

## CANCER

# Digital spatial profiling of intraductal papillary mucinous neoplasms: Toward a molecular framework for risk stratification

Matthew K. Iyer<sup>1</sup>, Chanjuan Shi<sup>2</sup>, Austin M. Eckhoff<sup>1</sup>, Ashley Fletcher<sup>1</sup>, Daniel P. Nussbaum<sup>1</sup>, Peter J. Allen<sup>1\*</sup>

The histopathologic heterogeneity of intraductal papillary mucinous neoplasms (IPMN) complicates the prediction of pancreatic ductal adenocarcinoma (PDAC) risk. Intratumoral regions of pancreaticobiliary (PB), intestinal (INT), and gastric foveolar (GF) epithelium may occur with either low-grade dysplasia (LGD) or high-grade dysplasia (HGD). We used digital spatial RNA profiling of dysplastic epithelium (83 regions) from surgically resected IPMN tissues (12 patients) to differentiate subtypes and predict genes associated with malignancy. The expression patterns of PB and GF lesions diverged from INT, suggesting that PB and GF arise from a common lineage. Transcriptional dysregulation within PB lesions mirrored that of PDAC, whereas INT and GF foci did not. Tumor necrosis factor/nuclear factor  $\kappa$ B (TNF-NF $\kappa$ B) and cell cycle (cycling S and cycling G<sub>2</sub>-M) programs occurred with relative prominence in PB and INT subtypes, respectively. Together, this study delineates markers of high-risk IPMN and insights into malignant progression.

## INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) remains a leading cause of cancer death, predominantly because of the lack of early detection strategies that enable identification of patients at a potentially curable stage (1). Intraductal papillary mucinous neoplasms (IPMN) are cystic lesions of the pancreas that represent a radiographically detectable precursor to pancreatic cancer (2). While the most IPMN do not progress to malignancy, the ability to accurately differentiate those lesions at low risk for progression [low-grade dysplasia (LGD)] from those at high risk [high-grade dysplasia (HGD) and early cancer] remains elusive (3). It is widely accepted that operative treatment of HGD is appropriate, whereas radiographic surveillance is appropriate for those with LGD. Consensus guidelines designed to predict risk of HGD through clinical, radiographic, laboratory, endoscopic, and cytologic parameters have an overall accuracy of approximately 60% (4–6). Hence, there has been intense interest in the development of more accurate biomarkers for high-risk IPMN.

Pathologic characterization of IPMN demonstrates multiple dysplastic histologic epithelial subtypes that often coexist within individual specimens. Epithelial subtypes observed histologically include pancreaticobiliary (PB), intestinal (INT), gastric foveolar (GF), and an oncocytic variant. The PB and INT subtypes comprise the overwhelming majority of IPMN with propensity for invasive cancer, and GF represents an indolent subtype associated with favorable prognosis. Studies comparing patient outcomes by subtype suggest that PB histology is more likely to harbor or progress to malignancy, and patients with invasive lesions from the PB subtype experience similar outcomes as patients with conventional PDAC (7–9). Genomic interrogation of IPMN has implicated DNA alterations in *KRAS*, *GNAS*, and *RNF43* as the most prevalent events within neoplasms, but these mutations may coexist and do not reliably

associate with histologic subtype or grade of dysplasia (10–12). Mutations such as *TP53*, *CDKN2A*, *SMAD4*, and others may be predictors of HGD or invasive carcinoma when present, but their prevalence is relatively low, and the sensitivity is limited. Numerous studies have proposed possible mRNA, microRNA, and protein biomarkers for high-risk IPMN, but to date, none has been incorporated into clinical use because of prognostic inaccuracies (13–18).

Intralesional heterogeneity further complicates the discovery of markers for high-risk IPMN. Studies of bulk tissue or cyst fluid convolute the mosaic of epithelial subtypes and grades of dysplasia that occur within individual patients, thereby obfuscating any underlying signal that may be present in disease foci. Microdissection can isolate and compare regions of dysplastic epithelium, but these approaches can be technically challenging and disrupt tissue quality (19, 20). Single-cell RNA sequencing (RNA-seq) of IPMN can characterize unique cell populations within bulk tissues, but do so at the expense of their spatial relationships, and are thus unable to distinguish differences between histologic subtype and pathologic grade (21).

The emergence of multiplex digital spatial profiling addresses the above challenges in deconvoluting heterogeneous tissues to delineate disease phenotypes (22, 23). This technology offers precise comparison of gene expression among user-defined disease regions without the need for cumbersome microdissection. We recently used this technology to explore the composition of immune cells within the IPMN tumor microenvironment (24). Here, we report spatial RNA profiling of ductal epithelium across subtypes to determine markers of dysplasia and identify biological processes that associate with malignant progression.

## RESULTS

### Patient characteristics

Digital spatial RNA profiling was performed on formalin-fixed, paraffin-embedded (FFPE) tissue specimens from 12 patients who

Copyright © 2023  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
License 4.0 (CC BY).

<sup>1</sup>Department of Surgery, Duke University, Durham, NC, USA. <sup>2</sup>Department of Pathology, Duke University, Durham, NC, USA.

\*Corresponding author. Email: peter.allen@duke.edu

underwent pancreatectomy for IPMN between 2017 and 2021 (Fig. 1A). Clinicopathologic details of the cohort are summarized in Table 1. A pancreatic pathologist (C.S.) characterized the specimens and prepared tissue blocks for profiling according to the following criteria: (i) predominantly INT ( $n = 6$ ) or PB ( $n = 6$ ) histology, (ii) presence of at least HGD within the specimen, and (iii) adequate areas of LGD and HGD within a single tissue block to facilitate a controlled comparison between grades of dysplasia (Fig. 1B). Because of the heterogeneity of the tumors, the specimens also contained numerous regions of GF epithelia with uniformly LGD. Invasive carcinoma occurred in four of six PB specimens and none of the INT specimens. To compensate for this potential bias, we prepared slides from areas of the specimen that entirely lacked invasive carcinoma.

### Targeted spatial transcriptome profiling

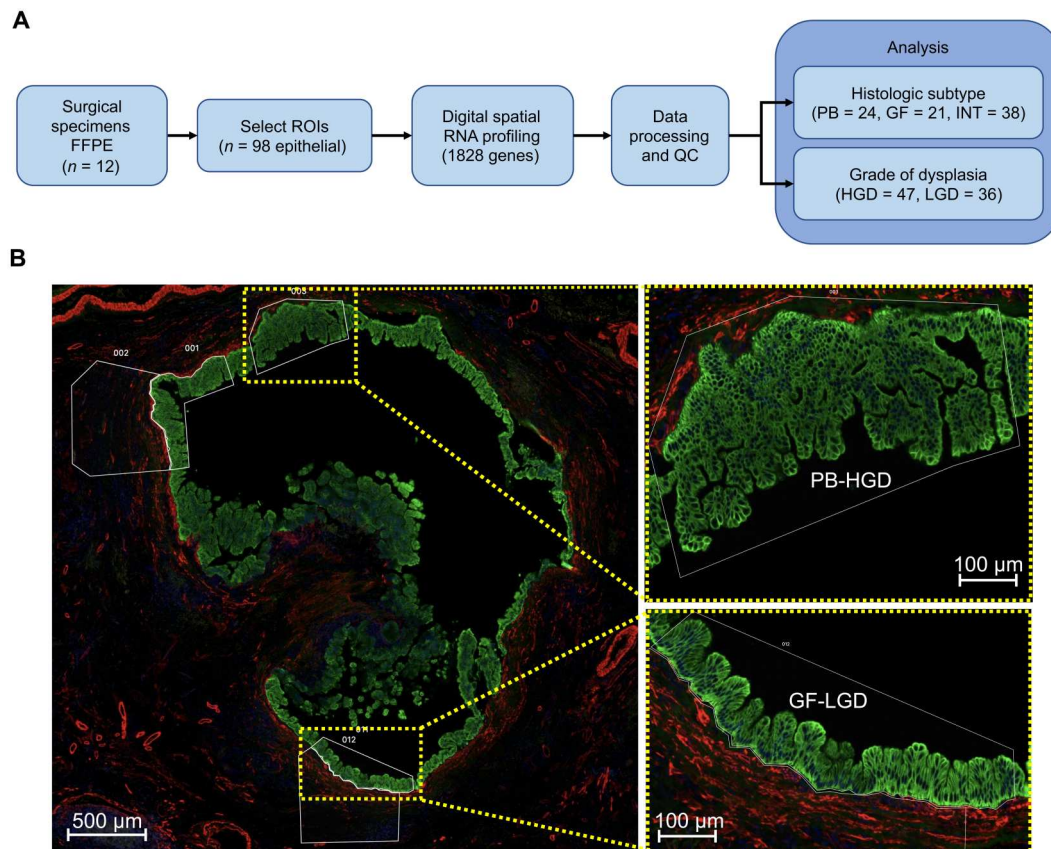
We used the Nanostring GeoMx Cancer Transcriptome Atlas (CTA) platform to profile a total of 98 epithelial areas of interest (AOIs) (50 HGD and 48 LGD) from all slides (6 to 10 per slide; table S1). The median nuclei count per AOI was 623 [interquartile range (IQR) 345 to 864]. High-throughput sequencing yielded a median of 608,651 aligned deduplicated reads per AOI (IQR 261,088 to 1,162,890). Sequencing yield was correlated with AOI nuclei count and surface area, suggesting a strong association with the quantity of in situ RNA (fig. S1, A and B). Of the 98 AOIs, 83 (84%) met quality control (QC) filtering criteria (fig.

S1C). Epithelial regions with HGD tended to contain greater cellular density than LGD AOIs, leading to increased sequencing counts and signal-to-noise area under the curve (snAUC) (Materials and Methods and fig. S1D). Of the epithelial subtypes, INT harbored the greatest cellular density, followed by PB and finally GF AOIs (fig. S1E). Sequencing yield and snAUC reflect these trends. The normalization strategy to account for this potential bias is discussed below.

A median of 48.3% of probes was expressed above background in each individual AOI (IQR 42.7 to 55.2%; table S2). Gene filtering retained 1288 of 1828 (70.4%) genes expressed at detectable levels above background in at least 20% of AOIs (fig. S2A and table S3). The raw count density distributions of the AOIs indicate that the filtered genes were of low abundance (fig. S2B). Background subtraction followed by quantile normalization was applied to account for differences in sequencing yield, enabling biological comparison of gene expression across AOIs (tables S4 and S5). The normalized count density distributions across individual AOIs were grossly similar (fig. S3A), and there was no observable bias by either pathologic grade or epithelial histology (fig. S3, B and C).

