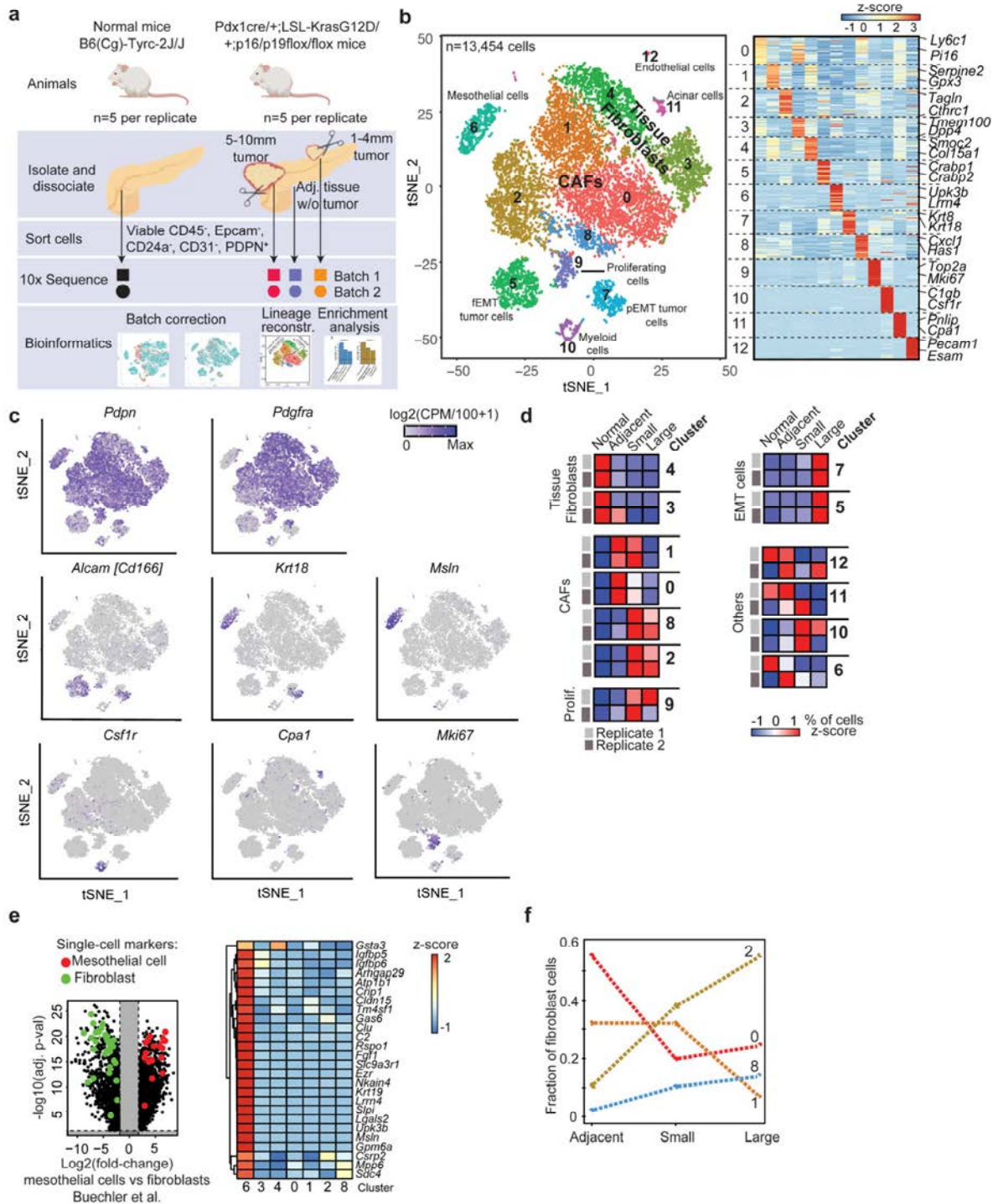1413   primary tumors, regression line in blue).

