



Published in final edited form as:

Nat Genet. 2015 October ; 47(10): 1168–1178. doi:10.1038/ng.3398.

Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma

Richard A. Moffitt¹, Raoud Marayati¹, Elizabeth L. Flate¹, Keith E. Volmar², S. Gabriela Herrera Loeza¹, Katherine A. Hoadley^{1,3}, Naim U. Rashid¹, Lindsay A. Williams^{1,4}, Samuel C. Eaton⁵, Alexander H. Chung⁵, Jadwiga K. Smyla¹, Judy M. Anderson⁶, Hong Jin Kim^{1,7}, David J. Bentrem⁸, Mark S. Talamonti⁹, Christine A. Iacobuzio-Donahue¹⁰, Michael A. Hollingsworth⁶, and Jen Jen Yeh^{1,5,7}

¹Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC

²University of North Carolina-Rex Healthcare, Chapel Hill, NC

³Department of Genetics, University of North Carolina, Chapel Hill, NC

⁴Department of Epidemiology, University of North Carolina, Chapel Hill, NC

⁵Department of Pharmacology, University of North Carolina, Chapel Hill, NC

⁶Eppley Cancer Institute, University of Nebraska, Lincoln, NE

⁷Department of Surgery, University of North Carolina, Chapel Hill, NC

⁸Department of Surgery and Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine Northwestern University, Chicago, IL

⁹Department of Surgery, NorthShore University HealthSystem, Evanston, IL

¹⁰Department of Pathology, David Rubenstein Center for Pancreatic Cancer Research, Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, NY

Abstract

Pancreatic ductal adenocarcinoma (PDAC) remains a lethal disease with a 5-year survival of 4%. A key hallmark of PDAC is extensive stromal involvement, which makes capturing precise tumor-specific molecular information difficult. Here, we have overcome this problem by applying blind source separation to a diverse collection of PDAC gene expression microarray data, which includes primary, metastatic, and normal samples. By digitally separating tumor, stroma, and normal gene expression, we have identified and validated two tumor-specific subtypes including a

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence to Jen Jen Yeh, jen_jen_yeh@med.unc.edu.

URLs: ICGC data portal: <http://dcc.icgc.org/>

TCGA data portal: <https://tcga-data.nci.nih.gov/tcga>

Accession codes: Expression data have been uploaded to GEO.

Author Contributions: Study design and manuscript writing: R.A.M. and J.J.Y.; Performed experiments and collected data: R.A.M., R.M., E.F., S.H.L., L.W., S.E., A.C., J.S.; Pathologist: K.V.; Data analysis: R.A.M., R.M., J.J.Y., K.H., N.R.; Provided samples: J.A., H.J.K., D.B., M.T., C.I-D, M.H., J.J.Y; Project oversight: J.J.Y.

“basal-like” subtype which has worse outcome, and is molecularly similar to basal tumors in bladder and breast cancer. Furthermore, we define “normal” and “activated” stromal subtypes which are independently prognostic. Our results provide new insight into the molecular composition of PDAC which may be used to tailor therapies or provide decision support in a clinical setting where the choice and timing of therapies is critical.

Rigorous sequencing studies have shown that few genetic alterations (*KRAS*, *CDKN2A*, *SMAD4*, and *TP53*) are prevalent in PDAC¹⁻³, but these and other analyses of PDAC tumors are hampered by limited tumor cellularity and the presence of abundant stroma intermixed with normal endocrine and exocrine cells. Additionally, metastatic samples often include cell types from the host organ. Thus, PDAC tumors are complex mixtures in which the malignant epithelial cells often represent a minority of the tissue compartment (Fig. 1a). Illustrating this limitation, an important study of exome and copy number in pancreatic cancer removed 43 of 142 patients due to low tumor cellularity affecting the sensitivity of mutation detection¹. While some studies use microdissection to improve tumor content⁴⁻⁷, microdissection for precision medicine approaches is not yet feasible. When considering gene expression of bulk tumors, normal pancreas and PDAC tissues often cluster together, separate from cell lines which are assumed to be purely neoplastic⁸.

Separating molecular signatures of tissue compartments from the measurements of bulk tumor belongs to the general class of problems called blind source separation. Previous studies have used samples of chronic pancreatitis, which is often fibrotic, to control for the presence of desmoplastic stroma in tumor samples⁹. In prostate cancer, Stuart et al. have used pathology assessments of cell types to train models of gene expression signatures of tumor, stroma, and normal tissue¹⁰. In a follow up study, they used their learned gene lists for *in silico* estimation of tissue components in a larger set of data¹¹. A similar approach has also been used to quantify stromal content across multiple data sets from the cancer genome atlas (TCGA)¹². Among source separation techniques, nonnegative matrix factorization (NMF) is especially well suited for biological data, because it constrains all sources to be positive in nature, reflecting the goal of identifying positive gene expression exemplars, rather than pairwise differences between tissue types. Briefly, we define NMF as modeling the matrix \mathbf{X} of expression for g genes and s samples, as the product of a matrix \mathbf{G} of g gene weights for k factors and a matrix \mathbf{S} of s sample weights for k factors. Alexandrov et al. have recently demonstrated that NMF is useful for a similar problem of identifying mutational signatures from the aggregate list of somatic mutations in human cancer samples^{13,14}. Similarly, Biton et al. have applied a related technique, Independent component analysis, to examine gene expression in bladder cancer¹⁵.

In this study, we have overcome the challenges of bulk tumor analysis where signal is averaged out between normal, tumor and stroma compartments, by using NMF to perform a virtual microdissection of primary and metastatic PDAC samples. This has allowed us to identify tumor-specific and stroma-specific subtypes with prognostic and biologic relevance. In addition, by focusing on tumor autonomous gene expression, we found that intra-patient tumor heterogeneity between primary and metastatic sites was unexpectedly low.

Results

Virtual microdissection of PDAC

We used NMF to analyze gene expression in a cohort of microarray data from 145 primary and 61 metastatic PDAC tumors, 17 cell lines, 46 pancreas and 88 distant site adjacent normal samples using Agilent (Agilent Technologies) human whole genome 4x44K DNA microarrays (106 primary tumors were previously used in a separate bulk analysis of gene expression (GSE21501¹⁶). To validate our findings, RNA sequencing was performed on 15 primary tumors, 37 pancreatic cancer patient-derived xenografts (PDX), 3 cell lines, and 6 cancer associated fibroblast (CAF) lines derived from deidentified patients with pancreatic cancer. Histology of all available samples was reviewed by a single blinded pathologist (KEV). Table 1 summarizes the demographic and clinical characteristics of patients in our cohorts.

NMF distinguishes normal and tumor compartments

A key obstacle in the analysis of gene expression data, particularly in PDAC, is the removal of confounding normal or stroma gene expression from local and distant organ sites. Supplementary Figure 1 shows example histology of samples with both tumor, normal, and stromal tissue. We used NMF to identify gene expression which we attribute to normal pancreas, liver, lung, muscle, and immune tissues. Expression of exemplar genes from these factors, i.e. genes with distinctly large weights in a single column of **G**, as well as factor weights for the samples, i.e. rows of **S**, showed excellent agreement with known tissue labels (Fig. 1b, c, Supplementary Fig. 2). Investigation of the exemplar genes from these factors further confirmed their role as confounding normal tissue. For example, using the Kolmogorov–Smirnov test, the top-weighted genes from the liver factor show significant ($p < 10^{-10}$) enrichment in the MSigDB term SU_LIVER (Supplementary table 1), and the highest weighted gene, fibrinogen beta (*FGB*) (Supplementary table 2), is specifically expressed in normal human liver tissue.

In addition to normal tissue from distant organs, we found two factors which were exclusive to pancreas tissue, but were differentiated from each other by their respective gene lists. One factor described endocrine function including expression of glucagon and insulin (*GCG* and *INS*) (Supplementary table 2), while the other factor described exocrine function including expression of digestive enzyme genes such as pancreatic lipase, *PNLIP* (Supplementary table 2). This unsupervised discovery of two molecularly distinct yet highly co-localized factors related to normal pancreatic function represents an important proof of concept in the use of NMF to identify novel features without pre-defined expression knowledge.