### Divergence of INT from PB and GF subtypes

An unbiased survey of expression patterns was performed using principal components analysis. The AOIs formed two groups that corresponded to histologic subtype, with INT AOIs forming one



**Fig. 1. Spatial RNA profiling of IPMN. (A)** Study overview flow diagram. **(B)** A representative high-resolution microscopy image from patient slide 6 depicting regions of HGD (PB) and LGD (GF) cyst epithelium chosen for profiling. ROIs, regions of interest; QC, quality control.

**Table 1. Patient characteristics.**

Patients	12
Median age: years (IQR)	70 (63–77)
Female sex	5 (42%)
<b>Race</b>	
Caucasian	8 (67%)
Asian	3 (25%)
Black	0 (0%)
Unidentified	1 (8%)
<b>Operation</b>	
Pancreaticoduodenectomy	9 (75%)
Distal pancreatectomy	2 (17%)
Other	1 (8%)
Median lesion size	5.2 cm
Invasive carcinoma within lesion	4 (33%)
Cancer stage	IA (2), IIA (2)
Median follow-up (months)	27
Cancer recurrence	2 (50%)
Death	3 (25%)
<b>Anatomic subtype</b>	
Main duct	5 (42%)
Branch duct	3 (25%)
Mixed	4 (33%)
<b>Predominant histologic subtype</b>	
Intestinal	6 (50%)
Pancreatobiliary	6 (50%)
Gastric foveolar	0 (0%)

group and PB-GF (non-INT type) AOIs forming the second group (Fig. 2A). The separation of AOIs by grade of dysplasia along the first two principal components varied among specimens (Fig. 2B). Of the INT patients, the AOIs from slides 4 and 12 grouped tightly together, suggesting homogeneity. Two patients (slides 1 and 3) that contained a mixture of GF (LGD) and INT (HGD) epithelium were notable because the GF and INT regions clustered with their respective histologic subtypes, rather than by patient. Notable separation between HGD and LGD AOIs occurred in INT slides 2 and 9. Invasive carcinoma, although excluded from slides prepared for digital spatial profiling, occurred in four specimens with PB histology (slides 5, 6, 8, and 11) (Fig. 2C). The HGD AOIs from these four patients grouped closely together, suggesting the existence of a high-risk phenotype. Of the two PB specimens lacking invasive carcinoma, one grouped with the other PB AOIs (slide 7) and other grouped more closely with GF AOIs (slide 10).

#### Identification of genes associated with epithelial subtype

DE analysis of PB, INT, and GF AOIs was performed in a pairwise fashion (PB versus GF, PB versus INT, and INT versus GF) to find subtype-specific gene expression (table S6). Intersections of the resulting DE genes formed PB-, GF-, and INT-specific signatures (fig. S3). The union of these sets comprised 127 histologic subtype-

specific genes (Fig. 3A). Hierarchical clustering of the AOIs produced distinct clusters corresponding to INT, GF, and PB histology. Clustering at the gene level further revealed three clusters denoted as C1-INT, C2-PB, and C3-PB-GF (non-INT). The AOIs from slides 2 and 10 did not cluster with their respective epithelial subtypes. Rather, slide 2 clustered with PB AOIs despite having a mix of INT and GF epithelium, and the PB-HGD AOIs from slide 10 clustered with GF (LGD).

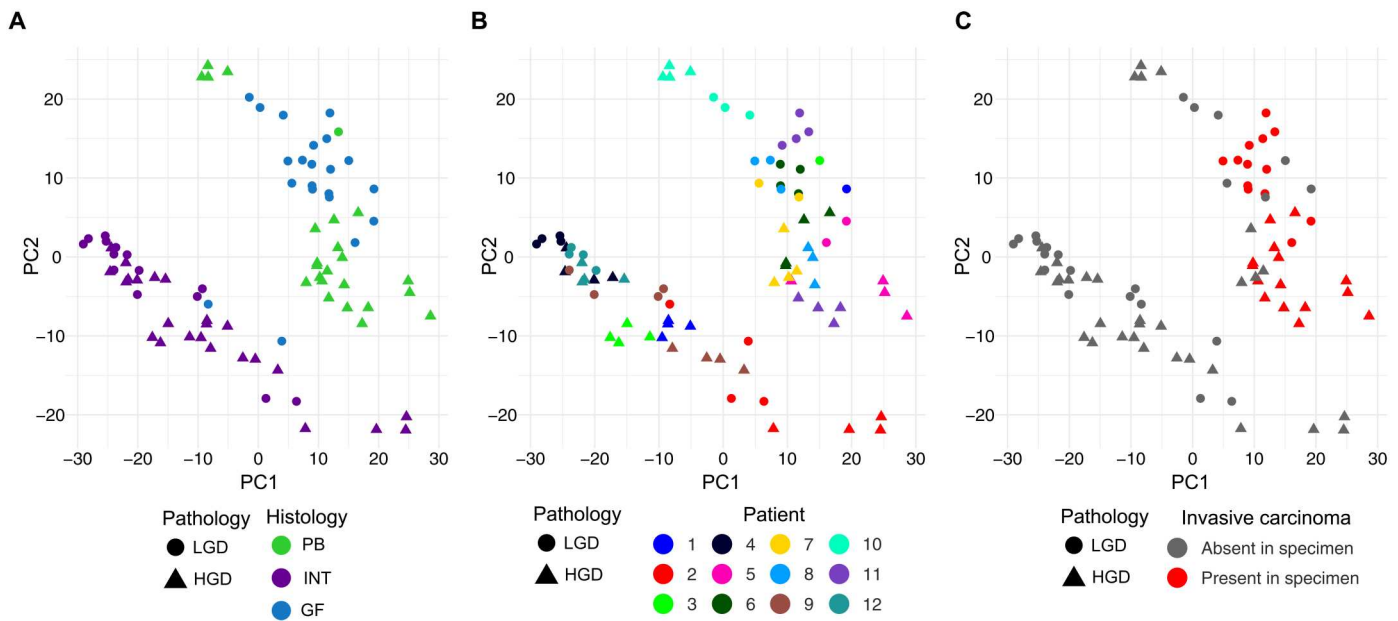
When describing histologic subtype marker genes across the three subtypes, the presence of known IPMN marker genes served as external validation for the DE results. Mucin genes, including *MUC1*, *MUC2*, *MUC4*, *MUC6*, and *MUC5AC*, have been proposed to discriminate between subtypes (14, 15), and the NanoString CTA probe set contained two of these mucin genes, *MUC1* and *MUC4*. Consistent with prior reports, *MUC1* was overexpressed in PB-GF (non-INT) AOIs, and *MUC4* was overexpressed in INT AOIs (Fig. 3, B and C) (14, 15, 25–27). At the RNA level, neither *MUC1* nor *MUC4* were found to be significantly altered in PB relative to GF AOIs. Notably, several genes outperformed *MUC1* and *MUC4* as IPMN subtype marker genes, including clusterin (*CLU*), retinol binding protein 4 (*RBP4*), and keratin 17 (*KRT17*), which showed marked overexpression in GF, INT, and PB AOIs, respectively (Fig. 3C).

#### Identification of genes associated with high-risk IPMN

Differential expression (DE) analysis of HGD versus LGD AOIs yielded 38 significant genes (30 overexpressed and 8 underexpressed) (Fig. 4A and table S6). Semisupervised clustering using these genes produced two distinct clusters of AOIs: (i) a “high-risk” cluster containing 33 of 83 AOIs (~40%) from 5 of 6 PB slides (5, 6, 7, 8, and 11) and 3 of 6 INT slides (1, 2, and 9) and (ii) a “low-risk” cluster containing all GF AOIs and HGD AOIs from slides 3, 4, 10, and 12.

Given that IPMN with HGD represents PDAC in situ, we postulated that gene expression changes essential to the progression to HGD should be retained in PDAC (8). To investigate this rationale, we compared DE genes in HGD versus LGD IPMN with PDAC gene sets curated from external sources: (i) Clinical Proteomic Tumor Analysis Consortium (CPTAC) RNA-seq and proteomics analysis of pancreatic cancer versus normal adjacent tissues, (ii) the Human Protein Atlas (HPA) analysis of The Cancer Genome Atlas (TCGA) RNA-seq data predicting prognostic genes in pancreatic cancer, (iii) Mao *et al.* analysis of bulk RNA-seq data from PDAC versus adjacent benign tissue, and (iv) Grützmann *et al.* meta-analysis of pancreatic cancer microarray experiments (see Materials and Methods) (28–31). The analyses from CPTAC and HPA (TCGA) were considered more robust as these were generated by large consortia using standardized protocols. Of the 30 genes overexpressed in HGD versus LGD IPMN, 6 were overexpressed in PDAC (CPTAC RNA-seq) and associated with unfavorable prognosis (HPA): *FOSL1*, *AREG*, *CCL20*, *SERPINB5*, *ITGA2*, and *LAMC2* (fig. S4). Additional genes found to be overexpressed in both IPMN and PDAC but not associated with prognosis included *PLAUR*, *IL1R2*, *KRT17*, *EPHA2*, *LIF*, *CD55*, and *CARD11*.

Gene Set Enrichment Analysis (GSEA) demonstrated significant enrichment [adjusted *P* value ( $P_{adj}$ ) < 0.05] against all curated gene sets (fig. S5 and table S7). Statistical significance largely depended on gene set size, with down-regulated genes exhibiting weaker enrichment than up-regulated genes. The most significant enrichment



**Fig. 2. Principal components analysis of normalized spatial RNA profiles.** Scatterplot of the first two principal components (PC1 and PC2). The shape of AOIs reflects the grade of dysplasia. (A to C) AOIs colored by (A) histologic subtype, (B) patient/slide ID, and (C) presence of carcinoma elsewhere in the specimen.

was observed with the RNA-seq analysis from CPTAC (Fig. 4B) and the HPA analysis of prognosis (Fig. 4C).