1414

1415   **Figure 7. *LRRC15* expression and its transcriptional signature predicts response**
1416   **to immunotherapy.** **a)** Boxplots comparing the distribution of the TGFB CAF (top) in
1417   excluded, inflamed and desert tumors from imvigor 210 between responders and non-
1418   responders. *** p<0.001, two-sided t-test; CR: Complete Response, PR: Partial
1419   Response, SD: Stable Disease, PD: Progressive disease **b)** Kaplan-Meier survival plot
1420   comparing survival probability (y-axis) and Follow-Up time for 134 patients with locally
1421   advanced or metastatic urothelial carcinoma (imvigor 210) receiving Atezolizumab
1422   treatment, restricted to tumors with excluded immune phenotype. Groups were split by
1423   high (red) or low (green) levels of TGFB CAF marker gene signature expression
1424   (median cutoff). **c)** Boxplots comparing the distributions of pairwise correlations of
1425   genes from the TGFB CAF and the F-TBRS signature in imvigor210 bulk RNA-seq data.
1426   **d)** Survival plot as in b), but here with a score based on the F-TBRS signature genes. **e)**
1427   Kaplan-Meier survival plot comparing survival probability (y-axis) and Follow-Up time for
1428   128 patients from the pcd trial receiving Atezolizumab treatment. Groups were split by
1429   high (red) or low (green) levels of TGFB CAF marker gene signature expression
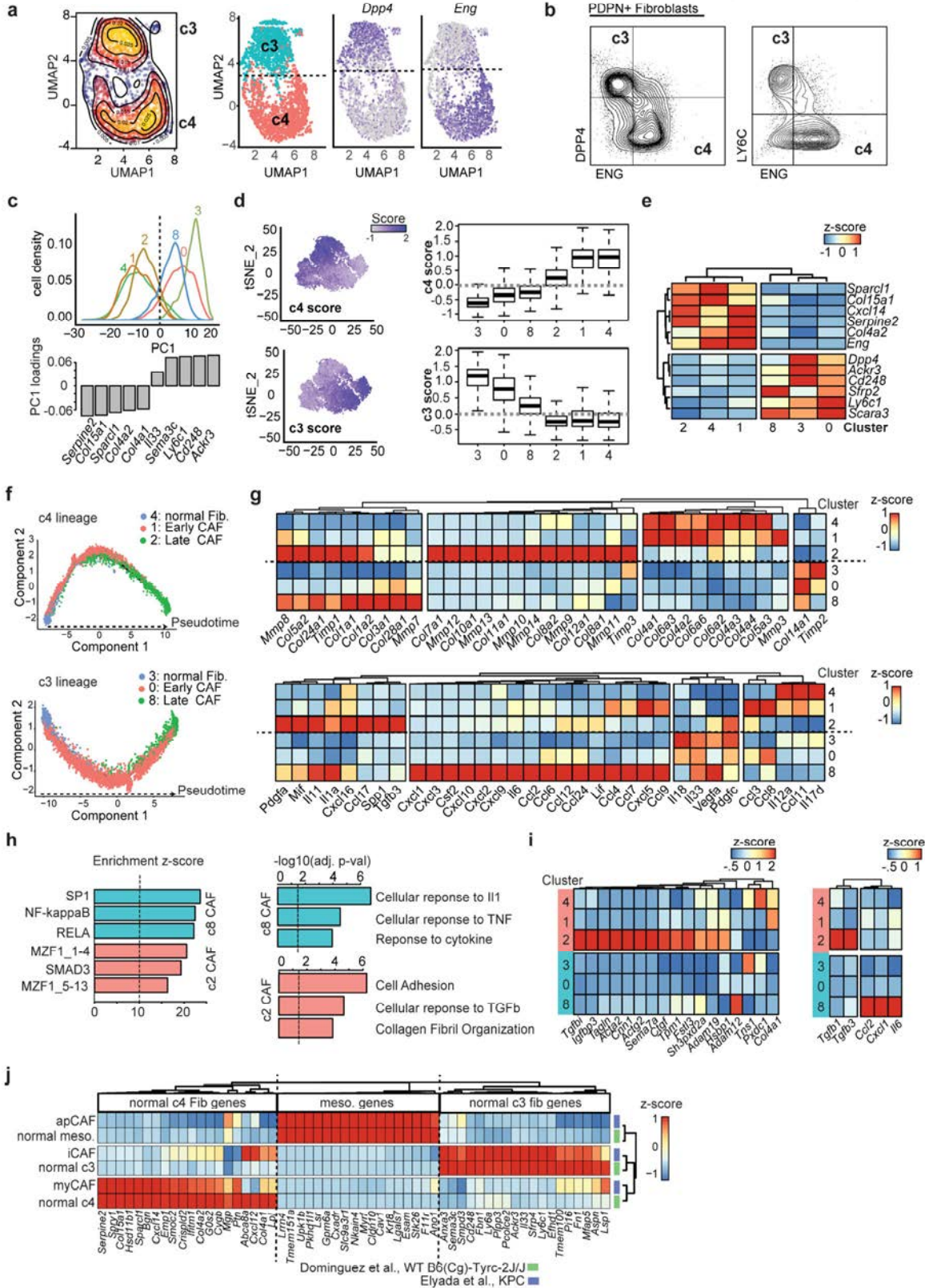1430   (>upper vs <lower quartile).

