Spatial Transcriptome Profiling of Intraductal Papillary Mucinous Neoplasms Establishes a Molecular Framework for Risk Stratification

Authors: Matthew K. Iyer, MD, PhD[1], Chanjuan Shi, MD, PhD[2], Austin M. Eckhoff, MD[1], Ashley Fletcher[1], Daniel P. Nussbaum[1], Peter J. Allen, MD[1*]

Affiliations:

[1]Department of Surgery, Duke University; Durham, North Carolina, USA.

[2]Department of Pathology, Duke University; Durham, North Carolina, USA.

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

One Sentence Summary: Spatial profiling of the ductal epithelial subtypes and grades of dysplasia found within IPMN yields new markers of high-risk disease and insights into malignant progression.

Abstract: A minority of patients with intraductal papillary mucinous neoplasm (IPMN) will develop pancreatic ductal adenocarcinoma (PDAC) while the majority harbor indolent disease. Histopathologic examination of resected specimens has demonstrated multiple sub-types including pancreaticobiliary (PB), intestinal (INT), and gastric foveolar (GF) that may harbor either low-grade (LGD) or high-grade dysplasia (HGD). Studies have suggested that PB epithelium and HGD confer an elevated risk of malignancy relative to other entities, however given the significant heterogeneity of subtype and grade within an individual patient, bulk tissue analyses have not clarified the gene expression programs corresponding to high-risk and low-risk disease. Here, we establish the transcriptomic framework of high-risk IPMN by leveraging digital spatial profiling of ductal epithelial regions (n = 83 after quality control) derived from surgically resected IPMN tissues (n = 12 patients) encompassing the spectrum of histologic subtypes (24 PB, 38 INT, 21 GF) and grades of dysplasia (36 LGD, 47 HGD). Globally, PB and GF expression patterns diverged from INT, suggesting two progenitor lineages. Transcriptional dysregulation within PB lesions mirrored that of conventional PDAC, whereas GF foci appeared indolent, and INT lesions constituted a proliferative phenotype that did not resemble PDAC. Biological processes implicated in neoplastic progression included inflammatory signaling and cell proliferation, which occurred with relative prominence in PB and INT subtypes, respectively. Taken together, the genes associated with high-risk IPMN nominated by this study may facilitate the development of an accurate molecular risk stratification assay for IPMN.

Main Text:

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) remains a leading cause of cancer death, predominantly due to the lack of early detection strategies enabling treatment 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 vast majority of IPMN do not progress to malignancy, the ability to accurately differentiate those lesions at low-risk for progression (low-grade dysplasia) from those at high-risk (high-grade dysplasia, early cancer) remains elusive*(3)*. It is widely accepted that operative treatment of high-grade dysplasia (HGD) is appropriate, whereas radiographic surveillance is appropriate for those with low-grade dysplasia (LGD). Consensus guidelines designed to predict risk of high-grade dysplasia 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 epithelial subtypes that often coexist within individual specimens. Epithelial subtypes observed histologically include pancreaticobiliary (PB), intestinal (INT), gastric foveolar (GF), and an oncocytic variant. Studies comparing patient outcomes by subtype suggest that PB histology is more likely to harbor or progress to malignancy, and patients with PDAC arising from a PB IPMN have a worse prognosis than those arising from an INT subtype*(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 frequently 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 rare and the specificity 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 due to 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-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 or 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 utilized 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 FFPE tissue specimens from 12 patients who underwent pancreatectomy for IPMN between 2017 and 2021 (**Fig. 1A**). Clinicopathologic details of the cohort are listed in **Table 1**. Radiographic subtypes within the cohort included main duct (5/12), branch duct (3/12), and mixed type (4/12). Tissue slides prepared from each specimen for spatial transcriptomics contained a mix of LGD and HGD, as exemplified in **Fig. 1B**. Predominant histologic subtypes were identified by a single pancreatic pathologist (CS) and included INT (6/12) and PB (6/12). Regions of GF histology contained solely LGD and coincided on slides with HGD INT and PB counterparts. Invasive carcinoma occurred in four of six PB specimens and none of the INT specimens. All foci of carcinoma were entirely excluded from tissue slides.

Spatial transcriptome profiling

A total of 98 epithelial AOIs (50 HGD, 48 LGD) were profiled (6-10 per slide, **Table S1**). The median nuclei count per AOI was 623 (IQR 345-864). High-throughput sequencing yielded a median of 608,651 aligned deduplicated reads per AOI (IQR 261,088 - 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. S1A, B**). Of the 98 AOIs, 83 (84%) met 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 AUC (snAUC) (**Methods**, **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 were expressed above background in each individual AOI (IQR 42.7% - 55.2%, **Table S2**). Gene filtering retained 1,288 / 1,828 (70.4%) genes expressed at

detectable levels above background in at least 20% of AOIs (**Fig. S2A, 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 individuals AOIs were grossly similar (**Fig. S3A**), and there was no observable bias by either pathologic grade or epithelial histology (**Fig. S3B, C**).