To search for molecular mechanisms underlying progression of IPMN, we performed exploratory GSEA against the Molecular Signatures Database (MSigDB) Hallmark gene set collection (Fig. 4D and table S8) (32). Genes were ranked by log fold change in HGD versus LGD IPMN. This resulted in 7 of 50 significantly enriched gene sets ( $P_{\text{adj}} < 0.01$ ). The gene set with the highest enrichment score represented genes regulated by nuclear factor  $\kappa$ B (NF $\kappa$ B) in response to tumor necrosis factor (TNF) signaling (TNF-NF $\kappa$ B). We interpreted this result as supportive of the known link between inflammatory signaling and progression in IPMN (33–36). An additional three of seven enriched gene sets pertained to cell proliferation. Leading edge analysis of the genes enriched in two or more gene sets yielded two unique clusters (Fig. 4E). The first cluster included *MKI67* and other genes associated with cell proliferation and division, and the second cluster involved genes associated with inflammatory signaling, hypoxia, and epithelial-to-mesenchymal transition.

### Genes that differentiate high-grade PB from INT IPMN

We next investigated neoplastic progression in PB and INT-predominant tumors. IPMN specimens were partitioned by their predominant histology subtype across the entire specimen. Accordingly, the AOIs formed four subgroups: PB-HGD ( $N = 23$ ), PB-LGD ( $N = 18$ ), INT-HGD ( $N = 24$ ), and INT-LGD ( $N = 23$ ). The PB-LGD subgroup included GF AOIs ( $N = 17$ ) and a single low-grade PB AOI. There were 50 DE genes (34 overexpressed and 16 underexpressed) specific to the PB-HGD versus PB-LGD analysis, and 6 genes (5 overexpressed and 1 underexpressed) specific to the INT-HGD versus INT-LGD analysis (Fig. 5A, fig. S6, and table S6). Carcinoembryonic antigen-related cellular adhesion molecule 6 (*CEACAM6*), which has been reported as a possible biomarker in IPMN and cholangiocarcinoma, was

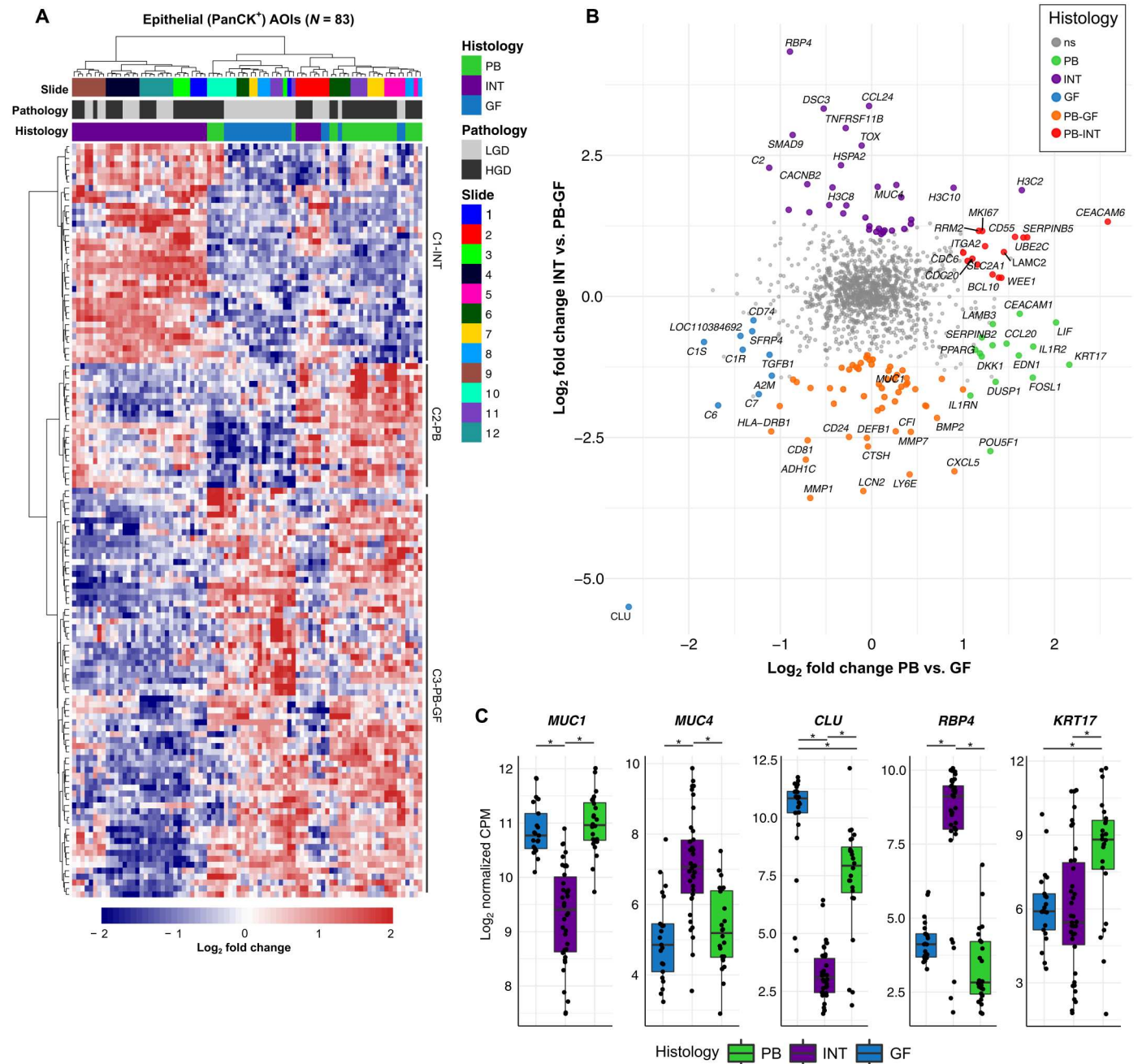
highly up-regulated in PB but not INT IPMN (Fig. 5B) (21, 37). In addition, laminin subunit beta-3 (*LAMB3*) and POU class 5 homeobox 1 (*POU5F1*) were specific to PB-HGD but not INT-HGD AOIs. Early growth response-1 (*EGR1*) was more specific to INT, although met significance thresholds in the grouped analysis of HGD versus LGD dysplasia.

### Association of IPMN subtypes with PDAC

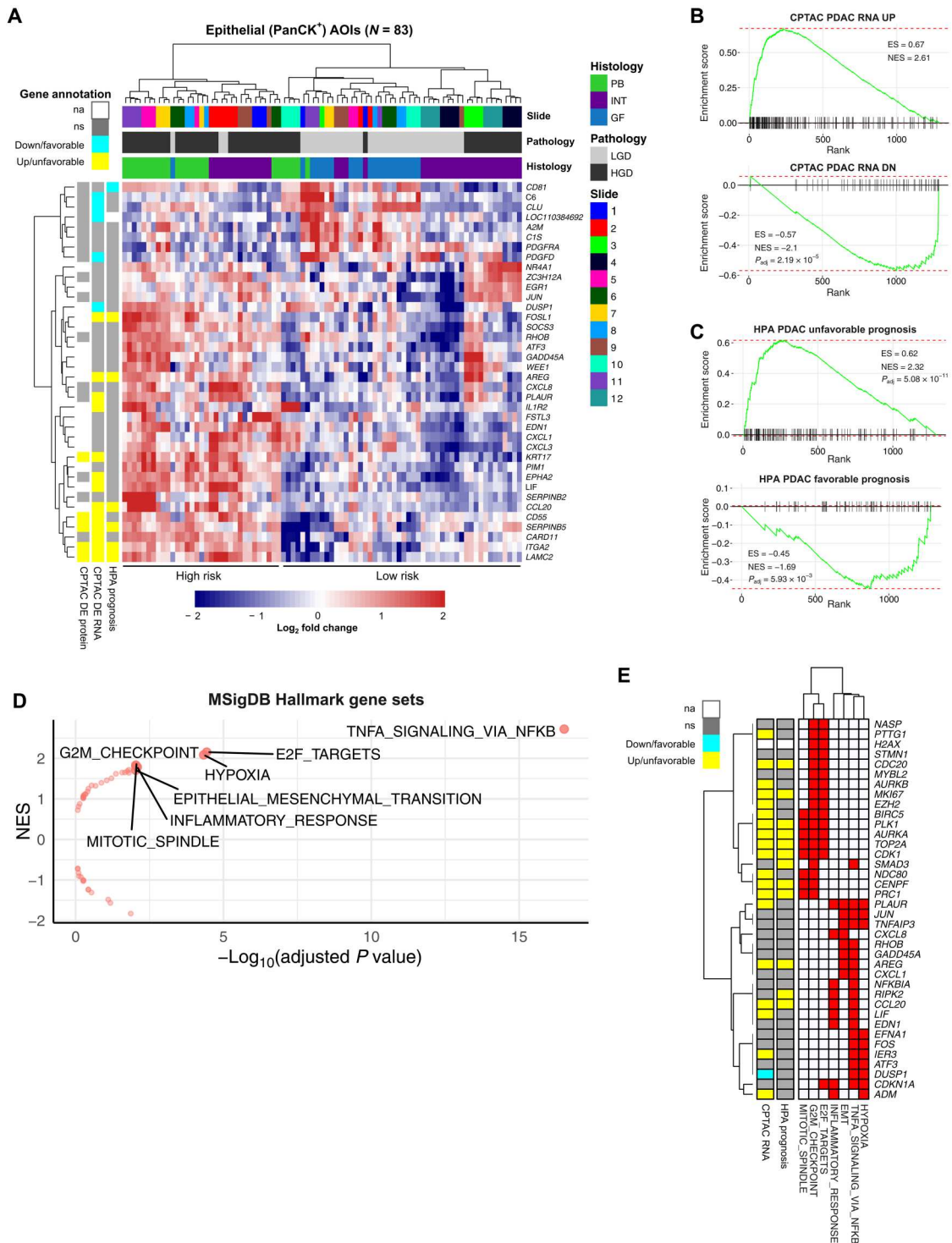
Subgroup DE analysis was performed on pairwise combinations of histopathologic groups (six comparisons), and the resulting ranked gene lists were tested against the curated PDAC gene sets (table S7). The PB-HGD versus PB-LGD analysis had the greatest absolute normalized enrichment score (NES) in 8 of 10 gene sets and outperformed analyses involving INT-HGD regions for every gene set (Fig. 5C). Of particular significance was the stark difference in NES for the HPA Prognosis gene sets. All comparisons involving PB-HGD AOIs were highly enriched, whereas none of the INT-HGD comparisons showed significant enrichment. This suggests that PB-HGD IPMN most closely resembles invasive carcinoma and may represent a direct precursor to malignancy in the dysplastic progression of IPMN.