To validate our normal expression signatures, all available samples were reviewed by a single pathologist (KEV) to independently assess the amount of tumor, normal, and stroma cellularity. We found that many factor weights were correlated or anti-correlated to tumor cellularity (Supplementary Fig. 3). Among normal and metastatic liver samples, for example, tumor-specific factor weights were correlated with cellularity whereas the normal-specific liver factor weight was inversely related to the tumor content of a sample (Fig. 1d).

These findings support our hypothesis that factor weights obtained from NMF are quantitatively indicative of underlying sample composition.

Identification of stroma-specific subtypes

Stroma is particularly important in PDAC. According to pathology assessments, stroma varies (Supplementary Fig. 1 c-e), and comprises on average 48% of our primary tumor samples with a standard deviation of 30%. Our analysis identified two factors, which describe gene expression from the stroma, which were distinctly different from the normal factors shown in Figure 1. Consensus clustering on exemplar genes from these two stroma factors divided tumor samples into two stromal subtypes, which we classified as “normal”, and “activated” (Fig. 2a). Patients with samples with an activated stroma subtype had worse median survival of 15 months and 60% 1-year survival when compared to patients with a normal stroma subtype (median 24 months, 1-year 82%, Fig. 2b). Both were notably absent in PDAC cell lines (Fig. 2c), which exhibited a distinct mitotic expression signature associated with mitotic checkpoints and DNA replication (Supplementary table 1)¹⁷. The fact that cell lines do not express these stromal factors, and that many metastatic samples express them at low levels (Supplementary Fig. 4), suggest that these genes are not expressed by the tumor epithelium. To further validate the stromal origin of these gene expression signatures, we isolated 6 CAF lines from primary tumors (Supplementary Fig. 5), and found that they robustly overexpressed our hypothesized stromal signatures compared to PDAC tumor cell lines which had no expression of the stromal signatures (Fig. 2c).

The vast majority of collagen gene expression was attributable to stromal compartments, with the lone exception being *COL17A1*, which was high in tumors. “Normal” stroma was characterized by relatively high expression of known markers for pancreatic stellate cells, smooth muscle actin, vimentin, and desmin, (*ACTA2*, *VIM*, and *DES*). Stellate cells have been shown to promote cancer cell survival *in vitro*¹⁸, but at the same time may restrain PDAC in mouse models^{19,20}. Targeting desmoplastic stroma by hedgehog pathway inhibition has been shown to both accelerate the development of disease²¹ and enhance delivery of chemotherapy²². In patients, the ratio of smooth muscle actin stained area to the collagen-stained area has been shown to be predictive of poor outcomes²³. We found that “activated” stroma was characterized by a more diverse set of genes associated with macrophages, such as the integrin *ITGAM*, and the chemokine ligands *CCL13* and *CCL18*. “Activated” stroma also expressed other genes which point to its role in tumor promotion, including the secreted protein *SPARC*, WNT family members *WNT2*, and *WNT5A*, gelatinase B (*MMP9*), and stromelysin 3 (*MMP11*). The presence of fibroblast activation protein (*FAP*), which has previously been related to worse prognosis, in the activated stroma suggests that an activated fibroblast state may be partially responsible for the poor outcomes for these patients²⁴. This lead us to hypothesize that the “normal” stroma factor may describe a “good” version of stroma and that “activated” stroma factor may describe the activated inflammatory stromal response that has been seen in previous studies to be responsible for disease progression²⁵⁻²⁷. Our factor analysis supports a complex, multi-gene model of stroma in PDAC, which may explain why single gene analysis has yielded mixed results.

Identification of tumor-specific subtypes

Independent of normal and stromal factors, we found that two tumor-specific factors define “classical” and “basal-like” subtypes of PDAC. When our samples were split into the two tumor subtypes (Fig. 3a), patients with basal-like subtype tumors had an overall worse median survival of 11 months and 44% 1-year survival compared to 19 months and 70% 1-year survival for those with classical subtype tumors ($p=0.007$, Fig. 3b). All cell lines assayed in this study ($p < 0.001$), as well as a majority of metastatic samples ($p = 0.002$), were classified as “basal-like”, suggesting that cell line models represent only one subset of PDAC. These subtypes as well as their prognostic value were independently validated within the recently published International Cancer Genome Consortium (ICGC) PDAC microarray data set (Fig. 3c, d)²⁸. Genes from the “basal-like” factor, including laminins and keratins, were also consistent with basal subtypes previously defined in bladder^{14,28,29} and breast³⁰ cancers (Fig. 3e-h). Interestingly, genes from our “basal-like” subtype reproduced subtype calls ($p<0.001$) in breast cancer, had prognostic value in breast cancer samples ($p<0.001$) and reproduced previous subtype calls in bladder cancer ($p<0.001$). Given these promising results, we developed a single-sample cross-platform classifier of basal-like subtype which was trained on our microarray, TCGA bladder, and Perou breast cancer data, with a 93% cross validation accuracy, and was able to classify TCGA breast cancer data with 92% accuracy during external validation (Supplementary Fig. 6)

Potential subtypes of PDAC have previously been described by Collisson et al.⁴. We used the published exemplar genes for “exocrine-like”, “classical”, and “quasimesenchymal” subtypes to cluster normal pancreas, cell lines, and primary PDAC tumors from our cohort (Supplementary Fig. 7a). The three previous classifications were also observed in our data, but did not hold prognostic power either by cluster label or by supervised classification with PAM³¹(Supplementary Fig. 7b). Furthermore, inclusion of the Collisson et al. subtypes into a multivariate Cox regression with our proposed tumor subtypes did not remove the predictive power of our subtyping ($p = 0.014$). By cross-referencing the genes from Collisson et al.'s model with our NMF model, we observed three key findings. First, “exocrine-like” genes overlapped with genes from our exocrine pancreas factor (17/17). Tumors in this cluster had expression indistinguishable from adjacent normal samples from our data set. Second, Collisson et al.'s “classical” genes overlapped with our “classical” subtype genes (20/22), for which we retain the naming convention “classical”. Third, the gene set associated with “quasimesenchymal” subtype appeared to be a mixed collection of genes from our “basal-like” tumor (6/20) and stromal subtypes (6/20). Thus, the appearance of stromal factors in the Collisson et al. list of “quasimesenchymal” class genes may explain the apparent mesenchymal-like gene expression that was observed.

“Basal-like” and “classical” tumors were found within both “normal” and “activated” stroma subtypes (Fig. 4a). Differential prognosis among tumor and stroma subtypes was cumulative, as “classical” subtype tumors with “normal” stroma subtypes ($n = 24$) had the lowest hazard ratio of 0.39, with a 95% CI of [0.21, 0.73], while the “basal-like” subtype tumors with “activated” stroma subtypes ($n = 26$) had the highest hazard ratio of 2.28 with a 95% CI of [1.34, 3.87] (Fig. 4b). In a multivariate Cox regression model which included tumor subtypes, stromal subtypes, and clinical variables (gender, race, T stage, N stage, margin

status, adjuvant therapy, histological grade, and age), both classifications were independently associated with survival (stroma subtypes: $p = 0.037$, tumor subtypes: $p = 0.003$).

Although basal-like subtype tumors have a worse prognosis, patients with basal-like subtype tumors showed a strong trend towards better response to adjuvant therapy ($p=0.072$, Fig. 4c,d). Among basal-like subtype patients, adjuvant therapy provided a hazard ratio of 0.38, (95% CI of [0.14, 1.09]), while in patients with classical subtype tumors, adjuvant therapy is associated with a hazard ratio of only 0.76 (95% CI [0.40, 1.43]). In our cohort, there was no association of most clinical variables (race, gender, T stage, N stage, differentiation, tumor cellularity) with survival, although positive nodal status trended towards significance, and positive margin status was significantly associated with worse survival (Table 1). Table 2 shows two-way associations of all subtype calls with clinical and pathological information from our cohort of PDAC patients. We found no association of tumor or stroma subtype with standard clinical or pathological variables, with the notable exception of mucinous features.