Figure 1

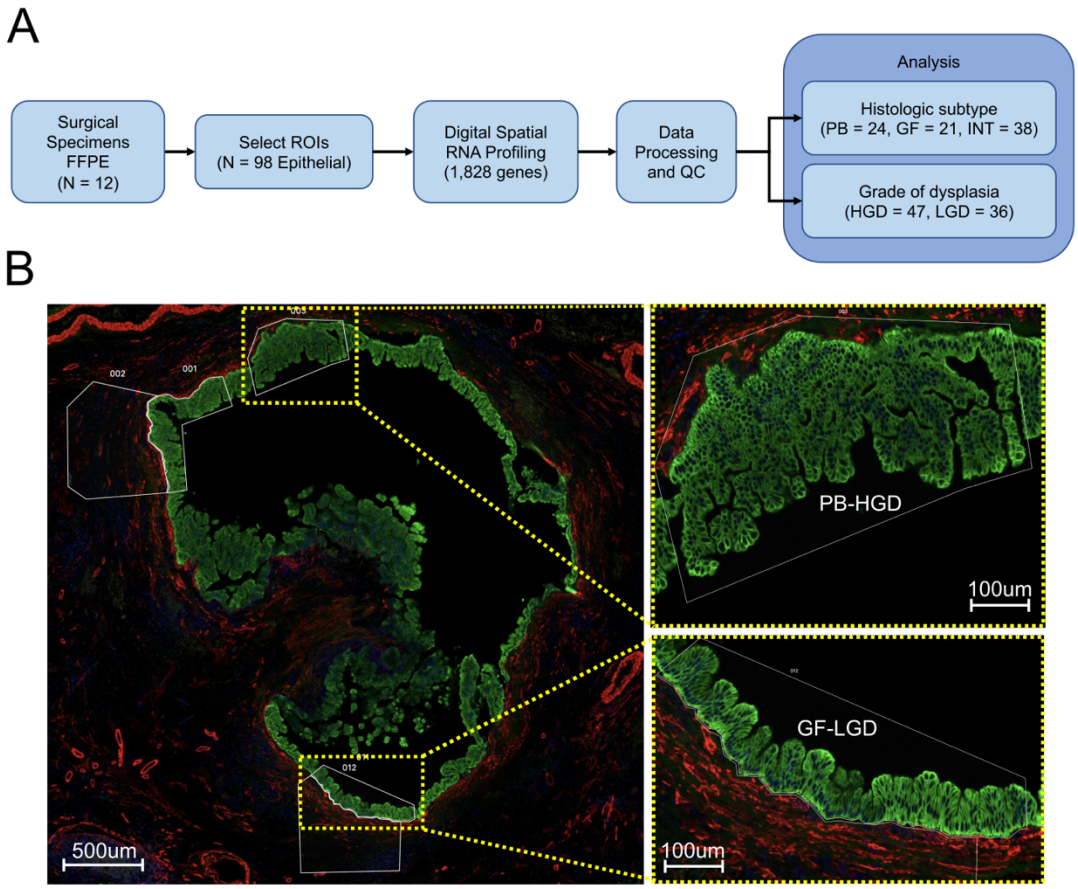

Fig. 1: Spatial RNA profiling of IPMN. (**A**) study overview flow diagram, (**B**) a representative high-resolution microscopy image from patient slide 6 depicting regions of HGD (PB) and LGD (GF) cyst epithelium chosen for profiling.

Dimensionality reduction analysis demonstrates grouping by epithelial subtype

An unbiased survey of global expression patterns was performed using principal component analysis. The AOIs formed two groups that corresponded to histologic subtype, with INT AOIs forming one group and PB-GF (non-INT type) AOIs forming the second (**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 homogeny. 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, though excluded from slides prepared for digital spatial profiling, occurred in four specimens with PB histology (slides 5, 6, 8, and 11) (**Fig. 2C**). The AOIs from these four patients, along with slide 7, grouped closely together, suggesting the existence of a shared high-risk phenotype. Interestingly, slide 10, prepared from a specimen devoid of carcinoma, strayed from the other PB and GF AOIs.