### Up-regulation of inflammatory signaling and cell proliferation during dysplastic progression

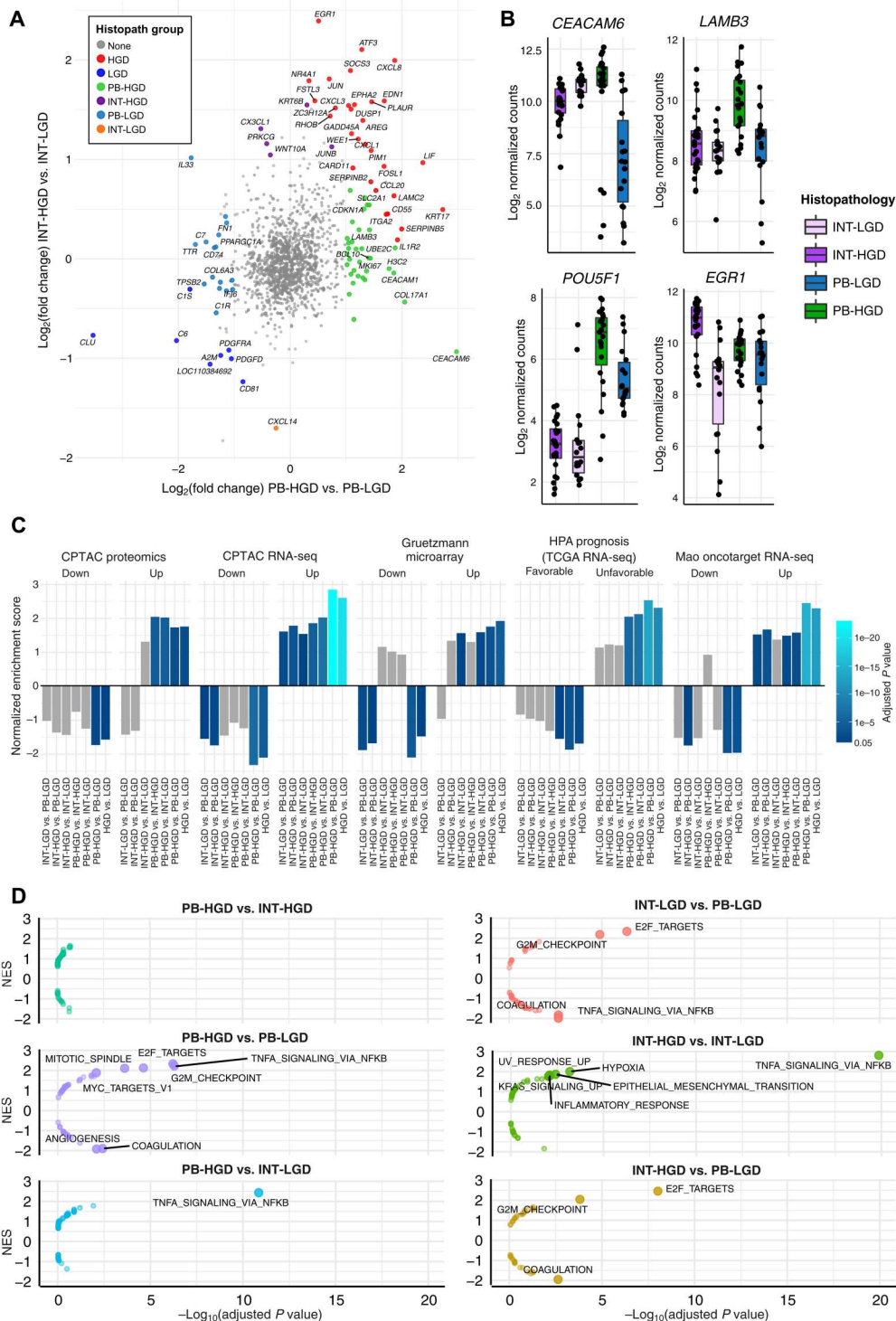
Exploratory GSEA against MSigDB Hallmark gene sets was replicated for each of the six pairs of histopathologic groups (Fig. 5D and table S8). Cell proliferation programs (S and G<sub>2</sub>-M phases) were up-regulated in INT-LGD relative to PB-LGD (GF) AOIs. This was concordant with the greater cell density observed within INT regions on microscopic examination and reflected in the initial QC analysis (figs. S1E and S8). By contrast, the TNF-NF $\kappa$ B transcriptional program was up-regulated in PB-LGD relative to INT-LGD AOIs. Both PB-HGD and INT-HGD regions overexpressed both TNF-NF $\kappa$ B and proliferation pathways relative to their LGD



**Fig. 3. Analysis of epithelial subtypes of IPMN.** (A) Heatmap plot of differentially expressed genes specific to PB, INT, and GF subtype [absolute  $\log_2$  fold change > 1, adjusted  $P$  value ( $P_{adj}$ ) < 0.05]. Columns represent individual AOs annotated by slide identifier, grade of dysplasia, and epithelial subtype. Rows represent individual genes with expression values scaled by z score. Hierarchical clustering of both columns and rows was performed. (B) Scatterplot showing  $\log_2$  fold change of PB versus GF genes (x axis) and INT versus PB-GF (non-intestinal) (y axis). Genes that are significantly differentially expressed (absolute  $\log_2$  fold change > 1,  $P_{adj}$  < 0.05) are colored by the histologic subtype that they represent. The top 10 differentially expressed genes from each analysis are labeled. The widely used marker gene *MUC1*, which was over-expressed in PB-GF relative to INT, is labeled for reference. (C) Boxplots showing the log-normalized gene expression of subtype marker genes *MUC1*, *MUC4*, *CLU*, *RBP4*, and *KRT17*. \* $P$  < 0.05. PanCK<sup>+</sup>, pan-cytokeratin positive; ns, not significant; CPM, counts per million.



**Fig. 4. Analysis by grade of dysplasia (HGD versus LGD).** (A) Heatmap plot of differentially expressed genes in AOs representing HGD versus LGD (absolute log<sub>2</sub> fold change > 1, *P*<sub>adj</sub> < 0.05). Gene expression values are scaled by z score. AOs (columns) are annotated by slide identifier, grade of dysplasia, and epithelial subtype. High-risk and low-risk AOI clusters are annotated below. Genes (rows) are further annotated by established markers PDAC, including datasets from the Human Protein Atlas (HPA Prognosis) and Clinical Proteomic Tumor Analysis Consortium (CPTAC) analysis (CPTAC DE RNA and CPTAC DE protein). (B and C) Gene Set Enrichment Analysis (GSEA) enrichment plots using external PDAC gene sets comparing AOs with HGD compared to LGD, ranked by log fold change. (D) Volcano plot of Molecular Signatures Database (MSigDB) Hallmark gene sets associated with HGD compared to LGD, with statistical significance plotted on the x axis and normalized enrichment score plotted on the y axis. Significantly enriched gene sets (*P*<sub>adj</sub> < 0.01) are shown with text labels. (E) Heatmap showing genes (rows) associated with two or more enriched Hallmark gene sets (columns). Rows are also annotated with evidence of dysregulation (CPTAC RNA) or prognostic relevance (HPA Prognosis) in PDAC.



**Fig. 5. Comparison of epithelial subtypes by grade of dysplasia.** (A) Scatterplot showing  $\log_2$  fold change of PB-HGD versus PB-LGD (x axis) and INT-HGD versus INT-LGD (y axis). Significant DE genes are colored by histopathologic subgroup. Top 10 DE genes from each analysis are labeled. (B) Boxplots showing the log-normalized gene expression of histopathologic marker genes. (C) Barplots depicting GSEA normalized enrichment score obtained from testing of histopathologic groups against several external PDAC gene sets. Bar colors depict the  $P_{adj}$  of each test. Gray bars are not statistically significant ( $P_{adj} > 0.05$ ). (D) Volcano plots of MSigDB Hallmark gene sets associated with each histopathologic subgroup (relative to its counterparts), with statistical significance plotted on the x axis and normalized enrichment score plotted on the y axis. Significant enrichment results ( $P_{adj} < 0.01$ ) are shown with text labels.

counterparts. Together, these results suggest that INT may arise as a primarily proliferative lesion before acquiring the TNF-NF $\kappa$ B program and progressing to HGD. By contrast, PB-LGD (GF) may arise in the setting of inflammatory signaling and acquire the capacity to proliferate during progression to HGD. No significant hallmark pathways distinguished PB-HGD from INT-HGD, suggesting that transcriptional programs not encompassed by MSigDB Hallmark gene sets must account for the marked differences between PB-HGD and INT-HGD.

### Unsupervised network analysis to delineate gene clusters associated with progression to carcinoma

To determine whether gene expression patterns could infer biological pathway activity and cancer risk without reliance on pathological annotation, we performed coexpression network analysis (Materials and Methods). This resulted in a network containing 300 genes connected by 791 edges (Fig. 6A and tables S9 and S10). Unsupervised clustering partitioned the network into 29 clusters of coexpressed genes. We discarded clusters with <10 genes (23 of 29) for which enrichment testing would be underpowered, leaving 6 communities for further investigation.

Hypergeometric enrichment testing was used to associate clusters with gene sets from the MSigDB Hallmark database and Gene Ontology Biological Processes (GO:BP; table S11) (38). After consolidating redundant biological processes, we selected the most statistically significant gene sets to represent each cluster (Fig. 6A). We then projected the standardized mean gene expression across PB-HGD and INT-HGD AOIs onto the network using color gradient overlays (Fig. 6B). To relate network clusters with PDAC, we performed enrichment testing against PDAC gene sets and visualized the CPTAC RNA-seq DE genes as a third network overlay (Fig. 6C).