Tumor-specific subtypes in patient-derived xenografts

To assess the tumor or stromal specificity of our signatures, we performed RNAseq on a group of 37 PDX tumors. PDX tumors are composed of human tumor cells surrounded by mouse stroma (Supplementary Fig. 8)²⁹. Genes from both of our tumor signatures were expressed as human transcripts, whereas genes from both of our stromal signatures were expressed as mouse transcripts (Fig. 2d, Supplementary Fig. 9a). Furthermore, we found that PDX RNAseq expression divided PDX into both classical and basal-like groupings (Supplementary Fig. 9b) while predominantly expressing an activated stromal signature (Fig. 2d). We found that while tumor-specific subtype was not predictive of graft success (Fig. 5a), patient tumors with an activated stroma subtype had significantly higher graft success rates than those with normal stroma subtype or low amounts of stroma (Fig. 5b) ($p=0.019$). Basal-like subtype tumors also exhibited faster growth rates than classical tumors ($p=0.032$) as measured by the length of time that tumors took to grow to 200 mm³ (TT200, Fig. 5c-d), a previously used metric for PDX growth³⁰. Retrospective analysis of patients who had matched PDX tumors found that a shorter TT200 was associated with an unfavorable recurrence-free survival ($p=0.035$, Fig. 5e), suggesting that PDX tumor growth rate may reflect patient biology.

We also measured both mouse and human-specific expression of the Collisson et al. genes in our PDX models. We found that while genes from the “classical” subtype were expressed by human cells in PDX, “quasimesenchymal” transcripts were expressed by a mixture of human and mouse cells, and “exocrine-like” transcripts were infrequently expressed (Supplementary Fig. 7c). This supports our hypothesis that while the “classical” subtype is a bona fide group, the “quasimesenchymal” subtype is partially driven by non-tumor contributions of stroma and the “exocrine-like” subtype by normal pancreas.

KRAS codon mutations, tumor-specific subtypes, and race

Studies of KRAS codon mutations have demonstrated that different codon mutations may have differential functions^{31,32} and in some clinical studies, have been shown to be

associated with differential outcome. Because PDX tumors are enriched for human-specific tumor cells, we evaluated *KRAS* codon mutations in our PDX cohort using manually curated RNAseq data. While the overall frequency of *KRAS* codon mutations was similar to a recent study of PDAC⁷, we noted that the *KRAS* G12D mutation was significantly overrepresented in our basal-like subtype while G12V was isolated to the classical subtype (Figure 5f, $p=0.030$). Furthermore, we found an overrepresentation of *KRAS* G12V mutations in African Americans (Figure 5g, $p<0.001$). In contrast to basal-like breast cancers which occur most frequently in African American women and have a worse prognosis³³, African American patients in our cohort tended to have mainly classical subtype tumors (13 vs 2). We found that similar to other cancers, African Americans had a worse prognosis after adjusting for tumor subtype (Figure 4e, $p=0.017$). African American patients with classical subtype tumors had a median survival of 13 months compared to Caucasian patients with classical subtype tumors who had a median survival of 19 months.

Other commonly mutated genes and altered pathways in PDAC

Previously, loss of SMAD4 has been shown to promote tumor growth^{34,35}. We found that, similar to previous PDX studies of PDAC, loss of SMAD4 was associated with graft success in PDX models³⁶ (Fig. 5h, Supplementary Fig. 10, $p=0.044$). Furthermore, in our PDX cohort, we found that SMAD4 expression was significantly higher in classical compared to basal-like subtype PDX tumors (Fig. 5i, $p=0.015$), consistent with the observation that SMAD4 loss confers a more aggressive phenotype.

Using mutation, genomic subtype³, and gene expression²⁸ data from publically available ICGC data in which we show recapitulation of our subtypes and prognosis, we evaluated significantly mutated genes and pathways in PDAC, including ones recently identified through whole-exome sequencing of microdissected primary PDAC tumors^{1-3,7}. We found no significant associations between our expression subtypes and these mutationally altered pathways, i.e. TGF β , RB, NOTCH, CTNNB1, SWI/SNF, and DNA repair (Supplementary Fig. 11). Furthermore, we found no overlap between our subtypes and recently identified genomic subtypes, or response to platinum therapy³. Consistent with this, a recent comprehensive study of somatic mutations in PDAC long-term survivors suggested that somatic mutations alone will not be sufficient to explain clinical outcome³⁷.

Given the overlap of our classical subtype with that of Collisson et al.⁵, it was not surprising to find that our classical subtype was also enriched for genes associated with GATA6 overexpression³⁸ (Supplementary Fig. 12a, Fig 4a). GATA6 has been found to promote epithelial cell differentiation^{38,39}. This prompted us to perform a more detailed histological markers of differentiation in our samples and found that samples with greater than 10% extracellular mucin, a marker of differentiation, comprised mostly of classical subtype tumors (88.5%, $n=23$) compared to only 11.5% ($n=3$) of basal-like subtype tumors (Supplementary Fig. 13, $p=0.042$, Table 2). Consistent with the increased presence of extracellular mucin, our classical subtype was enriched for genes upregulated in mucinous ovarian cancer (Supplementary Fig. 12b, WAMAUNYOKOLI_OVARIAN CANCER_GRADES_1_2_UP⁴⁰). Interestingly, our basal-like subtype was enriched for genes related to *KRAS* activation and *STK11* loss in a lung cancer mouse model where

STK11-deficient tumors demonstrated shorter latency and more frequent metastasis⁴¹ (Supplementary Fig. 12c). We found one sample with STK11 inactivation in the ICGC data; this sample was a basal-like subtype (Supplementary Fig. 11). Notably, our subtypes were not associated with other known signaling pathways in PDAC, including Fanconi anemia, DNA repair, chromatin remodeling, beta-catenin, RB, ARF, G1 (Fig. 4a). However, all of these pathways except for beta-catenin were considerably differentially expressed in cell lines compared to patient tumors, suggesting that gene expression in cell lines may be a deceptive representation of most tumors.

Low intra-patient heterogeneity in tumor-specific genes

We expect that only a subset of genes are relevant to the question of intra- and inter-patient heterogeneity in PDAC. Many methods exist to pre-select genes for supervised analysis³², but selection of the most differentially expressed genes is a common preprocessing step during unsupervised analysis³³. When clustering matched samples of metastatic and primary lesions using the 50 most differentially expressed genes among all matched samples, samples separated primarily by organ site instead of by patient (Fig. 6a, c). In contrast, when considering 25 top ranked exemplar genes each from the “basal-like” and “classical” factors, samples from the same patient clustered closer together, and were less dependent of organ site (Fig. 6b, d). This is further illustrated in a focused analysis of two patients (Fig. 6), whose tumor samples appear patient-specific when considering our tumor subtype gene list, but cluster by site when considering differentially expressed genes. Overall, we found that our tumor subtype gene list showed higher similarity (mean Pearson's $\rho=0.53$) between all other samples from the same patient than did the differentially expressed gene list ($\rho=0.32$, t -test $p<0.001$). Furthermore, our tumor subtype gene list produced much lower similarity among all other samples from the same organ site across different patients ($\rho = 0.04$) than the differentially expressed gene list ($\rho=0.34$, $p<0.001$). This observed similarity of tumor gene expression among tumors within the same patient suggests overall high inter-patient tumor heterogeneity and low heterogeneity between primary and metastatic sites. However, we did observe examples of intra-patient heterogeneity between metastatic sites. For example, lung metastases, even those from patients with “basal-like” tumors in other locations, clustered exclusively with the “classical” tumors, suggesting that some intra-patient heterogeneity may exist among metastatic sites, and supporting the previously reported divergent patterns of failure in PDAC³⁴.

Discussion

Our study represents the largest investigation of primary and metastatic PDAC gene expression to date. We have used NMF to identify novel prognostic subtypes of PDAC which may have been previously obscured by confounding normal and stromal tissue. Our identification of normal-, tumor-, and stroma-specific gene expression signatures is supported by both their overlap with previously identified gene lists and their expression in appropriate tissue types. Our tumor subtypes are further supported by their relationship to previously identified basal tumor subtypes in breast and bladder cancers and their prognostic relevance in external cohorts. Our findings of two different stroma subtypes may help explain the differential effects of stroma previously seen in preclinical models.