1431

Dominguez, et al., Figure 2

Dominguez, et al., Figure 3

Dominguez, et al., Figure 4

a

x-axis: Log FC early PDAC vs healthy
y-axis: Log FC late PDAC vs healthy

Lrrc15

- Sign. different both comparisons
- Sign. different only late PDAC
- Sign. different only early PDAC

b

KPP GEMM

DAPI PDPN EPCAM LRRC15

c

KPP Sub. Cutaneous

DAPI PDPN EPCAM LRRC15

d

subcutaneous KPP, gated on PDPN+ CAF

PDPN / isotype

PDPN / LRRC15

for all p<.05

% Lrrc15+ cells over is otype

PDPN CAF, DN, BEC, CD45, Tumor

for all p<.005

# of Lrrc15+ cells/g

PDPN CAF, DN, BEC, CD45, Tumor

e

day12

KPP alone | LRRC15+ CAF + KPP | Ly6C+ + KPP | c3 fibroblast +KPP | c4 fibroblast +KPP

x-axis: Days in coculture
y-axis: Total area um^2

****

- tumor KPP
- LRRC15+ c2 CAF + KPP
- Ly6c+ CAF c0/c3 + KPP
- c3 fibroblast + KPP
- c4 fibroblast + KPP

Dominguez et al, Figure 6

AA**C**R American Association
for Cancer Research
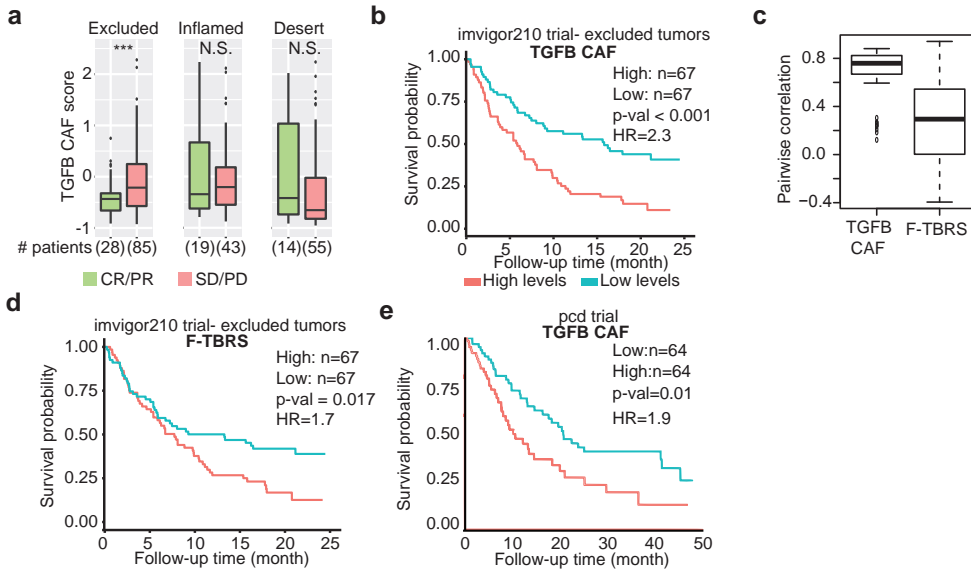
CANCER DISCOVERY

Single-cell RNA sequencing reveals stromal evolution into LRRC15+ myofibroblasts as a determinant of patient response to cancer immunotherapy

Claudia X Dominguez, Soren Muller, Shilpa Keerthivasan, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/2159-8290.CD-19-0644
Supplementary Material	Access the most recent supplemental material at: http://cancerdiscovery.aacrjournals.org/content/suppl/2019/11/07/2159-8290.CD-19-0644.DC1
Author Manuscript	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerdiscovery.aacrjournals.org/content/early/2019/12/13/2159-8290.CD-19-0644. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.