Clusters C4 and C7 warranted consideration as contributors to malignant progression based on significant enrichment for CPTAC RNA-seq PDAC genes (C4 22 of 46 genes,  $P_{\text{adj}} = 3.7 \times 10^{-6}$ ; C7 25 of 37 genes,  $P_{\text{adj}} = 8.3 \times 10^{-12}$ ). Cluster C4 was significantly enriched for TNF-NF $\kappa$ B signaling (17 of 46 genes,  $P_{\text{adj}} = 4.4 \times 10^{-6}$ ) and contained 15 of 30 genes overexpressed in HGD versus LGD IPMN. Genes within C4 tended to be expressed at higher levels in PB relative to INT IPMN. Accordingly, C4 harbored 15 genes overexpressed in PB-HGD versus PB-LGD AOIs, compared to only 6 genes overexpressed in INT-HGD versus INT-LGD AOIs. Cluster C7 was significantly enriched for cell cycle genes (28 of 37 genes,  $P_{\text{adj}} = 2.1 \times 10^{-18}$ ). Genes within C7 tended to be expressed at higher levels in INT relative to PB IPMN and contained more genes up-regulated in INT versus GF AOIs (22 of 37 genes) than PB versus GF AOIs (6 of 37 genes). In contrast to C4, C7 contained none of the DE genes in HGD versus LGD IPMN. Together, these results implicate C4 (TNF-NF $\kappa$ B) as a prominent transcriptional program in PB IPMN and C7 (cell cycle) as a prominent program in INT IPMN.

## DISCUSSION

Management of patients with IPMN presents an opportunity to prevent pancreatic cancer; however, current management strategies are limited in their ability to provide accurate recommendations because our ability to predict timing of progression is limited. Cytologic or pathologic confirmation of disease subtype and/or grade of dysplasia is difficult without operative resection, and resection is

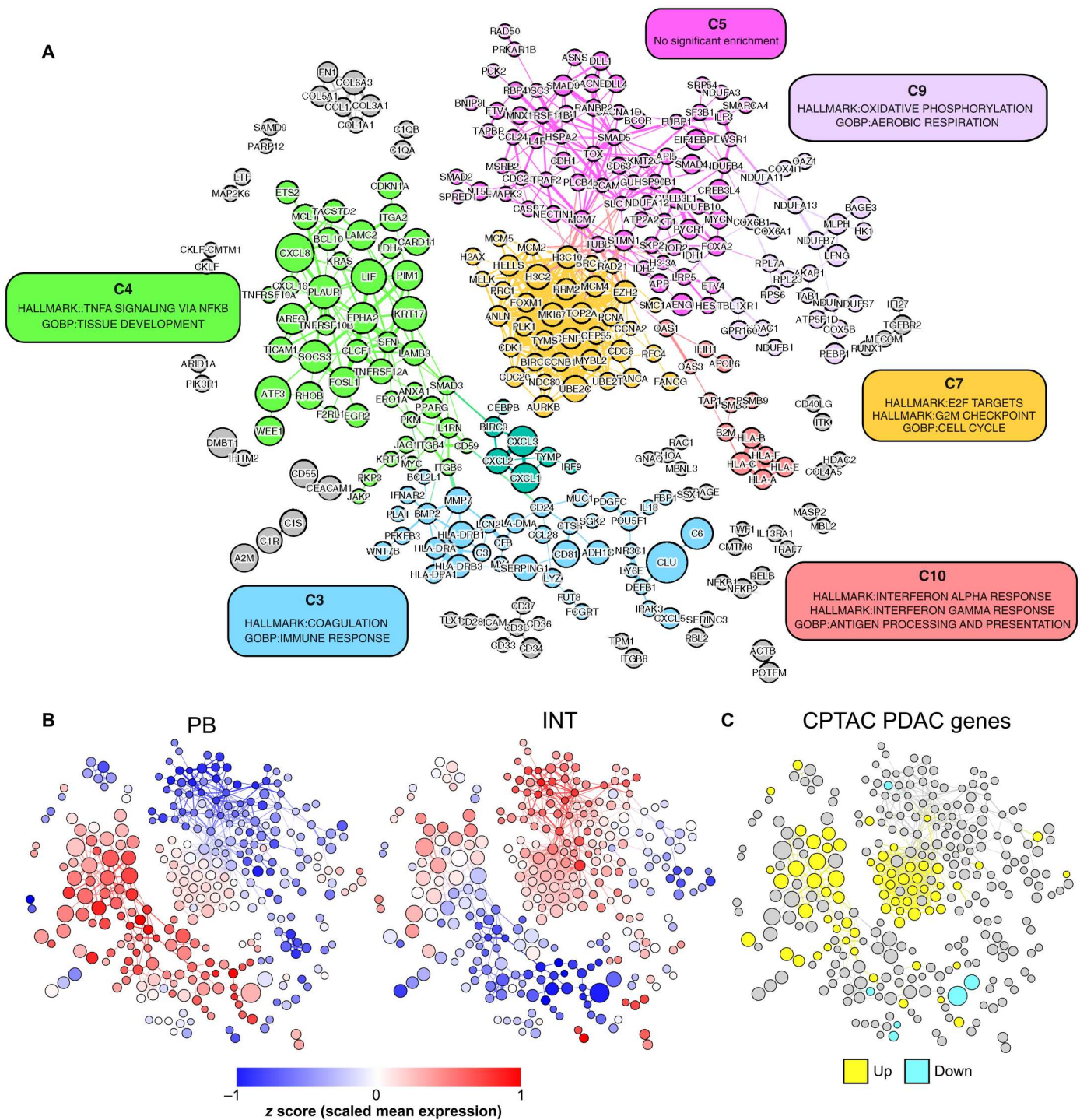
associated with substantial morbidity and even mortality. Asymptomatic patients who present without high-risk radiographic stigmata of carcinoma represent a clinical conundrum for which no accurate diagnostic modalities currently exist. For this patient population, the concept of a prognostic molecular assay holds great promise, but despite considerable investigation, no such assay has gained clinical traction.

Available retrospective evidence suggests an association between histopathology and clinical outcomes in patients with IPMN. Specifically, PB histologic subtype portends poor prognosis relative to INT, GF appears to represent an indolent entity, and the presence of HGD forecasts the development of invasive cancer (7, 8, 39–41). Numerous efforts have failed to translate these retrospective pathologic observations into predictive biomarkers for technical and disease-related reasons. The major caveat with “bulk” studies of IPMN tissue rests in the assignment of a single grade and subtype annotation to a specimen, effectively homogenizing the disease morphologies and degrees of dysplasia that occur within the affected pancreas. In an unsupervised clustering analysis of single-cell RNA-seq data from patients with IPMN, Bernard *et al.* (21) found subpopulations of cells from tissues designated LGD within clusters of HGD and carcinoma cells, corroborating histopathologic evidence that IPMN harbor a mixture of cells along a dysplastic spectrum. Ultimately, our understanding of the neoplastic progression of IPMN hinges upon the ability to characterize these tissues in a more granular fashion.

Until now, attempts to isolate neoplastic ductal epithelium required technically challenging tissue handling. Jury *et al.* (20) combined laser capture microdissection with microarray technology to study gene expression changes associated with IPMN progression. The investigators reported markers of pancreatic islet cells, including the hormones insulin, glucagon, and somatostatin, among the most significantly differentially expressed genes in their dataset, raising doubt that the microdissection procedure precisely isolated neoplastic epithelium. Sato *et al.* (19) used selective microdissection paired with microarrays to compare gene expression between normal ductal epithelia, noninvasive IPMN, and invasive IPMN. The authors reported several genes also found by our study, including serpin family B member 5 (*SERPINB5*), CD55 molecule (Cromer Blood Group) (*CD55*), and integrin subunit alpha 2 (*ITGA2*). However, the study exclusively examined regions of carcinoma in situ (HGD) and lacked classification by epithelial subtype, limiting its clinical translational potential. As an alternative to tissue isolation, IPMN cyst fluid—rich in DNA, RNA, and protein—can be obtained with minimally invasive fine needle aspiration. However, the acquired material constitutes a convolution of secreted molecules and sloughed off debris from normal and neoplastic pancreatic tissue. Therefore, biomarker identification from either cyst fluid or bulk tissue present similar challenges.

In the current study, we leveraged digital spatial RNA profiling using a targeted gene panel to characterize precise regions of IPMN histopathology across tissue slides and produce robust gene expression patterns by epithelial subtype and grade. Unsupervised dimensionality reduction analysis demonstrated distinct groups of INT and non-INT (PB-GF) AOIs. INT and GF AOIs clustered apart even when derived from the same patient and in the same tissue section. By contrast, the intimate association of PB and GF AOIs suggests that GF and PB share a common neoplastic cell lineage distinct from INT lesions. Given that GF regions are almost universally





**Fig. 6. Coexpression network analysis.** Gene coexpression network was produced by correlation analysis of all epithelial AOIs, where nodes represent individual genes and edges connect highly correlated gene pairs. Node size reflects the fold change in expression between HGD and LGD AOIs. **(A)** Unsupervised clustering of network with color-coded clusters. Clusters with >10 genes were tested for enrichment against MSigDB Hallmark gene sets, GO:BP gene sets, and PDAC gene sets. Cluster annotation based on the most significantly enriched gene sets are shown in colored boxes. **(B)** Overlay showing standardized mean gene expression (z score) across PB-HGD AOIs (left) and INT-HGD AOIs (right). **(C)** Overlay showing overexpressed (yellow) and underexpressed (cyan) genes from the CPTAC PDAC RNA-seq dataset.

deemed as low grade, we suggest that GF can essentially be considered a precursor to PB epithelium.

Now, mucin genes serve as a generally accepted differential marker of epithelial subtype. The two mucin genes (*MUC1* and *MUC4*) included in the CTA panel showed significant association with PB-GF and INT subtypes, respectively. However, despite the intended use of *MUC1* as a specific marker of PB IPMN, it did not discriminate between PB and GF AOIs in this dataset. Rather, *CLU*, *RBP4*, *KRT17*, and other candidate marker genes nominated by this analysis displayed superior potential to classify epithelial subtypes.

Comparison of gene expression by grade of dysplasia identified gene expression alterations that mirrored those reported in PDAC. Evaluation of gene expression enrichment against independent datasets also served as validation of the spatial profiling platform and our data analysis approach. Clustering the AOIs on the set of 38 genes dysregulated in HGD versus LGD partitioned the specimens into high-risk and low-risk groups. Eight of the 12 slides, including the 4 slides from patients with invasive carcinoma, harbored one or more high-risk AOIs. We envision that a gene expression classifier derived from this gene signature could serve as a risk stratification tool for patients with IPMN.

This study also corroborates evidence for individual candidate biomarkers of high-risk IPMN, including *CD55*, laminin subunit gamma 2 (*LAMC2*), amphiregulin (*AREG*), and others. As noted above, *CD55* was previously reported as a biomarker for IPMN by gene expression microarray experiments and was found to be associated with disease progression in a proteomic profiling study of IPMN cyst fluid (19, 42). An enzyme-linked immunosorbent assay of *LAMC2* in plasma of patients with PDAC augmented the accuracy of the widely used PDAC biomarker CA-19-9 (43). *LAMC2* has also been detected in pancreatic duct fluid exosomes in patients with IPMN and PDAC (44). *AREG* was found to be predictive of high-risk IPMN in a serum biomarker panel based on antibody microarray technology and has also been detected in pancreatic cyst fluid (45, 46). Validation studies using these and other candidate genes in pancreatic cyst fluid are warranted.

