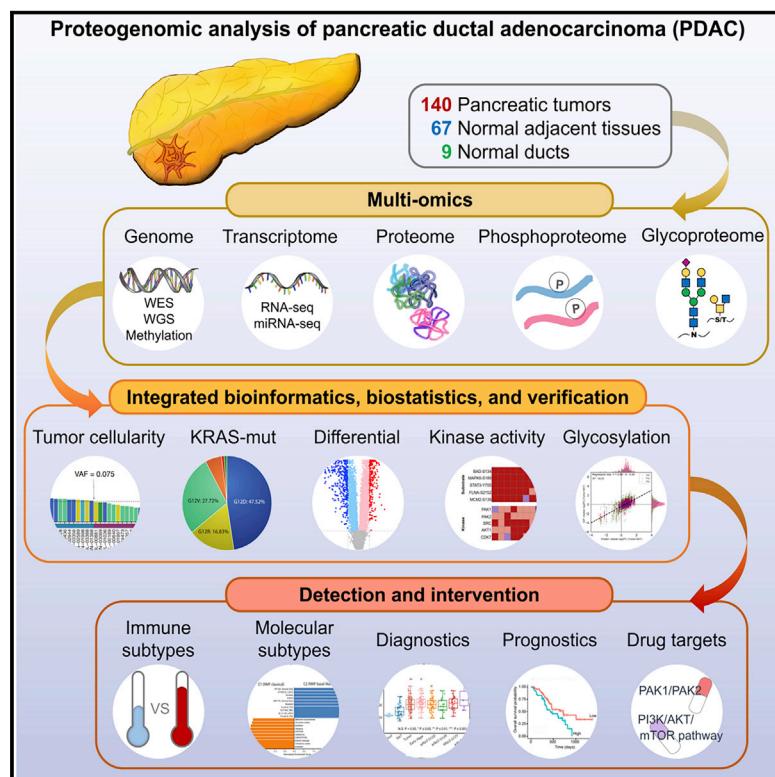


Proteogenomic characterization of pancreatic ductal adenocarcinoma

Graphical abstract



Authors

Liwei Cao, Chen Huang,
Daniel Cui Zhou, ..., Bing Zhang,
Hui Zhang, Clinical Proteomic Tumor
Analysis Consortium

Correspondence

lding@wustl.edu (L.D.),
bing.zhang@bcm.edu (B.Z.),
huizhang@jhu.edu (H.Z.)

In brief

Comparative multiomic analyses of pancreatic ductal adenocarcinoma tumors with normal adjacent and pancreatic ductal tissues provide insight into genomic, proteomic, and immune dysregulation in driving disease.

Highlights

- Proteogenomic characterization reveals the functional impact of genomic alterations
- Phosphoproteomics uncovers putative therapeutic targets downstream of KRAS
- Multiomics links endothelial cell remodeling and glycolysis to immune exclusion
- Proteomics and glycoproteomics reveal candidates for early detection or intervention



Resource

Proteogenomic characterization of pancreatic ductal adenocarcinoma

Liwei Cao,^{1,23} Chen Huang,^{2,3,23} Daniel Cui Zhou,^{4,5,23} Yingwei Hu,^{1,24} T. Mamie Lih,^{1,24} Sara R. Savage,^{2,3,24} Karsten Krug,^{6,24} David J. Clark,^{1,24} Michael Schnaubelt,¹ Lijun Chen,¹ Felipe da Veiga Leprevost,⁷ Rodrigo Vargas Eguez,¹ Weiming Yang,¹ Jianbo Pan,¹ Bo Wen,^{2,3} Yongchao Dou,^{2,3} Wen Jiang,^{2,3} Yuxing Liao,^{2,3} Zhiao Shi,^{2,3} Nadezhda V. Terekhanova,^{4,5} Song Cao,^{4,5} Rita Jui-Hsien Lu,^{4,5} Yize Li,^{4,5} Ruiyang Liu,^{4,5} Houxiang Zhu,^{4,5} Peter Ronning,^{4,5} Yige Wu,^{4,5} Matthew A. Wyczalkowski,^{4,5} Hariharan Easwaran,⁸ Ludmila Danilova,⁹ Arvind Singh Mer,¹⁰

(Author list continued on next page)

¹Department of Pathology, Johns Hopkins University, Baltimore, MD 21231, USA

²Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX 77030, USA

³Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

⁴Department of Medicine, Washington University in St. Louis, St. Louis, MO 631110, USA

⁵McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63108, USA

⁶Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA

⁷Department of Pathology, University of Michigan, Ann Arbor, MI 48109, USA

⁸The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21287, USA

⁹Department of Oncology, Johns Hopkins University, Baltimore, MD 21205, USA

¹⁰Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 1L7, Canada

¹¹Sema4, a Mount Sinai venture, Stamford, CT 06902, USA

¹²Institute for Systems Genetics, NYU Grossman School of Medicine, New York, NY 10016, USA

¹³Department of Biochemistry and Molecular Pharmacology, NYU Grossman School of Medicine, New York, NY 10016, USA

¹⁴Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada

¹⁵Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA

¹⁶Van Andel Research Institute, Grand Rapids, MI 49503, USA

¹⁷Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

¹⁸Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

¹⁹Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

(Affiliations continued on next page)

SUMMARY

Pancreatic ductal adenocarcinoma (PDAC) is a highly aggressive cancer with poor patient survival. Toward understanding the underlying molecular alterations that drive PDAC oncogenesis, we conducted comprehensive proteogenomic analysis of 140 pancreatic cancers, 67 normal adjacent tissues, and 9 normal pancreatic ductal tissues. Proteomic, phosphoproteomic, and glycoproteomic analyses were used to characterize proteins and their modifications. In addition, whole-genome sequencing, whole-exome sequencing, methylation, RNA sequencing (RNA-seq), and microRNA sequencing (miRNA-seq) were performed on the same tissues to facilitate an integrated proteogenomic analysis and determine the impact of genomic alterations on protein expression, signaling pathways, and post-translational modifications. To ensure robust downstream analyses, tumor neoplastic cellularity was assessed via multiple orthogonal strategies using molecular features and verified via pathological estimation of tumor cellularity based on histological review. This integrated proteogenomic characterization of PDAC will serve as a valuable resource for the community, paving the way for early detection and identification of novel therapeutic targets.

INTRODUCTION

With a 5-year survival rate below 10%, pancreatic ductal adenocarcinoma (PDAC) is one of the deadliest solid malignancies and is projected to become the second leading cause of cancer death by 2030 (Quante et al., 2016; Rahib et al., 2014). Due to the lack of early symptoms and dearth of reliable and effective

methods for screening and early detection, the majority of patients (80%–85%) present with locally advanced or distant metastatic disease and are unresectable (Hruban et al., 2019; Singhi et al., 2019). Combination cytotoxic chemotherapy serves as first-line treatment of metastatic PDAC and is responsible for only a modest improvement in survival (Roth et al., 2020). The median overall survival achieved in patients with metastatic



Seungyeul Yoo,¹¹ Joshua M. Wang,^{12,13} Wenke Liu,^{12,13} Benjamin Haibe-Kains,^{10,14} Mathangi Thiagarajan,¹⁵ Scott D. Jewell,¹⁶ Galen Hostetter,¹⁶ Chelsea J. Newton,¹⁶ Qing Kay Li,¹ Michael H. Roehrl,^{17,18} David Fenyö,^{12,13} Pei Wang,¹¹ Alexey I. Nesvizhskii,⁷ D.R. Mani,⁶ Gilbert S. Omenn,¹⁹ Emily S. Boja,²⁰ Mehdi Mesri,²⁰ Ana I. Robles,²⁰ Henry Rodriguez,²⁰ Oliver F. Bathe,^{21,25} Daniel W. Chan,^{1,8,25} Ralph H. Hruban,^{1,8,22,25} Li Ding,^{4,5,25,*} Bing Zhang,^{2,3,25,*} and Hui Zhang^{1,8,25,26,*} Clinical Proteomic Tumor Analysis Consortium

²⁰Office of Cancer Clinical Proteomics Research, National Cancer Institute, Bethesda, MD 20892, USA

²¹Departments of Surgery and Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

²²The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins University, Baltimore, MD 21231, USA

²³These authors contributed equally

²⁴These authors contributed equally

²⁵These authors contributed equally

²⁶Lead contact

*Correspondence: lding@wustl.edu (L.D.), bing.zhang@bcm.edu (B.Z.), huizhang@jhu.edu (H.Z.)

<https://doi.org/10.1016/j.cell.2021.08.023>

disease is still less than 12 months (Conroy et al., 2011; Von Hoff et al., 2013).

Over the last decade, comprehensive characterization of well-annotated tumor specimens has led to a better understanding of the key genomic alterations in PDAC, including somatic mutations in KRAS, TP53, CDKN2A, and SMAD4, and molecular classifications of tumors based on gene expression patterns (Thompson et al., 2020). In addition, these studies have revealed tumor-specific therapeutic targets (Balachandran et al., 2019; Golan et al., 2019). Unfortunately, targeted therapies are only available for a small number of pancreatic cancers, and resistance can emerge when single agents are given (Tao et al., 2020).

Somatic KRAS activating point mutations are the predominant genetic alteration in PDAC. However, KRAS is considered an undruggable target, except for a specific mutant form, G12C (Janes et al., 2018). In addition to downstream intracellular changes, it has been increasingly appreciated that KRAS mutations also have a broad impact on the tumor microenvironment, contributing to promotion and maintenance of cancer malignancy, responses to immunotherapy, and drug delivery. Since most pancreatic cancers are notoriously immunologically “cold,” molecular classifiers are needed to identify the small fraction of patients with relative “hot” cancers that may benefit from immunotherapy and reveal mechanisms driving immune exclusion in the majority of PDACs (Ho et al., 2020). Although genomic and transcriptomic features are needed to identify critical signaling pathways active in PDAC that can be targeted and have started to unravel the importance of stroma and the immune environment, they alone are insufficient to fully elucidate this cancer and support advancements in precision oncology. A proteogenomic approach that integrates proteomics and post-translational modification (PTM) analysis provides a more comprehensive view of pancreatic cancer biology and supports discovery of targets for early detection and treatment.

Here, we performed a comprehensive proteogenomic characterization of treatment-naïve PDACs, paired normal adjacent tissues (NATs), and macrodissected normal pancreatic duct tissues. We addressed the characteristic low neoplastic cellularity of pancreatic cancer by focusing on tumors with sufficient neoplastic cellularity as defined by several cross-validated methods, and we explored the proteogenomic features specific to neoplastic ductal epithelial cells by applying molecular and histological deconvolution. The use of tissue isolated from normal

pancreatic ducts allowed us to overcome the high acinar cell content of normal pancreatic parenchyma and compare directly neoplastic ductal epithelium with non-neoplastic ductal epithelium, confirming clinically relevant protein markers identified using PDACs and NATs. Moreover, integrated proteogenomic characterization revealed the phenotypic effects of genomic and epigenetic perturbations on proteins and protein modifications, and delineated PDAC molecular subtypes and cell microenvironment compositions. This dataset constitutes a rich resource for future studies focused on early detection and tumor-classification-based patient stratification to guide treatment selection.

RESULTS

Proteogenomic landscape of the PDAC cohort

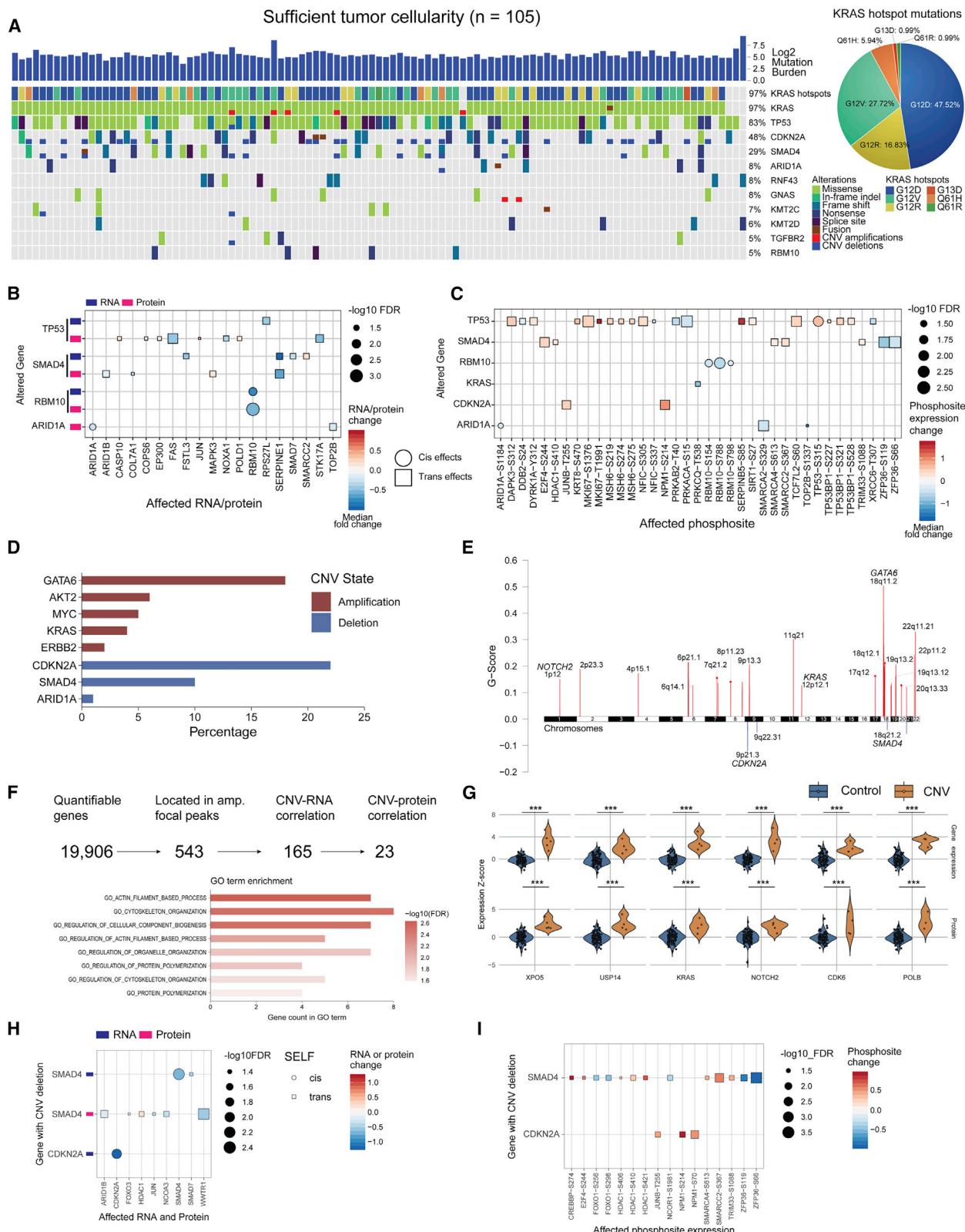
For proteogenomic characterization of PDAC, 140 treatment-naïve pancreatic tumors (135 PDACs and 5 pancreatic adenocarcinomas), 67 paired NATs, and 9 normal pancreatic duct tissues were collected and homogenized via cryopulverization for genomic, epigenomic, transcriptomic, and proteomic analyses. Clinical data are summarized in Table S1. Whole-exome sequencing (WES), whole-genome sequencing (WGS), RNA sequencing (RNA-seq), microRNA sequencing (miRNA-seq), DNA methylation, isobaric tandem mass tag (TMT)-labeling-based proteomics, phosphoproteomics, and glycoproteomics produced eight sets of omics data (Figure 1A). Proteomics, phosphoproteomics, and glycoproteomics analyses identified and quantified a total of 11,662 proteins, 51,469 phosphosites, and 34,024 glycopeptides (Table S1). We found high measurement reproducibility of the quality-control samples across the TMT plexes and no observable TMT-plex effect (Figure S1A). In this study, the median correlation between RNA and protein is 0.35 (Figure S1B), indicating a discrepancy between RNA and protein abundance, which was also observed in other cancer types (Vasaikar et al., 2019; Zhang et al., 2016a). When gene-wise correlation was performed separately for tumors and NATs (Figure S1B), a decreased median correlation within the NAT-only group relative to the tumor-only group was observed, which was noted in other tumor types (Clark et al., 2019; Gillette et al., 2020), potentially due to cell-type-specific translational regulation (Gonzalez et al., 2014).

A unique feature of this study is that samples were collected prospectively from multiple source sites worldwide, controlling



Figure 1. Proteogenomic landscape of the PDAC cohort

- (A) Sample numbers and omics data types of the cohort.
- (B) Country of origin, cancer stage, tumor site, and vital status proportions in the cohort.
- (C) Molecular and histology-based tumor estimates are used to classify samples into "sufficient" and "low" purity groups.
- (D) KRAS VAF distribution in the cohort colored by KRAS hotspot amino acid change. The sufficient neoplastic purity KRAS VAF cutoff, denoted by a dashed line, is 0.075 (15% neoplastic cellularity). The four samples with no KRAS mutations detected were also included in the sufficient tumor cellularity group since they had high mutation burden ($n > 25$), high CNV (index > 1), and/or additional driver events in TP53, CDKN2A, and SMAD4.



(legend on next page)

for ischemia time to ensure high-quality PTM analysis of proteins (Figure 1B). Our patient cohort reflected demographics of previous large-scale reports, with 75% of the resected tumors derived from the head of the pancreas (Becker et al., 2014; Kim et al., 2017; van Erning et al., 2018). As these were surgically resected cancers, the vast majority of patients were stage I–III, with only nine stage IV patients, and 42% of patients were alive at the time of the data freeze for the analysis reported in this study (Figure 1B).

To address the characteristic low neoplastic cellularity of pancreatic cancer, we identified 105 samples with sufficient neoplastic purity based on several criteria: KRAS variant allele fraction (VAF) ≥ 0.075 (equivalent to 15% neoplastic cellularity) or significant mutation burden and copy-number variations (CNVs) (Figures 1C and 1D). In addition, we used histology-based as well as DNA-methylation- and RNA-based molecular deconvolution methods (Onuchic et al., 2016) to estimate neoplastic cellularity using different data modalities. These estimates of neoplastic cellularity significantly correlated with KRAS VAF estimates (Figure S1C). In particular, KRAS VAF was highly correlated with DNA-methylation-based deconvolution (Spearman $r = 0.81$). We herein denote the remaining 35 tumors as “low purity” but emphasize that they do indeed contain neoplastic cells, as evidenced by presence of other significantly mutated gene (SMG) alterations, low KRAS VAF, and pathology review (Figure 1C). Principal-component analysis (PCA) at RNA, protein, phosphorylation, and glycosylation levels showed that high-purity tumors and NAT samples were separated but low-purity samples were spatially localized between the high-purity tumors and NAT samples, supporting our purity classification (Figure S1D). Low-purity samples were retained for selected analyses aimed at dissecting the tumor microenvironment and for tumor subtyping. We included nine normal macrodissected ductal tissues to serve as a true ductal/epithelial normal, in addition to NATs.

Impact of genomic alterations on transcriptome, proteome, and phosphorylation

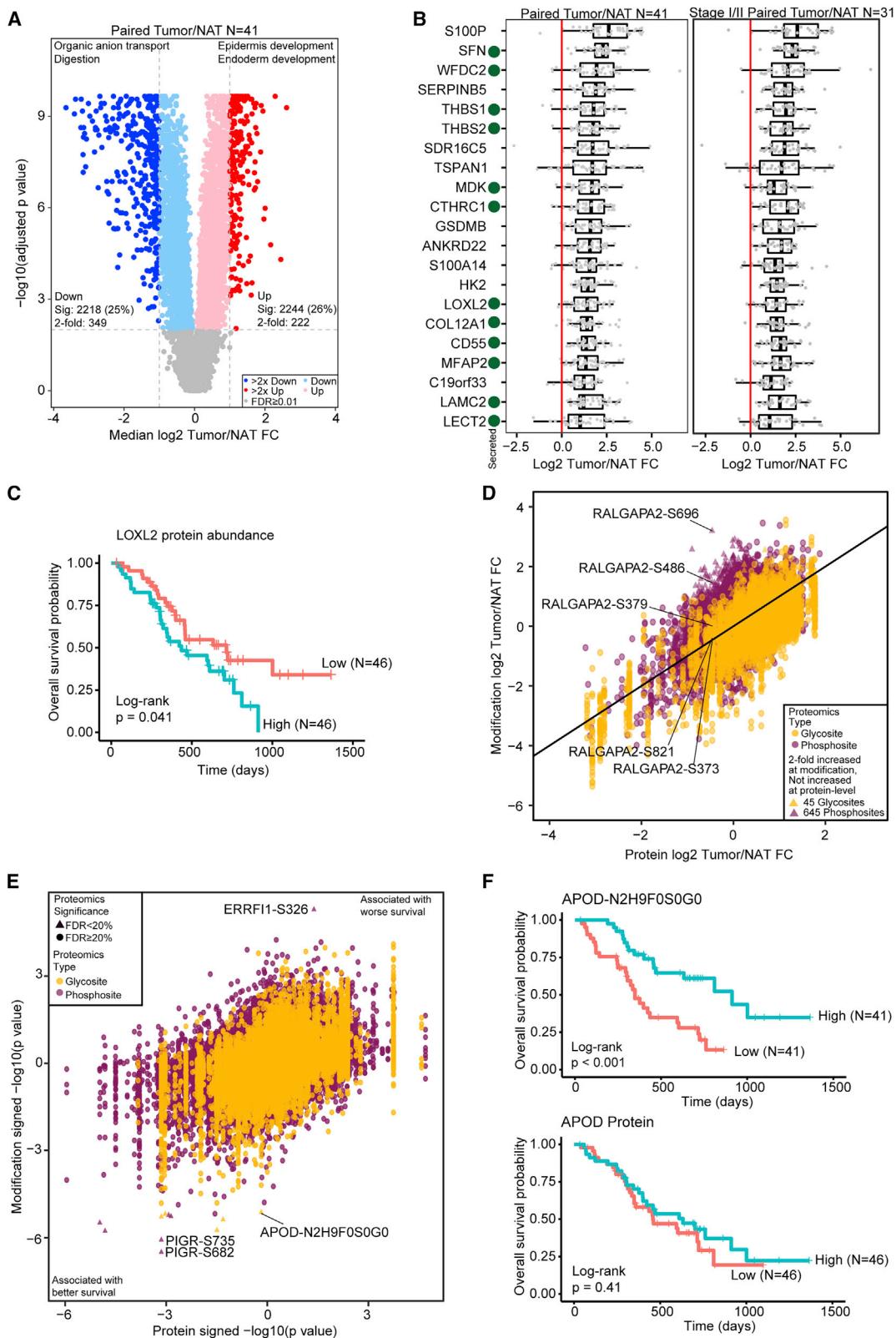
Among 105 tissue samples with sufficient tumor cellularity, somatic genomic alterations were detected in known pancreatic cancer driver genes KRAS, TP53, CDKN2A, and SMAD4 at rates of 97%, 83%, 48%, and 29%, respectively (Figure 2A). These rates are comparable to previous reports, with somewhat higher percentages of CDKN2A and SMAD4 alterations due to the in-

clusion of CNVs and fusions (Thompson et al., 2020). We found assessment and integration of CNVs to be critical, as evidenced by the presence of more CDKN2A focal deletions than intragenic mutations in our cohort (Caldas et al., 1994). Aside from these four major SMGs, we also detected ARID1A, RNF43, GNAS, KMT2C, KMT2D, TGFBR2, and RBM10 alterations in at least 5% of the tumors (Figure 2A).

We comprehensively characterized the impact of genetic alterations on RNA, protein, and phosphosite levels of the corresponding gene product (*cis*) or other interacting genes (*trans*) (Figures 2B and 2C) (STAR Methods). TP53 alterations had the most *trans*-effects at protein and phosphosite levels, with different targets identified at the RNA/protein levels and phosphosite levels, likely due to extensive post-translational regulation. Interestingly, mutations in TP53 were associated with an increase in phosphorylation of proteins involved in DNA damage repair pathways (e.g., MSH6, TP53, and TP53BP1), which suggests that these alterations play a role in maintaining genome integrity and preventing apoptosis (Figure 2C). In TP53 mutant tumors, we also observed higher phosphorylation of MKI67, a marker for cellular proliferation, which implies that these mutations may lead to increased cell growth rates (Herr et al., 2020). We further explored the effects of TP53 missense mutations compared to truncating mutations and observed a significantly greater *cis*-effect with higher TP53 protein expression and TP53-S315 phosphosite expression in the TP53 missense group compared to the wild-type group (Figure S2A), while there were no significant TP53 protein changes in *cis*-effects between the truncation and wild-type groups (Figure S2A). Interestingly, we observed similar *trans*-effects by both missense and truncation groups, which were associated with higher phosphorylation levels of the proteins in DNA damage repair pathways (Figure S2A). SMAD4 mutations were associated with downregulation of SERPINE1, a known transforming growth factor β (TGF β) pathway target (Dennler et al., 1998), at both the RNA and protein levels, as well as the upregulation of MAPK3 protein expression and downstream mitogen-activated protein kinase (MAPK) signaling (E2F4 phosphorylation). These associations have been reported *in vitro* (Chen et al., 2002) and in other cancer types (Gomis et al., 2006) (Figure 2B). We also identified several arm-level and focal-level CNVs, including amplifications in 9p, 11q, 18q, and 22q arms; GATA6 focal amplifications; and CDKN2A deletions (Figures 2D and 2E) that led to significant expression

Figure 2. Impact of genomic alterations on the transcriptome, proteome, and phosphoproteome

- (A) Genomic landscape of the cohort with sufficient tumor cellularity ($n = 105$) showing mutated genes with a frequency ≥ 0.05 . All mutation types are considered, including missense, frameshift, splice-site, copy-number alterations, and fusion events. G12D, G12R, and G12V are the most common KRAS driver mutations present in the cohort.
- (B) *Cis*- and *trans*-effects of genomic alterations on RNA and protein levels.
- (C) *Cis*- and *trans*-effects of genomic alterations on phosphosites. Protein levels are used as a covariate to remove protein abundance-related effects. In (B) and (C), *cis*-effects are denoted by circles, while *trans*-effects are denoted by squares.
- (D) Major gene copy-number amplification and deletion rates in the cohort. The log ratio cutoffs used are $[-0.4, 0.4]$.
- (E) Significant arm-level focal peaks detected using GISTIC.
- (F) CNV driver approach schematic. From all genes with copy-number events, 543 are located in the GISTIC focal peaks, of which 165 have RNA effects and 23 also have protein level effects.
- (G) Violin plots showing the impact of copy number in a select number of proteins from these 23 putative CNV drivers. *** $p < 0.001$. The control group in each comparison includes all samples without the copy-number event for each gene or protein.
- (H and I) The alterations of mRNA and proteins (H) and phosphosites (I) associated with CDKN2A and SMAD4 deletions. Samples with wild-type CDKN2A and SMAD4 serve as controls.



(legend on next page)

changes in genes, proteins, and phosphoproteins (Figure S2B). Since we observed a much larger number of amplifications, we focused on identifying putative new CNV drivers within the amplified foci (Figures 2E and 2F). Of 543 genes within amplification peaks, 165 showed significant correlation of copy number with corresponding RNA levels, including 23 that displayed concordant protein expression (Figure 2F). Proteins identified by this approach are representative of potentially novel *cis*-effects of CNV events and associated with actin filament process and cytoskeleton organization (Figures 2F and 2G), with reorganization of actin fibers having been previously implicated in tumorigenesis and metastasis (Manoli et al., 2019; Stevenson et al., 2012). In addition, most of the proteins regulated in *trans* by CNVs were located within chromosomes 7, 9, 17, and 18, while *trans*-effects of CNVs at the phospho level were sporadic (Figure S2C; Table S2). In particular, the CNV loss of SMAD4 and CDKN2A was associated with lower SMAD4 and CDKN2A mRNA levels in samples with SMAD4 deletion and CDKN2A deletion, respectively (Figure 2H). Regarding *trans*-effects of these deletions, higher phosphorylation of HDAC1 at S406, S410, and S421 and of SMARCA4 at S613 were associated with SMAD4 deletions, while CDKN2A deletions were associated with higher phosphorylation of NPM1 at S70 and S214 (Figure 2I).

To identify proteins possibly regulated by DNA methylation in tumors, we correlated RNA, protein, and phosphoprotein levels with promoter DNA methylation (Figure S2D). Among others, GSTM1 methylation resulted in downregulation of the corresponding RNA and protein, in agreement with reports that implicate GSTM1 in multiple cancers (Wang et al., 2016; Zhang et al., 2017). The extent of promoter DNA methylation was lower in NATs than in tumors (Figures S2E and S2F). We identified 86 epigenetically silenced genes, of which 22 were previously reported by the Cancer Genome Atlas (TCGA) (Table S2). Two genes (ZNF544 and THNSL2) that were epigenetically silenced in more than 10% of tumors were significantly associated with patient survival (Figures S2G–S2I). Two clusters (cluster M1 and M2) were identified by methylation-based subtyping, with cluster M2 harboring more extensive DNA hypermethylation (Table S2). We also observed a positive association between tumor cellularity and methylation status in these tumors (Table S2), in line with the TCGA study (Cancer Genome Atlas Research Network, 2017).