Tumor and stroma specific gene expression classified PDAC into four distinct subtypes with prognostic relevance. The orthogonal nature of tumor- and stroma-specific subtypes suggest an important interplay in patient tumors that will need to be taken into account as stroma and immune modulating therapies are studied. In our cohort, patients with basal-like tumors appeared to derive more benefit from adjuvant therapy. Whether basal-like and classical subtypes may be associated with response to specific therapies will need to be studied further as more effective therapies become available. One challenge will be defining preclinical model systems that recapitulate these subtypes as our results suggest that traditional cell lines are lacking in the classical subtype. Although we demonstrate that PDX models recapitulate tumor-specific subtypes, these models alone may not be sufficient due to either the lack of human stroma or overrepresentation of the activated stroma subtype in the tumors that are successfully grafted. Thus more detailed characterization of genetically engineered mouse models of PDAC models will be needed to determine which models best reflect both our tumor- and stroma-specific subtypes.

Recent exome sequencing studies have confirmed commonly mutated genes in PDAC but have not uncovered mutations that clearly confer survival differences^{2,3,7}. In fact, exome sequencing of a cohort of very long-term survivors of PDAC³⁷ found no differences in somatic mutations to explain the improved biology of tumors from these rare patients compared to the majority of patients with PDAC, suggesting that examining somatic mutations alone may not be sufficient to understand the biological and clinical differences in PDAC tumors. Furthermore exome sequencing studies and studies of microdissected samples are limited to the tumor compartment and overlook the stroma compartment which has been shown to be biologically critical in PDAC, with both tumor-promoting and tumor-inhibiting effects. Our results suggest that RNA subtypes may better capture the molecular landscape of PDAC and its reflection on patient outcome. We hypothesize that our RNA subtypes may reflect the broad effect of somatic mutations while also capturing the importance of the neoplastic stroma.

These results provide new insight into the molecular composition of PDAC which may be used for precision medicine. Furthermore, knowledge of these subtypes and their prognostic value may provide decision support in a clinical setting where the choice and timing of therapies is critical.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was partially supported by CA 14024 (JJY), the University Cancer Research Fund (JJY) and UNC Lineberger Comprehensive Cancer Center Postdoctoral Training Grant T32-CA009156 (RAM). We thank the UNC Tissue Procurement, Translational Pathology Laboratory, Animal Studies, Animal Histopathology and Center for Gastrointestinal Biology and Disease Histopathology core facilities, the PDX Program, and the Department of Pharmacology for tremendous technical support. We thank the patients and their families who generously donated their samples to research and in particular to the University of Nebraska Medical Center Rapid Autopsy Pancreatic Program and the Johns Hopkins Gastrointestinal Cancer Rapid Medical Donation Program.

References

1. Biankin AV, et al. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*. 2012; 491:399–405. [PubMed: 23103869]
2. Jones S, et al. Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. *Science*. 2008; 321:1801–1806. [PubMed: 18772397]
3. Waddell N, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*. 2015; 518:495–501. [PubMed: 25719666]
4. Yachida S, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*. 2010; 467:1114–1117. [PubMed: 20981102]
5. Collisson EA, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature medicine*. 2011; 17:500–503.
6. Crnogorac-Jurcevic T, et al. Expression profiling of microdissected pancreatic adenocarcinomas. *Oncogene*. 2002; 21:4587–4594. [PubMed: 12085237]
7. Witkiewicz AK, et al. Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nature communications*. 2015; 6
8. Iacobuzio-Donahue CA, et al. Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays. *The American journal of pathology*. 2003; 162:1151–1162. [PubMed: 12651607]
9. Logsdon CD, et al. Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer. *Cancer Research*. 2003; 63:2649–2657. [PubMed: 12750293]
10. Stuart RO, et al. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101:615–620. [PubMed: 14722351]
11. Wang Y, et al. In silico estimates of tissue components in surgical samples based on expression profiling data. *Cancer research*. 2010; 70:6448–6455. [PubMed: 20663908]
12. Yoshihara K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications*. 2013; 4
13. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell reports*. 2013; 3:246–259. [PubMed: 23318258]
14. Alexandrov LB, et al. Signatures of mutational processes in human cancer. *Nature*. 2013; 500:415–421. [PubMed: 23945592]
15. Biton A, et al. Independent Component Analysis Uncovers the Landscape of the Bladder Tumor Transcriptome and Reveals Insights into Luminal and Basal Subtypes. *Cell reports*. 2014; 9:1235–1245. [PubMed: 25456126]
16. Stratford JK, et al. A six-gene signature predicts survival of patients with localized pancreatic ductal adenocarcinoma. *PLoS medicine*. 2010; 7:e1000307. [PubMed: 20644708]
17. Whitfield ML, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Molecular biology of the cell*. 2002; 13:1977–2000. [PubMed: 12058064]
18. Froeling FE, et al. Retinoic Acid-Induced Pancreatic Stellate Cell Quiescence Reduces Paracrine Wnt- β -Catenin Signaling to Slow Tumor Progression. *Gastroenterology*. 2011; 141:1486–1497. e14. [PubMed: 21704588]
19. Özdemir BC, et al. Depletion of Carcinoma-Associated Fibroblasts and Fibrosis Induces Immunosuppression and Accelerates Pancreas Cancer with Reduced Survival. *Cancer Cell*. 2014; 25:719–734. [PubMed: 24856586]
20. Rhim AD, et al. Stromal elements act to restrain, rather than support, pancreatic ductal adenocarcinoma. *Cancer cell*. 2014; 25:735–747. [PubMed: 24856585]
21. Lee JJ, et al. Stromal response to Hedgehog signaling restrains pancreatic cancer progression. *Proceedings of the National Academy of Sciences*. 2014; 111:E3091–E3100.
22. Olive KP, et al. Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science*. 2009; 324:1457–1461. [PubMed: 19460966]