Figure 2

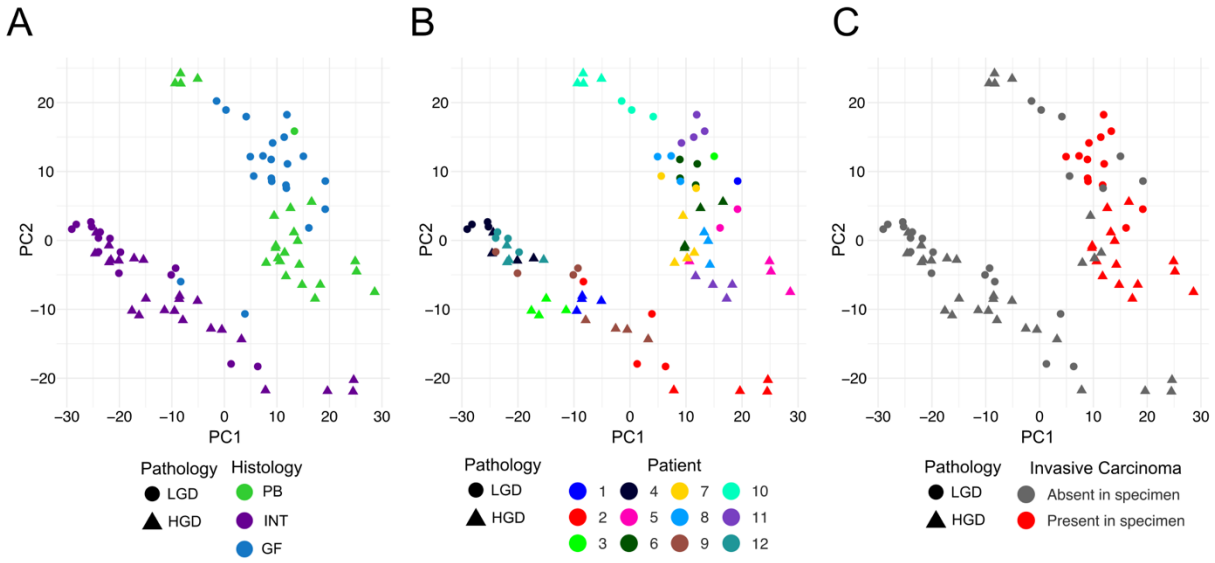

Fig. 2: Principal component analysis of normalized spatial RNA profiles. Scatter plot of the first two principal components. Shape of AOIs reflects grade of dysplasia. Panels depict AOIs colored by **(A)** Histologic subtype, **(B)** Patient / Slide ID, and **(C)** Presence of carcinoma elsewhere in the specimen.

Identification of genes associated with epithelial subtype

DE analysis of PB, INT, and GF AOIs was performed in a pairwise fashion (PB vs. GF, PB vs. INT, INT vs. GF) to discover 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 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*(14, 15, 25–27)* (**Fig. 3B and 3C**). 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**).