Discovery of specific molecular features of early-stage PDAC for tumor diagnosis and prognosis

Approximately 80% of PDAC tumors are unresectable, as patients are diagnosed at an advanced stage (Hruban et al., 2019). Thus, a panel of highly robust biomarkers for early detection

may improve survival as treatment modalities for these patients emerge. Proteins, phosphorylation sites, and glycosylation sites that are dysregulated in tumors relative to NATs represent putative candidates for early detection/prognosis. Relative to NATs, 2,218 and 2,244 proteins were significantly down- and upregulated, respectively, in PDACs (Figure 3A; Table S3). As expected, proteins with high abundance in NATs were related to normal pancreatic functions, while many of those upregulated in tumors were enriched for proteins involved in epidermal and endodermal development.

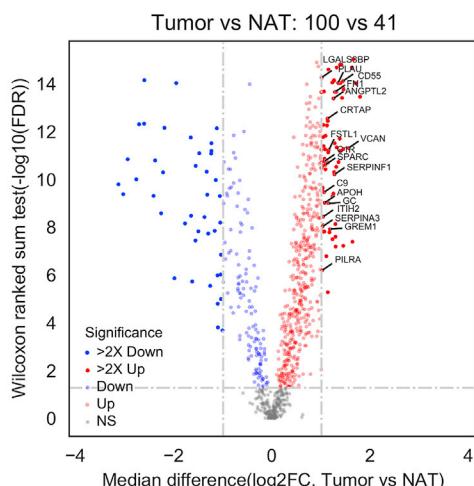
To identify proteins associated with PDACs, we focused on 222 proteins with >2-fold increase in abundance in tumors relative to NATs (Figure 3A). Twenty-seven proteins remained significantly upregulated by >2-fold in PDACs after adjusting for stromal and immune content. We additionally found that the differential expression between PDACs and NATs was similar to that between PDACs and normal ducts (Figure S3A), and 21 of 27 proteins were also upregulated >2-fold in PDACs compared to normal ducts (Figures 3B and S3B; Table S3). Importantly, these proteins were similarly upregulated in early-stage tumors (Figure 3B). In particular, 12 of these are secreted proteins and could serve as early detection markers in serum or pancreatic juice (González-Borja et al., 2019). Eleven proteins are reported as elevated in the Pancreatic Cancer Database, with six proteins (HK2, LOXL2, COL12A1, C19orf33, TSPAN1, and MDK) previously supported only by RNA or cell line proteomic evidence (Harsha et al., 2009; Thomas et al., 2014), with LOXL2 protein abundance associating with shorter overall survival (Figure 3C). Fourteen out of 21 tumor-associated proteins highlighted in Figure 3B as potential protein targets for early detection or prognostic markers were validated by the orthogonal method of data-independent acquisition (DIA) mass spectrometry analysis, indicating the reliability of our reported tumor-associated proteins (Table S3).

Compared to NATs, 4,908 phosphorylation sites and 1,727 *N*-linked glycosites showed significantly increased abundance in PDACs (Table S3; adjusted $p < 0.01$). In general, the differential abundance of PTMs was similar to the differential abundance at the protein level, while 45 *N*-linked glycosites and 645 phosphosites were upregulated >2-fold without a corresponding increase in protein abundance (Figure 3D). For example, while the protein abundance of RALGAPA2 was decreased in PDACs (Figure S3C), two phosphosites were increased >2-fold at S486 and S696, while three others were decreased or similar to NAT (Figure S3D). RALGAPA2 is related to KRAS signaling in pancreatic cancer (Beel et al., 2020), and exploring the function of these specific sites in future studies may be warranted.

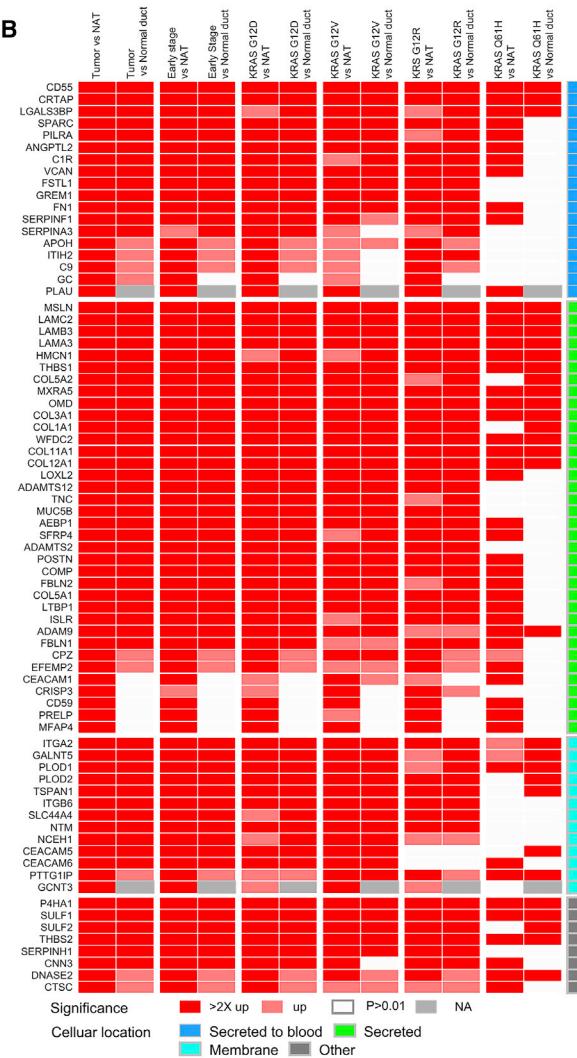
Figure 3. Identification of tumor-associated proteins and modification sites by comparison of tumor and normal tissues

- (A) Differential protein abundance between tumors and paired NATs. Selected Gene Ontology (GO) biological process terms for significantly increased and significantly decreased proteins are shown above the volcano plot.
- (B) Proteins with a median fold change >2 compared to matched NAT and with significantly increased abundance both compared to normal ducts and after adjusting for epithelial content for all samples and the subset of stage I/II samples. Secreted proteins are indicated with a green dot.
- (C) Kaplan-Meier curve for LOXL2 protein abundance association with overall survival. The two groups were separated by median LOXL2 abundance.
- (D) Median phosphosite and *N*-linked glycosylation site fold change compared to the protein fold change in tumor compared to matched NAT.
- (E) Cox regression signed p value for phosphosite and *N*-linked glycosylation site abundance association with survival compared to the protein association to survival.
- (F) Kaplan-Meier survival curves for an *N*-linked glycosylation site on APOD and APOD protein abundance.

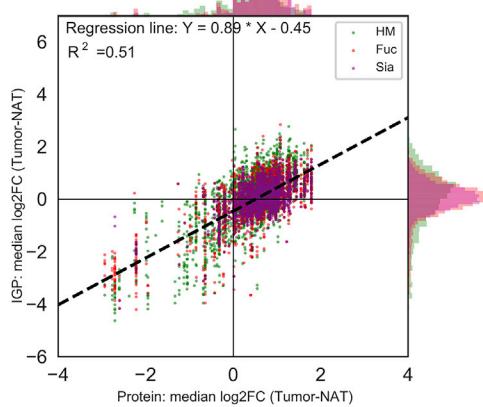
A



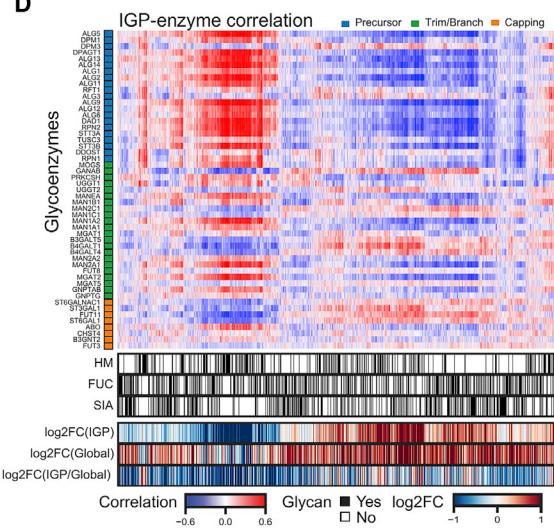
B



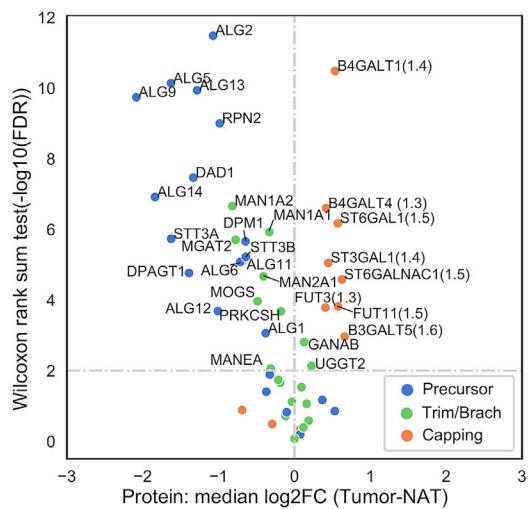
C



D



E



(legend on next page)

Finally, many of these PTMs, in addition to protein abundance, were associated with patient prognosis. Overall, the prognostic value of PTMs was similar to that of the protein (Figure 3E; Table S3). However, a particular *N*-linked glycosylation on APOD was associated with better overall survival, while total protein abundance was not (Figure 3F). Although decreased expression of APOD is associated with better prognosis in other cancer types, little is known about the role of this glycosylation site and its effect on APOD function (Ren et al., 2019). Additionally, two phosphosites on PIGR, which is involved in the epithelial-mesenchymal transition, were associated with better prognosis, while a site on ERRFI1, an ERBB signaling regulator, was associated with worse survival (Figure 3E).

Targeting glycoprotein biosynthesis for early detection and therapeutic intervention

Most of glycoproteins are membrane-bound or secreted proteins, making them a potential source of targets for immunotherapy and disease detection (Zhang et al., 2003; Hart and Copeland, 2010; Varki, 2017). Glycoproteomic analysis of PDACs and NATs identified 75 *N*-linked glycoproteins upregulated >2-fold in tumors in protein level (Figures 4A, S4A, and S4B; Table S4). Of these, 57 were reported in the Pancreatic Cancer Database (Harsha et al., 2009; Thomas et al., 2014), and 18 were newly identified in this study. Forty-eight out of 75 tumor-associated glycoproteins were further validated by DIA analysis (Table S4). In addition, mucin-type O-linked glycoproteins associated with CA19-9 antigen (Akagi et al., 2001; Hollingsworth and Swanson, 2004; Yue et al., 2011) were significantly upregulated in tumors and early-stage tumors relative to NATs and/or normal ducts (Figure S4C). We further discriminated tumor versus normal ductal tissue *N*-linked glycoprotein expression based on disparate hotspot KRAS mutations (G12D, G12V, G12R, and Q61H) (Figure 4B). Interestingly, CEACAM5 and CEACAM6 were significantly upregulated in tumors with KRAS G12D, G12V, and Q61H, but not G12R, mutations (Figures 4B and S4D). CEACAM6 is a poor prognostic marker for patients with PDAC, and CEACAM6 overexpression has been associated with low cytolytic T cell activity in PDAC (Pandey et al., 2019). A focused evaluation of *N*-linked glycoprotein expression in low-stage tumors revealed several candidates for early detection or treatment (Figure 4B; Table S4), including galectin-binding protein 3 (LGALS3BP) (Figure S4E). In addition to *N*-linked glycoprotein expression quantified from proteomic data, their glycosylated forms quantified from glycoproteomic data have provided unique expression patterns in tumors (Figure S4E).

The biosynthesis of *N*-linked glycoproteins is regulated mainly by two factors, the glycoprotein substrates and glycosylation en-

zymes for glycan synthesis and conjugation to glycoproteins. Although the alteration patterns of intact glycopeptides (IGPs) were mainly positively correlated to the protein abundances of glycoprotein substrates (Figure 4C), IGP and protein features were not always consistent, as we delineated heterogeneity of IGP abundances from the same protein displaying distinct glycan branching patterns across the pathological tissue types. Overall, *N*-linked glycoproteins upregulated in tumors were mainly modified by complex glycans with sialic acids and/or fucoses, and *N*-linked glycoproteins downregulated in tumors were mainly modified by oligomannose glycans (Figure 4C). These data indicate that focusing on sialylated and/or fucosylated glycans of the *N*-linked glycoproteins upregulated in PDAC may increase the specificity of markers for cancer. We next examined the intrinsic mechanism of these glycosylation alterations based on the abundance levels of glycosylation biosynthesis enzymes (Figure 4D). We found that the IGPs with glycosylation of sialylated or fucosylated glycans were positively correlated with the expression of glycosylation enzymes involved in glycan trimming/branching and capping (Figure 4D). Comparing the glycosylation enzyme expression revealed upregulation of glycosylation enzymes, including ST6GAL1, ST3GAL1, FUT3, FUT11, B4GALT1, B4GALT4, B3GALT5, and ST6GALNAC1, in tumors relative to NATs (Figure 4E). Of these, FUT3, FUT11, B4GALT4, and B4GALT1 were further validated by DIA analysis (Table S4). Some of these glycosylation enzyme changes, such as elevated ST6GAL1, ST3GAL1, and B4GALT1, in tumors were not observed at the transcriptomic level (Figure S4F), highlighting the added value of proteomics and glycoproteomics in our multiomic analysis. ST6GAL1 and ST3GAL1 regulate sialylation, while FUT3 and FUT11 are responsible for fucosylation, in line with our observation that PDAC upregulated proteins are mainly modified by sialylated and/or fucosylated glycans. Inhibition of these enzymes will likely attenuate increased sialylation and fucosylation glycan branching that was found on most tumor upregulated glycoproteins and serves as a potential therapeutic strategy for PDAC.

Kinase and substrate co-regulation reveals potential therapeutic targets

Since tumors with KRAS driver mutations are difficult to treat via targeted therapy, effective therapeutic intervention for PDAC, known to have a high frequency of KRAS mutations, has remained elusive (Upadhyay and Adjei, 2020). Protein phosphorylation is heavily involved in various signaling pathways during pancreatic carcinogenesis (Furuse and Nagashima, 2017; Ruckert et al., 2019). To investigate signal transduction pathways downstream of activated KRAS in search of alternative therapeutic

Figure 4. Glycoproteomic characterization identified *N*-linked glycoproteins and glycosylation enzymes for the early detection or therapeutic intervention

- (A) Differential expression analysis of *N*-linked glycoproteins in tumors to identify the most significant secreted (highlighted) and membrane *N*-linked glycoproteins elevated in tumors compared to NATs.
- (B) Upregulation of *N*-linked glycoproteins in all tumors, early-stage tumors, or tumors with different hotspot KRAS mutations relative to NATs and normal ducts at *N*-linked glycoprotein expression levels.
- (C) Comparative analysis of the expression of global proteomics and glycoproteomics. IGP, intact glycopeptides; HM, high-mannose-type glycopeptides; Fuc, fucosylated glycopeptides; Sia, sialylated glycopeptides.
- (D) Association of intact glycopeptide abundance and protein levels of glycosylation enzymes in tumors and NATs.
- (E) Differential protein expression of *N*-linked glycosylation enzymes between tumors and NATs.

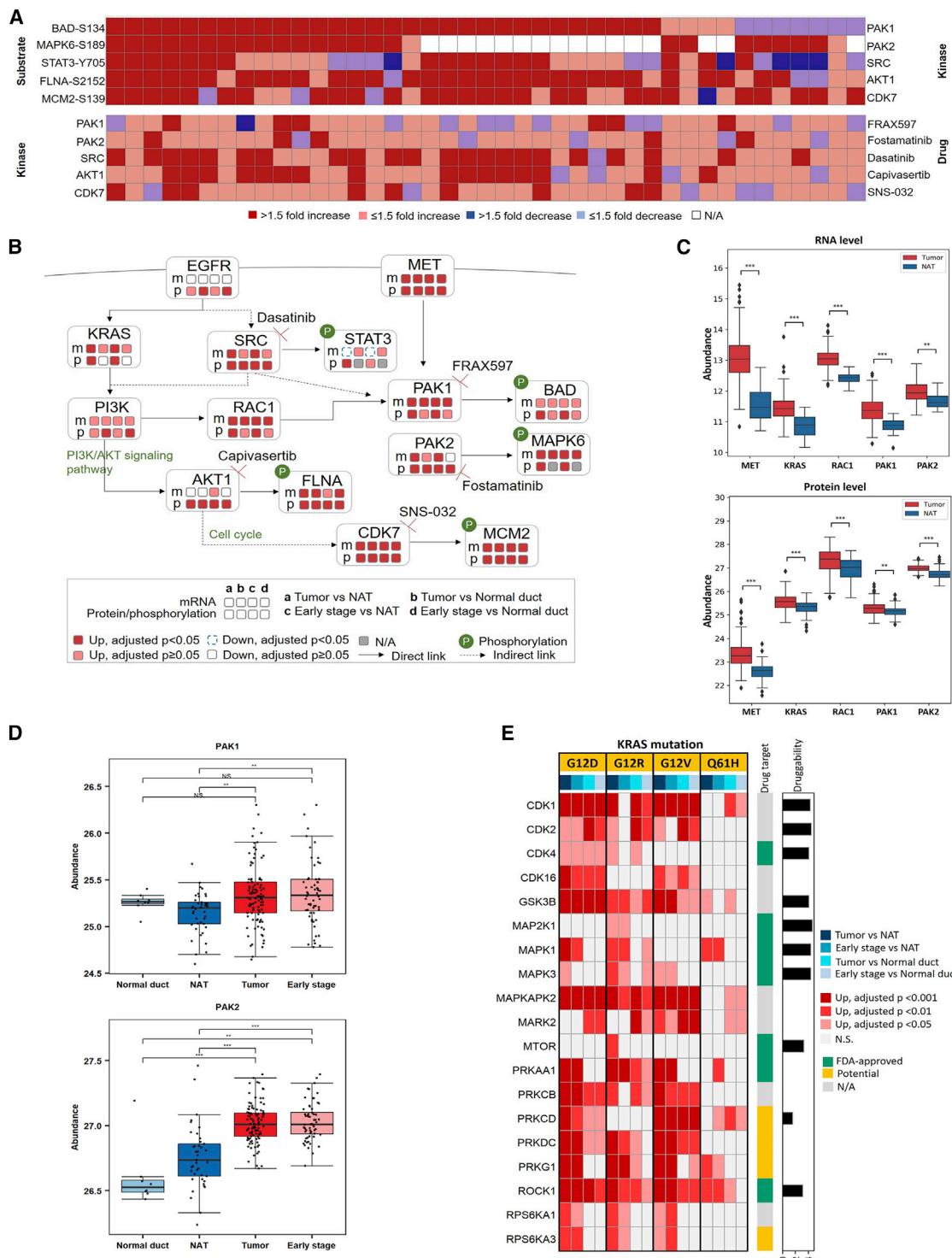


Figure 5. Kinase and substrate co-regulation

(A) Differential abundances between 41 tumor/NAT paired tissues of stratified phospho-substrates (top) and their associated kinases (bottom).
(B) Pathways based on the selected phospho-substrates and kinases, with relevant drugs. Expression changes on mRNA and/or protein/phosphosites between PDAC tumors and NATs/normal ducts are labeled.
(C) Expression profiles of PAK1- and PAK2-associated proteins at transcriptomic and proteomic levels.
(D) Expression profiles of PAKs in Normal duct, NAT, Tumor, and early stage.

(legend continued on next page)

targets, we analyzed protein phosphorylation events regulated by kinases on their respective phosphorylation substrates. By analyzing differential abundance of phosphopeptides between 41 tumor/NAT paired tissues, we stratified five phospho-substrates (MCM2, FLNA, BAD, MAPK6, and STAT3) corresponding to five kinases (CDK7, AKT1, PAK1, PAK2, and SRC), for which inhibitors are either US Food and Drug Administration (FDA) approved or under investigation (Wishart et al., 2018; Yeo et al., 2016) (Figure 5A). Previous studies have shown that elevated phosphorylated substrates are related to S-phase entry/progression (CDK7-MCM2) and that inhibition of CDK7 can result in cell-cycle arrest and suppress tumor progression (Clark et al., 2019; Montagnoli et al., 2006; Sava et al., 2020). AKT1 is a kinase downstream of KRAS (Figure 5B). The elevation of AKT1 expression in almost all tumors is a consequence of nearly universal KRAS mutations, which in turn stimulate the progression of G1/S transition, with consequent stimulation of proliferative activity (Cai et al., 2018; Jones et al., 2008; Pelosi et al., 2017). A class I p21-activated kinase (PAK), PAK1, showed higher expression in more than 70% of tumors, with its subsequent activity in PDAC tumors supported by elevated phosphorylation of its substrate (BAD-S134) (Figure 5A). Apoptosis induced by BAD is inhibited upon phosphorylation of BAD-S134 by PAK1, thus promoting cell proliferation and survival (Polzien et al., 2011; Ye and Field, 2012). PAK1 can be activated by direct interaction with RAC1 (Fan, 2020), and RAC1 was upregulated in most tumors (Figures 5B and 5C). PAK1 is an important effector of several receptor tyrosine kinases, such as MET (Rane and Minden, 2019; Zhou et al., 2014). We observed concordant upregulation of MET and PAK1 (Figure 5B), and we found that MET was concordantly upregulated with KRAS, RAC1, PAK1, and PAK2 at both the transcription and protein levels in tumors as well (Figure 5C). The MET/PAK1 signaling axis drives pancreatic carcinogenesis via regulation of cell proliferation, motility, and regulation of cytoskeletal remodeling (Zhou et al., 2014). Furthermore, constitutive activation of the SRC/STAT3 signaling axis enhances hepatocyte growth factor (HGF) promoter activity, which in turn activates PAK1 via HGF/MET signaling (Aznar et al., 2001; Lee et al., 2019; Wojcik et al., 2006; Yuan et al., 2015). Another member of the class I PAKs, PAK2, was also upregulated in almost 90% of tumors and likely responsible for elevated phosphorylation of MAPK6-S189 (Figures 5A–5C). The phosphorylation process is critical for the formation of the MAPK6-Prak complex for MAPK6 signaling, suggesting an important role for PAK2 activity in regulating atypical MAPK signaling associated with cell motility (De la Mota-Peynado et al., 2011). Expanding the phosphoproteomic analysis to include the normal ducts showed that the expression profiles of the class I PAKs and the other kinases, as well as their substrates, in PDAC tumors were substantially different from NATs and/or normal ducts, suggesting that these proteins were PDAC-associated kinases (Figures 5D, S5A, and S5B). Further-

more, the differential expression patterns of four of these kinases (PAK1, SRC, AKT1, and CDK7) were confirmed by DIA analysis (Table S5).

By evaluating phosphosite expression changes in tumors with different KRAS hotspot mutations relative to NATs (Figure S5C), we further stratified 19 kinases (Figure 5E; Table S5), including seven FDA-approved drug targets (Hobbs et al., 2020; Uhlen et al., 2015). These different patterns of kinase expression suggest alternative therapeutic targets associated with specific KRAS mutations. Given the importance of the class I PAKs to PDAC, combined inhibition of PAK1/2 and KRAS downstream pathways, such as MAPK/ERK and PI3K/AKT/mTOR, may increase therapeutic benefit by maximizing inhibition of tumor cell proliferation, motility, and signaling to the cytoskeleton.

Immune-cold PDACs associated with endothelial cell remodeling, glycolysis, and cell junction dysregulation

One limitation of molecular analyses of tumor and normal tissues is that they do not fully dissect the interaction between tumor-intrinsic biology and microenvironment dynamics. This knowledge gap is particularly consequential for PDAC, which is heavily driven by tumor microenvironmental features (Collisson et al., 2019). Here, we classified tumors based on microenvironmental cell signatures, with an emphasis on delineating the degree of immune infiltration, as targeting immune modulators has shown promise in the treatment of a variety of cancer types (Yang, 2015). Unlike other tumors, such as melanoma, PDACs are resistant to immune checkpoint inhibitors in general (Hilmi et al., 2018), and leveraging a comprehensive proteogenomic approach may provide insight into this phenomenon.

We used a transcriptomics-based deconvolution method (Aran et al., 2017) to delineate the cellular composition of all 140 tumors in this study (i.e., including tumors with low neoplastic cellularity as these more fully represent stromal components of the tumor), which was further validated by DNA-methylation-based tumor deconvolution (Figures 6A and S6A–S6C). Samples were classified into four clusters based on tumor/stromal/immune cell composition (Figure 6A). Of particular interest was a small group (cluster D) of tumors with higher CD8⁺ T cell infiltration accompanied by increased expression of cytotoxic enzymes and immune checkpoint molecules. We annotated samples in this cluster as “immune hot” tumors. Histologic review of these cases confirmed prominent inflammatory infiltrates associated with the tumor component (Figures S6D). Nevertheless, in one case (C3N-00303), the immune signature was likely a result of the inclusion of a lymph node in the tissue harvested, highlighting the critical importance of histologic review of biosamples used in the study of any cancer type (Figure S6E). Samples in clusters A, B, and C showed little immune infiltration. Because cluster A was enriched with non-neoplastic acinar and islet cells, as shown by the deconvolution and RNA subtyping, we considered only clusters B and C as true “immune

(E) Heatmap showing kinases elevated in different KRAS hotspot mutations. The kinases were identified based on their upregulated phospho-substrates. The drug target annotation is from the Human Protein Atlas (<https://www.proteinatlas.org/>) alongside the log-transformed druggability score based on the drug sensitivity evaluated in PDAC cell lines from Genomics of Drug Sensitivity in Cancer (<https://www.cancerrxgene.org/>). Normal duct, normal ductal tissues; NAT, normal adjacent tissues; tumor, all PDAC tumors; early stage, stage I and II PDAC tumors. Asterisks represent significant differences between two groups (Benjamini-Hochberg adjusted p): *p < 0.05; **p < 0.01; ***p < 0.001; N.S., not significant. The list of kinase inhibitors/drugs is not exhaustive.

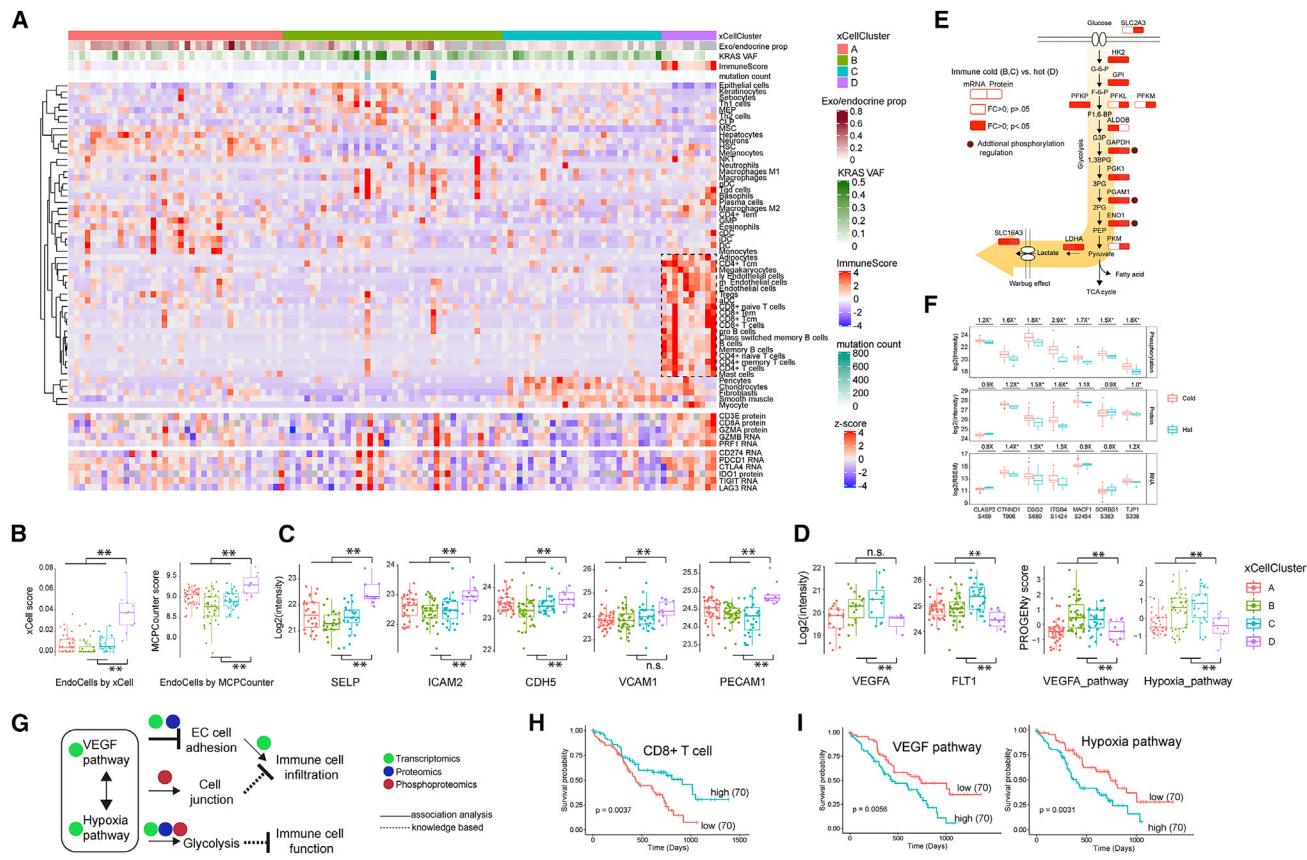


Figure 6. Delineation of the cellular composition of PDAC tumors and identification of biological events accounting for the immune-cold phenotype

(A) The 140 tumors were classified into four clusters based on tumor composition (upper heatmap). Cytotoxic T cells, together with endothelial cells enriched in cluster D, are highlighted by a rectangle. The expression of immune cytotoxic factor and checkpoint genes is shown in the sample order (lower heatmap).