23. Erkan M, et al. The activated stroma index is a novel and independent prognostic marker in pancreatic ductal adenocarcinoma. *Clinical Gastroenterology and Hepatology*. 2008; 6:1155–1161. [PubMed: 18639493]
24. Cohen SJ, et al. Fibroblast activation protein and its relationship to clinical outcome in pancreatic adenocarcinoma. *Pancreas*. 2008; 37:154–158. [PubMed: 18665076]
25. Hwang RF, et al. Cancer-associated stromal fibroblasts promote pancreatic tumor progression. *Cancer research*. 2008; 68:918–926. [PubMed: 18245495]
26. Vonlaufen A, et al. Pancreatic stellate cells: partners in crime with pancreatic cancer cells. *Cancer research*. 2008; 68:2085–2093. [PubMed: 18381413]
27. Herrera M, et al. Functional heterogeneity of cancer-associated fibroblasts from human colon tumors shows specific prognostic gene expression signature. *Clinical Cancer Research*. 2013; 19:5914–5926. [PubMed: 24052018]
28. Nones K, et al. Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. *International Journal of Cancer*. 2014; 135:1110–1118. [PubMed: 24500968]
29. Isella C, et al. Stromal contribution to the colorectal cancer transcriptome. *Nature genetics*. 2015; 47:312–319. [PubMed: 25706627]
30. Rubio-Viqueira B, et al. An in vivo platform for translational drug development in pancreatic cancer. *Clinical Cancer Research*. 2006; 12:4652–4661. [PubMed: 16899615]
31. Stolze B, Reinhart S, Bullinger L, Fröhling S, Scholl C. Comparative analysis of KRAS codon 12, 13, 18, 61, and 117 mutations using human MCF10A isogenic cell lines. *Scientific reports*. 2015; 5
32. Ihle NT, et al. Effect of KRAS oncogene substitutions on protein behavior: implications for signaling and clinical outcome. *Journal of the National Cancer Institute*. 2012; 104:228–239. [PubMed: 22247021]
33. Carey L, Winer E, Viale G, Cameron D, Gianni L. Triple-negative breast cancer: disease entity or title of convenience? *Nat Rev Clin Oncol*. 2010; 7:683–92. [PubMed: 20877296]
34. Bardeesy N, et al. Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. *Genes Dev*. 2006; 20:3130–46. [PubMed: 17114584]
35. Haeger SM, et al. Smad4 loss promotes lung cancer formation but increases sensitivity to DNA topoisomerase inhibitors. *Oncogene*. 2015
36. Garrido-Laguna I, et al. Tumor engraftment in nude mice and enrichment in stroma-related gene pathways predict poor survival and resistance to gemcitabine in patients with pancreatic cancer. *Clin Cancer Res*. 2011; 17:5793–800. [PubMed: 21742805]
37. Dal Molin M, et al. Very Long-term Survival Following Resection for Pancreatic Cancer Is Not Explained by Commonly Mutated Genes: Results of Whole-Exome Sequencing Analysis. *Clin Cancer Res*. 2015; 21:1944–50. [PubMed: 25623214]
38. Zhang Y, et al. A Gata6-Wnt pathway required for epithelial stem cell development and airway regeneration. *Nat Genet*. 2008; 40:862–70. [PubMed: 18536717]
39. Zhong Y, et al. GATA6 activates Wnt signaling in pancreatic cancer by negatively regulating the Wnt antagonist Dickkopf-1. *PLoS One*. 2015; 6:e22129. [PubMed: 21811562]
40. Wamunyokoli FW, et al. Expression profiling of mucinous tumors of the ovary identifies genes of clinicopathologic importance. *Clin Cancer Res*. 2006; 12:690–700. [PubMed: 16467078]
41. Ji H, et al. LKB1 modulates lung cancer differentiation and metastasis. *Nature*. 2007; 448:807–10. [PubMed: 17676035]
42. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:15545–15550. [PubMed: 16199517]
43. Neel NF, et al. Response to MLN8237 in pancreatic cancer is not dependent on RalA phosphorylation. *Molecular cancer therapeutics*. 2014; 13:122–133. [PubMed: 24222664]
44. Bachem MG, et al. Pancreatic carcinoma cells induce fibrosis by stimulating proliferation and matrix synthesis of stellate cells. *Gastroenterology*. 2005; 128:907–921. [PubMed: 15825074]

45. Conway T, et al. Xenome—a tool for classifying reads from xenograft samples. *Bioinformatics*. 2012; 28:i172–i178. [PubMed: 22689758]
46. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013; 14:R36. [PubMed: 23618408]
47. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*. 2012; 7:562–578. [PubMed: 22383036]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

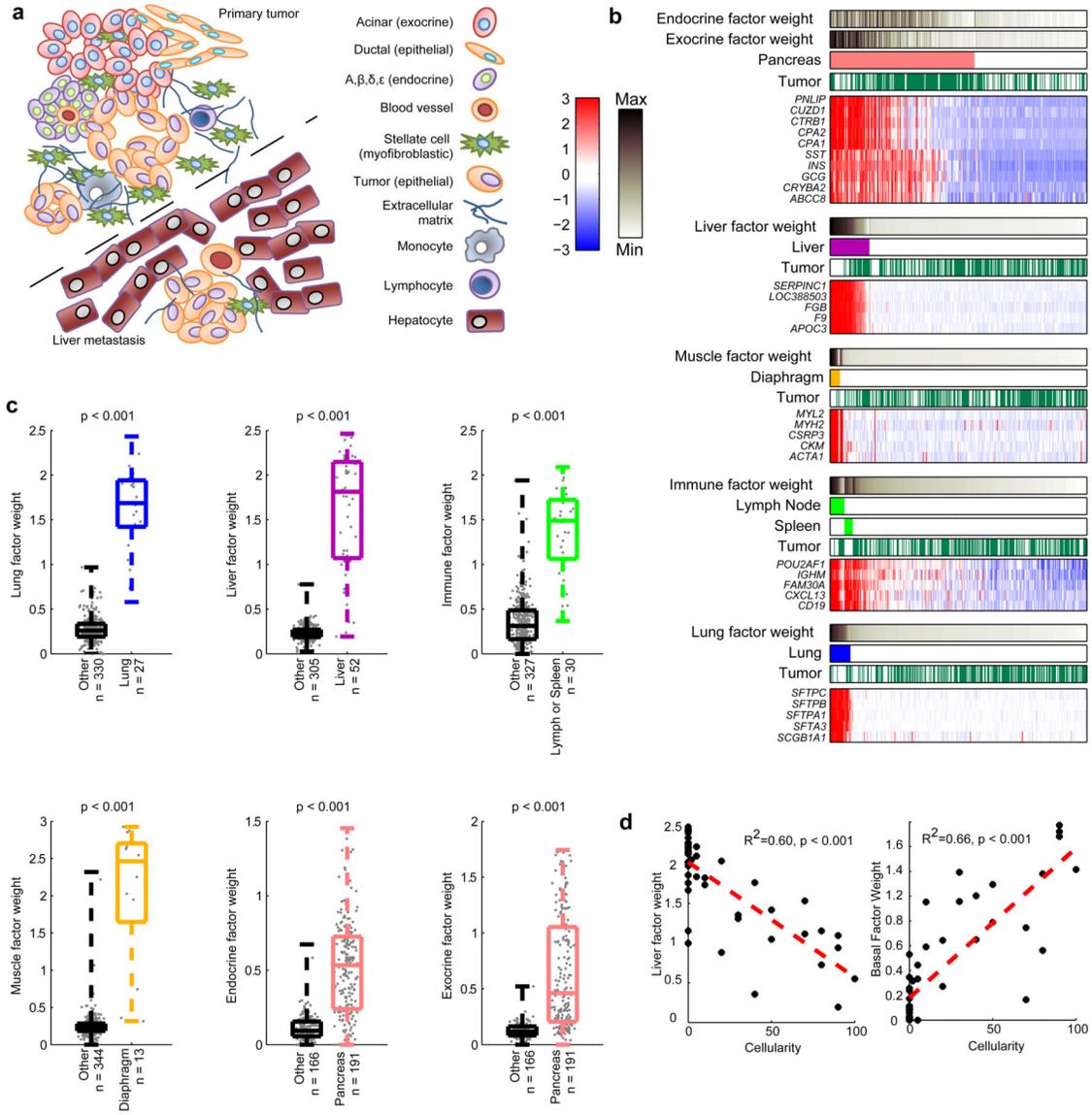


Figure 1. Successful Deconvolution of Normal Tissue with NMF. **(a)** Cartoon depicting the major cell types in primary tumor and liver metastasis samples. **(b)** (above) Overlap of sample types (solid colors) with factor weights (grayscale heat maps), and (below) heat maps of five exemplar genes for all tumors and adjacent normal tissues. Gene expression shown in the heat map has been Z-normalized. **(c)** Box and whiskers plots showing median, quartiles, and range comparing NMF factor weights across tissue types and corresponding t-test result. **(d)** Percent tumor cellularity versus NMF liver factor weight, and NMF basal tumor factor weight for metastases to the liver and adjacent liver samples. Linear regression lines are shown in red along with corresponding statistics.