Figure 3

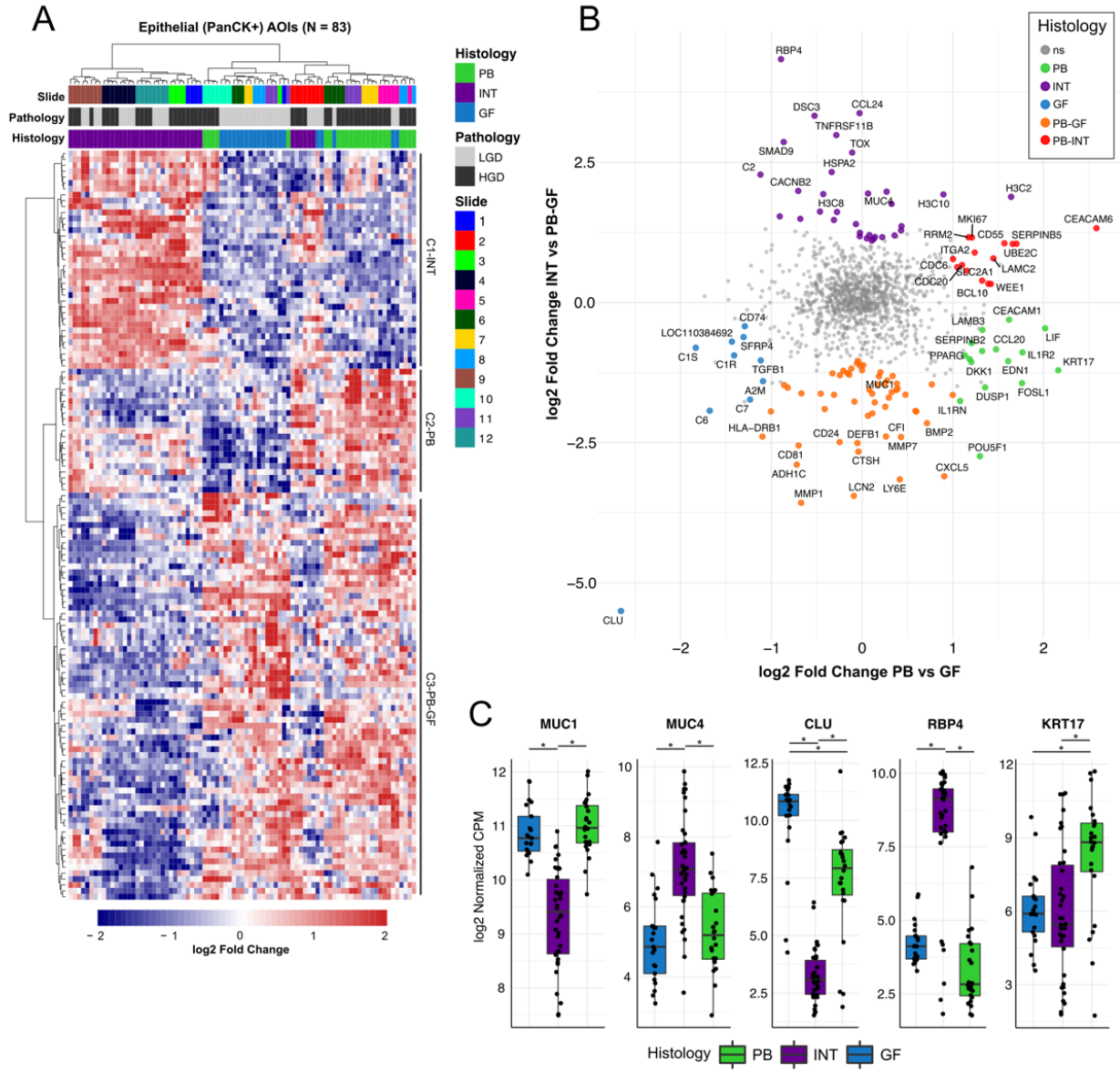

Fig. 3: Analysis of epithelial subtypes of IPMN. (**A**) Heatmap plot of differentially expressed genes specific to PB, INT, and GF subtype (absolute log2 fold change > 1, adjusted p-value < 0.05). Columns represent individual AOIs annotated by slide identifier, grade of dysplasia, and epithelial subtype. Rows represent individual genes with expression values scaled by *z*-score. Hierarchical clustering of both columns and rows was performed. (**B**) Scatter plot showing log2 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 log2 fold change > 1, adjusted p-value < 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 overexpressed 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*.

Characterization of genes associated with high-risk IPMN

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

To investigate the relevance of DE genes in the progression of IPMN to invasive carcinoma, we curated pancreatic cancer gene sets from several external sources: 1) CPTAC RNA-Seq and proteomics analysis of pancreatic cancer versus normal adjacent tissues, 2) the Human Protein Atlas (HPA) analysis of TCGA RNA-Seq data to determine prognostic genes in pancreatic cancer, 3) Mao *et al.* analysis of bulk RNA-Seq data from PDAC versus adjacent benign tissue, and 4) Grutzmann *et al*. meta-analysis of pancreatic cancer microarray experiments*(28–31)*. The analyses from CPTAC and HPA (TCGA) were considered more robust as these were generated by large consortia using standardized protocols. There were six DE genes in HGD IPMN meeting significance thresholds in both CPTAC and HPA datasets: *FOSL1*, *AREG*, *CCL20*, *SERPINB5*, *ITGA2*, *LAMC2* (**Fig. S4**). Additional genes found to be significant in the CPTAC dataset included *PLAUR*, *IL1R2*, *KRT17*, *EPHA2*, *LIF*, *CD55*, and *CARD11*.

Gene Set Enrichment Analysis (GSEA) demonstrated significant enrichment (adjusted p-value < 0.05) against all curated gene sets (**Fig. S5, Table S7**). Statistical significance largely depended on gene set size, with downregulated genes exhibiting weaker enrichment than upregulated genes.

The most significant enrichment 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 analyses against the MSigDB Hallmark gene set collection (**Fig. 4D**, **Table S8**)*(32)*. Genes were ranked by log fold change in HGD versus LGD IPMN. This resulted in 7/50 significantly enriched gene sets (padj < 0.01). The gene set with the highest enrichment score represented genes regulated by NF-κB in response to tumor necrosis factor (TNF) signaling (TNF-NF-κB). We interpreted this result as supportive of the known link between inflammatory signaling and the pathogenesis of IPMN*(33–36)*. An additional 3/7 enriched gene sets pertained to cell proliferation. Leading edge analysis of the genes enriched in two or more genes sets yielded two unique clusters (**Fig. 4E**). The first cluster included *MKI67* and other genes associated with cell proliferation and division, and the second involved genes associated with inflammatory signaling, hypoxia, and epithelial-to-mesenchymal transition (EMT).