(B) The comparison of endothelial cells between immune-hot and immune-cold samples based on the *in silico* deconvolution using either xCell or MCPCounter.

(C) Immune-cold tumors have reduced endothelial adhesion proteins.

(D) Immune-cold tumors have upregulated VEGF and hypoxia pathways.

(E) The immune-cold tumors had higher levels of glycolytic pathway components. Shown are the comparison of these components between immune cold versus hot at both the RNA and protein level. Some of the pathway components are identified with known functional phosphosites and are highlighted by brown circles.

(F) Phosphorylation pathway enrichment showed that the immune-cold samples have higher phosphorylation levels of cell junction proteins. Shown are immune cold versus hot fold changes for protein phosphorylation, protein expression and RNA expression. * $p < 0.05$, Student's t test.

(G) The possible working model associated with immune exclusion.

(H) The clinical outcome associated with CD8+ T cells.

(I) The clinical outcome associated with VEGF and hypoxia pathway activities.

In (B)–(D), ** $p < 0.01$; n.s., not significant (Student's t test). In (H) and (I), p values were derived from a log-rank test, and numbers in parentheses represent sample sizes for each group.

“cold” tumor groups (Figure S6F). Noticeably, immune-hot tumors were also enriched with endothelial cells, and the enrichment was supported by an independent deconvolution tool (Becht et al., 2016) (Figure 6B). In addition, cell-type association network analysis confirmed strong associations between endothelial cells and cytotoxic immune cells (Figure S6G).

Endothelial cells represent a physical connection between the circulatory system and tumor cells (Klein, 2018), and endothelial cell adhesion proteins are essential for immune cell recruitment and frequently downregulated in tumor-associated vasculature (Schaaf et al., 2018). Accordingly, immune-cold tumors in our cohort had reduced expression of endothelial adhesion proteins

(Figure 6C). Meanwhile, these tumors also showed elevated activity of VEGF and hypoxia pathways, as indicated by expression of VEGF and its receptor, as well as the inferred pathway activities (Schubert et al., 2018) (Figure 6D). Both VEGF and hypoxia pathways are integral to the remodeling of endothelial cells during tumorigenesis (Petrova et al., 2018). Together, these results support an association of endothelial cell remodeling and suppressed immune infiltration in immune-cold PDACs.

To further characterize the mechanisms underlying the immune-cold phenotype, we performed pathway analysis using RNA, protein, and phosphorylation data. Immune-cold samples had higher levels of glycolysis (Figures 6E and S6H), including

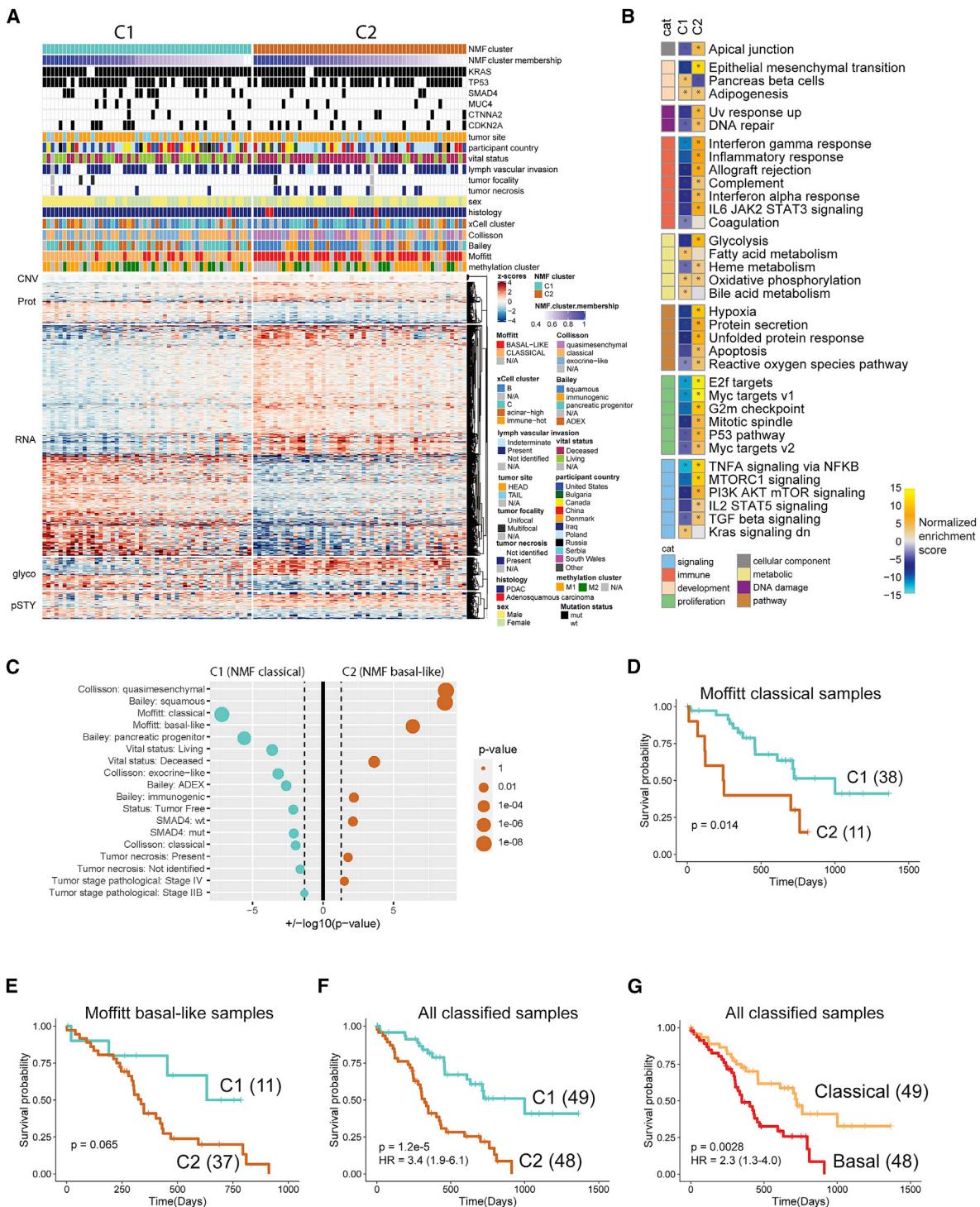


Figure 7. Proteogenomic subtyping of 105 high-purity tumors using gene copy number, mRNA, protein, phosphosite, and glycosite abundances largely separated tumors into two subtypes

(A) Heatmap depicting the Z-scored abundances of proteogenomic features separating the two clusters as determined by NMF. Cluster membership scores indicating the strength of association of each sample with a given cluster were calculated as proportional weights. The columns of the matrix are ordered by proteogenomic subtype and decreasing cluster membership score.

(B) Pathway-level analysis on proteogenomic subtypes. Shown are pathway activity scores of cancer hallmark gene sets derived from single-sample gene set enrichment analysis (ssGSEA) applied to the vector of feature weights characterizing each cluster. Asterisks indicate gene sets with false discovery rate (FDR) < 0.01. cat, category.

(legend continued on next page)

enrichment of enzymes responsible for the generation and secretion of lactate, a known immune suppressor in tumor microenvironment (de la Cruz-López et al., 2019). In addition, phosphorylation-specific pathway enrichment analysis showed that the immune-cold samples had higher phosphorylation levels of cell junction proteins (Figure S6l); this feature was not as robustly detected at the transcriptomic or proteomic levels (Figure 6F). Cell junction proteins play an important role in regulating endothelial cell permeability for small molecules and immune cell infiltration (Daneman and Prat, 2015; Radeva and Waschke, 2018). Here, our results suggest that the dysregulation of protein phosphorylation in the cell junction components might represent an additional mechanism of immune exclusion in PDAC tumors.

Together, these data suggest that endothelial cell remodeling, accompanied by elevated VEGF and hypoxia pathways, increased glycolysis, and cell junction dysregulation might collectively inhibit immune cell infiltration and function (Figure 6G). Inhibiting these biological processes, especially glycolysis and endothelial cell remodeling, both of which have been actively targeted in multiple cancer types (Annan et al., 2020; Pelicano et al., 2006), may be therapeutically exploited to boost antitumor immunity in immune-cold PDACs. This is supported by associating clinical outcomes with these processes. While CD8⁺ T cell infiltration was a favorable prognostic signature, elevated VEGF and hypoxia pathway signaling both were associated with decreased survival (Figures 6H and 6I).

Proteogenomic subtypes with strong prognostic relevance

We applied three main transcriptomics-based subtyping strategies for PDAC to the entire set of tumors to explore inter-sample heterogeneity (Figure S7A) (Bailey et al., 2016; Collisson et al., 2011; Moffitt et al., 2015). Consistent with a previous report (Cancer Genome Atlas Research Network, 2017), some of these molecular classifications overlapped significantly, such as “ADEX” (Bailey) and “exocrine-like” (Collisson), “Classical (Collisson)” and “pancreatic progenitor” (Bailey), and “squamous” (Bailey), “quasimesenchymal” (Collisson) and “basal-like” (Moffitt) ($p < 0.0001$, Fisher’s exact test). Notably, five adenosquamous carcinoma samples in our cohort were classified into “squamous” (Bailey), “quasimesenchymal” (Collisson), or “basal-like” (Moffitt) groups, in line with the understanding for this histological pancreatic cancer subtype (Boecker et al., 2020; Lenkiewicz et al., 2020; Moffitt et al., 2015). Non-negative matrix factorization (NMF)-based proteogenomics subtyping using multiomics data from all 140 tumors revealed four clusters with significant overlap with RNA subtypes (Figures S7B and S7C; Table S7A). Both RNA-based and multiomics-based subtyping results for the whole cohort were heavily confounded by tumor purity and cell-type composition (Figures S7B–S7G).

In order to partially mitigate the impact of tumor purity on subtyping, we further limited the NMF-based proteogenomics subtyping to the 105 PDAC tumors with sufficient tumor neoplastic cellularity. This analysis revealed two clusters (C1 and C2; Figure 7A), which showed significant overlap with Moffitt classical and Moffitt basal-like RNA subtypes, respectively. Since the Moffitt subtypes were derived by using a tumor-intrinsic gene signature (Moffitt et al., 2015), the difference between the two proteogenomic subtypes is more likely to reflect tumor-intrinsic biological signals, with the C2 subtype being associated with elevated activities of multiple proliferative signaling pathways and worse survival rates (Figures 7B and 7C).

Despite the overall concordance between proteogenomic subtypes derived from multiomics data and the Moffitt subtypes derived from RNA-seq data alone, the classification of 22 tumors was inconsistent (Figures 7A–7C). 11 Moffitt basal-like tumors were classified as proteogenomic classical, and 11 Moffitt classical tumors were classified as proteogenomic basal-like. Interestingly, splitting the Moffitt classical or basal tumors according to the proteogenomic clusters revealed a trend of distinct prognostic outcomes (Figures 7D and 7E). Concordantly, we observed that for the 97 PDAC samples with both proteogenomic and Moffitt assignments, the proteogenomics-dichotomized subtypes showed stronger prognostic separation than the Moffitt-dichotomized subtypes (Cox proportional-hazards [CoxPH] hazard ratio [HR], 3.4 versus 2.3 comparing favorable to adverse survival) (Figures 7F and 7G). We further interrogated the 392 proteins and 258 phosphosites with significant prognostic values (Tables S7E and S7F) and found that they were more likely to show differential abundance between the two proteogenomic subtypes (69% of the proteins and 78% of the phosphosites) than between the two Moffitt subtypes (39% of the proteins and 38% of the phosphosites).

To perform in-depth analysis of the two subtypes, we correlated molecular features, phosphorylation patterns of kinases, glycosylation enzymes, and therapeutic targets revealed by comprehensive proteogenomics with C1 and C2 subtypes (Tables S7G–S7P). The C2 subtype was associated with higher expression of most of kinases highlighted in Figure 5 and S5 (Table S7O). An association of chemotherapeutic drugs with C1 (e.g., docetaxel, vinblastine, and cabazitaxel) and kinase inhibitors with C2 (e.g., PP-242, CP466722, and sunitinib) was revealed by gene set enrichment analysis, which was further supported by elevated expression of inferred corresponding drug targets in C1 or C2 subtypes (Figure S7H; Table S7P). For instance, elevated mTOR, AKT, and ERK kinase expression and the enrichment of the PP-242 signature in C2 suggests that mTOR could be a potential therapeutic target in these patients.

Our binary proteogenomics subtyping focused only on the 105 samples with sufficient tumor purity to better understand

(C) Overrepresentation analysis of clinical variables, RNA subtypes, and somatic mutations in each proteogenomic subtype (p values derived from Fisher’s exact test). Size of the dots scale with the significance of association. Cyan dots indicate association with the C1 subtype, and orange dots indicate association with the C2 subtype. Vertical dashed lines correspond to a nominal p value of 0.05.

(D–G) Kaplan-Meier plots comparing the survival outcomes between Moffitt classical samples assigned into proteogenomic classical cluster (C1) and proteogenomic basal-like cluster (C2) (D), Moffitt basal-like samples assigned to the two proteogenomic clusters (E), the two proteogenomic clusters (F), and the two Moffitt subtypes (G). The p values were derived from log-rank test and numbers in parentheses represent sample sizes for each group. The hazard ratios (HRs) were derived from Cox proportional-hazards (CoxPH) regression and shown as “HR (95% confidence interval).”

tumor-intrinsic biology. On the other hand, the immune/microenvironment characterization was done using all 140 samples, as immune-hot samples generally have low neoplastic purity (Figure S7I). When 140 samples were used for proteogenomic subtyping, four subtypes were observed (Figure S7B). Interestingly, eight out of nine immune-hot samples comprised a subset of the C4 subtype (Figure S7I), which significantly overlapped with the Moffitt classical subtype (Figure S7B and S7C). By further comparing our microenvironment/immune profiling results to the studies focusing on the similar aspect of pancreatic cancer (Chan-Seng-Yue et al., 2020; Elyada et al., 2019; Maurer et al., 2019; Moffitt et al., 2015; Puleo et al., 2018), we identified that the microenvironmental features of immune-hot samples were more favorable for immune cell infiltration (Figure S6D).

Together, these results support the association of integrated proteogenomic subtyping with patient outcome. Further experimental investigation of the over activated proliferative and signaling pathways in the poor-prognosis proteogenomic basal-like subtype (Figure 7B) may facilitate the development of subtype-specific therapeutic strategies.

DISCUSSION

In this report, we describe a comprehensive proteogenomic investigation of PDAC that integrates multiomic profiles to provide insights into the impact of genomic and epigenomic perturbations on gene and protein expression, as well as PTMs. To ensure a robust comparison of PDACs with pair-matched NATs and normal ductal tissues, we leveraged molecular, histological, and computational approaches to annotate neoplastic cellularity of the tumors in our cohort and the high acinar content of normal tissues, with a major aim of including only high-quality samples in our analyses (Figures 1 and 2). This approach yielded robust identification of potential targets for early detection, diagnosis, or therapeutic intervention (Figures 3, 4, 5, 6, and 7).

We verified *KRAS* as the major driver gene in PDAC (Figures 1 and 2), in line with previous studies (Eser et al., 2014; Cancer Genome Atlas Research Network, 2017; Thompson et al., 2020). However, targeting the KRAS protein itself has failed due to its smooth surface topology and lack of a hydrophobic pocket for secure drug binding, leading to a dearth of approved KRAS-specific drugs, except for the compound MRTX849, for mutant *KRAS* G12C that is only present in <1% of PDACs (Berndt et al., 2011; Christensen et al., 2020; Vatansever et al., 2020). Interestingly, comparison of glycoprotein expression among tumors with different hotspot *KRAS* mutations revealed upregulation of CEACAM5 and CEACAM6 in PDACs with G12D, G12V, and Q61H mutations, but not with mutant G12R (Figure 4B). CEACAM5 and CEACAM6 belong to the immunoglobulin superfamily; mediate cell migration, cell invasion, and cell adhesion via homophilic and heterophilic binding to other proteins; and protect neoplastic cells from undergoing anoikis (Beauchemin and Arabzadeh, 2013; Blumenthal et al., 2005b). Although the effect of anti-CEACAM5/6 monoclonal antibodies (mAbs) on normal tissues remains to be determined, mAbs MN-15 and MN-3 can impede metastasis in preclinical studies by reducing adhesion of tumor cells to endothelial cells and extracellular matrix (Blumenthal et al., 2005a, 2005b; Govindan

et al., 2009; Strickland et al., 2009). Thus, anti-CEACAM5/6 mAb coupled to first-line chemotherapies may benefit patients with PDACs harboring *KRAS* G12D, G12V, and/or Q61H mutations. Alternatively, inhibition of critical downstream targets and nodes orchestrated by constitutively activated *KRAS* is an attractive strategy for PDAC treatment. The MAPK/ERK and PI3K/AKT/mTOR pathways represent major targets for therapeutic intervention of PDAC, and multiple inhibitors of each of the pathways are clinically available (Eser et al., 2014). While drugs that block these pathways are being tested in the clinic, new efforts are underway to exploit previously unrecognized vulnerabilities, such as altered signaling networks, for novel targeted therapies (Ducréux et al., 2019; Sapalidis et al., 2019). Our integration of proteomic and phosphoproteomic measurements revealed that PAK1/PAK2 kinases were upregulated in most PDACs in our cohort (Figure 5) and these kinases have been reported to be critical effectors/regulators of vital signaling pathways that mediate cellular cytoskeletal motility, proliferation, and survival (Zhou et al., 2014). Positioned downstream of oncogene *KRAS*, inhibitors of PAK1/PAK2 have potential as new ways of targeting *KRAS* and could be coupled with inhibitors that target the canonical *KRAS* downstream MAPK/ERK and phosphatidylinositol 3-kinase (PI3K)/AKT/mTOR pathways (Najahi-Missaoui et al., 2019; Semenova and Chernoff, 2017).

PDAC is characterized by a highly suppressive tumor microenvironment, and intratumoral infiltration by cytotoxic T cells is low for most patients (Elyada et al., 2019; Hessmann et al., 2020). Although immunotherapies that target cytotoxic T lymphocyte antigen-4 (CTLA-4) and programmed cell death protein-1 (PD-1) significantly benefit patients with several solid malignancies, they are ineffective in patients with PDACs, except for microsatellite-instability-high tumors, which account for <2% of PDACs (Goggins et al., 1998). The determinants of immune activation in PDAC are poorly understood, providing little therapeutic guidance (Balli et al., 2017). To dissect the tumor microenvironment, we leveraged our multiomics data and revealed that absence of endothelial cells associated with upregulation of VEGF and hypoxia pathway activities was, in turn, associated with immune cell exclusion in immune-cold tumors (Figure 6). Modifying tumor endothelial cells into a normal endothelial cell phenotype could possibly be achieved by antiangiogenic therapy, such as sorafenib and NGR-TNF, with upregulation of leukocyte-endothelial cell adhesion molecules and could possibly promote intratumoral immune cell infiltration (Allen et al., 2017; Elia et al., 2018; Ferrara et al., 2004). Hypoxia inducible factor-1 (HIF-1) is the main effector of the hypoxic microenvironment in pancreatic tumors and induces cell metabolism into glycolytic mode (Yuen and Diaz, 2014). Thus, therapies targeting HIF-1 activity, such as small molecules preventing the interactions of the HIF1- α and HIF1- β subunits, might also be beneficial for pancreatic immune-cold tumors (Petrova et al., 2018).

N-linked glycosylation occurs in the endoplasmic reticulum and Golgi apparatus and is mediated by the activity of a series of glycosidases and glycosyltransferases (Bieberich, 2014; Cao et al., 2018). Abnormal expressions of sialylated glycoproteins have been uncovered in various solid malignancies, including PDAC, and have been associated with invasiveness and metastatic potential (Hsieh et al., 2017; Suzuki, 2019; Vajaria

et al., 2016). Here, we have shown that most upregulated N-linked glycoproteins in PDAC are modified by sialylated glycans, consistent with upregulation of ST6GAL1 and ST3GAL1 in PDACs relative to NATs (Figure 4). These tumor upregulated N-linked glycoproteins were associated with vital signaling pathways involved in PDAC progression and metastasis (Figure S4B). Thus, inhibition of these sialyltransferases with more selective inhibitors may attenuate PDAC cell growth, survival, and metastasis via abrogation of the functions of these N-linked glycoproteins (Garnham et al., 2019; Vajaria et al., 2016; Macauley et al., 2014).

In total, this report exemplifies the unique and useful insights that can be gained when characterizing the disease state at multiple “omics” levels, enabling a deeper understanding of the functional consequences of genomic aberrations associated with PDAC. Integrating measurements of the transcriptome, proteome, phosphoproteome, and glycoproteome and comparative profiling of PDACs, NATs, and normal ducts enabled our detection of proteoforms associated with early-stage PDAC, as well as identification of potential therapeutic targets that may find utility in the clinical setting. Overall, our study delineates the molecular features that drive the PDAC phenotypes and provides a rich bioinformatic resource for future hypothesis-driven translational research.

LIMITATIONS OF THE STUDY

The objectives of this study were to comprehensively characterize PDAC tumors and NATs using multiomics as well as to provide proteogenomic resources to decipher the impacts of genomic alterations in gene expression, protein abundances, and PTMs. For these purposes, tissues collected by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) program are treatment-naïve and surgically resected. Consequently, there are inherent limitations to this study. First, although data on adjuvant patient treatment and outcome was sought, the present cohort comprises treatment-naïve samples, which limits extrapolation to metastatic disease treated with systemic therapy. While treatment data from therapeutic drug clinical trials are needed to investigate treatment outcomes related to the observed proteogenomic subtypes, currently, such data are limited in that clinical trials only generate transcriptomic data (O'Kane et al., 2020). Second, proteogenomic data provide rich resources for correlating different molecular alterations that are essential for hypothesis generation to decipher molecular functions or prediction of treatment options. However, causal effects of the correlations cannot be determined from this study. The biological hypothesis or treatment prediction would need further validation using cell lines, patient-derived xenograft (PDX) models, or clinical trials. Third, proteogenomic measurements of this study are deployed using bulk tumor and NAT tissues, where the impact of heterogeneity in cellularity and tumor microenvironment cannot be fully accounted for. Here, we addressed this limitation by selecting a subset of tissue samples with sufficient tumor cellularity for focused analyses. However, enrichment of tumor cellularity using laser capture microdissection or characterization of tissues through single-cell analyses would be beneficial (Elyada et al., 2019; Maurer et al., 2019).

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Human Subjects
 - Clinical Data Annotation
- METHOD DETAILS
 - Specimen Acquisition
 - Genomic, Epigenomic, and Transcriptomic Sample Preparation and Data Acquisition
 - PCR-Free Whole Genome Sequencing
 - Illumina Infinium Methylation EPIC Bead Chip Array
 - RNA Sequencing
 - Genomic, Epigenomic, and Transcriptomic Data Processing
 - Proteomic, Phosphoproteomic, and Glycoproteomic Sample Preparation and Data Acquisition
 - Proteomic, Phosphoproteomic, and Glycoproteomic Data Processing
 - Data Quality Control
 - Integrated Analysis
 - Kinase and Substrate Co-regulation
 - RNA Subtyping
 - Multi-omics Clustering
 - Inference of subtype-specific drug signatures (DSigDB GSEA)
 - Methylation-based Deconvolution
 - Tumor Microenvironment Inference
 - Adjustment for Epithelial Content
 - Over-representation Analysis
 - Survival Analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2021.08.023>.

CONSORTIA

The members of the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium are Mital Amin, Eunkyoung An, Christina Ayad, Oliver F. Bathe, Thomas Bauer, Chet Birger, Michael J. Birrer, Simina M. Boca, William Bociuk, Emily S. Boja, Melissa Borucki, Shuang Cai, Liwei Cao, Song Cao, Steven A. Carr, Sandra Cerda, Daniel W. Chan, Huan Chen, Lijun Chen, Steven Chen, David Chesla, Arul M. Chinnaian, David J. Clark, Antonio Colaprico, Sandra Cottingham, Daniel Cui Zhou, Felipe da Veiga Leprevost, Ludmila Danilova, Magdalena Derejiska, Saravana M. Dhanasekaran, Li Ding, Marcin J. Domagalski, Yongchao Dou, Brian J. Druker, Elizabeth Duffy, Maureen A. Dyer, Hariharan Easwaran, Nathan J. Edward, Matthew J. Ellis, Jennifer Eschbacher, David Fenyö, Alicia Francis, Jesse Francis, Stacey Gabriel, Nikolay Gabrovski, Johanna Gardner, Gad Getz, Michael A. Gillette, Charles A. Goldthwaite Jr, Pamela Grady, Shuai Guo, Benjamin Haibe-Kains, Pushpa Hariharan, Tara Hiltke, Barbara Hindenach, Katherine A. Hoadley, Galen Hostetter, Ralph H. Hruban, Yingwei Hu, Chen Huang, Jasmine Huang, Scott D. Jewell,

Wen Jiang, Corbin D. Jones, Karen A. Ketchum, Christopher R. Kinsinger, Jennifer M. Koziak, Karsten Krug, Katarzyna Kusnierz, Qing Kay Li, Yize Li, Yuxing Liao, T. Mamie Lih, Ruiyang Liu, Tao Liu, Wenke Liu, Jiang Long, Rita Jui-Hsien Lu, David Mallery, D. R. Mani, Sailaja Mareedu, Ronald Matteotti, Nicolle Maunganidze, Peter B. McGarvey, Arvind Singh Mer, Mehdi Mesri, Parham Minoo, Alexey I. Nesvizhskii, Chelsea J. Newton, Gilbert S. Omenn, Oxana V. Paklina, Jianbo Pan, Amanda G. Paulovich, Samuel H. Payne, Olga Potapova, Barbara Pruetz, Liqun Qi, Ana I. Robles, Nancy Roche, Karin D. Rodland, Henry Rodriguez, Michael H. Roehrl, Daniel C. Rohrer, Peter Ronning, Sara R. Savage, Eric E. Schadt, Michael Schnaubelt, Alexey V. Shabunin, Troy Shelton, Zhiao Shi, Yvonne Shutack, Shilpi Singh, Michael Smith, Richard D. Smith, James Suh, Nadezhda V. Terekhanova, Ratna R. Thangudu, Mathangi Thagarajan, Shirley X. Tsang, Ki Sung Um, Dana R. Valley, Negin Vatanian, Joshua M. Wang, Pei Wang, Wenyi Wang, Bo Wen, George D. Wilson, Maciej Wiznerowicz, Yige Wu, Matthew A. Wyczalkowski, Weiming Yang, Seungyeul Yoo, Bing Zhang, Hui Zhang, Zhen Zhang, Grace Zhao, and Houxiang Zhu.

ACKNOWLEDGMENTS

This work was supported by the National Cancer Institute (NCI) Clinical Proteomic Tumor Analysis Consortium (CPTAC), National Institutes of Health (NIH) grants U24CA210955, U24CA210985, U24CA210986, U24CA210954, U24CA210967, U24CA210972, U24CA210979, U24CA210993, U01CA214114, U01CA214116, and U01CA214125. A part of this research was funded through NIH grants P30CA008748 and P30ES017885, the Troper Wojcicki Foundation, the Rolfe Pancreatic Cancer Foundation, and the Evelyn Grollman Glick Scholar Award.

AUTHOR CONTRIBUTIONS

Conceptualization, H.R., O.F.B., D.W.C., R.H.H., L.D., B.Z., and H.Z.; investigation, L.W.C., Y.W.H., T.M.L., D.J.C., M.S., L.J.C., R.V.E., W.M.Y., and H.Z.; software, C.H., D.C.Z., Y.W.H., T.M.L., S.R.S., K.K., M.S., J.B.P., B.W., Y.C.D., F.d.V.L., Y.X.L., Z.A.S., D.F., P.W., D.R.M., A.I.N., L.D., B.Z., and H.Z.; data quality control, C.H., D.C.Z., Y.W.H., M.S., and B.W.; formal analysis, L.W.C., C.H., D.C.Z., Y.W.H., T.M.L., S.R.S., K.K., M.S., W.M.Y., W.J., N.V.T., S.C., R.J.-H.L., Y.Z.L., R.Y.L., H.X.Z., P.R., Y.G.W., M.A.W., H.E., L.D., A.S.M., S.Y., J.M.W., W.K.L., B.H.-K., F.d.V.L., Q.K.L., M.H.R., D.R.M., A.I.N., G.S.O., O.F.B., D.W.C., R.H.H., L.D., B.Z., and H.Z.; writing – original draft, L.W.C., C.H., D.C.Z., Y.W.H., T.M.L., S.R.S., K.K., D.J.C., R.H.H., B.Z., and H.Z.; writing – review & editing, all authors; supervision, D.R.M., A.I.N., O.F.B., D.W.C., R.H.H., L.D., B.Z., and H.Z.; project administration, E.S.B., M.M., A.I.R., and H.R.; funding acquisition, D.R.M., A.I.N., D.W.C., L.D., B.Z., and H.Z.