To explore the biological underpinnings of IPMN, we performed a combination of supervised and unsupervised analyses. The supervised analysis leveraged annotation of AOIs by a pancreatic pathologist, DE testing, and GSEA to find biological associations. The unsupervised analysis used gene expression correlation information to construct a coexpression network and a community detection algorithm to partition the network into clusters of highly correlated genes. Clusters were then assessed for biological significance through hypergeometric enrichment testing. Ultimately, the two analysis approaches led to similar conclusions. Two transcriptional programs appear to be driving neoplastic progression in IPMN: inflammatory signaling (TNF- $\text{NF}\kappa\text{B}$ ) and cell proliferation (S and G<sub>2</sub>-M phases). Activation of cell proliferation was more prominent in INT relative to PB lesions, whereas inflammatory signaling was more pronounced in PB than INT. Malignant potential, assessed by enrichment of gene alterations shared with PDAC datasets, was predominantly associated with PB epithelium. The unsupervised network analysis corroborated these findings, yielding a 46-gene cluster associated with the constellation of PB epithelium, inflammatory signaling, and PDAC genes, without a priori knowledge of pathologic annotations. Measurement of the transcriptional activity of this gene signature could transcend the histopathologic

designations that currently serve as surrogate measures of malignancy risk.

This study has several important limitations. First, the CTA probe panel used in this study measures only ~10% of human protein coding genes. A targeted panel reduces the power of GSEA and hypergeometric testing alike to detect significant biological pathway enrichment. Genes not measured by the CTA panel may outperform the marker genes nominated by this study or provide evidence of other transcriptional programs with relevance in IPMN. Incorporation of a more comprehensive panel will be important in future studies.

Second, our modest cohort size of 12 specimens may not embody the breadth of disease biology across the common epithelial subtypes (PB, GF, and INT) and omits rare entities such as oncoytic IPMN. In addition, the co-occurrence of invasive carcinoma in most of the PB cohort and none of the INT cohort confounds our comparison of the two subtypes. Our attempt to mitigate this confounder by requiring all tissue slides to be devoid of invasive carcinoma may or may not be compensatory. Certainly, the available clinical outcome data support our findings: Patients with invasive carcinoma derived from INT fare far better than patients with PB-derived carcinomas (7). Expanded profiling that includes specimens with and without invasive carcinoma from both subtypes will be needed to fully resolve this issue.

A third and related limitation of the cohort design was the requirement that every specimen in the study has regions of HGD. It is conceivable that regions of LGD from specimens lacking HGD/invasive carcinoma could be different from areas of LGD found in conjunction with HGD/invasive carcinoma. The idea that high-risk IPMN could be detected before the development of HGD would certainly alter our clinical approach to the disease. Expanded spatial profiling that includes regions of normal ductal epithelium and invasive carcinoma could address this intriguing possibility.

In summary, our findings offer several refinements to our understanding of IPMN. First, GF epithelium likely represents a precursor to PB rather than a common progenitor to either PB or INT. Second, the activation of inflammatory signaling associates with high-risk IPMN and occurs predominantly in PB lesions. This finding lends credence to ongoing clinical trials of anti-inflammatory therapies in the prevention of IPMN progression. Last, the incorporation of subtype-specific and high-risk marker genes nominated by this study may facilitate the development of an accurate risk stratification assay in IPMN.

## MATERIALS AND METHODS

### Patient recruitment

Archival biospecimens from patients who had undergone pancreatic resection for IPMN at Duke University Hospital System between 2017 and 2021 were considered for spatial RNA profiling. The Institutional Review Board approved the use of deidentified patient specimens for retrospective molecular profiling. Informed consent was not required because of the retrospective nature of the study with minimal risk and deidentification of the specimens. Clinicopathological data were collected by study coordinators and securely stored in the REDCap database. Archived FFPE specimens were procured and reviewed by a board-certified pathologist specializing in pancreatic pathology to confirm diagnosis (C.S.). Specimen

blocks were cut into 5- $\mu\text{m}$ -thick serial sections. One section was stained with hematoxylin and eosin and imaged using a Nikon TE2000-E microscope for pathology review. Sections containing regions of both LGD and HGD were selected for spatial transcriptomics and mounted on a positively charged slide for this application.

### Spatial RNA profiling

Digital spatial RNA profiling was conducted using the NanoString GeoMx Digital Spatial Profiler (DSP) (22). Our pathologist (C.S.) selected regions of interest (ROIs) annotated by histologic subtype (PB, GF, and INT) and grade of dysplasia (LGD or HGD). We selected ROIs that encompassed the spectrum of subtype-grade combinations present on each slide and included multiple biological replicates of each combination. The GeoMx DSP imposes a maximum ROI diameter of 700  $\mu\text{m}$ . Individual ROIs were drawn to maximize the number of epithelial cells contained while adhering to the size constraint. Segmental profiling of individual cell populations within each ROI was performed by staining the tissues with fluorescently conjugated antibodies: CD45 for immune cells, smooth muscle actin (SMA) for stromal fibroblasts, and anti-pan-cytokeratin (PanCK) for epithelial cells. DSP tissue slides were incubated with the fluorescently conjugated antibodies to CD45, PanCK, and SMA along with a cocktail of photocleavable oligonucleotide probes from the GeoMx CTA kit. Segmentation thus produced multiple AOIs from each ROI. Libraries were prepared according to the NanoString GeoMx Library Preparation Manual and pooled to equimolar concentration. RNA was sequenced under standard conditions on an Illumina NovaSeq 6000 to a depth of 30 read pairs/ $\mu\text{m}^2$ .

### Quality control and normalization

Ultraviolet-cleaved barcode sequencing reads were processed by the GeoMx DSP Analysis Server. Processing steps included read trimming, alignment, and deduplication. A tabulated matrix of probe counts for each AOI was exported from the DSP server and subsequently analyzed using R version 4.2.0.

QC metrics for each AOI were computed, including the receiver operating characteristic curve and the associated AUC metric of gene probes ( $N = 8584$ ) versus negative probes ( $N = 75$ ) (47). This was used to denote the snAUC. A limit of detection (LOD) for each AOI was set to the 90th percentile count of negative probes. A probe count below this LOD was considered undetectable. The following QC criteria were used to filter AOIs: (i) >100,000 aligned deduplicated total counts and (ii) snAUC > 0.65. AOIs that did not meet these QC criteria were removed.

The CTA kit features multiple probes per gene (8584 independent probes targeting 1829 unique gene identifiers). Probes with undetectable expression in >80% of AOIs were removed, and the geometric mean of representative probe counts was computed to produce a single-expression value for each gene. Raw gene counts were normalized in two steps to account for differences in background levels and sequencing output across AOIs. The geometric mean of negative probe counts, a measure of background level, was subtracted from each AOI. Background-subtracted counts were scaled by total library size and subjected to quantile normalization (48).

### Data analysis

Analysis of filtered, normalized gene expression data was performed in the R language with Bioconductor (49). Dimensionality reduction analysis was performed with principal components analysis. DE analysis was conducted using the limma package (50). Criteria for calling DE genes included absolute  $\log_2$  fold change > 1.0 and  $P_{\text{adj}} < 0.05$ . Heatmap plots were generated using the pheatmap R package with row and column clustering using the "ward.D2" method (51). GSEA was conducted using the Bioconductor package fgsea (52–54). Genes were ranked by their  $\log_2$  fold change in each DE analysis (e.g., HGD versus LGD). Gene sets from MSigDB version 7.5.1. were obtained from the msgidbr package (54, 55).

### Curation of external datasets

External RNA-seq analysis results from the National Cancer Institute's CPTAC proteogenomic characterization of pancreatic adenocarcinoma were obtained from table S3 (28). Genes were merged by official gene symbol. Genes with an absolute  $\log_2$  fold change  $\geq 1.0$  and  $P_{\text{adj}} < 0.05$  were considered differentially expressed and included in gene sets. RNA-seq analysis from Mao *et al.* (29) comparing PDAC versus matched normal pancreas was curated into gene sets ( $\log_2$  fold change  $\geq 1.0$ ,  $P_{\text{adj}} < 0.01$ ). Prognostic pathology information from the HPA resource was downloaded from the HPA website ([www.proteinatlas.org/download/pathology.tsv.zip](http://www.proteinatlas.org/download/pathology.tsv.zip)) (30). Genes with prognostic significance were merged by official gene symbol and curated into gene sets.

### Coexpression network analysis

Gene-gene coexpression analysis was performed by computing the Spearman correlation matrix of all genes, as well as a null distribution of correlation coefficients from 10,000 random permutations of the gene expression values. Multiple testing correction was applied using the qvalue package in Bioconductor (56). Coexpressed gene pairs were designated as having absolute correlation coefficient  $\geq 0.7$  and  $q < 0.01$ . Community detection was performed using Leiden clustering with a resolution parameter of 0.5 (57). Visualizations of the resulting correlation network were produced by Gephi using the ForceAtlas 2 layout algorithm (58). Edge weights were set to the correlation coefficient raised to the fourth power. Cluster enrichment analysis was performed using the "enricher" function from the clusterProfiler package (59).

### Supplementary Materials

#### This PDF file includes:

Fig S1 to S8  
Legend for data file S1

#### Other Supplementary Material for this manuscript includes the following:

Data file S1

### REFERENCES AND NOTES

1. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2022. *CA Cancer J. Clin.* **72**, 7–33 (2022).
2. R. H. Hruban, A. Maitra, S. E. Kern, M. Goggins, Precursors to pancreatic cancer. *Gastroenterol. Clin. North Am.* **36**, 831–849 (2007).
3. R. Salvia, A. Burelli, G. Perri, G. Marchegiani, State-of-the-art surgical treatment of IPMNs. *Langenbecks Arch. Surg.* **406**, 2633–2642 (2021).