Figure 4

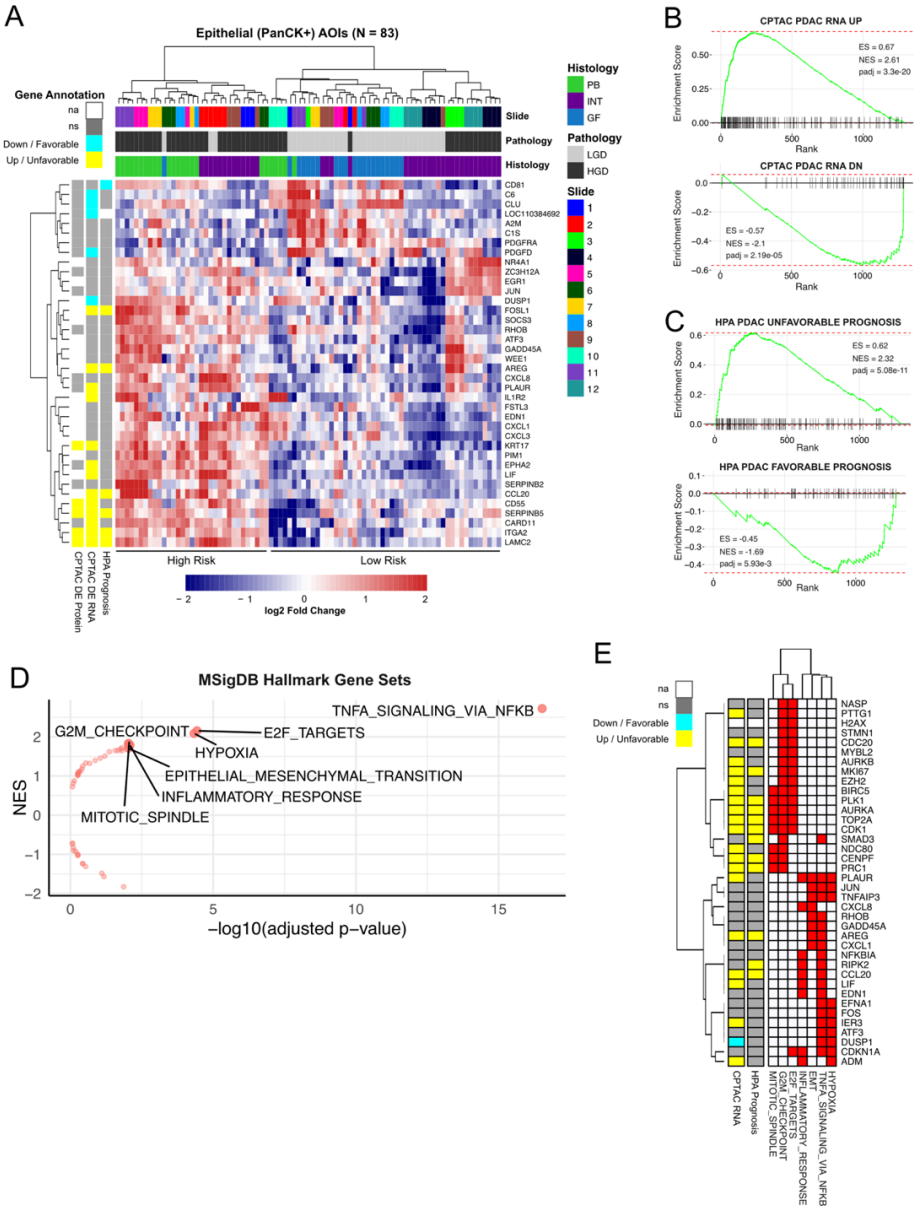

Fig. 4: Analysis by grade of dysplasia (HGD versus LGD). (**A)** Heatmap plot of differentially expressed genes in AOIs representing HGD versus LGD (absolute log2 fold change > 1, adjusted p-value < 0.05). Gene expression values are scaled by *z*-score. AOIs (columns) are annotated by slide identifier, 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 CPTAC consortium analysis (CPTAC DE RNA, CPTAC DE Protein). **(B-C)** GSEA enrichment plots using external PDAC gene sets comparing AOIs with HGD compared to LGD, ranked by log fold change. **(D)** Volcano plot of 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 (adjusted p-value < 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.

High grade PB and INT subtypes express distinct gene signatures

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 ROI. There were 50 DE genes (34 overexpressed, 16 underexpressed) specific to the PB-HGD versus PB-LGD analysis, and 6 genes (5 overexpressed, 1 underexpressed) specific to the INT-HGD versus INT-LGD analysis (**Fig. 5A**, **Fig. S6**, **Table S6**). Carcinoembryonic Antigen-Related Cellular Adhesion Molecule 6 (*CEACAM6*), which has been reported as a possible biomarker in IPMN and cholangiocarcinoma, was highly upregulated in PB but not INT IPMN*(21, 37)* (**Fig. 5B**). Additionally, 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, though met significance thresholds in the grouped analysis of HGD versus LGD dysplasia.

The transcriptomes of PB IPMNs, but not INT IPMNs, resemble PDAC

Subgroup DE analysis was performed on pairwise combinations of histopathologic groups (6 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 8/10 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.

Upregulation of inflammatory signaling and cell proliferation occurs during dysplastic progression

Exploratory GSEA against MSigDB Hallmark gene sets was replicated for each of the 6 pairs of histopathologic groups (**Fig. 5D**, **Table S8**). Cell proliferation was upregulated 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 (**Fig. S1E, Fig. S8**). By contrast, inflammatory signaling genes were upregulated in PB-LGD relative to INT-LGD AOIs. Both PB-HGD and INT-HGD regions overexpressed both inflammatory and proliferative pathways relative to their LGD counterparts. Taken together, these results suggest that INT may arise as a primarily proliferative lesion before acquiring an inflammatory phenotype 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 the presence of a final common phenotype in HGD. Yet, this does not fully explain the marked enrichment for PDAC signatures seen only in PB transcriptomes.