DECLARATION OF INTERESTS

R.H.H. has the potential of receiving royalty payments from Thrive Earlier Diagnosis for the GNAS invention in a relationship overseen by Johns Hopkins University. The remaining authors declare no competing interests.

Received: October 8, 2020

Revised: March 19, 2021

Accepted: August 18, 2021

Published: September 16, 2021

REFERENCES

- Akagi, J., Takai, E., Tamori, Y., Nakagawa, K., and Ogawa, M. (2001). CA19-9 epitope a possible marker for MUC-1/Y protein. *Int. J. Oncol.* **18**, 1085–1091.
- Allen, E., Jabouille, A., Rivera, L.B., Lodewijckx, I., Missiaen, R., Steri, V., Feyen, K., Tawney, J., Hanahan, D., Michael, I.P., and Bergers, G. (2017). Combined antiangiogenic and anti-PD-L1 therapy stimulates tumor immunity through HEV formation. *Sci. Transl. Med.* **9**, eaak9679.
- Annan, D.A., Kikuchi, H., Maishi, N., Hida, Y., and Hida, K. (2020). Tumor Endothelial Cell-A Biological Tool for Translational Cancer Research. *Int. J. Mol. Sci.* **21**, 3238.
- Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220.
- Arntfield, M.E., and van der Kooy, D. (2011). beta-Cell evolution: How the pancreas borrowed from the brain: The shared toolbox of genes expressed by neural and pancreatic endocrine cells may reflect their evolutionary relationship. *Bioessays* **33**, 582–587.
- Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* **526**, 68–74.
- Aznar, S., Valerón, P.F., del Rincon, S.V., Pérez, L.F., Perona, R., and Lacal, J.C. (2001). Simultaneous tyrosine and serine phosphorylation of STAT3 transcription factor is involved in Rho A GTPase oncogenic transformation. *Mol. Biol. Cell* **12**, 3282–3294.
- Babiceanu, M., Qin, F., Xie, Z., Jia, Y., Lopez, K., Janus, N., Facemire, L., Kumar, S., Pang, Y., Qi, Y., et al. (2016). Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res.* **44**, 2859–2872.
- Bailey, P., Chang, D.K., Nones, K., Johns, A.L., Patch, A.M., Gingras, M.C., Miller, D.K., Christ, A.N., Bruxner, T.J., Quinn, M.C., et al.; Australian Pancreatic Cancer Genome Initiative (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52.
- Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al.; MC3 Working Group; Cancer Genome Atlas Research Network (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371–385.e18.
- Balachandran, V.P., Beatty, G.L., and Dougan, S.K. (2019). Broadening the Impact of Immunotherapy to Pancreatic Cancer: Challenges and Opportunities. *Gastroenterology* **156**, 2056–2072.
- Balli, D., Rech, A.J., Stanger, B.Z., and Vonderheide, R.H. (2017). Immune Cytolytic Activity Stratifies Molecular Subsets of Human Pancreatic Cancer. *Clin. Cancer Res.* **23**, 3129–3138.
- Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112.
- Beauchemin, N., and Arabzadeh, A. (2013). Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev.* **32**, 643–671.
- Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elaroui, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., and de Reyniès, A. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218.
- Becker, A.E., Hernandez, Y.G., Frucht, H., and Lucas, A.L. (2014). Pancreatic ductal adenocarcinoma: risk factors, screening, and early detection. *World J. Gastroenterol.* **20**, 11182–11198.
- Beel, S., Kolloch, L., Apken, L.H., Jürgens, L., Bolle, A., Sudhof, N., Ghosh, S., Wardemann, E., Meisterernst, M., Steinestel, K., and Oeckinghaus, A. (2020). κ B-Ras and Ral GTPases regulate acinar to ductal metaplasia during pancreatic adenocarcinoma development and pancreatitis. *Nat. Commun.* **11**, 3409.
- Benelli, M., Pescucci, C., Marseglia, G., Severgnini, M., Torricelli, F., and Magi, A. (2012). Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics* **28**, 3232–3239.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. B Stat. Methodol.* **57**, 289–300.
- Berndt, N., Hamilton, A.D., and Sebti, S.M. (2011). Targeting protein prenylation for cancer therapy. *Nat. Rev. Cancer* **11**, 775–791.

- Bieberich, E. (2014). Synthesis, Processing, and Function of *N*-glycans in *N*-glycoproteins. *Adv. Neurobiol.* 9, 47–70.
- Blumenthal, R.D., Hansen, H.J., and Goldenberg, D.M. (2005a). Inhibition of adhesion, invasion, and metastasis by antibodies targeting CEACAM6 (NCA-90) and CEACAM5 (Carcinoembryonic Antigen). *Cancer Res.* 65, 8809–8817.
- Blumenthal, R.D., Osorio, L., Hayes, M.K., Horak, I.D., Hansen, H.J., and Goldenberg, D.M. (2005b). Carcinoembryonic antigen antibody inhibits lung metastasis and augments chemotherapy in a human colonic carcinoma xenograft. *Cancer Immunol. Immunother.* 54, 315–327.
- Boecker, W., Tiemann, K., Boecker, J., Toma, M., Muders, M.H., Löning, T., Buchwalow, I., Oldhafer, K.J., Neumann, U., Feyerabend, B., et al. (2020). Cellular organization and histogenesis of adenosquamous carcinoma of the pancreas: evidence supporting the squamous metaplasia concept. *Histochem. Cell Biol.* 154, 97–105.
- Brunet, J.P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. USA* 101, 4164–4169.
- Cai, J., Lu, W., Du, S., Guo, Z., Wang, H., Wei, W., and Shen, X. (2018). Tenascin-C Modulates Cell Cycle Progression to Enhance Tumour Cell Proliferation through AKT/FOXO1 Signalling in Pancreatic Cancer. *J. Cancer* 9, 4449–4462.
- Caldas, C., Hahn, S.A., da Costa, L.T., Redston, M.S., Schutte, M., Seymour, A.B., Weinstein, C.L., Hruban, R.H., Yeo, C.J., and Kern, S.E. (1994). Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nat. Genet.* 8, 27–32.
- Cancer Genome Atlas Research Network. Electronic address: andrew_aguirre@dfci.harvard.edu; Cancer Genome Atlas Research Network (2017). Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* 32, 185–203.e13.
- Cao, L., Diedrich, J.K., Ma, Y., Wang, N., Pauthner, M., Park, S.R., Delahunt, C.M., McLellan, J.S., Burton, D.R., Yates, J.R., and Paulson, J.C. (2018). Global site-specific analysis of glycoprotein *N*-glycan processing. *Nat. Protoc.* 13, 1196–1212.
- Chan-Seng-Yue, M., Kim, J.C., Wilson, G.W., Ng, K., Figueira, E.F., O’Kane, G.M., Connor, A.A., Denroche, R.E., Grant, R.C., McLeod, J., et al. (2020). Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat. Genet.* 52, 231–240.
- Chen, C.R., Kang, Y., Siegel, P.M., and Massagué, J. (2002). E2F4/5 and p107 as Smad cofactors linking the TGF β receptor to c-myc repression. *Cell* 110, 19–32.
- Christensen, J.G., Olson, P., Briere, T., Wiel, C., and Bergo, M.O. (2020). Targeting Kras g12c -mutant cancer with a mutation-specific inhibitor. *J. Intern. Med.* 288, 183–191.
- Cho, K.C., Clark, D.J., Schnaubelt, M., Teo, G.C., Leprevost, F.D.V., Bocik, W., Boja, E.S., Hiltke, T., Nesvizhskii, A.I., and Zhang, H. (2020). Deep Proteomics Using Two Dimensional Data Independent Acquisition Mass Spectrometry. *Anal. Chem.* 92, 4217–4225.
- Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219.
- Clark, D.J., Hu, Y., Bocik, W., Chen, L., Schnaubelt, M., Roberts, R., Shah, P., Whiteley, G., and Zhang, H. (2018). Evaluation of NCI-7 Cell Line Panel as a Reference Material for Clinical Proteomics. *J. Proteome Res.* 17, 2205–2215.
- Clark, D.J., Dhanasekaran, S.M., Petralia, F., Pan, J., Song, X., Hu, Y., da Veiga Leprevost, F., Reva, B., Lih, T.M., Chang, H.Y., et al.; Clinical Proteomic Tumor Analysis Consortium (2019). Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* 179, 964–983.e31.
- Collisson, E.A., Bailey, P., Chang, D.K., and Biankin, A.V. (2019). Molecular subtypes of pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* 16, 207–220.
- Collisson, E.A., Sadanandam, A., Olson, P., Gibb, W.J., Truitt, M., Gu, S., Cooc, J., Weinkle, J., Kim, G.E., Jakkula, L., et al. (2011). Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* 17, 500–503.
- Conroy, T., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.L., Gourgou-Bourgade, S., de la Foucardière, C., et al.; Groupe Tumeurs Digestives de l’Institut National du Cancer; PRODIGE Intergroup (2011). FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N. Engl. J. Med.* 364, 1817–1825.
- da Veiga Leprevost, F., Haynes, S.E., Avtonomov, D.M., Chang, H.Y., Shanmugam, A.K., Mellacheruvu, D., Kong, A.T., and Nesvizhskii, A.I. (2020). Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods* 17, 869–870.
- Daneman, R., and Prat, A. (2015). The blood-brain barrier. *Cold Spring Harb. Perspect. Biol.* 7, a020412.
- de la Cruz-López, K.G., Castro-Muñoz, L.J., Reyes-Hernández, D.O., García-Carrancá, A., and Manzo-Merino, J. (2019). Lactate in the Regulation of Tumor Microenvironment and Therapeutic Approaches. *Front. Oncol.* 9, 1143.
- De la Mota-Peynado, A., Chernoff, J., and Beeser, A. (2011). Identification of the atypical MAPK Erk3 as a novel substrate for p21-activated kinase (Pak) activity. *J. Biol. Chem.* 286, 13603–13611.
- Dennler, S., Itoh, S., Vivien, D., ten Dijke, P., Huet, S., and Gauthier, J.M. (1998). Direct binding of Smad3 and Smad4 to critical TGF beta-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. *EMBO J.* 17, 3091–3100.
- Djomehri, S.I., Gonzalez, M.E., da Veiga Leprevost, F., Tekula, S.R., Chang, H.Y., White, M.J., Cimino-Mathews, A., Burman, B., Basrur, V., Argani, P., et al. (2020). Quantitative proteomic landscape of metaplastic breast carcinoma pathological subtypes and their relationship to triple-negative tumors. *Nat. Commun.* 11, 1723.
- Duan, G., Li, X., and Köhn, M. (2015). The human DEPhosphorylation database DEPOD: a 2015 update. *Nucleic Acids Res.* 43, D531–D535.
- Ducreux, M., Seufferlein, T., Van Laethem, J.L., Laurent-Puig, P., Smolenschi, C., Malka, D., Boige, V., Hollebecque, A., and Conroy, T. (2019). Systemic treatment of pancreatic cancer revisited. *Semin. Oncol.* 46, 28–38.
- Elia, A.R., Grioni, M., Basso, V., Curnis, F., Freschi, M., Corti, A., Mondino, A., and Bellone, M. (2018). Targeting Tumor Vasculature with TNF Leads Effector T Cells to the Tumor and Enhances Therapeutic Efficacy of Immune Checkpoint Blockers in Combination with Adoptive Cell Therapy. *Clin. Cancer Res.* 24, 2171–2181.
- Elyada, E., Bolisetty, M., Laise, P., Flynn, W.F., Courtois, E.T., Burkhardt, R.A., Teinor, J.A., Belleau, P., Biffi, G., Lucito, M.S., et al. (2019). Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer Discov.* 9, 1102–1123.
- Eser, S., Schnieke, A., Schneider, G., and Saur, D. (2014). Oncogenic KRAS signalling in pancreatic cancer. *Br. J. Cancer* 111, 817–822.
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 46 (D1), D649–D655.
- Fan, G. (2020). FER mediated HGF-independent regulation of HGFR/MET activates RAC1-PAK1 pathway to potentiate metastasis in ovarian cancer. *Small GTPases* 11, 155–159.
- Ferrara, N., Hillan, K.J., Gerber, H.P., and Novotny, W. (2004). Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat. Rev. Drug Discov.* 3, 391–400.
- Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T.M., Young, G., Fennell, T.J., Allen, A., Ambrogio, L., et al. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* 12, R1.
- Furuse, J., and Nagashima, F. (2017). Emerging protein kinase inhibitors for treating pancreatic cancer. *Expert Opin. Emerg. Drugs* 22, 77–86.
- Gao, Q., Liang, W.W., Foltz, S.M., Mutharasu, G., Jayasinghe, R.G., Cao, S., Liao, W.W., Reynolds, S.M., Wyczalkowski, M.A., Yao, L., et al.; Fusion Analysis Working Group; Cancer Genome Atlas Research Network (2018). Driver

- Fusions and Their Implications in the Development and Treatment of Human Cancers. *Cell Rep.* 23, 227–238.e3.
- Garnham, R., Scott, E., Livermore, K.E., and Munkley, J. (2019). ST6GAL1: A key player in cancer. *Oncol. Lett.* 18, 983–989.
- Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 11, 367.
- Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia, F., Li, Y., Liang, W.W., Revu, B., et al.; Clinical Proteomic Tumor Analysis Consortium (2020). Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. *Cell* 182, 200–225.e35.
- Goggins, M., Offerhaus, G.J., Hilgers, W., Griffin, C.A., Shekher, M., Tang, D., Sohn, T.A., Yeo, C.J., Kern, S.E., and Hruban, R.H. (1998). Pancreatic adenocarcinomas with DNA replication errors (RER+) are associated with wild-type K-ras and characteristic histopathology. Poor differentiation, a syncytial growth pattern, and pushing borders suggest RER+. *Am. J. Pathol.* 152, 1501–1507.
- Golan, T., Hammel, P., Reni, M., Van Cutsem, E., Macarulla, T., Hall, M.J., Park, J.O., Hochhauser, D., Arnold, D., Oh, D.Y., et al. (2019). Maintenance Olaparib for Germline BRCA-Mutated Metastatic Pancreatic Cancer. *N. Engl. J. Med.* 381, 317–327.
- Gomis, R.R., Alarcón, C., Nadal, C., Van Poznak, C., and Massagué, J. (2006). C/EBPbeta at the core of the TGFbeta cytostatic response and its evasion in metastatic breast cancer cells. *Cancer Cell* 10, 203–214.
- Gonzalez, C., Sims, J.S., Hornstein, N., Mela, A., Garcia, F., Lei, L., Gass, D.A., Amendolara, B., Bruce, J.N., Canoll, P., and Sims, P.A. (2014). Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors. *J. Neurosci.* 34, 10924–10936.
- González-Borja, I., Viúdez, A., Goñi, S., Santamaría, E., Carrasco-García, E., Pérez-Sanz, J., Hernández-García, I., Sala-Elarre, P., Arrazubi, V., Oyaga-Iriarte, E., et al. (2019). Omics Approaches in Pancreatic Adenocarcinoma. *Cancers (Basel)* 11, 1052.
- Govindan, S.V., Cardillo, T.M., Moon, S.J., Hansen, H.J., and Goldenberg, D.M. (2009). CEACAM5-targeted therapy of human colonic and pancreatic cancer xenografts with potent labetuzumab-SN-38 immunoconjugates. *Clin. Cancer Res.* 15, 6052–6061.
- Haas, B.J., Dobin, A., Li, B., Stransky, N., Pochet, N., and Regev, A. (2019). Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol.* 20, 213.
- Hänelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 14, 7.
- Harsha, H.C., Kandasamy, K., Ranganathan, P., Rani, S., Ramabadran, S., Gollapudi, S., Balakrishnan, L., Dwivedi, S.B., Telikicherla, D., Selvan, L.D., et al. (2009). A compendium of potential biomarkers of pancreatic cancer. *PLoS Med.* 6, e1000046.
- Hart, G.W., and Copeland, R.J. (2010). Glycomics hits the big time. *Cell* 143, 672–676.
- Herr, P., Boström, J., Rullman, E., Rudd, S.G., Vesterlund, M., Lehtiö, J., Helleday, T., Maddalo, G., and Altun, M. (2020). Cell Cycle Profiling Reveals Protein Oscillation, Phosphorylation, and Localization Dynamics. *Mol. Cell. Proteomics* 19, 608–623.
- Hessmann, E., Buchholz, S.M., Demir, I.E., Singh, S.K., Gress, T.M., Ellnerrieder, V., and Neesse, A. (2020). Microenvironmental Determinants of Pancreatic Cancer. *Physiol. Rev.* 100, 1707–1751.
- Hilmi, M., Bartholin, L., and Neuzillet, C. (2018). Immune therapies in pancreatic ductal adenocarcinoma: Where are we now? *World J. Gastroenterol.* 24, 2137–2151.
- Ho, W.J., Jaffee, E.M., and Zheng, L. (2020). The tumour microenvironment in pancreatic cancer - clinical challenges and opportunities. *Nat. Rev. Clin. Oncol.* 17, 527–540.
- Hobbs, G.A., Baker, N.M., Miermont, A.M., Thurman, R.D., Pierobon, M., Tran, T.H., Anderson, A.O., Waters, A.M., Diehl, J.N., Papke, B., et al. (2020). Atypical KRAS^{G12R} Mutant Is Impaired in PI3K Signaling and Macropinocytosis in Pancreatic Cancer. *Cancer Discov.* 10, 104–123.
- Hollingsworth, M.A., and Swanson, B.J. (2004). Mucins in cancer: protection and control of the cell surface. *Nat. Rev. Cancer* 4, 45–60.
- Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 43, D512–D520.
- Hruban, R.H., Gaida, M.M., Thompson, E., Hong, S.M., Noë, M., Brosens, L.A., Jongepieter, M., Offerhaus, G.J.A., and Wood, L.D. (2019). Why is pancreatic cancer so deadly? The pathologist's view. *J. Pathol.* 248, 131–141.
- Hsieh, C.C., Shyr, Y.M., Liao, W.Y., Chen, T.H., Wang, S.E., Lu, P.C., Lin, P.Y., Chen, Y.B., Mao, W.Y., Han, H.Y., et al. (2017). Elevation of β-galactoside α2,6-sialyltransferase 1 in a fructoseresponsive manner promotes pancreatic cancer metastasis. *Oncotarget* 8, 7691–7709.
- Hu, Y., Shah, P., Clark, D.J., Ao, M., and Zhang, H. (2018). Reanalysis of Global Proteomic and Phosphoproteomic Data Identified a Large Number of Glycopeptides. *Anal. Chem.* 90, 8065–8071.
- Huang, K.L., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., et al.; Cancer Genome Atlas Research Network (2018). Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell* 173, 355–370.e14.
- Huang, C., Chen, L., Savage, S.R., Eguez, R.V., Dou, Y., Li, Y., da Veiga Leprevost, F., Jaehnig, E.J., Lei, J.T., Wen, B., et al.; Clinical Proteomic Tumor Analysis Consortium (2021). Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell* 39, 361–379.e16.
- Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121.
- Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* 166, 740–754.
- Janes, M.R., Zhang, J., Li, L.S., Hansen, R., Peters, U., Guo, X., Chen, Y., Barber, A., Firdaus, S.J., Darjania, L., et al. (2018). Targeting KRAS Mutant Cancers with a Covalent G12C-Specific Inhibitor. *Cell* 172, 578–589.e17.
- Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., et al. (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321, 1801–1806.
- Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al.; Genome Aggregation Database Consortium (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443.
- Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383–5392.
- Kim, H., and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* 23, 1495–1502.
- Kim, J., Bamlet, W.R., Oberg, A.L., Chaffee, K.G., Donahue, G., Cao, X.J., Chari, S., Garcia, B.A., Petersen, G.M., and Zaret, K.S. (2017). Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19-9 blood markers. *Sci. Transl. Med.* 9, eaah5583.
- Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., and Saunders, C.T. (2018). Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* 15, 591–594.
- Klein, D. (2018). The Tumor Vascular Endothelium as Decision Maker in Cancer Therapy. *Front. Oncol.* 8, 367.
- Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576.

- Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* 14, 513–520.
- Krug, K., Mertins, P., Zhang, B., Hornbeck, P., Raju, R., Ahmad, R., Szucs, M., Mundt, F., Forestier, D., Jane-Valbuena, J., et al. (2019). A Curated Resource for Phosphosite-specific Signature Analysis. *Mol. Cell. Proteomics* 18, 576–593.
- Krug, K., Jaehnig, E.J., Satpathy, S., Blumenberg, L., Karpova, A., Anurag, M., Miles, G., Mertins, P., Geffen, Y., Tang, L.C., et al.; Clinical Proteomic Tumor Analysis Consortium (2020). Proteogenomic Landscape of Breast Cancer Tumorigenesis and Targeted Therapy. *Cell* 183, 1436–1456.e31.
- Lee, H., Jeong, A.J., and Ye, S.K. (2019). Highlighted STAT3 as a potential drug target for cancer therapy. *BMB Rep.* 52, 415–423.
- Lenkiewicz, E., Malasi, S., Hogenson, T.L., Flores, L.F., Barham, W., Phillips, W.J., Roesler, A.S., Chambers, K.R., Rajbhandari, N., Hayashi, A., et al. (2020). Genomic and Epigenomic Landscaping Defines New Therapeutic Targets for Adenosquamous Carcinoma of the Pancreas. *Cancer Res.* 80, 4324–4334.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, M., Xie, X., Zhou, J., Sheng, M., Yin, X., Ko, E.A., Zhou, T., and Gu, W. (2017). Quantifying circular RNA expression from RNA-seq data using model-based framework. *Bioinformatics* 33, 2131–2139.
- Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 47 (W1), W199–W205.
- Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1, 417–425.
- Lih, T.M., Clark, D.J., and Zhang, H. (2019). Omic-Sig: Utilizing Omics Data to Explore and Visualize Kinase-Substrate Interactions. *bioRxiv*. <https://doi.org/10.1101/746123>.
- Liu, J., Bell, A.W., Bergeron, J.J., Yanofsky, C.M., Carrillo, B., Beaudrie, C.E., and Kearney, R.E. (2007). Methods for peptide identification by spectral comparison. *Proteome Sci.* 5, 3.
- Lurie, E., Liu, D., LaPlante, E.L., Thistlithwaite, L.R., Yao, Q., and Milosavljevic, A. (2020). Histoepigenetic analysis of the mesothelin network within pancreatic ductal adenocarcinoma cells reveals regulation of retinoic acid receptor gamma and AKT by mesothelin. *Oncogenesis* 9, 62.
- Macaulay, M.S., Arlian, B.M., Rillahan, C.D., Pang, P.C., Bortell, N., Marcondes, M.C., Haslam, S.M., Dell, A., and Paulson, J.C. (2014). Systemic blockade of sialylation in mice with a global inhibitor of sialyltransferases. *J. Biol. Chem.* 289, 35149–35158.
- Manoli, S., Coppola, S., Duranti, C., Lulli, M., Magni, L., Kuppala, N., Nielsen, N., Schmidt, T., Schwab, A., Beccetti, A., and Arcangeli, A. (2019). The Activity of Kv 11.1 Potassium Channel Modulates F-Actin Organization During Cell Migration of Pancreatic Ductal Adenocarcinoma Cells. *Cancers (Basel)* 11, 135.
- Maurer, C., Holmstrom, S.R., He, J., Laise, P., Su, T., Ahmed, A., Hibshoosh, H., Chabot, J.A., Oberstein, P.E., Sepulveda, A.R., et al. (2019). Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes. *Gut* 68, 1034–1043.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12, R41.
- Mertins, P., Tang, L.C., Krug, K., Clark, D.J., Gritsenko, M.A., Chen, L., Clauer, K.R., Clauss, T.R., Shah, P., Gillette, M.A., et al. (2018). Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. *Nat. Protoc.* 13, 1632–1661.
- Moffitt, R.A., Marayati, R., Flate, E.L., Volmar, K.E., Loeza, S.G., Hoadley, K.A., Rashid, N.U., Williams, L.A., Eaton, S.C., Chung, A.H., et al. (2015). Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat. Genet.* 47, 1168–1178.
- Montagnoli, A., Valsasina, B., Brotherton, D., Troiani, S., Rainoldi, S., Tenca, P., Molinari, A., and Santocanale, C. (2006). Identification of Mcm2 phosphorylation sites by S-phase-regulating kinases. *J. Biol. Chem.* 281, 10281–10290.
- Najahi-Missaoui, W., Quach, N.D., Jenkins, A., Dabke, I., Somanath, P.R., and Cummings, B.S. (2019). Effect of P21-activated kinase 1 (PAK-1) inhibition on cancer cell growth, migration, and invasion. *Pharmacol. Res. Perspect.* 7, e00518.
- Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75, 4646–4658.
- O'Kane, G.M., Grünwald, B.T., Jang, G.H., Masoomian, M., Picardo, S., Grant, R.C., Denroche, R.E., Zhang, A., Wang, Y., Lam, B., et al. (2020). GATA6 Expression Distinguishes Classical and Basal-like Subtypes in Advanced Pancreatic Cancer. *Clin. Cancer Res.* 26, 4901–4910.
- Onuchic, V., Hartmaier, R.J., Boone, D.N., Samuels, M.L., Patel, R.Y., White, W.M., Garovic, V.D., Oesterreich, S., Roth, M.E., Lee, A.V., and Milosavljevic, A. (2016). Epigenomic Deconvolution of Breast Tumors Reveals Metabolic Coupling between Constituent Cell Types. *Cell Rep.* 17, 2075–2086.
- Pandey, R., Zhou, M., Islam, S., Chen, B., Barker, N.K., Langlais, P., Srivastava, A., Luo, M., Cooke, L.S., Weterings, E., and Mahadevan, D. (2019). Carcinoembryonic antigen cell adhesion molecule 6 (CEACAM6) in Pancreatic Ductal Adenocarcinoma (PDA): An integrative analysis of a novel therapeutic target. *Sci. Rep.* 9, 18347.
- Pellicano, H., Martin, D.S., Xu, R.H., and Huang, P. (2006). Glycolysis inhibition for anticancer treatment. *Oncogene* 25, 4633–4646.
- Pelosi, E., Castelli, G., and Testa, U. (2017). Pancreatic Cancer: Molecular Characterization, Clonal Evolution and Cancer Stem Cells. *Biomedicines* 5, 65.
- Perfetto, L., Briganti, L., Calderone, A., Cerquone Perpetuini, A., Iannuccelli, M., Langone, F., Licata, L., Marinkovic, M., Mattioni, A., Pavlidou, T., et al. (2016). SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res.* 44 (D1), D548–D554.
- Petrova, V., Annicchiarico-Petruzzelli, M., Melino, G., and Amelio, I. (2018). The hypoxic tumour microenvironment. *Oncogenesis* 7, 10.
- Polzien, L., Baljuls, A., Albrecht, M., Hekman, M., and Rapp, U.R. (2011). BAD contributes to RAF-mediated proliferation and cooperates with B-RAF-V600E in cancer signaling. *J. Biol. Chem.* 286, 17934–17944.
- Puleo, F., Nicolle, R., Blum, Y., Cros, J., Marisa, L., Demetter, P., Quertinmont, E., Svrcik, M., Elarouci, N., Iovanna, J., et al. (2018). Stratification of Pancreatic Ductal Adenocarcinomas Based on Tumor and Microenvironment Features. *Gastroenterology* 155, 1999–2013.e3.
- Quante, A.S., Ming, C., Rottmann, M., Engel, J., Boeck, S., Heinemann, V., Westphalen, C.B., and Strauch, K. (2016). Projections of cancer incidence and cancer-related deaths in Germany by 2020 and 2030. *Cancer Med.* 5, 2649–2656.
- Radeva, M.Y., and Waschke, J. (2018). Mind the gap: mechanisms regulating the endothelial barrier. *Acta Physiol. (Oxf.)* 222. <https://doi.org/10.1111/apha.12860>.
- Rahib, L., Smith, B.D., Aizenberg, R., Rosenzweig, A.B., Fleshman, J.M., and Matrisian, L.M. (2014). Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* 74, 2913–2921.
- Rane, C.K., and Minden, A. (2019). P21 activated kinase signaling in cancer. *Semin. Cancer Biol.* 54, 40–49.