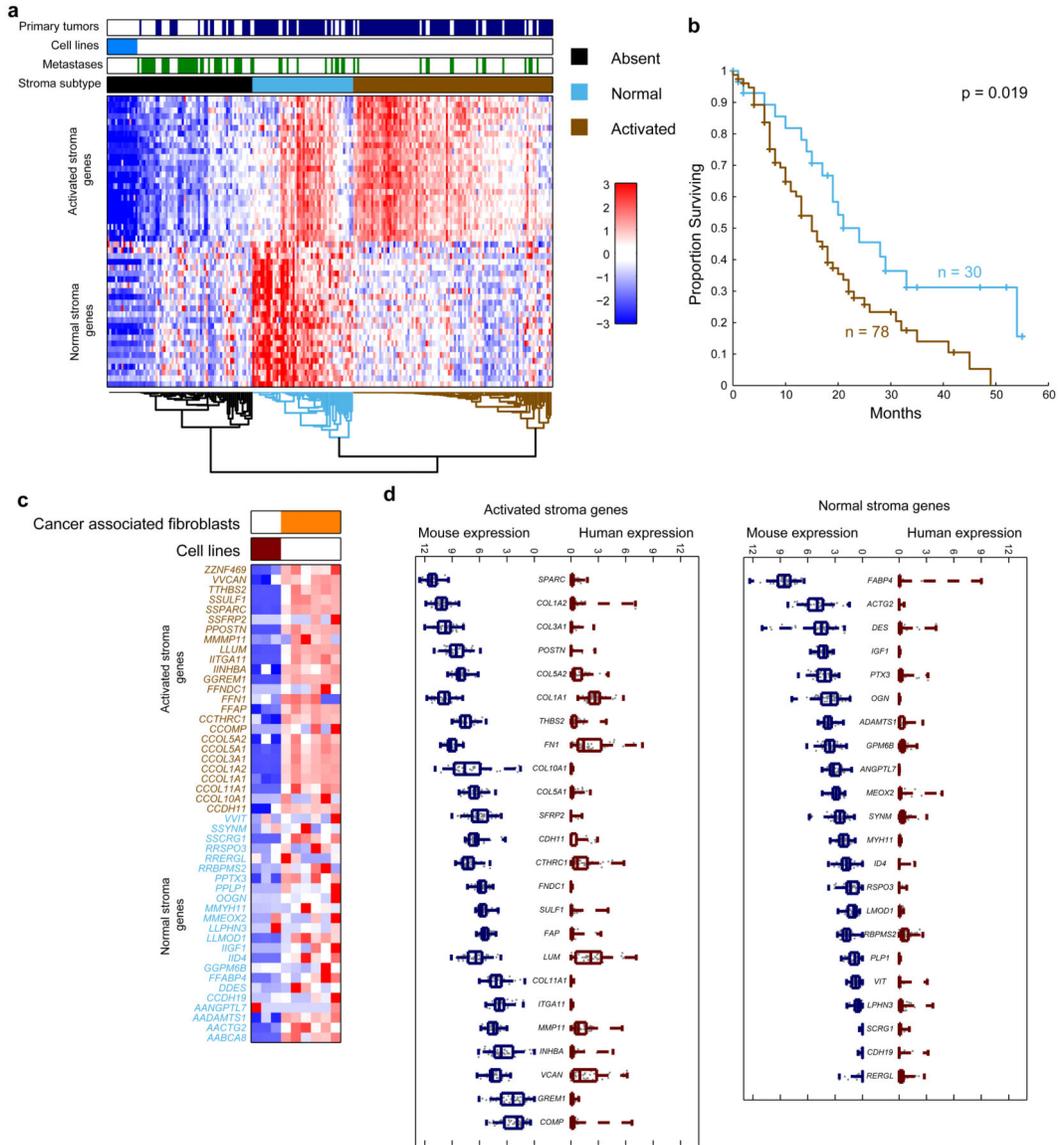


Figure 2. Dual action of stroma is described by distinct gene expression patterns which are not expressed in cell lines. **(a)** Consensus clustered heat map of UNC primary tumor samples, metastases, and cell lines using genes from stromal factors. Samples clustered into 3 groups, describing samples with activated stroma, normal stroma, and samples with low or absent stromal gene expression. **(b)** Kaplan-Meier survival analysis of resected PDAC patients from the activated and normal stromal clusters shows that samples in the activated stroma group have worse prognosis, with a hazard ratio of 1.94 (CI = [1.11,3.37], p = 0.019). **(c)** Gene expression of stromal signatures are overexpressed in CAFs as compared to tumor cell lines. **(d)** Genes from both stromal signatures are specifically overexpressed by the mouse stroma in PDX tumors, and not expressed by the human tumor cells.

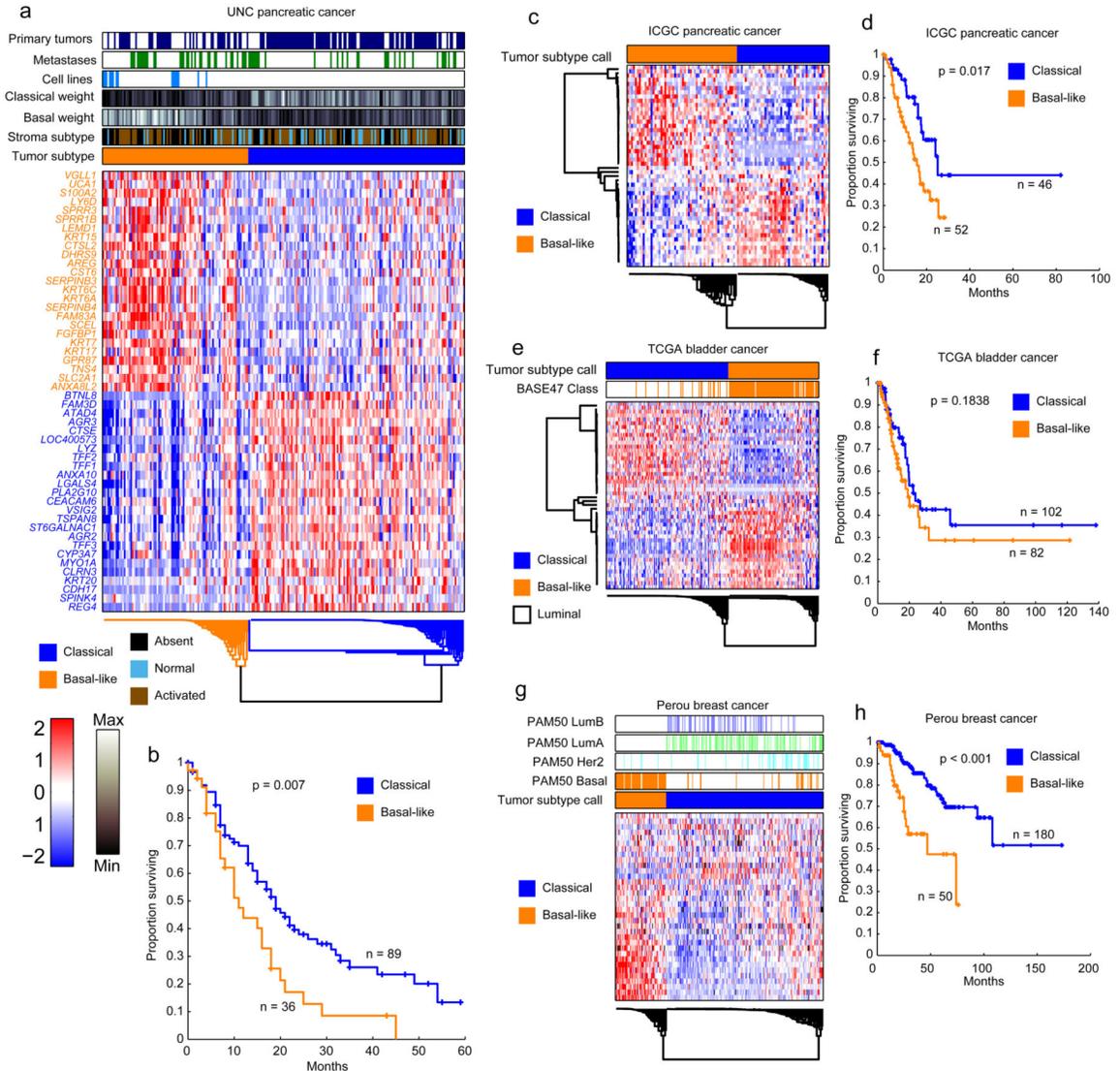


Figure 3.