Figure 5

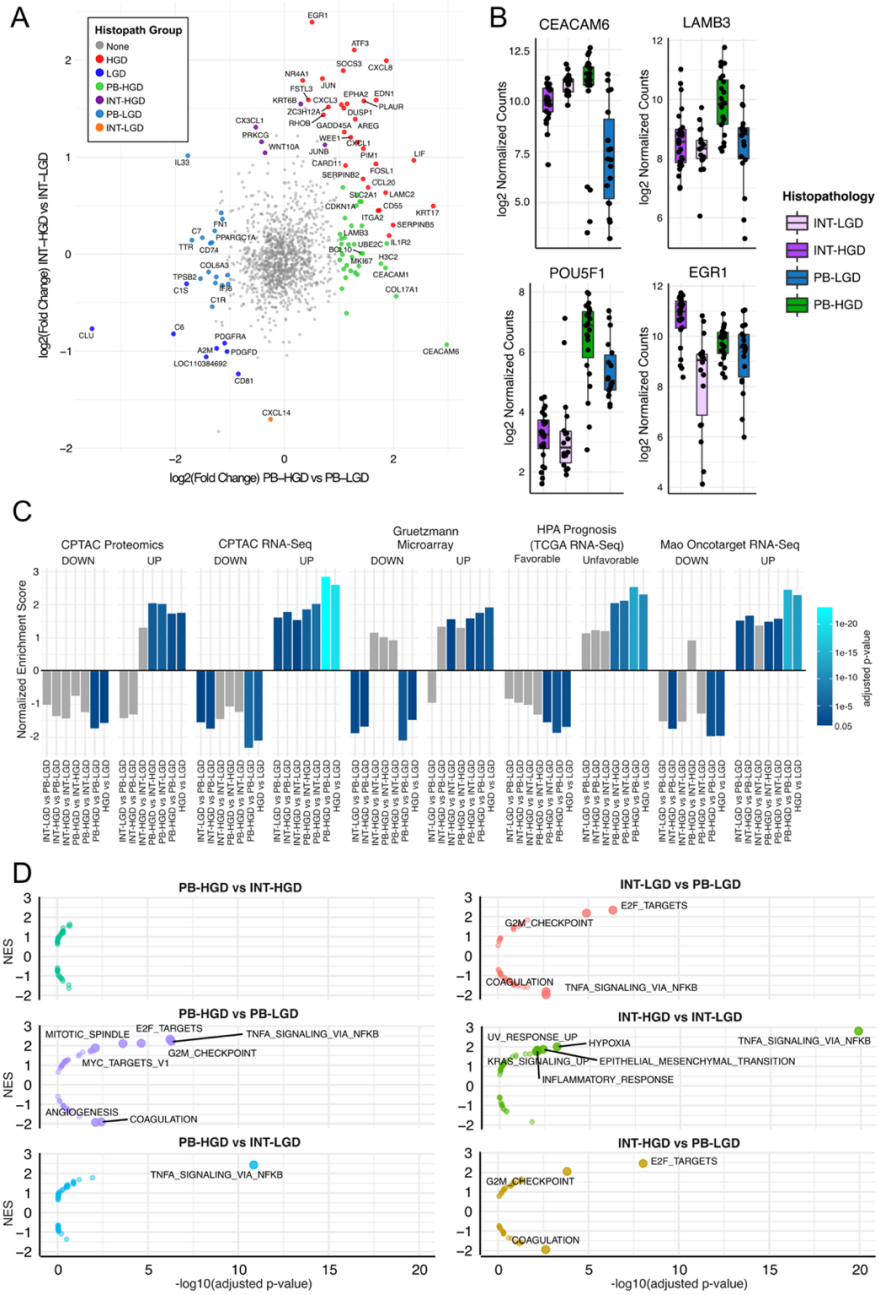

Fig. 5: Comparison of epithelial subtypes by grade of dysplasia. (**A**) Scatter plot showing log2 fold change of PB-HGD versus PB-LGD (*x* axis) and INT-HGD versus INT-LGD (*y* axis). Significant DE genes are colored by histopathologic subgroup. Top 10 DE genes from each analysis are labeled. (**B**) Boxplots showing the log-normalized gene expression of histopathologic marker genes. (**C**) Bar plots depicting GSEA normalized enrichment score obtained from testing of histopathologic groups against several external PDAC gene sets. Bar colors depict the adjusted p-value of each test. Gray bars are not statistically significant (padj > 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 (adjusted p-value < 0.01) are shown with text labels.