- Ranzinger, R., Herget, S., von der Lieth, C.W., and Frank, M. (2011). Glyco-meDB—a unified database for carbohydrate structures. *Nucleic Acids Res.* 39, D373–D376.
- Ren, L., Yi, J., Li, W., Zheng, X., Liu, J., Wang, J., and Du, G. (2019). Apolipoproteins and cancer. *Cancer Med.* 8, 7032–7043.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
- Roth, M.T., Cardin, D.B., and Berlin, J.D. (2020). Recent advances in the treatment of pancreatic cancer. *F1000Res.* 9, 131.
- Ruckert, M.T., de Andrade, P.V., Santos, V.S., and Silveira, V.S. (2019). Protein tyrosine phosphatases: promising targets in pancreatic ductal adenocarcinoma. *Cell. Mol. Life Sci.* 76, 2571–2592.
- Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.* 38, D497–D501.
- Sapalidis, K., Kosmidis, C., Funtanidou, V., Katsaounis, A., Barmpas, A., Koimtzis, G., Mantalobas, S., Alexandrou, V., Aidoni, Z., Koulouris, C., et al. (2019). Update on current pancreatic treatments: from molecular pathways to treatment. *J. Cancer* 10, 5162–5172.
- Sava, G.P., Fan, H., Coombes, R.C., Buluwela, L., and Ali, S. (2020). CDK7 inhibitors as anticancer drugs. *Cancer Metastasis Rev.* 39, 805–823.
- Schaaf, M.B., Garg, A.D., and Agostinis, P. (2018). Defining the role of the tumor vasculature in antitumor immunity and immunotherapy. *Cell Death Dis.* 9, 115.
- Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., Garnett, M.J., Blüthgen, N., and Saez-Rodriguez, J. (2018). Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* 9, 20.
- Scott, A.D., Huang, K.L., Weerasinghe, A., Mashl, R.J., Gao, Q., Martins Rodrigues, F., Wyczalkowski, M.A., and Ding, L. (2019). CharGer: clinical Characterization of Germline variants. *Bioinformatics* 35, 865–867.
- Semenova, G., and Chernoff, J. (2017). Targeting PAK1. *Biochem. Soc. Trans.* 45, 79–88.
- Shteynberg, D.D., Deutsch, E.W., Campbell, D.S., Hoopmann, M.R., Kusebauch, U., Lee, D., Mendoza, L., Midha, M.K., Sun, Z., Whetton, A.D., and Moritz, R.L. (2019). PTMProphet: Fast and Accurate Mass Modification Localization for the Trans-Proteomic Pipeline. *J. Proteome Res.* 18, 4262–4272.
- Singhi, A.D., Koay, E.J., Chari, S.T., and Maitra, A. (2019). Early Detection of Pancreatic Cancer: Opportunities and Challenges. *Gastroenterology* 156, 2024–2040.
- Stevenson, R.P., Veltman, D., and Machesky, L.M. (2012). Actin-bundling proteins in cancer progression at a glance. *J. Cell Sci.* 125, 1073–1079.
- Strickland, L.A., Ross, J., Williams, S., Ross, S., Romero, M., Spencer, S., Erickson, R., Sutcliffe, J., Verbeke, C., Polakis, P., et al. (2009). Preclinical evaluation of carcinoembryonic cell adhesion molecule (CEACAM) 6 as potential therapy target for pancreatic adenocarcinoma. *J. Pathol.* 218, 380–390.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Sun, S., Hu, Y., Ao, M., Shah, P., Chen, J., Yang, W., Jia, X., Tian, Y., Thomas, S., and Zhang, H. (2019). *N*-GlycositeAtlas: a database resource for mass spectrometry-based human N-linked glycoprotein and glycosylation site mapping. *Clin. Proteomics* 16, 35.
- Suzuki, O. (2019). Glycosylation in lymphoma: Biology and glycotherapy. *Pathol. Int.* 69, 441–449.
- Tao, H., Liu, S., Huang, D., Han, X., Wu, X., Shao, Y.W., and Hu, Y. (2020). Acquired multiple secondary *BRCA2* mutations upon PARPi resistance in a metastatic pancreatic cancer patient harboring a *BRCA2* germline mutation. *Am. J. Transl. Res.* 12, 612–617.
- The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45 (D1), D158–D169.
- Thomas, J.K., Kim, M.S., Balakrishnan, L., Nanjappa, V., Raju, R., Marimuthu, A., Radhakrishnan, A., Muthusamy, B., Khan, A.A., Sakamuri, S., et al. (2014). Pancreatic Cancer Database: an integrative resource for pancreatic cancer. *Cancer Biol. Ther.* 15, 963–967.
- Thompson, E.D., Roberts, N.J., Wood, L.D., Eshleman, J.R., Goggins, M.G., Kern, S.E., Klein, A.P., and Hruban, R.H. (2020). The genetics of ductal adenocarcinoma of the pancreas in the year 2020: dramatic progress, but far to go. *Mod. Pathol.* 33, 2544–2563.
- Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. *Science* 356, eaal3321.
- Toghi Eshghi, S., Yang, W., Hu, Y., Shah, P., Sun, S., Li, X., and Zhang, H. (2016). Classification of Tandem Mass Spectra for Identification of *N*- and *O*-linked Glycopeptides. *Sci. Rep.* 6, 37189.
- Tsou, C.C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A.C., and Nesvizhskii, A.I. (2015). DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* 12, 258–264.
- Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* 13, 966–967.
- Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardindoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419.
- Upadhyay, D., and Adjei, A.A. (2020). KRAS: From undruggable to a druggable Cancer Target. *Cancer Treat. Rev.* 89, 102070.
- Vajaria, B.N., Patel, K.R., Begum, R., and Patel, P.S. (2016). Sialylation: an Avenue to Target Cancer Cells. *Pathol. Oncol. Res.* 22, 443–447.
- van Erning, F.N., Mackay, T.M., van der Geest, L.G.M., Groot Koerkamp, B., van Laarhoven, H.W.M., Bonsing, B.A., Wilmink, J.W., van Santvoort, H.C., de Vos-Geelen, J., van Eijck, C.H.J., et al.; Dutch Pancreatic Cancer Group (2018). Association of the location of pancreatic ductal adenocarcinoma (head, body, tail) with tumor stage, treatment, and survival: a population-based analysis. *Acta Oncol.* 57, 1655–1662.
- Varki, A. (2017). Biological roles of glycans. *Glycobiology* 27, 3–49.
- Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B., Dou, Y., Zhang, Y., Shi, Z., Arshad, O.A., et al.; Clinical Proteomic Tumor Analysis Consortium (2019). Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* 177, 1035–1049.e19.
- Vatansever, S., Erman, B., and Gümus, Z.H. (2020). Comparative effects of oncogenic mutations G12C, G12V, G13D, and Q61H on local conformations and dynamics of K-Ras. *Comput. Struct. Biotechnol. J.* 18, 1000–1011.
- Von Hoff, D.D., Ervin, T., Arena, F.P., Chiorean, E.G., Infante, J., Moore, M., Seay, T., Tjulandin, S.A., Ma, W.W., Saleh, M.N., et al. (2013). Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *N. Engl. J. Med.* 369, 1691–1703.
- Wang, S.C., Huang, C.C., Shen, C.H., Lin, L.C., Zhao, P.W., Chen, S.Y., Deng, Y.C., and Liu, Y.W. (2016). Gene Expression and DNA Methylation Status of Glutathione S-Transferase Mu1 and Mu5 in Urothelial Carcinoma. *PLoS ONE* 11, e0159102.
- Wang, L.B., Karpova, A., Gritsenko, M.A., Kyle, J.E., Cao, S., Li, Y., Rykunov, D., Colaprico, A., Rothstein, J.H., Hong, R., et al.; Clinical Proteomic Tumor Analysis Consortium (2021). Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* 39, 509–528.e20.
- Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M.; Cancer Genome Atlas Research Network (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120.
- Wenger, C.D., and Coon, J.J. (2013). A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. *J. Proteome Res.* 12, 1377–1386.

- Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26, 1572–1573.
- Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46 (D1), D1074–D1082.
- Wojcik, E.J., Sharifpoor, S., Miller, N.A., Wright, T.G., Watering, R., Tremblay, E.A., Swan, K., Mueller, C.R., and Elliott, B.E. (2006). A novel activating function of c-Src and Stat3 on HGF transcription in mammary carcinoma cells. *Oncogene* 25, 2773–2784.
- Xi, R., Lee, S., Xia, Y., Kim, T.M., and Park, P.J. (2016). Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants. *Nucleic Acids Res.* 44, 6274–6286.
- Yang, Y. (2015). Cancer immunotherapy: harnessing the immune system to battle cancer. *J. Clin. Invest.* 125, 3335–3337.
- Yang, G., Hu, Y., Sun, S., Ouyang, C., Yang, W., Wang, Q., Betenbaugh, M., and Zhang, H. (2018a). Comprehensive Glycoproteomic Analysis of Chinese Hamster Ovary Cells. *Anal. Chem.* 90, 14294–14302.
- Yang, W., Ao, M., Hu, Y., Li, Q.K., and Zhang, H. (2018b). Mapping the O-glycoproteome using site-specific extraction of O-linked glycopeptides (EXoO). *Mol. Syst. Biol.* 14, e8486.
- Yang, W., Song, A., Ao, M., Xu, Y., and Zhang, H. (2020). Large-scale site-specific mapping of the O-GalNAc glycoproteome. *Nat. Protoc.* 15, 2589–2610.
- Ye, D.Z., and Field, J. (2012). PAK signaling in cancer. *Cell. Logist.* 2, 105–116.
- Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25, 2865–2871.
- Yeo, D., He, H., Patel, O., Lowy, A.M., Baldwin, G.S., and Nikfarjam, M. (2016). FRAX597, a PAK1 inhibitor, synergistically reduces pancreatic cancer growth when combined with gemcitabine. *BMC Cancer* 16, 24.
- Yoo, M., Shin, J., Kim, J., Ryall, K.A., Lee, K., Lee, S., Jeon, M., Kang, J., and Tan, A.C. (2015). DSigDB: drug signatures database for gene set analysis. *Bioinformatics* 31, 3069–3071.
- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612.
- Yu, F., Teo, G.C., Kong, A.T., Haynes, S.E., Avtonomov, D.M., Geisler, D.J., and Nesvizhskii, A.I. (2020). Identification of modified peptides using localization-aware open search. *Nat. Commun.* 11, 4065.
- Yuan, J., Zhang, F., and Niu, R. (2015). Multiple regulation pathways and pivotal biological functions of STAT3 in cancer. *Sci. Rep.* 5, 17663.
- Yue, T., Partyka, K., Maupin, K.A., Hurley, M., Andrews, P., Kaul, K., Moser, A.J., Zeh, H., Brand, R.E., and Haab, B.B. (2011). Identification of blood-protein carriers of the CA 19-9 antigen and characterization of prevalence in pancreatic diseases. *Proteomics* 11, 3665–3674.
- Yuen, A., and Díaz, B. (2014). The impact of hypoxia in pancreatic cancer invasion and metastasis. *Hypoxia (Auckl)* 2, 91–106.
- Zhang, H., Li, X.J., Martin, D.B., and Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* 21, 660–666.
- Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, B., McDermott, J.E., Zhou, J.Y., Petyuk, V.A., Chen, L., Ray, D., et al.; CPTAC Investigators (2016a). Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* 166, 755–765.
- Zhang, J., White, N.M., Schmidt, H.K., Fulton, R.S., Tomlinson, C., Warren, W.C., Wilson, R.K., and Maher, C.A. (2016b). INTEGRATE: gene fusion discovery using whole genome and transcriptome data. *Genome Res.* 26, 108–118.
- Zhang, J., Wu, Y., Hu, X., Wang, B., Wang, L., Zhang, S., Cao, J., and Wang, Z. (2017). GSTT1, GSTP1, and GSTM1 genetic variants are associated with survival in previously untreated metastatic breast cancer. *Oncotarget* 8, 105905–105914.
- Zhou, W., Jubb, A.M., Lyle, K., Xiao, Q., Ong, C.C., Desai, R., Fu, L., Gnad, F., Song, Q., Haverty, P.M., et al. (2014). PAK1 mediates pancreatic cancer cell migration and resistance to MET inhibition. *J. Pathol.* 234, 502–513.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Sodium chloride	Santa Cruz Biotechnology	Catalog: sc-295833
Tris(hydroxymethyl)aminomethane	Invitrogen	Catalog: AM9855G
Ethylenediaminetetraacetic acid	Sigma	Catalog: E7889
Aprotinin	Sigma	Catalog: A6103
Leupeptin	Roche	Catalog: 11017101001
Phenylmethylsulfonyl fluoride	Sigma	Catalog: 93482
Sodium fluoride	Sigma	Catalog: S7920
Phosphatase Inhibitor Cocktail 2	Sigma	Catalog: P5726
Phosphatase Inhibitor Cocktail 3	Sigma	Catalog: P0044
Urea	Sigma	Catalog: U0631
PUGNAc	Sigma	Catalog: A7229
Dithiothreitol	Thermo Fisher Scientific	Catalog: 20291
Iodoacetamide	Thermo Fisher Scientific	Catalog: A3221
Lysyl endopeptidase, Mass Spectrometry Grade	Wako Chemicals	Catalog: 125-05061
Sequencing grade modified trypsin	Promega	Catalog: V511X
Formic acid	Fisher Chemical	Catalog: A117-50
C18 SepPak	Waters	Catalog: WAT054925
Oasis MAX 1 cc Vac Cartridge, 30 mg Sorbent per Cartridge	Waters	Catalog: 186000366
Tandem mass tags – 10plex	Thermo Fisher Scientific	Catalog: 90406
TMT11-131C label reagent	Thermo Fisher Scientific	Catalog: A34807
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	Alfa Aesar	Catalog: J63218
Acetonitrile, Optima LC/MS	Fisher Chemical	Catalog: A955-4
Water, Optima LC/MS	Fisher Chemical	Catalog: W6-4
Anhydrous acetonitrile	Sigma	Catalog: 271004
Hydroxylamine solution	Sigma	Catalog: 467804
Ammonium Hydroxide solution	Sigma	Catalog: 338818
Polymer-Strong Cation Exchange	Thomas Scientific	Catalog: 1141Z03
CDS Empore C18 Extraction Disks	Chromatographic Specialties	Catalog: CD2215
Ni-NTA agarose beads	QIAGEN	Catalog: 30410
Iron (III) chloride	Sigma	Catalog: 451649
Trifluoroacetic acid	Sigma	Catalog: 302031
Triethylammonium acetate buffer	Sigma	Catalog: 90358
Critical commercial assays		
BCA Protein Assay Kit	Thermo Fisher Scientific	Catalog: 23225
Deposited data		
Proteomic Data Commons	National Institutes of Health	PDC: https://pdc.cancer.gov/pdc/
Genomic Data Commons	National Institutes of Health	GDC: https://gdc.cancer.gov/
LinkedOmics	Zhang Lab	LinkedOmics: http://www.linkedomics.org/data_download/CPTAC-PDAC/
PhosphositePlus	Hornbeck et al., 2015	https://www.phosphosite.org/homeAction.action

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
The Human Protein Atlas	(Thul et al., 2017)	www.proteinatlas.org
UniprotKB	The UniProt Consortium, 2017	https://www.uniprot.org/
Genomics of Drug Sensitivity in Cancer	Iorio et al., 2016	https://www.cancerrxgene.org/
Software and algorithms		
MS-PyCloud	https://www.biorxiv.org/content/10.1101/320887v1.full	https://bitbucket.org/mschnau1/ms-pycloud/src/main/
GPQuest2.1	Hu et al., 2018	https://github.com/huizhanglab-jhu/GPQuest
R v3.6	R Development Core Team	https://www.R-project.org
Bioconductor v3.9	Huber et al., 2015	https://bioconductor.org/
Python v3.7	Python Software Foundation	https://www.python.org/
BIC-Seq2	Xi et al., 2016	http://compbio.med.harvard.edu/BIC-seq/
GISTIC2 v2.0.22	Mermel et al., 2011	https://github.com/broadinstitute/gistic2
Strelka v2.9.2	Kim et al., 2018	https://github.com/Illumina/strelka
VarScan v2.3.8	Koboldt et al., 2012	https://dkoboldt.github.io/varscan/
Pindel v0.2.5	Ye et al., 2009	https://github.com/genome/pindel
MuTect v1.1.7	Cibulskis et al., 2013	https://github.com/broadinstitute/mutect
somaticwrapper v1.3	Ding Lab	https://github.com/ding-lab/somaticwrapper
Samtools v1.2	Li et al., 2009	https://www.htslib.org/
GATK v4.0.0.0	McKenna et al., 2010	https://github.com/broadgsa/gatk
bam-readcount v0.8	McDonnell Genome Institute	https://github.com/genome/bam-readcount
germlinewrapper v1.1	Ding Lab	https://github.com/ding-lab/germlinewrapper
EricScript v0.5.5	Benelli et al., 2012	https://sites.google.com/site/bioericscript/
INTEGRATE v0.2.6	Zhang et al., 2016b	https://sourceforge.net/projects/integrate-fusion/
STAR-Fusion v1.5.0	Haas et al., 2019	https://github.com/STAR-Fusion/STAR-Fusion
Terra	Broad Institute data science platform.	https://terra.bio/
NMF (R-package)	Gaujoux and Seoighe, 2010	https://cran.r-project.org/web/packages/NMF/index.html
PANOPLY	Broad Institute	https://github.com/broadinstitute/PANOPLY/
ssGSEA2.0	Krug et al., 2019	https://github.com/broadinstitute/ssGSEA2.0
Omic-Sig	Lih et al., 2019	https://github.com/hzhangjhu/Omic-Sig
MSFragger-3.0	Kong et al., 2017	https://msfragger.nesvilab.org/
Philosopher-v3.2.8	da Veiga Leprevost et al., 2020	https://philosopher.nesvilab.org/
TMT-Integrator-v1.0.10	Djomehri et al., 2020	http://tmt-integrator.nesvilab.org/
PeptideProphet	Keller et al., 2002	http://tools.proteomecenter.org/wiki/index.php?title=Main_Page
PTMProphet	Shteynberg et al., 2019	http://tools.proteomecenter.org/wiki/index.php?title=Main_Page
ProteinProphet	Nevzihskii et al., 2003	http://tools.proteomecenter.org/wiki/index.php?title=Main_Page
xCell	Aran et al., 2017	https://xcell.ucsf.edu/
MCPCounter	Becht et al., 2016	https://github.com/ebecht/MCPcounter
ConsensusClusterPlus	Wilkerson and Hayes, 2010	https://www.bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
EDec	Onuchic et al., 2016	https://github.com/BRL-BCM/EDec
Webgestalt	Liao et al., 2019	http://www.webgestalt.org/
OmicsEV	Zhang Lab	https://github.com/bzhanglab/OmicsEV

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Hui Zhang (huizhang@jhu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The raw proteomic data files generated during this study are available at the Proteomic Data Commons (PDC: <https://pdc.cancer.gov/pdc/>). Genomic, epigenomic, and transcriptomic data generated for this publication are available at the Genomic Data Commons (GDC: <https://gdc.cancer.gov/>). All processed data tables are available at PDC (PDC: <https://pdc.cancer.gov/pdc/publications>) and LinkedOmics (LinkedOmics: http://www.linkedomics.org/data_download/CPTAC-PDAC/).

The workflow described under ‘Multi-omics clustering’ has been implemented as a module for PANOPLY (<https://github.com/broadinstitute/PANOPLY/>) running on Broad’s cloud platform Terra (<https://app.terra.bio/>). The docker containers encapsulating the source code and required R-packages for NMF clustering and ssGSEA are available on Dockerhub (broadcptacdev/pgdac_mo_nmf:15, broadcptac/pgdac_ssgsea:5). The data evaluation tool has been implanted as a R package available in OmicsEV (<https://github.com/bzhanglab/OmicsEV>). The codes for genomics data processing pipelines are available in <https://github.com/ding-lab/>.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Subjects

A total of 140 cases of patients with pancreatic cancer (135 PDACs and 5 pancreatic adenosquamous carcinoma) were carefully evaluated histologically and included in this study. Institutional review boards at tissue source sites, reviewed protocols and consent documentation adhering to the Clinical Proteomic Tumor Analysis Consortium (CPTAC) guidelines for study participation.

Clinical Data Annotation

A total of 140 participants (74 males, 66 females between the age group of 31-85) were collected for this study by 11 different tissue source sites (TSS) from 7 different countries. Clinical data were obtained from tissue source sites and aggregated by an internal database called the CDR (Comprehensive Data Resource) that synchronizes with the CPTAC DCC (Data Coordinating Center). Clinical data can be accessed and downloaded from the DCC. Demographics, histopathologic information, treatment and patient outcome information were collected and reviewed for consistency before deposition into the Proteomic Data Commons (PDC) and Genomic Data Commons (GDC). All histologic and radiologic details can be accessed from The Cancer Imaging Archive (TCIA) Public Access portal. The genotypic, clinical, geographical and other associated metadata is summarized in Table S1. The cohort consists of 53% male (n = 74) and 47% female (n = 66), in line with the previous observation of a slightly higher incident rate in men than in women (Kim et al., 2017; Cancer Genome Atlas Research Network, 2017). Age distributions [31-40 (2.9%), 41-50 (9.3%), 51-60 (16.4%), 61-70 (42.9%), 71-80 (25.7%), and 81-90 (2.9%)] and stage distributions [I (16.4%), II (42.9%), III (30.0%), and IV (6.4%)] of the patients reflect the general incidence of surgically resected PDAC (Cancer Genome Atlas Research Network, 2017).

METHOD DETAILS

Specimen Acquisition

The tumor, normal adjacent tissue (NAT), and whole blood samples used in this manuscript were prospectively collected for the CPTAC project. Treatment naive patients scheduled for surgical treatment for a pancreatic mass suspected to be pancreatic ductal adenocarcinoma were considered. Patients who underwent cancer treatment more than ten years prior were included if the cancer was at a site other than the pancreas. Only histopathologically-defined adult pancreatic ductal adenocarcinoma (135) and adenosquamous carcinoma (5) were considered for analysis. 67 out of the 140 had matched normal tissue from non-neoplastic pancreatic

tissue as an acceptable normal. To supplement the normal cohort, 9 additional normal macrodissected main pancreatic ductal tissues were collected from unmatched patients who underwent surgery for benign neoplasms. The tumor specimen weights ranged from 150 to 1000 mg. The average tissue mass was 258 mg. For most cases, three to four tumor specimens were collected. Each tissue specimen endured cold ischemia for less than 30 minutes prior to freezing in liquid nitrogen; the average ischemic time was 20 minutes from resection/collection to freezing. Specimens were flash frozen in liquid nitrogen. Histologic sections obtained from top and bottom portions from each case were reviewed by a board-certified pathologist and a disease specific expert pathologist to confirm the assigned pathology. Although there was no tumor nuclei cutoff, for samples to be deemed acceptable, the top and bottom sections had to contain an average of less than 20% necrosis as assessed histologically. Specimens were shipped overnight from the tissue source sites to the biospecimen core resource (BCR) located at Van Andel Research Institute, Grand Rapids, MI using a cryoport that maintained an average temperature of less than -140°C . At the biospecimen core resource, specimens were confirmed for pathology qualification and prepared for genomic, transcriptomic, and proteomic analyses. Selected specimens were cryopulverized using a Covaris CryoPREP instrument and material aliquoted for subsequent molecular characterization. Genomic DNA and total RNA were extracted and sent to the genome sequencing centers. The whole exome and whole genome DNA sequencing and methylation EPIC array analyses were performed at the Broad Institute, Cambridge. Total RNA and miRNA sequencing were performed at the University of North Carolina, Chapel Hill, NC. Material for proteomic analyses were sent to the Proteomic Characterization Center (PCC) at the Johns Hopkins University, Baltimore, MD.

Genomic, Epigenomic, and Transcriptomic Sample Preparation and Data Acquisition

Sample Processing for Genomic DNA and Total RNA Extraction

Each primary tumor was obtained from a single sample from surgical resections, with a requirement of a minimum of 125 mg of tumor tissue and 50 mg of adjacent normal tissue. DNA and RNA were extracted from the tumor and normal specimens using a co-isolation protocol (QIAGEN's QIAAsymphony DNA Mini Kit and QIAAsymphony RNA Kit). Genomic DNA was also isolated from peripheral blood (3–5 mL) to serve as matched normal reference material. The Qubit dsDNA BR Assay Kit was used with the Qubit® 2.0 Fluorometer to determine the concentration of dsDNA in an aqueous solution. Only samples with enough DNA yield that passed quality control were sent for genomic characterization. RNA quality was quantified using the NanoDrop 8000 and had its quality assessed using Agilent Bioanalyzer. Only samples that passed RNA quality control and had a minimum RIN (RNA integrity number) score of 7 underwent RNA sequencing. Identity matching for germline, normal adjacent tissue, and tumor tissue was assayed at the BCR using the Illumina Infinium QC array.

Whole Exome Sequencing

Library Construction. Library construction was performed as described in Fisher et al. (2011), with the following modifications: the initial genomic DNA input shearing was reduced from 3 μg to 20–250 ng in 50 μL of solution. For adaptor ligation, Illumina paired-end adapters were replaced with palindromic forked adapters, purchased from IDT, with unique dual-indexed molecular barcode sequences to facilitate downstream pooling. Kapa HyperPrep reagents in 96-reaction kit format were used for end repair/A-tailing, adaptor ligation, and library enrichment PCR. During post-enrichment SPRI cleanup, elution volume was reduced to 30 μL to maximize library concentration and a vortexing step was added to maximize the amount of template eluted.

In-solution Hybrid Selection. Libraries were pooled into groups of up to 96 samples. Hybridization and capture were performed using the relevant components of Illumina's Nextera Exome Kit, following the manufacturer's suggested protocol, except for a few modifications. The following modifications were made: all libraries within a library construction plate were pooled prior to hybridization, the Midi plate from Illumina's Nextera Exome Kit was replaced with a skirted PCR plate to facilitate automation, and all hybridization and capture steps were automated on the Agilent Bravo liquid handling system.

Preparation of Libraries for Cluster Amplification and Sequencing. After post-capture enrichment, library pools were quantified using qPCR (automated assay on the Agilent Bravo) using a kit purchased from KAPA Biosystems with probes specific to the ends of the adaptors. Based on qPCR quantification, libraries were normalized to 2 nM.

Cluster Amplification and Sequencing. Cluster amplification of DNA libraries was performed according to the manufacturer's protocol (Illumina) using exclusion amplification chemistry and flowcells. Flowcells were sequenced utilizing sequencing-by-synthesis chemistry. Flowcells were then analyzed using RTA v.2.7.3 or later. Each pool of whole exome libraries was sequenced on paired 76 cycle runs with two 8 cycle index reads across the number of lanes needed to meet coverage for all libraries in the pool. Pooled libraries were then run on HiSeq 4000 paired-end runs targeting a depth of coverage of 300x for 122 tumor sample libraries and 150x for the remaining 18 tumor sample libraries and the NAT and blood normal sample libraries. The raw Illumina sequence data were demultiplexed and converted to fastq files with adaptor and low-quality sequences trimmed out.

PCR-Free Whole Genome Sequencing

Preparation of Libraries for Cluster Amplification and Sequencing. An aliquot of genomic DNA (350 ng in 50 μL) was used as the input into DNA fragmentation. Shearing was performed acoustically using a Covaris focused-ultrasonicator, targeting 385 bp fragments. Following fragmentation, additional size selection was performed using a SPRI cleanup. Library preparation was performed using a kit from KAPA Biosystems (KAPA Hyper Prep without the amplification module) and with palindromic forked adapters with unique 8-base index sequences embedded within the adaptor (purchased from IDT). Following sample preparation, libraries were quantified

using qPCR (KAPA Biosystems) with probes specific to adaptor ends. This assay was automated using Agilent's Bravo liquid handling platform. Based on qPCR quantification, libraries were normalized to 1.7 nM and pooled into 24-plexes.

Cluster Amplification and Sequencing (HiSeq X). Sample pools were combined with HiSeq X Cluster Amp Reagents EPX1, EPX2, and EPX3 into single wells on a strip tube using the Hamilton Starlet Liquid Handling system. Cluster amplification of the templates was performed according to the manufacturer's protocol (Illumina) with the Illumina cBot. Flowcells were sequenced with a target of 15x to 30x depth on HiSeq X utilizing sequencing-by-synthesis kits to produce 151bp paired-end reads.

Illumina Infinium Methylation EPIC Bead Chip Array

The MethylationEPIC array uses an 8-sample version of the Illumina Beadchip capturing > 850,000 DNA methylation sites per sample. 250 ng of DNA was used for the bisulfite conversation using the Infinium MethylationEPIC BeadChip Kit. The EPIC array includes sample plating, bisulfite conversion, and methylation array processing. After scanning, the data was processed through an automated genotype calling pipeline.

RNA Sequencing

Quality Assurance and Quality Control of RNA Analytes. All RNA analytes were assayed for RNA integrity, concentration, and fragment size. Samples for total RNA-seq were quantified on a TapeStation system (Agilent, Inc. Santa Clara, CA). Samples with RINs > 8.0 were considered high quality.

Total RNA-seq Library Construction. Total RNA-seq library construction was performed from the RNA samples using the TruSeq Stranded RNA Sample Preparation Kit and bar-coded with individual tags following the manufacturer's instructions (Illumina). Libraries were prepared on an Agilent Bravo Automated Liquid Handling System. Quality control was performed at every step and the libraries were quantified using the TapeStation system.