Tumor specific gene expression suggests two subtypes of PDAC with similarities to other tumor types. **(a)** Consensus clustered heat map of primary tumors, metastatic tumors, and cell line models of PDAC using correlation as the underlying distance function shows two subtypes of PDAC. **(b)** Kaplan-Meier survival analysis of resected primary patients from each tumor subtype (36 basal-like, 89 classical) in **a** shows differential prognosis among subtypes with a hazard ratio of 1.89, and a 95% CI of [1.19, 3.02]. **(c)** Consensus clustered heat map of tumors in the ICGC PDAC cohort split by basal and classical factor gene expression into basal-like (n=56) and classical (n=47) tumors. **(d)** Basal-like tumors in the ICGC data set has a hazard ratio of 2.11, with a 95% CI of [1.14, 3.89]. Median follow up was 20 months **(e)** Consensus clustered heat map of TCGA Bladder cancer samples split by basal and classical factor gene expression into basal-like (n=128) and classical-like (n=95) tumors strongly agrees with BASE47 basal calls shown above the heat map. **(f)** Subtyping in the TCGA BLCA data set had a hazard ratio of 1.43, with a 95% CI of [0.84, 2.42] **(g)**

Consensus clustered heat map of the Perou breast cancer data set as split by basal factor genes (n=72 basal-like, n=223 not basal) strongly agrees with the division of samples into previously published basal and non-basal subtypes. **(h)** Basal-like breast cancer, as defined by our labeling, had a hazard ratio of 3.52, with a 95% CI of [1.94, 6.38].

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

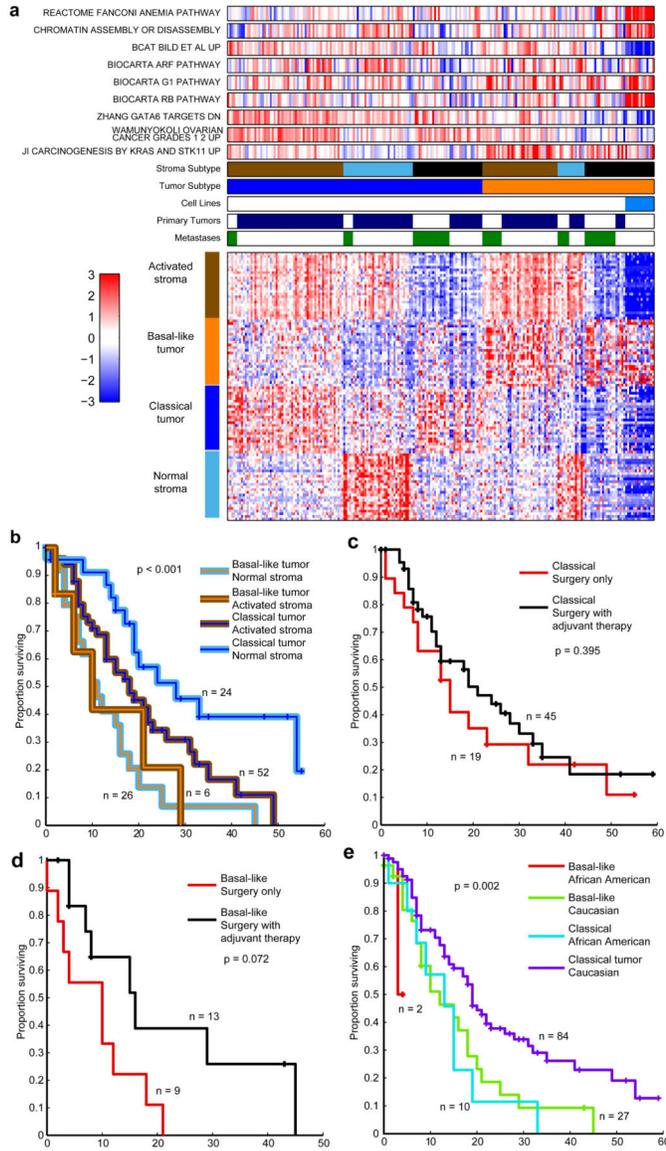


Figure 4. Multivariate survival analysis of tumor and stromal subtypes. **(a)** Heat map of tumor samples using 25 genes from each of the tumor and stromal factors, with samples sorted horizontally by classification. Signature scores for selected gene sets appear above for each sample. **(b)** Combined Kaplan-Meier survival analysis of resected primary patients from basal-like or classical tumor types and normal or activated stroma subtypes with differential survival ($p < 0.001$ log-rank test). Differential prognosis among subtypes shows complementarity. Classical tumors with normal stroma subtypes ($n=24$) had the lowest hazard ratio of 0.39, and a 95% CI of [0.21, 0.73], while basal-like tumors with activated stroma subtypes ($n=26$) had the highest hazard ratio of 2.28 with a 95% CI of [1.34, 3.87]. **(c)** Kaplan-Meier survival analysis shows that patients with classical subtype tumors show less response to adjuvant therapy (HR = 0.76, 95% CI [0.40, 1.43]) compared to **(d)** basal-like tumors (HR of 0.38, and a 95% CI of [0.14, 1.09]). **(e)** Kaplan-Meier survival analysis shows that African-

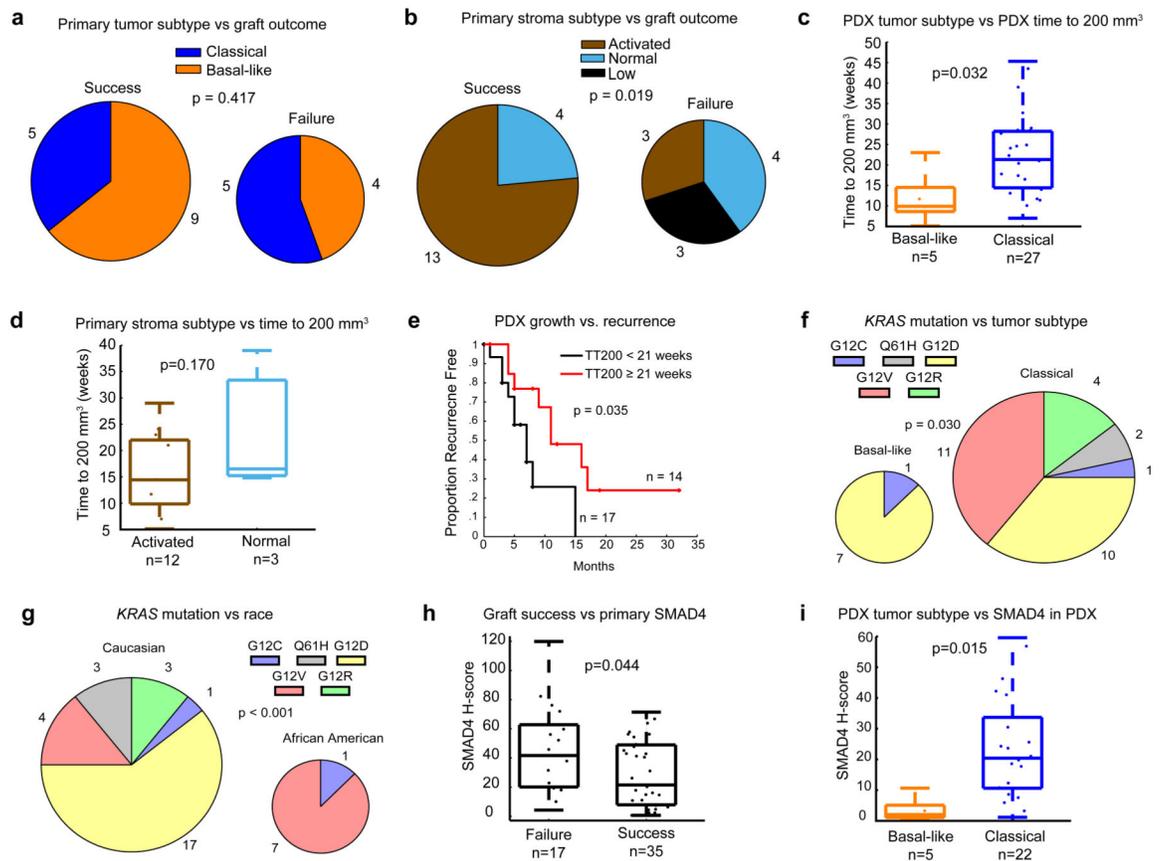
Americans have worse overall survival in both basal-like and classical subtypes, with a Hazard ratio of 2.28 and a 95% CI of [1.16,4.5].