Coexpression network analysis identifies gene signatures associated with progression to carcinoma

To eliminate the possibility of pathologic annotation bias impacting the analysis, a completely unsupervised approach was undertaken. A coexpression network was constructed with 300 highly correlated genes connected by 791 edges (**Fig. 6A, Tables S9 and S10**). Unsupervised clustering of the network produced 29 communities, of which 6 had ≥10 nodes. Hypergeometric enrichment testing was then used to associate communities with gene sets from the MSigDB Hallmark database as well as Gene Ontology Biological Processes (GO:BP, **Table S11**)*(38)*. Redundant biological processes were manually consolidated and the most significant results were selected to represent each community. Colored overlays using standardized PB-HGD and INT-HGD gene expression were produced to visually aid in associating network communities with epithelial histopathology (**Fig. 6B**). The CPTAC RNA-Seq DE genes were included in a separate overlay (**Fig. 6C**).

Two communities, C4 and C7, stood out as clear contributors to neoplastic progression and merit careful consideration. The C4 community was significantly enriched for genes regulated by NF-κB in response to TNF (17 / 46 genes, adjusted p-value = 4.4e-6) and contained 15 of the 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 just 6 genes overexpressed in INT-HGD versus INT-LGD AOIs. The C7 community was significantly enriched for cell cycle genes (28 / 37 genes, adjusted p-value = 2.1e-18). Genes within C7 tended to be expressed at higher levels in INT relative to PB IPMN, and contained more genes upregulated in INT versus GF AOIs (22 / 37 genes) than PB

versus GF AOIs (6 / 37 genes). In contrast to C4, C7 contained none of the DE genes in HGD versus LGD IPMN. Both communities were highly enriched for the CPTAC RNA-Seq PDAC genes (C4 22 / 46 genes, adjusted p-value 4.97e-4, C7 25 / 37 genes, adjusted p-value 5.39e-6). Taken together, these associations point to C4 as an inflammatory transcriptional program associated with high-risk IPMN, and C7 as a proliferative transcriptional program essential for neoplasia but not necessarily implicated in malignant transformation.