4. J. Wu, Y. Wang, Z. Li, H. Miao, Accuracy of Fukuoka and American gastroenterological association guidelines for predicting advanced neoplasia in pancreatic cyst neoplasm: A meta-analysis. *Ann. Surg. Oncol.* **26**, 4522–4536 (2019).
5. M. Tanaka, C. Fernández-del Castillo, T. Kamisawa, J. Y. Jang, P. Levy, T. Ohtsuka, R. Salvia, Y. Shimizu, M. Tada, C. L. Wolfgang, Revisions of international consensus Fukuoka guidelines for the management of IPMN of the pancreas. *Pancreatology* **17**, 738–753 (2017).
6. S. S. Vege, B. Ziring, R. Jain, P. Moayyedi, M. A. Adams, S. D. Dorn, S. L. Dudley-Brown, S. L. Flamm, Z. F. Gellad, C. B. Gruss, L. R. Kosinski, J. K. Lim, Y. Romero, J. H. Rubenstein, W. E. Smalley, S. Sultan, D. S. Weinberg, Y.-X. Yang, American gastroenterological association institute guideline on the diagnosis and management of asymptomatic neoplastic pancreatic cysts. *Gastroenterology* **148**, 819–822 (2015).
7. M. Distler, S. Kersting, M. Niedergethmann, D. E. Aust, M. Franz, F. Rückert, F. Eehalt, C. Pilarsky, S. Post, H.-D. Saeger, R. Grützmann, Pathohistological subtype predicts survival in patients with intraductal papillary mucinous neoplasm (IPMN) of the pancreas. *Ann. Surg.* **258**, 324–330 (2013).
8. M. Mino-Kenudson, C. F. Castillo, Y. Baba, N. P. Valsangkar, A. S. Liss, M. Hsu, C. Correa-Gallego, T. Inggakul, R. P. Johnston, B. G. Turner, V. Androustopoulos, V. Deshpande, D. McGrath, D. V. Sahani, W. R. Brugge, S. Ogino, M. B. Pitman, A. L. Warshaw, S. P. Thayer, Prognosis of invasive intraductal papillary mucinous neoplasm depends on histological and precursor epithelial subtypes. *Gut* **60**, 1712–1720 (2011).
9. A. C. Yopp, N. Katabi, M. Janakos, D. S. Klimstra, M. I. D'Angelica, R. P. DeMatteo, Y. Fong, M. F. Brennan, W. R. Jarnagin, P. J. Allen, Invasive carcinoma arising in intraductal papillary mucinous neoplasms of the pancreas. *Ann. Surg.* **253**, 968–974 (2011).
10. M. Noë, N. Niknafs, C. G. Fischer, W. M. Hackeng, V. Beleva Guthrie, W. Hosoda, M. Debeljak, E. Papp, V. Adleff, J. R. White, C. Luchini, A. Pea, A. Scarpa, G. Butturini, G. Zamboni, P. Castelli, S.-M. Hong, S. Yachida, N. Hiraoka, A. J. Gill, J. S. Samra, G. J. A. Offerhaus, A. Hoorens, J. Verheij, C. Jansen, N. V. Adsay, W. Jiang, J. Winter, J. Albores-Saavedra, B. Terris, E. D. Thompson, N. J. Roberts, R. H. Hruban, R. Karchin, R. B. Scharpf, L. A. A. Brosens, V. E. Velculescu, L. D. Wood, Genomic characterization of malignant progression in neoplastic pancreatic cysts. *Nat. Commun.* **11**, 4085 (2020).
11. M. C. Tan, O. Basturk, A. R. Brannon, U. Bhanot, S. N. Scott, N. Bouvier, J. LaFemina, W. R. Jarnagin, M. F. Berger, D. Klimstra, P. J. Allen, GNAS and KRAS mutations define separate progression pathways in intraductal papillary mucinous neoplasm-associated carcinoma. *J. Am. Coll. Surg.* **220**, 845–854.e1 (2015).
12. E. Amato, M. D. Molin, A. Mafficini, J. Yu, G. Malleo, B. Rusev, M. Fassan, D. Antonello, Y. Sadakari, P. Castelli, G. Zamboni, A. Maitra, R. Salvia, R. H. Hruban, C. Bassi, P. Capelli, R. T. Lawlor, M. Goggins, A. Scarpa, Targeted next-generation sequencing of cancer genes dissects the molecular profiles of intraductal papillary neoplasms of the pancreas. *J. Pathol.* **233**, 217–227 (2014).
13. A. V. Maker, V. Hu, S. S. Kadkol, L. Hong, W. Brugge, J. Winter, C. J. Yeo, T. Hackert, M. Büchler, R. T. Lawlor, R. Salvia, A. Scarpa, C. Bassi, S. Green, Cyst fluid biosignature to predict intraductal papillary mucinous neoplasms of the pancreas with high malignant potential. *J. Am. Coll. Surg.* **228**, 721–729 (2019).
14. A. V. Maker, N. Katabi, M. Gonen, R. P. DeMatteo, M. I. D'Angelica, Y. Fong, W. R. Jarnagin, M. F. Brennan, P. J. Allen, Pancreatic cyst fluid and serum mucin levels predict dysplasia in intraductal papillary mucinous neoplasms of the pancreas. *Ann. Surg. Oncol.* **18**, 199–206 (2011).
15. A. V. Maker, S. Carrara, N. B. Jamieson, M. Pelaez-Luna, A. M. Lennon, M. D. Molin, A. Scarpa, L. Frulloni, W. R. Brugge, Cyst fluid biomarkers for intraductal papillary mucinous neoplasms of the pancreas: A critical review from the international expert meeting on pancreatic branch-duct-intraductal papillary mucinous neoplasms. *J. Am. Coll. Surg.* **220**, 243–253 (2015).
16. N. V. Peters, J. W. Kunstman, Clinical implications of the molecular characterization of intraductal papillary mucinous neoplasms of the pancreas. *J. Cancer Metastasis Treat.* **7**, 32 (2021).
17. M. T. Yip-Schneider, R. A. Carr, H. Wu, M. C. Schmidt, Prostaglandin E2: A pancreatic fluid biomarker of intraductal papillary mucinous neoplasm dysplasia. *J. Am. Coll. Surg.* **225**, 481–487 (2017).
18. K. K. Das, H. Xiao, X. Geng, C. Fernandez-del-Castillo, V. Morales-Oyarvide, E. Daglilar, D. G. Forcione, B. C. Bounds, W. R. Brugge, M. B. Pitman, M. Mino-Kenudson, K. M. Das, mAb Das-1 is specific for high-risk and malignant intraductal papillary mucinous neoplasm (IPMN). *Gut* **63**, 1626–1634 (2014).
19. N. Sato, N. Fukushima, A. Maitra, C. A. Iacobuzio-Donahue, N. T. van Heek, J. L. Cameron, C. J. Yeo, R. H. Hruban, M. Goggins, Gene expression profiling identifies genes associated with invasive intraductal papillary mucinous neoplasms of the pancreas. *Am. J. Pathol.* **164**, 903–914 (2004).
20. R. P. Jury, B. J. Thibodeau, L. E. Fortier, T. J. Geddes, S. Ahmed, B. L. Pruetz, M. A. Farinola, G. D. Wilson, Gene expression changes associated with the progression of intraductal papillary mucinous neoplasms. *Pancreas* **41**, 611–618 (2012).
21. V. Bernard, A. Semaan, J. Huang, F. A. San Lucas, F. C. Mulu, B. M. Stephens, P. A. Guerrero, Y. Huang, J. Zhao, N. Kamyabi, S. Sen, P. A. Scheet, C. M. Taniguchi, M. P. Kim, C.-W. Tzeng, M. H. Katz, A. D. Singhi, A. Maitra, H. A. Alvarez, Single-cell transcriptomics of pancreatic cancer precursors demonstrates epithelial and microenvironmental heterogeneity as an early event in neoplastic progression. *Clin. Cancer Res.* **25**, 2194–2205 (2019).
22. C. R. Merritt, G. T. Ong, S. E. Church, K. Barker, P. Danaher, G. Geiss, M. Hoang, J. Jung, Y. Liang, J. McKay-Fleisch, K. Nguyen, Z. Norgaard, K. Sorg, I. Sprague, C. Warren, S. Warren, P. J. Webster, Z. Zhou, D. R. Zollinger, D. L. Dunaway, G. B. Mills, J. M. Beechem, Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat. Biotechnol.* **38**, 586–599 (2020).
23. P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, J. Frisén, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
24. A. M. Eckhoff, A. A. Fletcher, K. Landa, M. Iyer, D. P. Nussbaum, C. Shi, S. K. Nair, P. J. Allen, Multidimensional immunophenotyping of intraductal papillary mucinous neoplasms reveals novel t cell and macrophage signature. *Ann. Surg. Oncol.* **29**, 7781–7788 (2022).
25. V. M. Castellano-Megías, C. I. Andrés, G. López-Alonso, F. Colina-Ruizdelgado, Pathological features and diagnosis of intraductal papillary mucinous neoplasm of the pancreas. *World J. Gastrointest. Oncol.* **6**, 311–324 (2014).
26. I. Kitazono, M. Higashi, S. Kitamoto, S. Yokoyama, M. Horinouchi, M. Osako, T. Shimizu, M. Tabata, S. K. Batra, M. Goto, S. Yonezawa, Expression of MUC4 mucin is observed mainly in the intestinal type of intraductal papillary mucinous neoplasm of the pancreas. *Pancreas* **42**, 1120–1128 (2013).
27. S. Detlefsen, M. Jakobsen, M. F. B. Nielsen, G. Klöppel, M. B. Mortensen, Expression of CD117, CK17, CK20, MUC4, villin and mismatch repair deficiency in pancreatic intraductal papillary mucinous neoplasm. *Pathol. Res. Pract.* **217**, 153312 (2021).
28. L. Cao, C. Huang, D. C. Zhou, Y. Hu, T. M. Lih, S. R. Savage, K. Krug, D. J. Clark, M. Schnaubelt, L. Chen, F. de Veiga Leprevost, R. V. Eiguez, W. Yang, J. Pan, B. Wen, Y. Dou, W. Jiang, Y. Liao, Z. Shi, N. V. Terekhanova, S. Cao, R. J.-H. Lu, Y. Li, R. Liu, H. Zhu, P. Ronning, Y. Wu, M. A. Wyczalkowski, H. Easwaran, L. Danilova, A. S. Mer, S. Yoo, J. M. Wang, W. Liu, B. Haibe-Kains, M. Thiagarajan, S. D. Jewell, G. Hostetter, C. J. Newton, Q. K. Li, M. H. Roehrl, D. Fenyö, P. Wang, A. I. Nesvizhskii, D. R. Mani, G. S. Omenn, E. S. Boja, M. Mesri, A. I. Robles, H. Rodriguez, O. F. Bathe, D. W. Chan, R. H. Hruban, L. Ding, B. Zhang, H. Zhang; Clinical Proteomic Tumor Analysis Consortium, Proteomic characterization of pancreatic ductal adenocarcinoma. *Cell* **184**, 5031–5052.e26 (2021).
29. Y. Mao, J. Shen, Y. Lu, K. Lin, H. Wang, Y. Li, P. Chang, M. G. Walker, D. Li, RNA sequencing analyses reveal novel differentially expressed genes and pathways in pancreatic cancer. *Oncotarget* **8**, 42537–42547 (2017).
30. M. Uhlen, C. Zhang, S. Lee, E. Sjöstedt, L. Fagerberg, G. Bidkhori, R. Benfeitas, M. Arif, Z. Liu, F. Edfors, K. Sanli, K. von Felilitzen, P. Oksvold, E. Lundberg, S. Hober, P. Nilsson, J. Mattsson, J. M. Schwenk, H. Brunnström, B. Glimelius, T. Sjöblom, P.-H. Edqvist, D. Djureinovic, P. Micke, C. Lindskog, A. Mardinoglu, F. Ponten, A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507 (2017).
31. R. Grützmann, H. Boriss, O. Ammerpohl, J. Lüttges, H. Kalthoff, H. K. Schackert, G. Klöppel, H. D. Saeger, C. Pilarsky, Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes. *Oncogene* **24**, 5079–5088 (2005).
32. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
33. E. Sadot, O. Basturk, D. S. Klimstra, M. Gönen, L. Anna, R. K. G. Do, M. I. D'Angelica, R. P. DeMatteo, T. P. Kingham, W. R. Jarnagin, P. J. Allen, Tumor-associated neutrophils and malignant progression in intraductal papillary mucinous neoplasms: An opportunity for identification of high-risk disease. *Ann. Surg.* **262**, 1102–1107 (2015).
34. H. Kim, W. Jung, Y. Chan Shin, I. W. Han, Y. Byun, H. W. Lee, J. S. Heo, D. W. Choi, C.-S. Lim, The diagnostic and prognostic values of inflammatory markers in intraductal papillary mucinous neoplasm. *HPB* **23**, 1623–1628 (2021).
35. M. A. Al Efishat, M. A. Attiyeh, A. A. Eaton, M. Gönen, D. Prosser, A. E. Lokshin, C. F. Castillo, K. D. Lillemoe, C. R. Ferrone, I. Pergolini, M. Mino-Kenudson, N. Rezaee, M. Dal Molin, M. J. Weiss, J. L. Cameron, R. H. Hruban, M. I. D'Angelica, T. P. Kingham, R. P. DeMatteo, W. R. Jarnagin, C. L. Wolfgang, P. J. Allen, Multi-institutional validation study of pancreatic cyst fluid protein analysis for prediction of high-risk intraductal papillary mucinous neoplasms of the pancreas. *Ann. Surg.* **268**, 340–347 (2018).
36. G. Gemenetzi, F. Bagante, J. F. Griffin, N. Rezaee, A. A. Javed, L. L. Manos, A. M. Lennon, L. D. Wood, R. H. Hruban, L. Zheng, A. Zaheer, E. K. Fishman, N. Ahuja, J. L. Cameron, M. J. Weiss, J. He, C. L. Wolfgang, Neutrophil-to-lymphocyte ratio is a predictive marker for invasive malignancy in intraductal papillary mucinous neoplasms of the pancreas. *Ann. Surg.* **266**, 339–345 (2017).
37. J. B. Rose, C. Correa-Gallego, Y. Li, J. Nelson, A. Alseidi, W. S. Helton, P. J. Allen, M. I. D'Angelica, R. P. DeMatteo, Y. Fong, T. P. Kingham, K. V. Kowdley, W. R. Jarnagin,