Total RNA Sequencing. Indexed libraries were prepared and run on HiSeq 4000 paired end 75 base pairs to generate a minimum of 120 million reads per sample library with a target of greater than 90% mapped reads, typically with four-sample pools. The raw Illumina sequence data were then demultiplexed and converted to FASTQ files, and adaptor and low-quality sequences were quantified. Reads were mapped to the hg38 human genome reference and underwent significant QC steps: estimating the total number of reads that mapped, amount of RNA mapping to coding regions, amount of rRNA in sample, number of genes expressed, and relative expression of housekeeping genes. Samples passing this QA/QC were then clustered with other expression data from similar and distinct tumor types to confirm expected expression patterns. Atypical samples were then SNP typed from the RNA data to confirm source analyte.

miRNA-seq Library Construction. miRNA-seq library construction was performed from the RNA samples using the NEXTflex Small RNA-Seq Kit (v3, PerkinElmer) and bar-coded with individual tags following the manufacturer's instructions. Libraries were prepared on the Sciclone Liquid Handling Workstation and quality control checks were performed at every step. Libraries were quantified using a TapeStation system and an Agilent Bioanalyzer using the Small RNA analysis kit. Pooled libraries were then size selected according to NEXTflex Kit specifications using a Pippin Prep system (Sage Science).

miRNA Sequencing. Indexed libraries were loaded on the Hiseq 4000 to generate a minimum of 10 million reads per library with a minimum of 90% reads mapped. The raw Illumina sequence data were demultiplexed and converted to FASTQ files for downstream analysis. Samples were assessed for the number of miRNAs called, species diversity, and total abundance.

Genomic, Epigenomic, and Transcriptomic Data Processing

Genome alignment

WGS, WES, and RNA-Seq data were harmonized by NCI Genomic Data Commons (GDC) <https://gdc.cancer.gov/about-data/gdc-data-processing>, to the hg38 human reference genome, version GRCh38.d1.vd1.

Whole-genome copy number variation

Copy number variation (CNV) was detected using BIC-Seq2 (module versions NBICseq-seg v0.7.2 and NBICseq-norm v0.2.4) (Xi et al., 2016) from WGS tumor and normal paired BAMs. A bin size of 100bp and a lambda of 3 for segmentation smoothing was used. We used the mean of overlapping segment data to further summarize the CNV data into gene-level copy number changes. We also used GISTIC2 v2.0.22 (Mermel et al., 2011) to integrate results from individual patients and identify focal genomic regions recurrently amplified or deleted in our samples. The threshold for gene or arm-level CNV status was 0.4 for copy number gains and -0.4 for losses.

Somatic Variant Calling

Somatic variants were called from WES tumor and normal paired BAMs using SomaticWrapper v1.6. SomaticWrapper merges and filters variant calls from four callers: Strelka v2.9.2 (Kim et al., 2018), VarScan v2.3.8 (Koboldt et al., 2012), Pindel v0.2.5 (Ye et al., 2009), and MuTect v1.1.7 (Cibulskis et al., 2013). SNV calls were obtained from Strelka, VarScan, and Mutect. Indel calls were obtained from Strelka2, VarScan, and Pindel. The following filters were applied to get variant calls of high confidence:

- Normal VAF ≤ 0.02 and tumor VAF ≥ 0.01 . The tumor VAF cutoff is set lower to account for the unique low neoplastic cellularity in PDAC.
- Read depth in tumor ≥ 14 and normal ≥ 8

- Indel length < 100 bp
- All variants must be called by 2 or more callers
- All variants must be exonic
- Exclude variants in dbSNP but not in COSMIC

Germline Variant Calling

Germline variant calling was performed using the GermlineWrapper v1.1 pipeline, which implements multiple tools for the detection of germline INDELS and SNVs. Germline SNVs were identified using VarScan v2.3.8 (with parameters: -min-var-freq 0.10 -p value 0.10, -min-coverage 3-strand-filter 1) operating on a mpileup stream produced by samtools v1.2 (with parameters: -q 1 -Q 13) and GATK v4.0.0.0 (McKenna et al., 2010) using its haplotype caller in single-sample mode with duplicate and unmapped reads removed and retaining calls with a minimum quality threshold of 10. All resulting variants were limited to the coding region of the full-length transcripts obtained from Ensembl release 95 plus additional two base pairs flanking each exon to cover splice donor/acceptor sites. We required variants to have allelic depth ≥ 5 reads for the alternative allele in both tumor and normal samples. We used bam-readcount v0.8 for reference and alternative alleles quantification (with parameters: -q 10 -b 15) in both normal and tumor samples. Additionally, we filtered all variants with $\geq 0.05\%$ frequency in gnomAD v2.1 (Karczewski et al., 2020) and The 1000 Genomes Project (Auton et al., 2015).

Pathogenic Germline Variant Classification

To predict the pathogenicity of germline variants, we annotate each variant with Variant Effect Predictor (VEP) and process them using the CharGer pipeline with the parameters from a previous pan-cancer TCGA study (Huang et al., 2018; Scott et al., 2019). Briefly, the CharGer pipeline considers pathogenic peptide changes from ClinVar, hotspot variants, minor allele frequency from ExAC, and several *in silico* analyses (such as Sift and PolyPhen). Each predicted pathogenic variant was then manually reviewed.

DNA Methylation Microarray Processing

Raw methylation idat files were downloaded from CPTAC DCC and GDC. Beta values of CpG loci were reported after functional normalization, quality check, common SNP filtering, and probe annotation using Li Ding Lab's methylation pipeline v1.1 (https://github.com/ding-lab/cptac_methylation). To derive the gene-level methylation, we focused on the probes located in the promoter region and simultaneously located in annotated CpG island and aggregate their levels by median (Clark et al., 2019). Resulting beta values of methylation were used for downstream analysis.

mRNA and Circular RNA Quantification

The hg38 reference genome and RefSeq annotations were used for the RNAseq data analysis and were downloaded from the UCSC table browser. First, CIRI (v2.0.6) was used to call circular RNA with default parameters and BWA (version 0.7.17-r1188) was used as the mapping tool. The cutoff of supporting reads for circRNAs was set to 10. Then we used a pseudo-linear transcript strategy to quantify gene and circular RNA expression (Li et al., 2017). In brief, for each sample, linear transcripts of circular RNAs were extracted and 75bp (read length) from the 3' end was copied to the 5' end. The modified transcripts were called pseudo-linear transcripts. Transcripts of linear genes were also extracted and mixed with pseudo-linear transcripts. RSEM (version 1.3.1) with Bowtie2 (version 2.3.3) as the mapping tool was used to quantify gene and circular RNA expression based on the mixed transcripts. After quantification, the upper quantile method was applied for normalization. The normalized matrix was log2-transformed and separated into gene and circular RNA expression matrices.

Gene Fusion Detection

We used three callers, STAR-Fusion v1.5.0 (Haas et al., 2019), INTEGRATE v0.2.6 (Zhang et al., 2016b), and EricScript v0.5.5 (Benelli et al., 2012), to call consensus fusion/chimeric events in our samples. Calls by each tool using tumor and normal RNA-Seq data were then merged into a single file and extensive filtering is done. As STAR-Fusion has higher sensitivity, calls made by this tool with higher supporting evidence (defined by fusion fragments per million total reads, or FFPM > 0.1) were required, or a given fusion must be reported by at least 2 callers. We then removed fusions present in our panel of blacklisted or normal fusions, which included uncharacterized genes, immunoglobulin genes, mitochondrial genes, and others, as well as fusions from the same gene or paralog genes and fusions reported in TCGA normal samples (Gao et al., 2018), GTEx tissues (reported in STAR-Fusion output), and non-cancer cell studies (Babiceanu et al., 2016).

miRNA Quantification

miRNA-Seq FASTQ files were downloaded from GDC. We reported the mature miRNA and precursor miRNA expression in TPM (Transcripts Per Million) after adaptor trimming, quality check, alignment, annotation, reads counting using Li Ding Lab's miRNA pipeline https://github.com/ding-lab/CPTAC_miRNA. The mature miRNA expression was calculated irrespective of its gene of origin by summing the expression from its precursor miRNAs.

Proteomic, Phosphoproteomic, and Glycoproteomic Sample Preparation and Data Acquisition

Sample Processing for Protein Extraction and Tryptic Digestion

All samples for the current study were prospectively collected for the CPTAC PDAC project as described above and processed for mass spectrometry (MS) analysis at Johns Hopkins University. Tissue lysis and downstream sample preparation for global proteomic, phosphoproteomic, and glycoproteomic analysis were carried out as previously described (Mertins et al., 2018; Yang et al., 2018a). Each of cryo-pulverized pancreatic cancer tissues, normal adjacent tissues (NAT), and normal ductal tissues was lysed in

lysis buffer (8 M urea, 75 mM NaCl, 50 mM Tris (pH 8.0), 1 mM EDTA, 2 µg/mL aprotinin, 10 µg/mL leupeptin, 1 mM PMSF, 10 mM NaF, Phosphatase Inhibitor Cocktail 2 and Phosphatase Inhibitor Cocktail 3 [1:100 dilution], and 20 µM PUGNAC) by vortexing for 20 s. The lysed tissue was placed on ice for 15 min. This process was repeated one time. Cell debris was removed by centrifugation at 20,000 × g for 10 min at 4°C. The protein-containing supernatant was collected and measured by BCA assay (Pierce). The sample concentration was adjusted to 8 mg/ml with the lysis buffer, and an appropriate volume of the protein solution was used for the downstream reduction, alkylation and digestion. Proteins were reduced and alkylated with dithiothreitol (DTT, Thermo Fisher, 5 mM, 37°C, 1h) and iodoacetamide (IAM, Sigma, 10 mM, room temperature (RT) in the dark, 45 min), respectively. The reduced proteins were diluted 1:4 with 50 mM Tris-HCl (pH 8.0) to reduce urea concentration and digested with LysC (Wako Chemicals, an enzyme-to-substrate ratio of 1 mAU:50 mg, RT, 2h) followed by trypsin (Promega, an enzyme-to-substrate ratio of 1:50, RT, 16h). The proteolytic reaction was quenched by adjusting pH to < 3 with 50% of formic acid (FA, Fisher Chemicals). The peptides were desalted on reversed phase C18 SPE columns (Waters) and dried using Speed-Vac (Thermo Scientific).

Tandem Mass Tag (TMT) Labeling of Peptides

Dried peptides from each sample were labeled with 11-plex TMT reagents (Thermo Fisher Scientific). Peptides (300 µg) were dissolved in 60 µL of 100 mM HEPES (pH 8.5). A reference sample was created by pooling an aliquot from 129 pancreatic cancer tissues and 52 NAT tissues (representing ~85% of the sample cohort), and was included in all TMT 11-plex sets as a reference channel. An internal quality control (QC) sample that was an NCI-7 Cell Line Panel sample ([Clark et al., 2018](#)) was interspersed among TMT 11-plex sets. 140 pancreatic cancer tissues and 67 NATs were co-randomized to 24 TMT 11-plex sets, while 8 normal ductal tissues were assigned to the 25th TMT 11-plex set. TMT reagents were dissolved in 250 µL of anhydrous acetonitrile (Sigma), and 20 µL of each TMT reagent was added to the corresponding aliquot of peptides. The reaction was incubated at RT for 1h with shaking and quenched with 5% hydroxylamine at RT for 15 min. The labeled peptides were desalted on reversed phase C18 SPE columns (Waters) and dried using Speed-Vac (Thermo Scientific).

Peptide Fractionation by Basic Reversed-phase Liquid Chromatography (bRPLC)

The desalted peptides from each TMT set were dissolved in 900 µL of 5 mM ammonium formate (pH 10) and 2% acetonitrile (ACN) and fractionated with a 4.6 mm x 250 mm Zorbax Extend-C18 analytical column (3.5 µm beads, Agilent) lined up with an Agilent 1220 Series HPLC. Buffer A and B were 5 mM ammonium formate in 2% ACN (pH 10) and 5 mM ammonium formate in 90% ACN, respectively. Peptides were separated by using a non-linear gradient: 0% buffer B (7 min), 0% to 16% buffer B (6 min), 16% to 40% buffer B (60 min), 40% to 44% buffer B (4 min), 44% to 60% buffer B (5 min) and then held at 60% buffer B for 14 min. The flow rate was set at 1 mL/min. Collected fractions were concatenated into 24 fractions as described previously ([Mertins et al., 2018](#)). Eight percent of each of the 24 fractions was aliquoted, cleaned up with strong cation exchange (SCX) stage tip, and dried in a Speed-Vac. Samples were resuspended in 3% ACN, 0.1% FA prior to ESI-LC-MS/MS analysis. The remaining sample was utilized for phosphopeptide enrichment.

Enrichment of Phosphopeptides by Fe-IMAC

The remaining 92% of the sample was further concatenated into 12 fractions before being subjected to phosphopeptide enrichment using immobilized metal affinity chromatography (IMAC) as previously described ([Mertins et al., 2018](#)). Ni-NTA agarose beads (QIAGEN) were conditioned and incubated with a 10mM FeCl₃ aqueous solution at RT for 1h to prepare Fe³⁺-NTA agarose beads. Peptides from each fraction were reconstituted in 80% ACN, 0.1% trifluoroacetic acid (TFA) and incubated with 10 µL of the Fe³⁺-IMAC beads at RT for 30 min. The bead solution was spun down at 1,000 g for 1 min. The supernatant containing unbound peptides was separated from beads and collected for subsequent glycopeptide enrichment. The beads were resuspended in 80% ACN (0.1% TFA) and loaded onto conditioned C18 stage tip. The tip was washed twice with 80% ACN (0.1% TFA) followed by 1% FA. The flowthroughs were collected and combined with the unbound peptides for subsequent glycopeptide enrichment. Peptides were eluted from beads to C18 stage tip with 70 µL of 500 mM dibasic potassium phosphate, pH 7.0 three times. The tip was then washed twice with 1% FA to remove salts. Peptides were eluted twice with 80% ACN (0.1% FA), dried down, and redissolved in 3% ACN, 0.1% FA prior to ESI-LC-MS/MS analysis.

Enrichment of Glycopeptides

All unbound peptides from phosphopeptide enrichment were desalted on reversed phase C18 SPE column (Waters). The glycopeptides were enriched with OASIS MAX solid-phase extraction (Waters). The MAX cartridge was conditioned with 3 × 1 mL ACN, then 3 × 1 mL of 100 mM triethylammonium acetate buffer, followed by 3 × 1 mL of water, and finally 3 × 1 mL of 95% ACN (1% TFA). The peptides were loaded twice. The cartridge was washed with 4 × 1 mL of 95% ACN (1% TFA) to remove non-glycosylated peptides. The glycopeptide fraction was eluted with 50% ACN (0.1% TFA), dried down, and reconstituted in 3% ACN, 0.1% FA prior to ESI-LC-MS/MS analysis.

ESI-LC-MS/MS for Global Proteome, Phosphoproteome, and Glycoproteome Analysis

The TMT-labeled global proteome, phosphoproteome, and glycoproteome fractions were analyzed using Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific). Approximately 0.8 µg of peptides were separated on an in-house packed 28 cm x 75 mm diameter C18 column (1.9 mm ReproSil-Pur C18-AQ beads (Dr. Maisch GmbH); Picofrit 10 mm opening (New Objective)) lined up with an Easy nLC 1200 UHPLC system (Thermo Scientific). The column was heated to 50°C using a column heater (Phoenix-ST). The flow rate was set at 200 nL/min. Buffer A and B were 3% ACN (0.1% FA) and 90% ACN (0.1% FA), respectively. The peptides were separated with a 6%-30% B gradient in 84 min. Peptides were eluted from the column and nanosprayed directly into the mass spectrometer. The mass spectrometer was operated in a data-dependent mode. Parameters for global proteomic samples

were set as follows: MS1 resolution – 60,000, mass range – 350 to 1800 m/z, RF Lens – 30%, AGC Target – 4.0e5, Max injection time – 50 ms, charge state include – 2–6, dynamic exclusion – 45 s. The cycle time was set to 2 s, and within this 2 s the most abundant ions per scan were selected for MS/MS in the orbitrap. MS2 resolution – 50,000, high-energy collision dissociation activation energy (HCD) – 37, isolation width (m/z) – 0.7, AGC Target – 2.0e5, Max injection time – 105 ms. Parameters for phosphoproteomic samples were set as follows: MS1 resolution – 60,000, mass range – 350 to 1800 m/z, RF Lens – 30%, AGC Target – 4.0e5, Max injection time – 50 ms, charge state include – 2–6, dynamic exclusion – 45 s. The cycle time was set to 2 s, and within this 2 s the most abundant ions per scan were selected for MS/MS in the orbitrap. MS2 resolution – 50,000, high-energy collision dissociation activation energy (HCD) – 34, isolation width (m/z) – 0.7, AGC Target – 2.0e5, Max injection time – 100 ms. Parameters for glycoproteomic samples were set as follows: MS1 resolution – 60,000, mass range – 500 to 2000 m/z, RF Lens – 30%, AGC Target – 5.0e5, Max injection time – 50 ms, charge state include – 2–6, dynamic exclusion – 45 s. The cycle time was set to 2 s, and within this 2 s the most abundant ions per scan were selected for MS/MS in the orbitrap. MS2 resolution – 50,000, high-energy collision dissociation activation energy (HCD) – 35, isolation width (m/z) – 0.7, AGC Target – 1.0e5, Max injection time – 100 ms.

ESI-LC-MS/MS for Global Proteome Data-Independent Acquisition (DIA) Analysis

Unlabeled, digested peptide material from individual tissue samples (PDAC and NAT) was spiked with index Retention Time (iRT) peptides (Biognosys) and subjected to DIA analysis. Approximately 1 µg of peptides were separated on an in-house packed 28 cm x 75 mm diameter C18 column (1.9 mm ReproSil-Pur C18-AQ beads (Dr. Maisch GmbH); Picofrit 10 mm opening (New Objective)) lined up with an Easy nLC 1200 UHPLC system (Thermo Scientific). The column was heated to 50°C using a column heater (Phoenix-ST). The flow rate was set at 200 nL/min. Buffer A and B were 3% ACN (0.1% FA) and 90% ACN (0.1% FA), respectively. The peptides were separated with a 7–30% B gradient in 118 min. Peptides were eluted from the column and nanosprayed directly into Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific). The mass spectrometer was operated in a data-independent mode. The DIA segment consisted of one MS1 scan (350–1650 m/z range, 120K resolution) followed by 30 MS2 scans (variable m/z range, 30K resolution). Additional parameters were as follows: MS1: RF Lens – 30%, AGC Target 3.0e6, Max IT – 60 ms, charge state include – 2–6; MS2: isolation width (m/z) – 0.7, AGC Target – 3.0e6, Max IT – 120 ms.

Spectral Library generation for Data-Independent Acquisition Analysis

For spectral library generation, an aliquot (2 mg) of unlabeled, digested peptide material from individual tissue samples (PDAC and NAT) was pooled and subjected to bRPLC. Ninety-six fractions were collected into a 96-well plate. These fractions were pooled every eight fraction (e.g., combining fractions #1, #9, #17, #25, #33, #41, #49, #57, #65, #73, #81, and #89; #2, #18, #26, #34, #42, #58, #66, #74, #82, and #90; and so on). The resulting 8 fractions were dried in a speed vacuum centrifuge, resuspended in 3% ACN, 0.1% formic acid, and spiked with iRT peptides prior to ESI-LC-MS/MS analysis. Parameters were the same as previously described for ESI-LC-MS/MS for TMT-labeled global proteome analysis with a high-energy collision dissociation activation energy (HCD) – 34.

Proteomic, Phosphoproteomic, and Glycoproteomic Data Processing

Proteomic and Phosphoproteomic Data Processing

MS/MS spectra were searched using the MSFragger version 3.0 ([Kong et al., 2017](#)) against a CPTAC3 RefSeq human protein sequence database appended with an equal number of decoy sequences. For the analysis of whole proteome data, MS/MS spectra were searched using a precursor-ion mass tolerance of 10 ppm, and allowing C12/C13 isotope errors (–1/0/1/2/3). MS and MS/MS mass calibration, MS/MS spectral deisotoping, and parameter optimization were enabled ([Yu et al., 2020](#)). Cysteine carbamidomethylation (+57.0215), lysine TMT labeling (+229.1629), and peptide N-terminal TMT labeling were specified as fixed modifications. Methionine oxidation (+15.9949) and serine TMT labeling (+229.1629) were specified as variable modifications. The search was restricted to tryptic and semi-trypic peptides, allowing up to two missed cleavage sites. For phosphopeptide enriched data, the set of variable modifications also included phosphorylation (+79.9663) of serine, threonine, and tyrosine residues, but excluded the serine TMT labeling, and with C12/C13 isotope errors parameter set to (0/1/2).

The post-processing of the search results was done using the Philosopher toolkit version v3.2.8 ([da Veiga Leprevost et al., 2020](#)). MSFragger output files (in pepXML format) were processed using PeptideProphet ([Keller et al., 2002](#)) (with the high-mass accuracy binning and semi-parametric mixture modeling options) to compute the posterior probability of correct identification for each peptide to spectrum match (PSM). In the phosphopeptide-enriched dataset, PeptideProphet files were additionally processed using PTMProphet ([Shteynberg et al., 2019](#)) to localize the phosphorylation sites. The resulting pepXML files from PeptideProphet (or PTMProphet) from all 25 TMT 11-plex experiments were then processed together to assemble peptides into proteins (protein inference) and to create a combined file (in protXML format) of high confidence protein groups.

The combined protXML file and the individual PSM lists for each TMT 11-plex were further processed using the Philosopher filter command as follows. Each peptide was assigned either as a unique peptide to a particular protein group or set as a razor peptide to a single protein group with the most peptide evidence. The protein groups assembled by ProteinProphet ([Nesvizhskii et al., 2003](#)) were filtered to 1% protein-level False Discovery Rate (FDR) using the best peptide approach (allowing both unique and razor peptides) and applying the picked FDR target-decoy strategy. In each TMT 11-plex, the PSM lists were filtered using a sequential FDR strategy, retaining only those PSMs with PeptideProphet probability of 0.9 or higher (which in these data corresponded to less than 1% PSM-level FDR) and mapped to proteins that also passed the global 1% protein-level FDR filter. For each PSM that passed these filters, the corresponding precursor ion MS1 intensity was extracted using the Philosopher label-free quantification module, using 10 p.p.m mass tolerance and 0.4 min retention time window for extracted ion chromatogram peak tracing.

Also, for all PSMs corresponding to a TMT-labeled peptide, eleven TMT reporter ion intensities were extracted from the MS/MS scans (using 0.002 Da window). The precursor ion purity scores were calculated using the intensity of the sequenced precursor ion and that of other interfering ions observed in MS1 data (within a 0.7 Da isolation window). All supporting information for each PSM, including the accession numbers and names of the protein/gene selected based on the protein inference approach with razor peptide assignment and quantification information (MS1 precursor-ion intensity and the TMT reporter ion intensities), was summarized in the output PSM.tsv files, one file for each TMT 11-plex experiment.

To generate summary reports on different levels (gene, peptide, and protein for global and phosphopeptide enriched data; additional modification site report for phosphopeptide data), all PSM.tsv files were processed together using TMT-Integrator (Djomehri et al., 2020). Each PSM in a PSM.tsv file that passed the following criteria were kept for creating integrated reports, including (1) having a TMT label at peptide N terminus, (2) having non-zero intensity in the reference channel, (3) precursor-ion purity above 50%, (4) summed reported ion intensity (across all channels) not in the lower 5% of all PSMs (2.5% for phosphopeptide enriched data), (5) fully tryptic peptides, (6) peptide with phosphorylation (for phosphopeptide enriched data). For a peptide with redundant PSMs in the same MS run, only the PSM with the highest summed TMT intensity was kept for later analysis. PSMs mapping to common external contaminant proteins was excluded, and both unique and razor peptides were used for quantification. Next, the reporter ion intensities of each PSM were log2 transformed and normalized by the reference channel intensity (i.e., subtracted log2 reference intensity from those log2 report ion intensities), therefore the intensities were converted into a log2-based ratio (denoted as 'ratios' in the following paragraphs). After converting the intensities to ratios, the PSMs were grouped based on the predefined level (i.e., gene, protein, peptide, and site-level). The interquartile range (IQR) algorithm was then applied to remove the outliers in each PSM group, and the remaining ratios were median centered. The ratios were converted back to abundances using the weighted sum of the MS1 intensities of the top three most intense peptide ions, with the weighting factor (computed for each PSM) taken as the ratio of the reference channel intensity to the summed reporter ion intensity (across all channels). In generating the site-level reports (phosphopeptide-enriched data), sites with PTMProphet computed localization probability equal or greater than 0.75 were considered as confidently localized. Additional details regarding these steps can be found in Clark et al. (2019).

Glycoproteomic Data Processing

Glycoproteomic and phosphoproteomic raw data files were converted to universal format mzML files using the msconvert tool from ProteoWizard, and searched with the GPQuest search engine (version 2.1) with the following modifications: dynamic oxidation (+15.9949 Da) on Met, and static carbamidomethylation (+57.021464 Da) on Cys residues. GPQuest was applied to identify intact N-linked glycopeptides to MS/MS spectra using two approaches: searching spectra containing oxonium ions ('oxo-spectra') and identifying intact N-linked glycopeptides. The oxonium ions were used as the signature features of the glycopeptides from the MS/MS spectra, which were caused by the fragmentation of glycans attached to intact glycopeptides in the mass spectrometer. In this study, the MS/MS spectra containing the oxonium ions (m/z 204.0966) in the top 10 abundant peaks (N-linked glycopeptide search) and top 1000 abundant peaks (O-linked glycopeptide search) after removing TMT reporter ions were considered as the potential glycopeptide candidates. The intact N-linked glycopeptides were identified by using GPQuest to search against the glycopeptide database of glycositeatlas (Sun et al., 2019) and a glycan database collected from the public database of GlycomeDB (Ranzinger et al., 2011). Each tandem mass spectrum was first processed in a series of preprocessing procedures, including removing reporter ions, spectrum de-noising, intensity square root transformation (Liu et al., 2007), oxonium ions evaluation and glycan type prediction (Toghi Eshghi et al., 2016). The top 100 peaks in each preprocessed spectrum were matched to the fragment ion index generated from a peptide sequence database to identify all the candidate peptides. All the qualified (≥ 6 fragment ions matchings) candidate peptides were compared with the spectrum again to calculate the Morpheus scores (Wenger and Coon, 2013) by considering all the peptide fragments, glycopeptide fragments, and their isotope peaks. The peptide having the highest Morpheus score was then assigned to the spectrum. The mass gap between the assigned peptide and the precursor mass was searched in the glycan database to find the associated glycan. The best hits of all 'oxo-spectra' were filtered by precursor isotopes distribution fitting score and then ranked by the Morpheus score in descending order, in which those with FDR < 1% and covering > 10% total intensity of each tandem spectrum were reserved as qualified identifications. The precursor mass tolerance was set as 10 ppm, and the fragment mass tolerance was 20 ppm.

For the identification of O-linked glycopeptides, the LC-MS/MS data were searched against a peptide database generated from 2,225 O-linked glycoproteins and 84 Functional Glycomics Gateway (CFG) O-linked glycan database. The 2,225 O-linked glycoproteins were collected from glycoproteins identified using EXoO method (Yang et al., 2018b, 2020) and O-glycoprotein database (<http://www.oglyp.org/>). An IR score of over 0.2 and Morpheus score of at least 7 were used to filter the data and the decoy identification was used to calculate the FDR for the identification of O-linked glycopeptides. The identified O-linked glycopeptides were compared to the list of N-linked glycopeptides identified in this study, and the overlapped glycopeptides were removed from the final list of O-linked glycopeptides.

Peptide-spectrum matches (PSMs) were quantified using the MS-PyCloud proteomics pipeline (<https://bitbucket.org/mschnau1/ms-pycloud/src/main/>). TMT correction factors were applied in order to correct the MS2 intensity of each PSM. Only fully-trypic peptides with up to two missed cleavages were retained. Glycopeptide (peptide + glycan) false discovery rate (FDR) was restricted to less than or equal to 1 percent by applying a PSM-level FDR filter of less than or equal to 0.25 percent, requiring a minimum of two PSMs per peptide, and a minimum of one peptide per protein. Sample log2 ratios were calculated for each PSM relative to the pooled reference for that sample's TMTplex after median normalizing each sample in the TMTplex to the pooled reference. PSM log2 ratios were

then rolled up to glycopeptide-level by taking the median of PSMs that map to the same glycopeptide. The glycopeptide log2 ratio matrix was then median normalized across all samples. The glycopeptide abundance matrix was derived from the log2 ratio matrix by adding the median log2 value of all TMTplex pooled reference summed MS2 intensities to each sample log2 ratio for a given glycopeptide.

Protein Database Searching and Quantification of Global DIA Data

Raw mass spectrometry files from DIA and data dependent acquisition (DDA) platforms were processed using the DIA-Umpire (Tsou et al., 2015) based pipeline to generate a combined spectral library that integrated DDA and DIA search results (Cho et al., 2020; Clark et al., 2019). The combined library was then converted to Spectronaut (Biognosys) format and loaded into Spectronaut. The DIA data was searched using default settings of Spectronaut (Cho et al., 2020; Clark et al., 2019), and the results were exported without normalization. The protein abundances were further grouped by unique gene names using sum of all the protein abundances belonging to the identical gene name. The protein abundances in the protein expression matrix were log2-transformed. The missing value excluded median abundance $M_i = \text{median}(A_{ij}, j = 1, \dots, p)$ of all p proteins in each sample i were calculated. The median abundance of the first sample (C3L-01124-T) was selected as the reference M_0 . The abundances in each sample were median centered to M_0 (normalized $A_{ij} = A_{ij} - M_i + M_0$).