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Figure 5.**

Associations between tumor and stroma subtypes, PDX tumors, *KRAS* mutations and SMAD4 expression. **(a)** Tumor subtype was not associated with PDX graft success rate ($p=0.417$). **(b)** Activated stromal subtype samples engraft with higher success rates than low or normal stromal subtype samples ($p=0.019$) **(c)** Basal-like tumor subtype PDX reach 200 mm³ faster than classical subtype PDX ($p=0.032$). **(d)** PDX from samples with activated stroma subtype or normal stroma subtype do not have significantly different times to reach 200 mm³ ($p=0.170$). **(e)** PDX tumors with faster growth rates were associated with earlier recurrences in patients (HR = 0.31, 95% CI [0.10, 0.92]). **(f)** *KRAS* mutation type is not uniformly distributed among race or subtype. *KRAS* G12D mutations are more prevalent in basal-like subtype tumors than classical tumors ($p=0.030$). **(g)** African Americans have more G12V mutations, while Caucasians have more G12D mutations ($p<0.001$). **(h)** SMAD4 staining in primary tumors is predictive of successful PDX engraftment ($p=0.044$). **(i)** Basal-like subtype PDX exhibit weaker SMAD4 staining than classical subtype PDX ($p=0.015$).

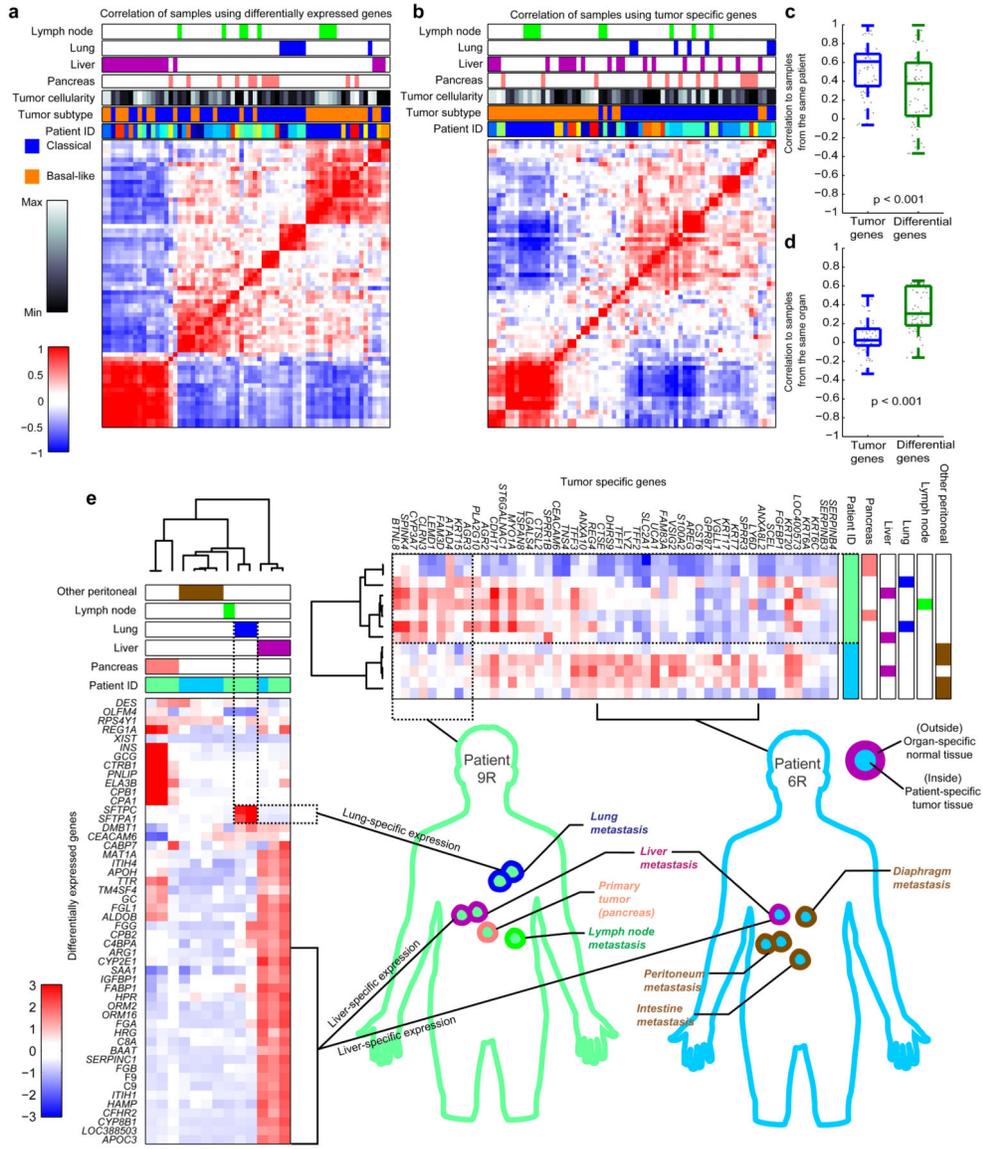


Figure 6. Overcoming tumor cellularity reveals true heterogeneity among matched primary and metastatic sites. **(a)** Sample-sample correlations of matched primary and metastatic tumors using the 50 most differentially expressed genes across all samples (“DE50”) causes samples to group by organ location. **(b)** Sample-sample correlations using 25 genes each from classical and basal-like tumor lists (“T50”) caused samples to cluster instead by tumor subtype and patient of origin. **(c)** Correlation of samples within the same patient is higher when using T50 genes than when using DE50 genes. **(d)** Correlation of samples originating in the same organ was higher when using DE50 than when using T50. **(e)** Clustering of multiple samples from two patients using the DE50 divides samples by organ. Genes expressed highly in lung and liver tissue are noted with brackets. Clustering of the same samples using T50 genes separates samples by patient. Brackets note genes which differentiate the two patients. A diagram of sampled locations for these patients indicated by

concentric circles, illustrating how samples simultaneously exhibit both patient (inner color) and organ (outer color) specific gene expression.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Demographics and Univariate Cox analysis

	All	Resected with Survival	Univariate Cox p-value	Microarray Primary	RNAseq Primary	RNAseq PDX
Race						
Caucasian	128	121	0.507	99	9	25
African American	23	18	0.333	10	3	8
Other	8	7	0.821	5	0	3
Gender						
F	90	83	0.348	67	5	23
M	80	68	0.348	55	8	14
T Stage						
T1	4	4	0.420	2	1	2
T2	22	20	0.530	20	2	5
T3	131	122	0.743	91	9	28
T4	1	1	0.115	1	0	0
N Stage						
N0	49	43	0.068	36	7	10
N1	112	106	0.068	80	5	25
M Stage						
M0	160	149		115	12	35
M1	3	0		2	0	1
Adjuvant Therapy						
Yes	74	70	0.055	44	5	21
No	30	28	0.055	27	3	7
Differentiation						
Well	16	13	0.940	16	0	1
Moderate	49	47	0.398	49	1	3
Poor	34	31	0.407	34	1	2
PDX						
Graft Success	44	37	0.164	11	8	37
Graft Failure	18	12	0.164	9	3	0
Margin						
Positive	58	52	0.026	34	5	17
Negative	93	88	0.026	75	7	17
Total	193	163		143	15	37

Table 2

Summary of associations with clinical covariates and subtypes

Covariate	Tumor Subtype		Fisher's Exact		Stroma Subtype		Fisher's Exact
	Classical	Basal-like	p-value	Normal	Activated	p-value	
Gender	Male	16	0.849	15	36		1
	Female	64		17	43		
Race	Caucasian	27	0.521	26	65		1
	African American	13		3	7		
T Stage	T2	6	0.59	5	14		1
	T3	87		25	59		
N Stage	N0	9	0.532	11	22		0.649
	N1	72		21	54		
Margin	Positive	8	0.385	7	22		0.629
	Negative	65		22	49		
Adjuvant Therapy	Yes	13	0.437	10	30		0.769
	No	9		5	19		
Differentiation	Poor	11	0.479	11	18		0.203
	Well	16		13	44		
Mucin	Low Mucin	24	0.042	18	43		0.792
	High Mucin	3		6	19		
Stroma	Normal	8	0.144				
	Activated	31					