- F. G. Rocha, The role of biliary carcinoembryonic antigen-related cellular adhesion molecule 6 (CEACAM6) as a biomarker in cholangiocarcinoma. *PLOS ONE* **11**, e0150195 (2016).
38. The Gene Ontology Consortium, The gene ontology resource: Enriching a GOLD mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).
  39. A. C. Yopp, P. J. Allen, Prognosis of invasive intraductal papillary mucinous neoplasms of the pancreas. *World J. Gastrointest. Surg.* **2**, 359–362 (2010).
  40. Y. X. Koh, H. L. Zheng, A.-Y. Chok, C. S. Tan, W. Wyone, T. K. H. Lim, D. M. Y. Tan, B. K. P. Goh, Systematic review and meta-analysis of the spectrum and outcomes of different histologic subtypes of noninvasive and invasive intraductal papillary mucinous neoplasms. *Surgery* **157**, 496–509 (2015).
  41. X. Qi, X. Zhao, J. Su, M. Xu, W. Zhang, H. Sheng, Z. Li, J. Wang, Malignant transformation and overall survival of morphological subtypes of intraductal papillary mucinous neoplasms of the pancreas: A network meta-analysis. *Eur. J. Intern. Med.* **26**, 652–657 (2015).
  42. M. Do, H. Kim, D. Shin, J. Park, H. Kim, Y. Han, J.-Y. Jang, Y. Kim, Marker identification of the grade of dysplasia of intraductal papillary mucinous neoplasm in pancreatic cyst fluid by quantitative proteomic profiling. *Cancers* **12**, 2383 (2020).
  43. A. Chan, I. Prassas, A. Dimitromanolakis, R. E. Brand, S. Serra, E. P. Diamandis, I. M. Blasutig, Validation of biomarkers that complement CA19.9 in detecting early pancreatic cancer. *Clin. Cancer Res.* **20**, 5787–5795 (2014).
  44. J. Zheng, J. M. Hernandez, A. Doussot, L. Bojmar, C. P. Zambirinis, B. Costa-Silva, E. J. A. H. van Beek, M. T. Mark, H. Molina, G. Askan, O. Basturk, M. Gonen, T. P. Kingham, P. J. Allen, M. I. D'Angelica, R. P. DeMatteo, D. Lyden, W. R. Jarnagin, Extracellular matrix proteins and carcinoembryonic antigen-related cell adhesion molecules characterize pancreatic duct fluid exosomes in patients with pancreatic cancer. *HPB* **20**, 597–604 (2018).
  45. M. T. Tun, R. K. Pai, S. Kwok, A. Dong, A. Gupta, B. C. Visser, J. A. Norton, G. A. Poultides, S. Banerjee, J. Van Dam, A. M. Chen, S. Friedland, B. A. Scott, R. Verma, A. W. Lowe, W. G. Park, Diagnostic accuracy of cyst fluid amphiregulin in pancreatic cysts. *BMC Gastroenterol.* **12**, 15 (2012).
  46. S. Roth, P. Bose, M. S. S. Alhamdani, S. A. Mustafa, C. Tjaden, K. Zamzow, U. Hinz, C. W. Michalski, J. P. Neoptolemos, J. D. Hoheisel, M. W. Büchler, T. Hackert, Noninvasive discrimination of low and high-risk pancreatic intraductal papillary mucinous neoplasms. *Ann. Surg.* **273**, e273–e275 (2021).
  47. X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, M. Müller, pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
  48. B. M. Bolstad, R. A. Irizarry, M. Åstrand, T. P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
  49. W. Huber, V. J. Carey, R. Gentleman, S. Anders, M. Carlson, B. S. Carvalho, H. C. Bravo, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K. D. Hansen, R. A. Irizarry, M. Lawrence, M. I. Love, J. MacDonald, V. Obenchain, A. K. Oleś, H. Pagès, A. Reyes, P. Shannon, G. K. Smyth, D. Tenenbaum, L. Waldron, M. Morgan, Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
  50. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
  51. Kolde, Raivo, pheatmap: Pretty heatmaps (2019); <https://CRAN.R-project.org/package=pheatmap>.
  52. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
  53. G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, M. N. Artyomov, A. Sergushichev, Fast gene set enrichment analysis. *bioRxiv* 060012 [Preprint] 1 February 2021. <https://doi.org/10.1101/060012>.
  54. A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, J. P. Mesirov, Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
  55. I. Dolgalev, MSigDB gene sets for multiple organisms in a tidy data format (2022); <https://CRAN.R-project.org/package=msigdb>.
  56. J. Storey, A. Bass, A. Dabney, qvalue: Q-value estimation for false discovery rate control (2022); <http://github.com/jdstorey/qvalue>.
  57. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: Guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
  58. M. Bastian, S. Heymann, M. Jacomy, Gephi: An Open Source Software for Exploring and Manipulating Networks, in *Proceedings of the International AAAI Conference on Weblogs and Social Media* (2009), vol. 3, pp. 361–362.
  59. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).

**Acknowledgments:** We thank H. K. Iyer for critical review of statistical and bioinformatics methodology. We also thank the Duke University BioRepository and Precision Pathology Center (Duke BRPC; supported by P30CA014236) and the National Cancer Institute's Cooperative Human Tissue Network (CHTN; supported at Duke University by UM1CA239755) for the provision of samples. **Funding:** This work was supported by National Institutes of Health grant R01 (CA182076): Biomarker validation for intraductal papillary mucinous neoplasms of the pancreas; National Institutes of Health grant T-32 (T32-CA093245): Translational research in surgical oncology; and Cancer Center Support grant P-30 (P30-CA014236): Duke Cancer Institute. **Author contributions:** Conceptualization: P.J.A. Methodology: M.K.I., C.S., A.M.E., A.F., and P.J.A. Investigation: M.K.I., C.S., A.M.E., A.F., D.P.N., and P.J.A. Visualization: M.K.I., D.P.N., A.M.E., and P.J.A. Funding acquisition: A.M.E. and P.J.A. Project administration: A.F. and P.J.A. Supervision: P.J.A. Writing—original draft: M.K.I. Writing—review and editing: M.K.I., D.P.N., C.S., A.M.E., and P.J.A. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Code and data have been deposited at Zenodo: DOI 10.5281/zenodo.7479047. Code is also available at GitHub: [https://github.com/mkiyer/spatial\\_rna\\_ipmn](https://github.com/mkiyer/spatial_rna_ipmn). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 17 August 2022  
 Accepted 10 February 2023  
 Published 17 March 2023  
 10.1126/sciadv.ade4582