Data Quality Control

Different normalization methods for global proteomics, phosphoproteomics and glycoproteomics data were evaluated using OmicsEV (<https://github.com/bzhanglab/OmicsEV/>) and an optimal normalization method was then selected for each data type. After the data were normalized, batch effect was also evaluated using OmicsEV both visually by correlation heatmaps ordered by TMT-plex and by PCA. For each PC, the Pearson correlation coefficient to the batch covariate was calculated and significance was assessed by using one-way ANOVA. None of the first 3 PCs were significantly correlated to the TMT-plex, indicating the lack of a batch effect. Pairwise comparisons between replicate samples and samples within TMT plexes were conducted using the square of the Pearson correlation coefficient (R^2) based on the data generated using a virtual reference-based method. In the correlation analysis, only features without any missing value were used. The virtual reference of proteomic, phosphoproteomic, *N*-linked glycoproteomic, and O-linked glycoproteomic data was calculated as the median PSM intensity from all channels in the TMTplex, with zero value intensities being omitted.

In addition to extensive QC during data acquisition, RNA sequencing data quality was assessed using FastQC. To detect potential sample swaps or mislabeling across data types, genome-wide correlations at all omics levels (e.g., RNA-Protein) were used to determine sample identity concordance. Of note, the gender of one case (C3N-02295) was predicted to be male based on mRNA data, which was inconsistent with the clinical data provided. We decided to include this case in our cohort since we did not perform any gender-related analysis.

Integrated Analysis

Mutation Impact on the RNA, Proteome, and Phosphoproteome

We examined the *cis*- and *trans*-effects of 11 genes with somatic mutations that were significantly mutated in previous large scale PDAC studies (Bailey et al., 2018; Weinstein et al., 2013) on the RNA, proteome, and phosphoproteome. We collected a set of interacting proteins partners from OmniPath (downloaded on 2018-03-29) (Türel et al., 2016), DEPOD (downloaded on 2018-03-29) (Duan et al., 2015), CORUM (downloaded on 2018-06-29) (Ruepp et al., 2010), Signor2 (downloaded on 2018-10-29) (Perfetto et al., 2016), and Reactome (downloaded on 2018-11-01) (Fabregat et al., 2018). We used this interaction set to assess the *trans*-effects of these genes. After excluding silent mutations, samples were separated into mutated and WT groups for each gene of interest, removing samples with missing values. We used the Wilcoxon rank-sum test to report differentially expressed features (RNA, proteins, or phosphosites) between the two groups, requiring at least 3 samples in each comparison group. Differentially enriched features passing an FDR < 0.05 cut-off were separated into two categories based on *cis*- and *trans*-effects.

Copy Number Impacts on Gene and Protein Levels

To infer focal-level significant copy number alterations (SCNA) we used GISTIC2 (Mermel et al., 2011) with the default parameters except for increased thresholds for amplifications and deletions (i.e., -ta and -td parameters of GISTIC2), that were set to 0.4, and confidence level set to 0.95. This analysis was performed on the segment-level SCNA data for the autosomes.

We first filtered all the genes to those with quantifiable copy number, gene expression, and proteomics ($N = 11,623$). Next, we also filtered genes for those occurring in the focal amplified regions identified by GISTIC2 with Q value < 0.25 ($N = 543$). Finally, we filtered the genes by their CN-mRNA correlation and CN-protein correlation to keep the genes with significant CN *cis*-effect (FDR < 0.05, Spearman's correlation). The resulting set of genes ($N = 23$) was used for the gene set enrichment analysis to identify significantly enriched GO-biological processes (Subramanian et al., 2005).

DNA Methylation Associations with RNA, Protein, and Phosphorylation

To investigate the association between methylation and proteomics expression, for each gene, we first calculated Z scores for its mRNA expression, protein, and phosphorylation levels and beta values for DNA methylation. We then calculated Pearson correlation scores with its associated significance between methylation and gene expression, protein, and phosphorylation levels for all pairs of genes, respectively.

The 69 tumors and 9 normal adjacent tissues (NATs) covered in both RNA and DNA methylation datasets were involved in this identification of epigenetically-silenced genes. Probes that were located in CpG Islands (CpGIs) and transcript start sites (TSS) were selected except those located on X and Y chromosomes. Hierarchical clustering analysis (Scipy 1.5.2, Python package) showed that the tumor and NAT tissues can be separated based on the beta value of DNA methylation except one NAT sample ([Figure S2E](#)). The approach of identification of epigenetically-silenced genes is similar to the TCGA project ([Cancer Genome Atlas Research Network, 2017](#)). Level 3 RNA-seq RSEM data were log2-transformed [$\log_2(RSEM+1)$] and used to assess the expression levels associated with DNA methylation changes. DNA methylation and gene expression data were merged by Entrez Gene IDs. We removed the CpG sites that were methylated in the NATs (mean β -value > 0.2). We then dichotomized the DNA methylation data using a β -value of > 0.3 to definite positive DNA methylation, and further eliminated CpG sites methylated in fewer than 3% of the tumor samples. For each probe/gene pair, we applied the following algorithm: 1) classify the tumors as either methylated ($\beta > 0.3$) or unmethylated ($\beta \leq 0.3$); 2) compute the mean expression in the methylated and unmethylated groups; 3) compute the standard deviation of the expression in the unmethylated group. We then selected probes for which the mean expression in the methylated group was lower than 1.64 (10% of one-sided Z distribution) standard deviations of the mean expression in the unmethylated group and the NAT tissues. We labeled each individual tumor sample as epigenetically silenced for a specific probe/gene pair selected from above if: a) it belonged to the methylated group and b) the expression of the corresponding gene was lower than the mean of the unmethylated group of samples. If there were multiple probes associated with the same gene, a sample identified as epigenetically silenced at more than or equal to half the probes for the corresponding gene was also labeled as epigenetically silenced at the gene level. The methylation status of ZNF544 was also found significantly correlated with survival time by using the Python package lifelines (version 0.25.4, <https://doi.org/10.5281/zenodo.4002777>).

Differential Abundance Analysis

Paired differential abundance analysis between tumor and NATs was performed using the Wilcoxon signed-rank test. At least 50% of the paired samples were required to have non-missing values. Significance was determined to be Benjamini-Hochberg corrected p value < 0.01 and fold change was calculated as the median log2 fold change. Unpaired differential abundance analysis was performed using the Wilcoxon rank sum test. At least 4 samples in both groups were required to have non-missing values. Adjusted p values and fold changes were calculated as above. Immunohistochemistry data were collected from the Human Protein Atlas ([Uhlen et al., 2015](#)) for pancreatic cancer samples. The list of secretable proteins was also collected from the Human Protein Atlas.

Glycoproteomics Analysis

Tumor related glycoproteins identification. The Wilcoxon rank sum test was used to compare the global protein expression difference of each protein containing at least one glycopeptide identified in the glycoproteomics data (termed glycoprotein) in tumors and NATs. At least 50% of all samples were required to have non-missing values. At least 4 samples in each group were required to have non-missing values. The p values were corrected to FDR values using the Benjamini-Hochberg method (statsmodels.stats.multitest.multipletests, version 0.12.0, Python 3.7). The statistical significant up-/downregulations were determined by using FDR < 0.01 , while the median log2 fold changes (\log_2FC) = 1 or -1 were applied to further dichotomize the significant changes to '2x up/down' and 'up/down' respectively ([Figure 4A](#)). The secreted-to-blood glycoproteins were annotated with their gene names. The protein subcellular location information ([Table S4](#)) was collected from two resources: The Human Protein Atlas (<http://www.proteinatlas.org/>) ([Thul et al., 2017](#)) and UniprotKB (<https://www.uniprot.org/>) ([The UniProt Consortium, 2017](#)). The sunburst plot (Plotly, Python package) of the up-/downregulated glycoproteins and their corresponding cellular locations were shown in [Figure S4A](#). The gene set enrichment analysis of altered glycoproteins was achieved by Webgestalt (<http://www.webgestalt.org/>).

Impact of early Stage and KRAS hotspot mutations on N-linked glycoprotein expression. We compared the glycoprotein expression measured in tumors (including subsets of early stage: stage I and II, and four most common KRAS hotspot mutations: G12D, G12R, G12V, and Q61H) and normal tissues (including NATs and normal duct tissues). The Wilcoxon rank sum test was applied on each pair of comparison (stats, R package) to investigate the secreted glycoproteins significantly upregulated in tumors ([Figure 4B](#); [Table S4](#)). The P values were adjusted by Benjamini-Hochberg procedures. If the p value was < 0.01 , the result was annotated as " $>2x$ up" when the fold change > 2 , otherwise it was annotated as "up." The Wilcoxon tests were also applied in the investigation of tumor (all tumors, early stage tumors, and four KRAS mutant subsets: G12D, G12V, G12R, and Q61H) and normal (NATs or normal duct tissues) comparison for other proteins, including MUC family proteins ([Figure S4C](#)), CEACAM5 and CEACAM6 ([Figure S4D](#)), and LGALS3BP, HPX, COL6A1 and their corresponding glycopeptides ([Figure S4E](#)).

Protein glycosylation comparison on protein level and intact glycopeptide level. The log2 fold change (FC) of the intact glycopeptides and the corresponding global protein expression were shown in [Figure 4C](#). The associated glycans on the intact glycopeptides were classified to three groups of oligomannose (HM), fucose (Fuc), and sialic acid (Sia) based on the composition of the glycans. The projection of the distribution of log2 FC values were shown in the top and right side for protein and intact glycopeptides, respectively.

Correlation between the glycosylation enzymes and intact glycopeptide expression and investigation of glycosylation biosynthetic pathways. The intact glycopeptide expression was hypothesized to be influenced at least by the expression of substrate glycoproteins and glycosylation enzymes. The Spearman's rank correlation coefficient was used to measure the correlation between the abundance (log2 ratio values) of intact glycopeptides and the abundance of glycosylation enzymes identified from the global proteomic data in this study. The correlation matrix was further arranged by the order of enzymes in the glycosylation synthetic pathways and visualized in [Figure 4D](#). The glycan compositions were linked to the intact glycopeptides in the middle panel of [Figure 4D](#). The result of Wilcoxon ranked sum tests on the tumor/NAT abundance comparison of the glycosylation enzymes was shown in [Figure 4E](#).

to illustrate the overall trend of downregulated precursor pathway and upregulated capping pathways of glycosylation in tumors on protein level. The same approach was applied on mRNA data and shown in [Figure S4F](#).

Kinase and Substrate Co-regulation

To discover the phosphorylation events that were relevant to PDAC, we utilized phosphosite abundance data to examine the relationship between phospho-substrates and their associated kinases. The kinase-substrate association was first extracted from Phospho-SitePlus ([Hornbeck et al., 2015](#)) to eliminate phosphosites that were not reported as well as those without associated kinases identified in our global proteome data. Next, we inspected any substantial differences among 41 tumor/NAT pairs, especially those showing higher changes in tumors, by calculating the fold change (log2 scale) between each paired sample as well as groups (median log2 fold change). Finally, we ranked each tumor (> 1.5 fold increase) among different kinase-substrate pairs to obtain the high ranked phospho-substrate events in the majority of tumors. We identified five phospho-substrate events of five kinases with inhibitors that are either FDA-approved or under investigation. Data was analyzed using Omic-Sig (<https://github.com/hzhangjh/Omic-Sig>) ([Lih et al., 2019](#)). Kinases enriched in different hotspot KRAS mutations were stratified from the phospho-substrates (at least 2 substrates) showing elevated expression profiles (> 2 fold increase with adjusted p < 0.05) in the differential analysis between KRAS mutant tumors and NATs. The differential analysis was conducted using Wilcoxon rank-sum test (unpaired samples) and Wilcoxon signed-rank test (paired samples) in transcriptomics, global proteomics, or phosphoproteomics data (at least 50% of all samples were required to have non-missing values) between PDAC tumors and NATs/normal ductal tissues as well as between early stage PDAC tumors and NATs/normal ductal tissues (proteins/phosphosites quantified in at least 4 samples in both groups). The p value was adjusted using the Benjamini Hochberg method. The druggability score was calculated by summing the number of PDAC cell lines with positive drug response from Genomics of Drug Sensitivity in Cancer (<https://www.cancerrxgene.org/>) and then log transformed.

RNA Subtyping

The RNA subtyping was performed similarly to the TCGA PDAC study ([Cancer Genome Atlas Research Network, 2017](#)). Specifically, for the three RNA subtyping schemes reported previously (i.e., Collisson, Bailey, and Moffitt), the gene signatures were obtained from the original publications ([Bailey et al., 2016](#); [Collisson et al., 2011](#); [Moffitt et al., 2015](#)). The harmonized RNA expression matrices (with zero counts less than 50% for all the genes) for these gene signatures (Collisson N = 61; Bailey N = 488 and Moffitt N = 49) were normalized by z-scoring in the gene-wise manner. Next, we applied consensus clustering ([Wilkerson and Hayes, 2010](#)) to these subsetted RNA matrices to identify sample groups with distinct expression patterns for these signature genes. We chose the K value (i.e., cluster number) equal to the reported subtype numbers of each subtyping scheme, after we checked the Consensus Cumulative Distribution Function (CDF) plot and the Delta Area plot to ensure that these Ks indeed represented the best cohort partition ([Wilkerson and Hayes, 2010](#)). These clusters were further labeled by interrogating their expression of the signature genes used at the first place.

Multi-omics Clustering

Non-negative matrix factorization (NMF)-based multi-omics clustering was performed similar to as previously described ([Gillette et al., 2020](#); [Huang et al., 2021](#); [Krug et al., 2020](#); [Wang et al., 2021](#)). Briefly, NMF was used to perform unsupervised clustering of tumor samples using gene copy number aberrations, mRNA and protein expression, and phosphorylation and glycosylation sites abundances. To enable integrative multi-omics clustering, we required all data types (and converted if necessary) to represent ratios to either a common reference measured in each TMT plex (proteome, phosphorylation and glycosylation sites) or an *in-silico* common reference calculated as the median abundance across all samples. All data tables were then concatenated and only features quantified in all tumors were used for subsequent analysis (no missing values were allowed). Features with the lowest standard deviation (bottom 5th percentile) across all samples were deemed uninformative and were removed from the dataset. Each column in the data matrix was further scaled and standardized such that all features from different data types were represented as z-scores. Since NMF requires a non-negative input matrix, the data matrix of z-scores was further converted into a non-negative matrix as follows:

1. Create one data matrix with all negative numbers zeroed.
2. Create another data matrix with all positive numbers zeroed and the signs of all negative numbers removed.
3. Concatenate both matrices resulting in a data matrix twice as large as the original, but with positive values only and zeros and hence appropriate for NMF.

The resulting matrix was then subjected to NMF analysis leveraging the NMF R-package ([Gaujoux and Seoighe, 2010](#)) and using the factorization method described in [Brunet et al. \(2004\)](#). Given a factorization rank k (where k is the number of clusters), NMF decomposes a $p \times n$ data matrix V into two matrices W and H such that multiplication of W and H approximates V . Matrix H is a $k \times n$ matrix whose entries represent weights for each sample (1 to N) to contribute to each cluster (1 to k), whereas matrix W is a $p \times k$ matrix representing weights for each feature (1 to p) to contribute to each cluster (1 to k). Matrix H was used to assign samples to clusters by choosing the k with maximum score in each column of H . For each sample, we calculated a cluster membership score as the maximal fractional score of the corresponding column in matrix H .

To determine the optimal factorization rank k (number of clusters) for the multi-omic data matrix, a range of clusters between $k = 2$ and 10 was tested. For each k we factorized matrix V using 50 iterations with random initializations of W and H . To determine the optimal factorization rank we calculated two metrics for each k : 1) cophenetic correlation coefficient measuring how well the intrinsic structure of the data was recapitulated after clustering and 2) the dispersion coefficient of the consensus matrix as defined in [Kim and Park \(2007\)](#) measuring the reproducibility of the clustering across 50 iterations. The optimal k was defined as the maximum of the product of both metrics for cluster numbers between $k = 2$ and 10. Having determined the optimal factorization rank k , and in order to achieve robust factorization of the multi-omics data matrix, the NMF analysis was repeated using 500 iterations with random initializations of W and H .

Matrix W containing the weights of each feature in a certain cluster was used to derive a list of representative features separating the clusters using the method proposed in [Kim and Park \(2007\)](#). Cluster-specific features were further subjected to a 2-sample moderated t test ([Ritchie et al., 2015](#)) comparing the feature abundance between the respective cluster and all other clusters. Derived p values were adjusted for multiple hypothesis testing using the method proposed by [Benjamini and Hochberg \(1995\)](#).

In order to functionally characterize the clustering results, normalized enrichment scores (NES) of cancer-relevant gene sets were calculated by projecting the matrix of signed multi-omic feature weights (W_{signed}) onto Hallmark pathway gene sets ([Liberzon et al., 2015](#)) using ssGSEA ([Barbie et al., 2009](#)). To derive a single weight for each gene measured across multiple omics data types (protein, RNA, phosphorylation site, acetylation site) we retained the weight with maximal absolute amplitude. We used the ssGSEA implementation available on <https://github.com/broadinstitute/ssGSEA2.0> using the following parameters:

- gene.set.database = "h.all.v6.2.symbols.gmt"
- sample.norm.type = "rank"
- weight = 1
- statistic = "area.under.RES"
- output.score.type = "NES"
- nperm = 1000
- global.fdr = TRUE
- min.overlap = 5
- correl.type = "z.score"

The association between the resulting clusters and inferred phenotypes (e.g., RNA subtypes) and clinical variables, either a Fisher's exact test (R function *fisher.test*) for discrete variables or a Wilcoxon rank-sum test (*ggpubr* R-package) in case of continuous variables was used to assess overrepresentation in tumors assigned to each cluster.

Inference of subtype-specific drug signatures (DSigDB GSEA)

Gene Set Enrichment Analysis (GSEA) implemented in the WebGestaltR R-package ([Liao et al., 2019](#)) was used to infer signatures of approved drugs (D1, 1,202 gene sets) and kinase inhibitors (D2, 1,220 gene sets) available in the drug signatures database (DSigDB; [Yoo et al., 2015](#); <http://dsigdb.tanlab.org/DSigDBv1.0/>). Based on the signed multi-omic feature weights (W_{signed}) for the two proteogenomic clusters described above, a single weight for each protein was derived by retaining the weight with maximal amplitude. Negative weights indicated in cluster C1, positive weights proteins with specific expression in cluster C2, respectively. The resulting vector of protein weights ($n = 5,773$) was used as ranking in WebGestaltR. Normalized enrichment scores and p values were based on 1,000 permutations. Other relevant parameters in WebGestaltR were set as follows: sigMethod = "top," topThr = 10, minNum = 5, fdrThr = 0.05, sigMethod = "fdr," fdrMethod = "BH."

Methylation-based Deconvolution

We used an established methylation-based deconvolution method, EDec ([Onuchic et al., 2016](#)) to dissect the composition of different cell types within the whole bulk tumor. In brief, EDec assumes that the methylation observed from the whole bulk tumor is a linear combination of the methylation from each constituent cell type, weighted by their proportion within the whole bulk tumor. Mathematically, EDec applies the NMF algorithm to the methylation matrix (i.e., a # methylation feature by # samples matrix, with entries being beta values) profiled from the whole bulk tumor and generates a # methylation feature by # cell-type matrix (with entries being beta vale) and a # cell type by # samples matrix (with entries being cell type proportions). To ensure that the second matrix reflects the cell type composition, EDec uses only the methylation features (i.e., probe-level methylation) that are known to have differential levels across the presumed cell types within the tumor. We selected such methylation features from cell lines or physically purified tissues that are available in the public database ([Table S6](#); adapted from [Lurie et al., 2020](#)). Based on the robustness of NMF matrix decomposition, the methylation-based deconvolution resulted into four cell types: tumor epithelial cells, immune cells, stromal cells and mature exocrine and endocrine cells.

Tumor Microenvironment Inference

The RNA-based tumor microenvironment inference tool ESTIMATE ([Yoshihara et al., 2013](#)) was used to derive the overall immune score and stromal score for each sample. In addition, two established RNA-based immune cell inference tools, xCell ([Aran et al.,](#)

2017) and MCPCounter (Becht et al., 2016), were used to dissect the relative level of different immune cell infiltration. For these tools, we used RNA expression quantified as upper-quantile normalized RSEM and kept only genes with zero counts less than 50% as the input. We found that the xCell results were sparse for some immune cells and further filtered out immune cell types with zero readout in > 80% of samples. The remaining cell types were used to derive the microenvironment-based grouping using consensus clustering (Wilkerson and Hayes, 2010). To further explore the relationship between these tumor microenvironment components, a correlation network was built by performing the Pearson's correlation for all pairs of xCell components and linking the pairs with adjusted p value < 0.05 with edges. The network module discovery was performed by edge betweenness implanted in the R package 'igraph'.

To contextualize our immune profiling results with current understanding in the field, we extracted the gene signatures from relevant publications, including the "normal stromal" and "activated stromal" genes from Moffitt et al., 2015, "immune" and "ECM" genes from Maurer et al., 2019, and 'classical' (signature 1 and 6) and 'basal-like' (signature 2 and 10) genes from Chan-Seng-Yue et al., 2020 and used single sample GSEA (ssGSEA that was implemented in GSVA R package (Hänzelmann et al., 2013) to infer the corresponding microenvironment activities. For the ssGSEA scores derived from Maurer et al., 2019, we applied the data deconvolution method reported in the publication and used stroma-specific gene expression for the analysis. In addition, we generated the subtypes reported by Puleo et al., 2018 by centroid-based subtyping using the reported gene signatures from the publication.

Adjustment for Epithelial Content

For the fifteen patients with an ESTIMATE score for both tumor and NAT samples, a linear mixed model was used to correct for non-epithelial content. The ImerTest package (doi = {10.18637/jss.v082.i13}) in R was used. Proteomics data had to be available for both the tumor and normal samples for at least 10 of the patients. The tumor type and z-scored ESTIMATE score were fixed effects and the patient was a random effect. P value for expression was adjusted using Benjamini-Hochberg, with 0.01 considered significant. A beta value > 1 for expression was used to filter tumor-associated proteins.

Over-representation Analysis

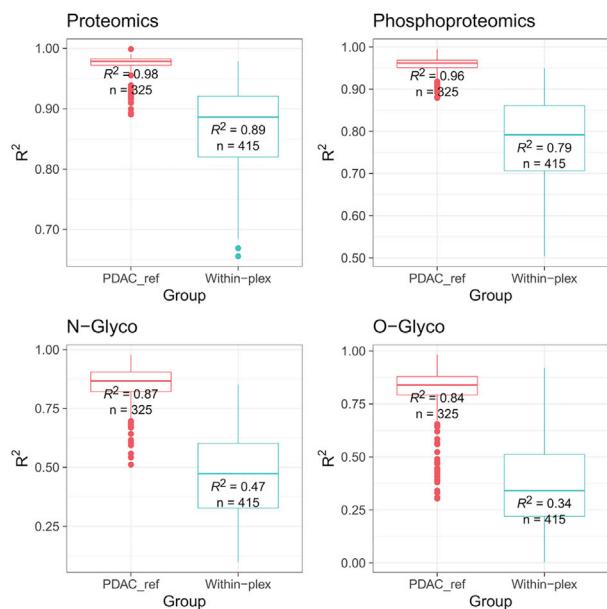
Over-representation analysis of Gene Ontology Biological Process terms was performed using WebGestaltR (Liao et al., 2019) with the > 2-fold increased or decreased proteins in tumors versus NAT compared to a background of all quantified proteins (proteins non-missing in at least 50% of the paired samples). For PTMs, proteins containing at least one PTM that was > 2-fold increased or decreased in tumors versus NAT were compared to a background of proteins containing at least one quantified site (non-missing in at least 50% of the paired samples). A Benjamini-Hochberg corrected p value of 0.01 was considered significant.

Survival Analysis

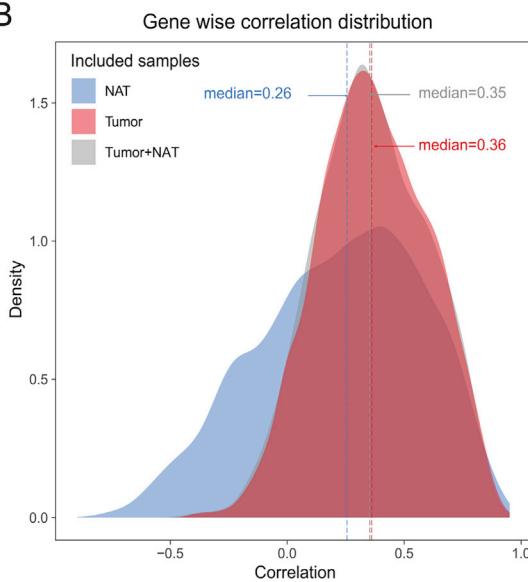
Cox proportional hazards regression (from the R package *survival*) on overall survival was performed to test the association between survival outcomes to continuous variables. Logrank test (from the R package *survminer*) was used to test the differential survival outcomes between categorical variables. Samples with a death event within 30 days of surgery were excluded. For the survival association analysis for the tumor proteomics data, the proteins were filtered to keep the ones with no-missing values for at least 10 patients.

Supplemental figures

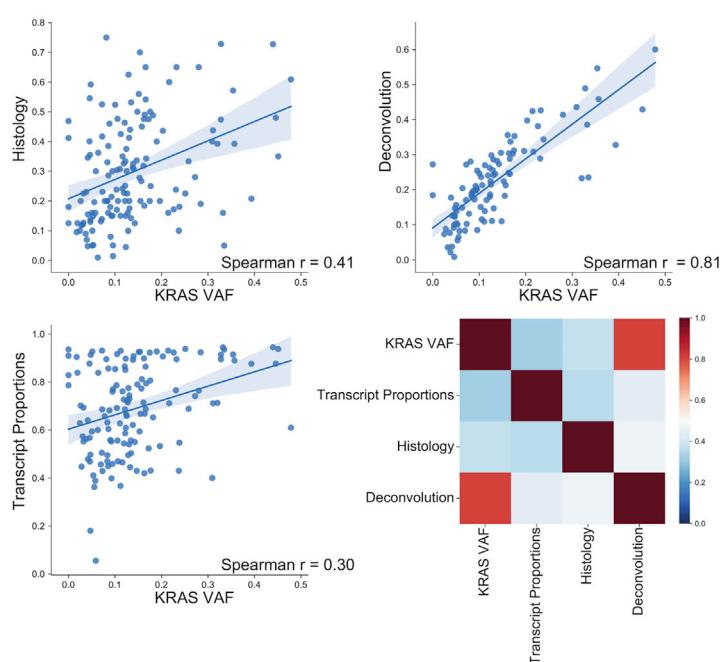
A



B



C



D

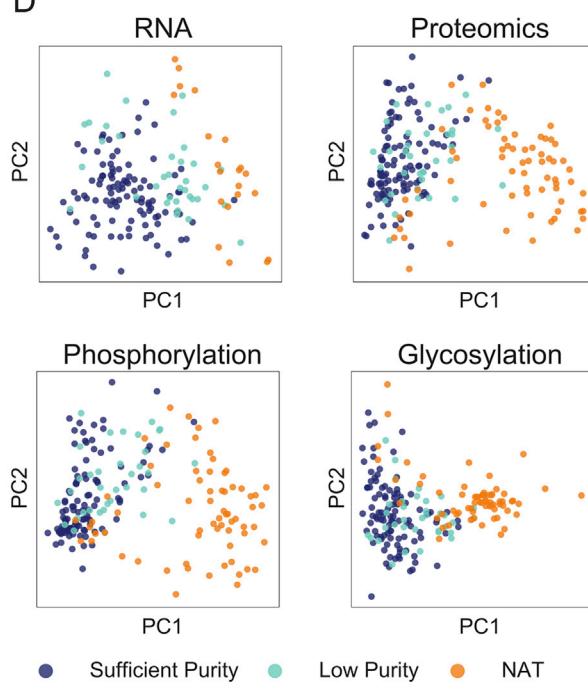
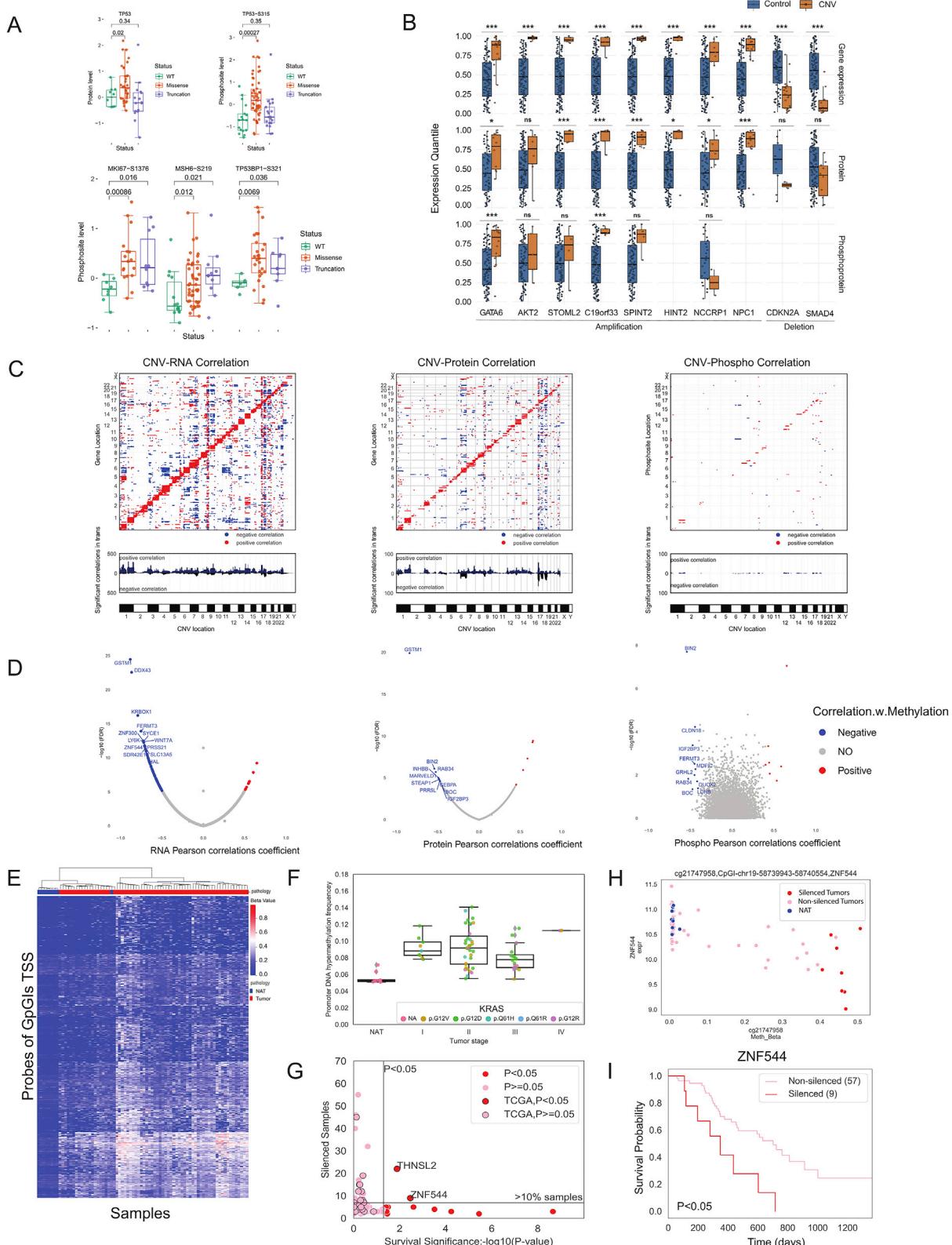


Figure S1. Characterization of PDAC cohort and proteogenomic data, related to Figure 1

(A) Pairwise comparisons between replicate samples and samples within TMT plex. (B) Gene-wise correlation of mRNA and protein expression in NATs, tumors, and all tumors and NATs combined. (C) Estimation of neoplastic cellularity by histology, DNA methylation-based deconvolution, and transcript proportion-based methods and their correlation with KRAS VAF tumor fraction estimates. (D) PCAs at the RNA, protein, phosphorylation, and N-linked glycosylation levels of sufficient purity, low purity, and NAT samples.



(legend on next page)

Figure S2. Impacts of genomic and epigenetic alterations, related to Figure 2

(A) *Cis-* and *trans*-effects of the *TP53* truncation and missense mutation groups. Samples carrying a frameshift insertion or deletion, splice-site mutation, nonsense mutation, or CNV deletion in *TP53* were included in the truncation group. The missense group was composed of samples with missense mutations in *TP53*. Samples with wild-type *TP53* serve as the control. (B) CNV impacts on gene, protein, and phosphoprotein of the most common proteins with CNV events. The control group in each comparison includes all samples without the copy number event for each gene or protein. (C) Functional impacts of CNVs on gene/protein/phosphorylation expression. The data in the top panel show the correlations of CNVs to mRNA, protein, and phosphorylation levels with significant positive correlations in red and significant negative correlations in blue (adjusted p value < 0.05). The data in the bottom panel show the summed number of significant positive and negative correlations at mRNA, protein, and phosphopeptide. (D) Correlation and anti-correlation of gene, protein, and phosphoprotein levels with DNA methylation levels. (E) The two dimensional hierarchical clustering of methylation probes (y axis) and tumor and NAT samples (x axis) on PDAC DNA methylation data. (F) DNA methylation status between tumors with different pathological stages and NATs. (G) Scatterplot of epigenetically silenced genes on survival significance and silenced samples counting. (H) Scatterplot of beta values of probe cg21747958 and RNA expression of ZNF544 in tumors and NATs. (I) Kaplan-Meier Plot of ZNF544 methylation association with overall survival. The two groups were separated by whether ZNF544 is silenced in the sample.

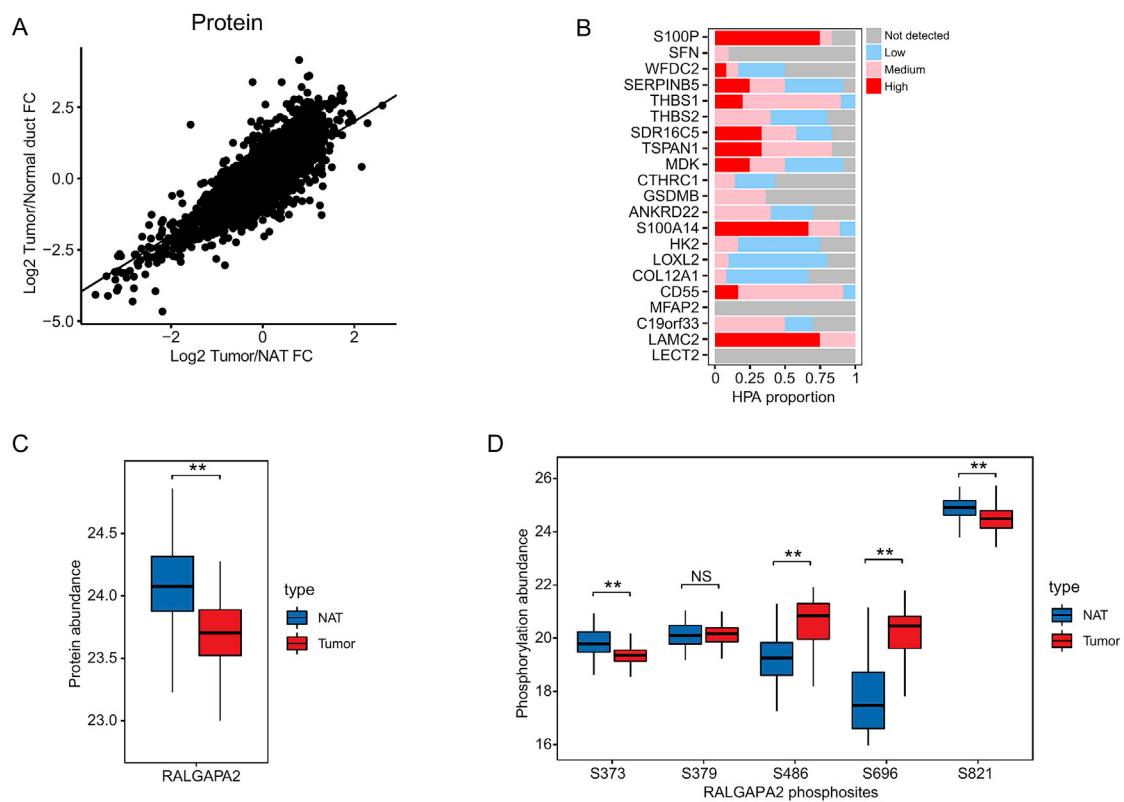
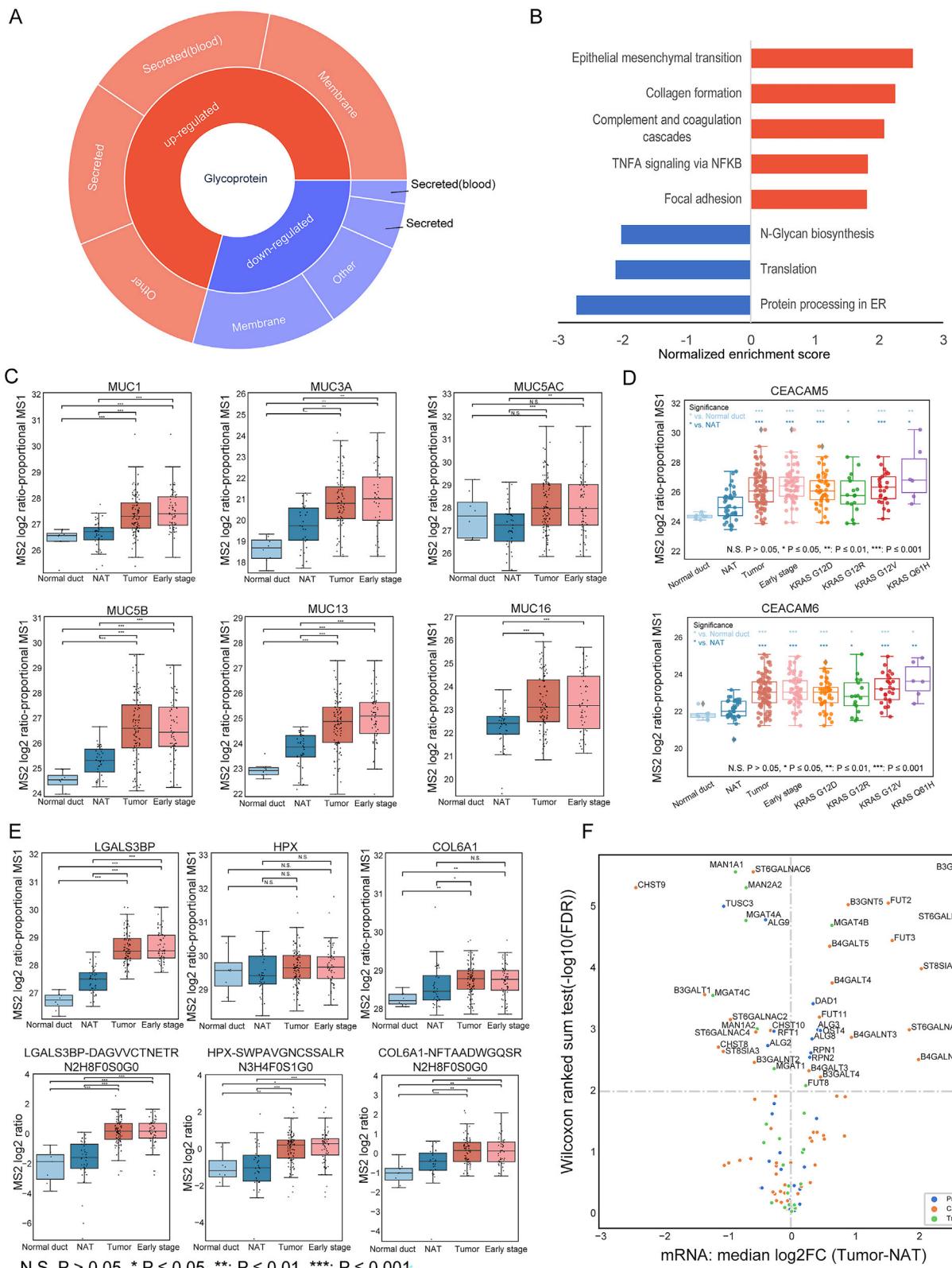


Figure S3. Tumor-associated protein and phosphosites, related to Figure 3

(A) Median log₂ protein fold change between tumor and paired NAT or unpaired normal ductal tissue. (B) Immunohistochemistry levels for tumor-associated proteins in pancreatic cancer samples in Human Protein Atlas. (C) RALGAPA2 protein abundance in tumor and NAT. (D) RALGAPA2 phosphosite abundance in tumor and NAT. **p < 0.001 Wilcoxon signed rank test, NS = not significant.



N.S. P > 0.05, * P ≤ 0.05, **: P ≤ 0.01, ***: P ≤ 0.001.

(legend on next page)

Figure S4. Differential expression of glycoproteins and glycosylation enzymes between PDAC tumors, early-stage PDAC tumors, NATs, and normal duct tissues, related to Figure 4

(A) Cellular distribution of tumor upregulated and downregulated N-linked glycoproteins. (B) Analysis of significantly differentially regulated pathways between tumors and NATs at the N-linked glycoprotein expression level. (C) Expression pattern of mucin 1 (MUC1), mucin 3A (MUC3A), mucin 5AC (MUC5AC), mucin 5B (MUC5B), mucin 13 (MUC13), and mucin 16 (MUC16) in tumors, early stage tumors, NATs, and normal duct tissues. The significance is defined as: N.S. $p > 0.05$, * $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ (D) Expression pattern of CEA cell adhesion molecule 5 (CEACAM5) and CEA cell adhesion molecule 6 (CEACAM6) in tumors, early stage tumors, tumors with different KRAS hotspot mutations, NATs, and normal duct tissues. (E) Expression pattern of galectin binding protein 3 (LGALS3BP), hemopexin (HPX), and collagen type VI alpha 1 chain (COL6A1) in tumors, early stage tumors, NATs, and normal duct tissues. (F) Differential gene expression of N-linked glycosylation enzymes between tumors and NATs.

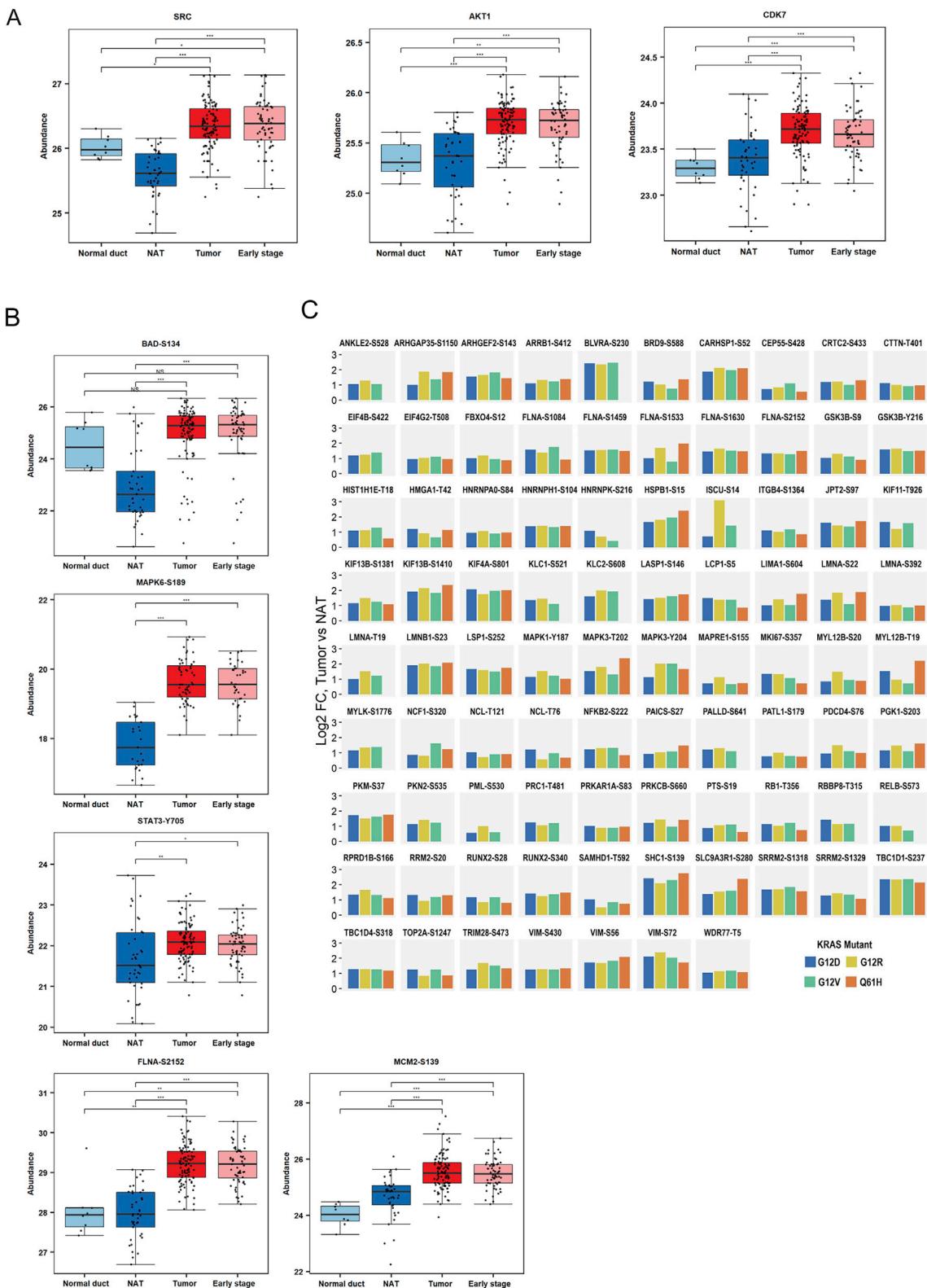


Figure S5. Kinase-substrate pairs and phosphosites enriched in KRAS mutant tumors, related to Figure 5

(A) Expression profiles of SRC, AKT1, and CDK7 in Normal duct, NAT, Tumor, and Early stage. (B) Expression profiles of BAD-S134, MAPK6-S189, STAT3-Y705, FLNA-S2152, and MCM2-S139 in Normal duct, NAT, Tumor, and Early stage. (C) Bar charts showing the phospho-substrates enriched (> 2-fold increase with

(legend continued on next page)

adjusted p value < 0.05) in the *KRAS* mutant tumors relative to NATs that were used for identifying kinases enriched in different *KRAS* hotspot mutant tumors. Normal duct: normal ductal tissues; NAT: normal adjacent tissues; Tumor: all PDAC tumors; Early stage: Stage I and II PDAC tumors. Asterisks represent significant differences between two groups (Benjamini-Hochberg adjusted p): *p < 0.05; **p < 0.01; ***p < 0.001; N.S., not significant.

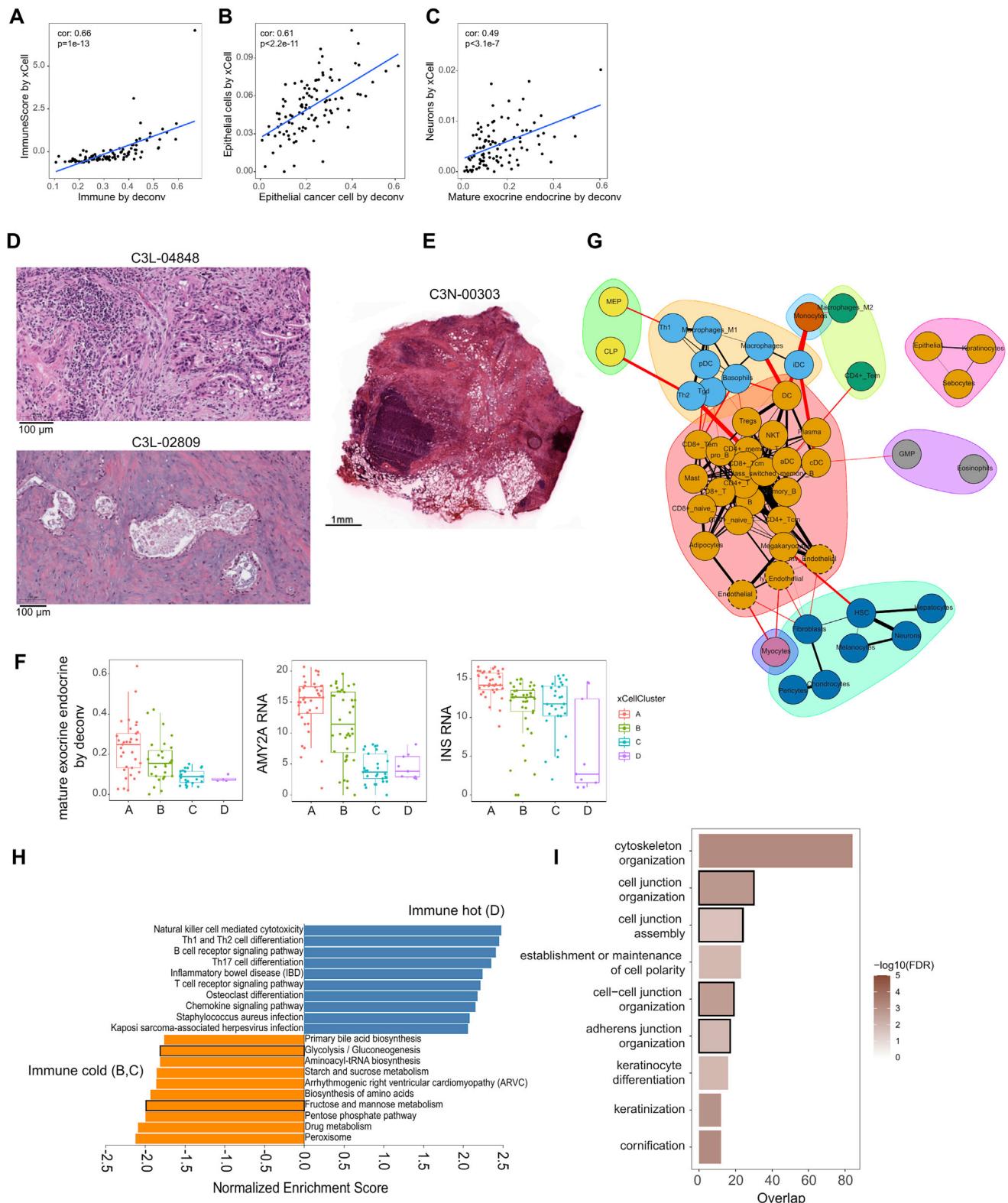
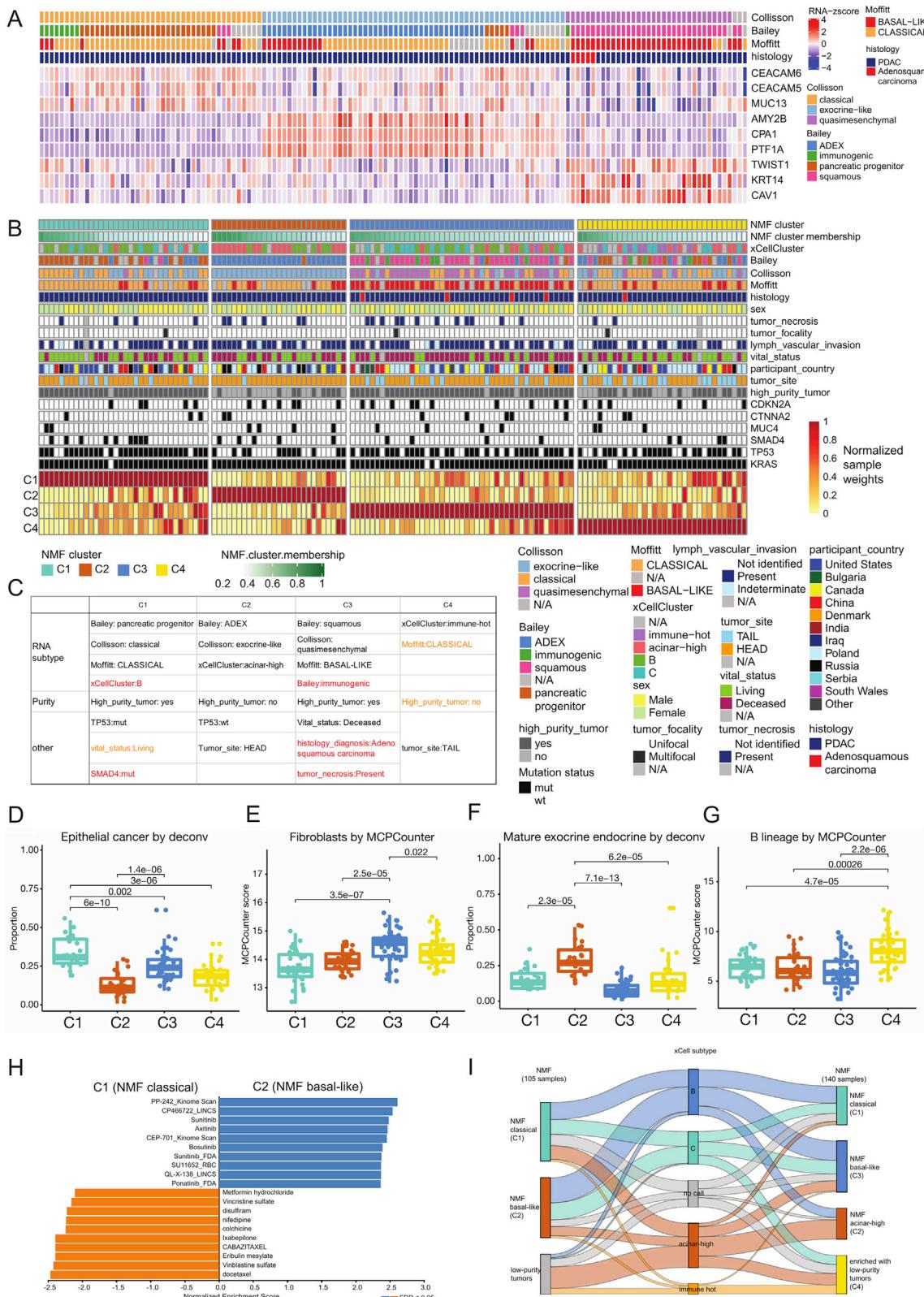


Figure S6. Identification of biological processes associated with immune-cold tumors, related to Figure 6

(A-C) The relative levels of tissue components (A: immune; B: Epithelial; C: exo/endocrine) inferred by xCell correlate with the counterparts derived from methylation-based deconvolution. Note that xCell gene signatures do not explicitly contain exo/endocrine cells and the neurons were instead chosen as they are

(legend continued on next page)

similar to exo/endocrine cells in terms of the regulatory factors utilized during development and differentiation (Arntfield and van der Kooy, 2011). Shown are Pearson's correlation coefficients and p values. (D) Representative histological images showing samples with immune cell infiltration (C3L-04848, 'immune hot') and samples with immune cell depletion (C3L-02809, 'immune cold'). (E) Representative histological image showing an example of contamination of lymph nodes (C3N-00303, to be excluded from 'immune hot'). (F) Boxplots comparing the abundance of exo/endocrine cells (from methylation-based deconvolution) and gene expression of amylase (AMY) and insulin (INS). (G) Correlation network showing the relationship among all cell types inferred by xCell. Several network modules are identified to reflect the close connection for certain cell types and denoted by different colors. The endothelial cells are highlighted by dashed circles. (H) Pathway enrichment using the differentially expressed proteins between immune hot (cluster D) and immune cold (cluster B and C) samples. Shown are pathways that are statistically significant ($FDR < 0.05$) for either directions. (I) Pathway enrichment of genes with phospho-specific regulation in the immune cold samples ($FDR < 0.05$). Pathways highlighted by black borders are the ones discussed in the main text.



(legend on next page)

Figure S7. Unsupervised, non-negative matrix factorization-based proteogenomic subtyping of 140 tumors using gene copy number, mRNA and protein expression, and phosphosite and glycosylation site abundances and clinical relevance of binary proteogenomic subtypes from the 105 tumors and the 4 proteogenomic subtypes from all 140 tumor samples, related to Figure 7

(A) Heatmap showing the clustering of samples by the three published RNA subtyping schemes. Shown are also the histological diagnosis (PDAC versus pancreatic adenosquamous carcinoma) and selected signature gene expression. (B) Heatmap depicting the contribution of samples (x axis) to each of the four clusters (C1-C4, y axis). Sample weights have been normalized by the maximum weight in each column. Cluster membership scores indicating the strength of association of each sample with a given cluster were calculated as proportional weights. The columns of the matrix are ordered by NMF subtype and decreasing cluster membership score. (C) Overrepresentation analysis of clinical variables and RNA-subtypes in each proteogenomic cluster (Fisher's exact test) shown in panel A). All clusters showed significant overlap with RNA-subtypes (Bailey, Collisson, and Moffitt). Low purity tumors were enriched in C2 and C4. (D-G) Distribution of inferred cell types (y axis) across NMF-subtypes (x axis). (H) Cluster-specific drug signatures inferred by Gene Set Enrichment Analysis of DSigDB gene sets using NMF weights of protein features (derived from binary proteogenomic subtypes from the 105 tumors) as a ranking. NMF classical cluster (C1) showed enrichment of gene sets targeted by various therapeutics, while the NMF basal-like (C2) cluster was enriched for gene sets of several kinase inhibitors targeting mTOR, ATM, PDGFR, VEGF or BTK (FDR < 0.05). (I) Sankey diagram depicting the association of samples classified into xCell-based subtypes (center) with binary proteogenomic subtypes from 105 tumors (left) and the 4 proteogenomic subtypes from all 140 tumor samples (right). Nodes are colored according to the NMF or xCell subtype, links between nodes are colored by xCell subtype.