Figure 6

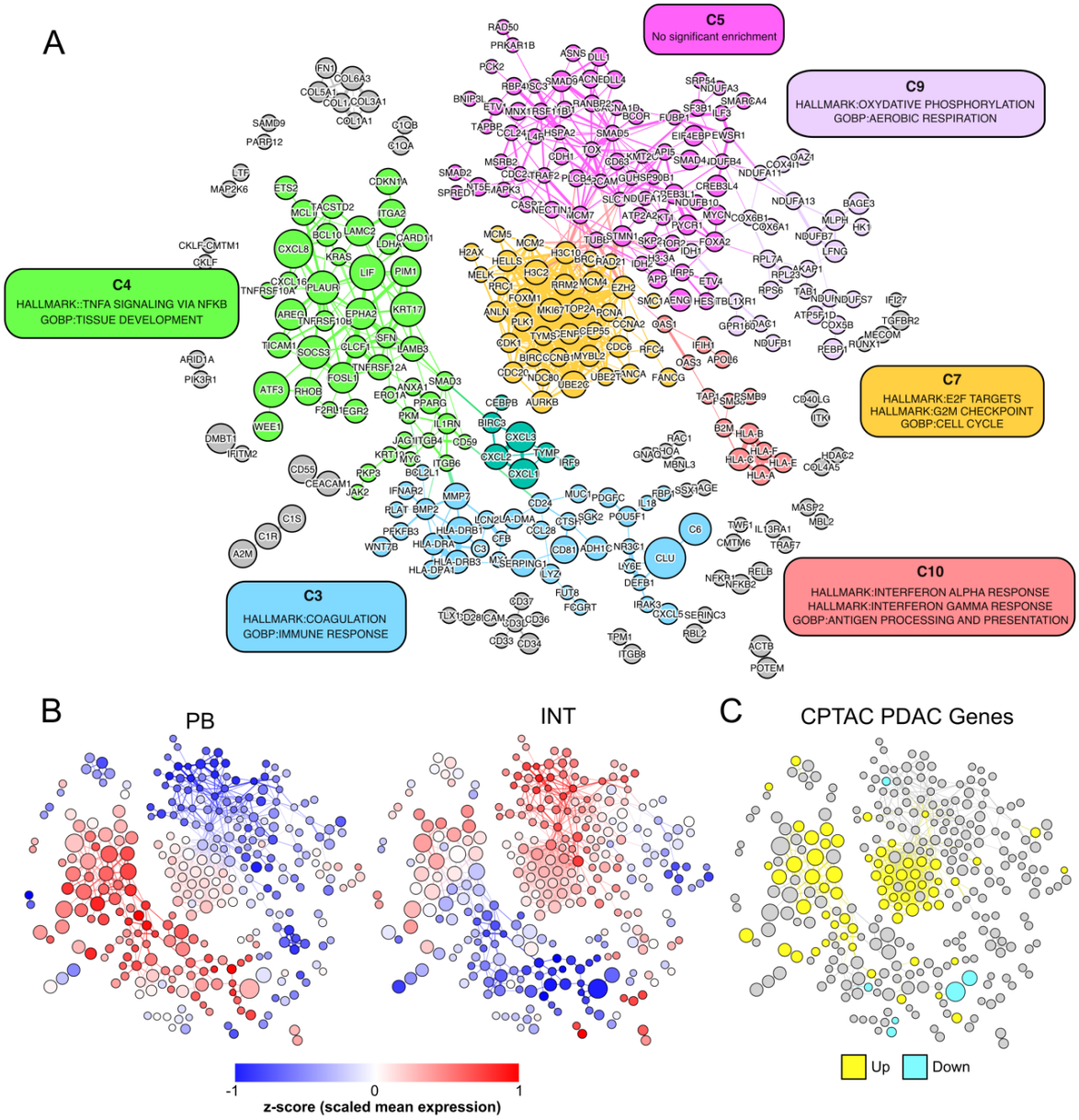

Fig. 6: Coexpression network analysis. Gene coexpression network was produced by correlation analysis of all epithelial AOIs, where nodes represent individual genes and edges connect highly correlated gene pairs. Node size reflects the fold change in expression between HGD and LGD AOIs. (**A**) Unsupervised clustering of network with color-coded communities. Communities with >10 genes were tested for enrichment against MSigDB Hallmark Gene Sets and GO:BP Gene Sets. Significantly enriched gene sets within each community 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.

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. Cytologic or pathologic confirmation of disease subtype and/or grade of dysplasia is difficult without operative resection, and resection is associated with significant 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 significant 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 epithelial architecture 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, 39–42)*. 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.* found subpopulations of cells from tissues designated LGD within clusters of HGD and carcinoma cells, corroborating histopathologic evidence that IPMNs harbor a mixture of cells along a dysplastic spectrum*(21)*. 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*. combined laser capture microdissection (LCM) with microarray technology to study gene expression changes associated with IPMN progression*(20)*. 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 significant doubt that the microdissection procedure precisely isolated neoplastic epithelium. Sato *et al.* employed selective microdissection paired with microarrays to compare gene expression between normal ductal epithelia, non-invasive IPMN, and invasive IPMN*(19)*. 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 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. Importantly, 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 deemed low-grade, we suggest that GF can essentially be considered a precursor to PB epithelium.

Currently, mucin genes serve as a generally accepted differential marker of epithelial subtype. The two mucin genes (*MUC1* and *MUC4*) included in the CTA probe set 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, clusterin (*CLU*), retinol binding protein 4 (*RBP4*), keratin 17 (*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 technology 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 twelve slides, including the four 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, 43)*. An enzyme-linked immunosorbent assay (ELISA) of *LAMC2* in plasma of PDAC patients augmented the accuracy of the widely used PDAC biomarker CA-19-9*(44)*. *LAMC2* has also been detected in pancreatic duct fluid exosomes in patients with IPMN and PDAC*(45)*. *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*(46, 47)*. Validation studies 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, differential expression testing, and GSEA to discover biological associations. The unsupervised analysis utilized gene expression correlation information to construct a coexpression network, and a community detection algorithm to partition the network into groups of highly correlated genes. Communities 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 and cell proliferation. 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 signature associated with the constellation of PB epithelium, pro-inflammatory signaling, and PDAC genes, without *a priori* knowledge of

biological conditions. 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 measured approximately 10% of human protein coding genes. Several of the mucin marker genes, such as *MUC2*, *MUC5AC*, and others, were not measured by this probe panel. Furthermore, a limited probe panel reduces the statistical power of GSEA and hypergeometric tests of biological pathway enrichment. We anticipate that incorporation of the recently released GeoMx Human Whole Transcriptome Atlas (WTA) panel, which measures ~18,000 genes, will provide an enhanced characterization of IPMN progression. Second, our cohort of twelve specimens lacks the breadth to capture rare histopathologic entities such as oncocytic IPMN. Third, every specimen in this study contained areas of LGD and HGD. It is conceivable that LGD from specimens completely devoid of HGD could differ from the areas of LGD adjacent to HGD or invasive carcinoma. Finally, the addition of regions of normal ductal epithelium and invasive carcinoma could lead to a more comprehensive characterization of IPMN.

In summary, our findings offer several refinements to our understanding of IPMN. First, GF epithelium may represent a low-grade precursor to PB rather than a distinct epithelial subtype. 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. Finally, 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

Eligible patients per Institutional Review Board (IRB) approved protocols who had undergone pancreatic resection for IPMN at Duke University Hospital System (DUHS) between 2017 and 2021 were initially considered for spatial transcriptomic analysis. Clinicopathological data was collected by study coordinators and securely stored in a RedCap database. Archived FFPE specimens were procured and reviewed by a board-certified pathologist specializing in pancreatic pathology to confirm diagnosis (CS). Specimen blocks were cut into 5 µm thick serial sections. One section was stained with hematoxylin and eosin (H&E) 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)*. Regions of interest (ROIs) were selected by CS and annotated by histologic subtype (PB, GF, INT) and grade of dysplasia (LGD or HGD). 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 photo-

cleavable-oligonucleotide probes from the GeoMX Cancer Transcriptome Atlas (CTA) kit. Segmentation thus produced multiple Areas of Interest (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/μm^2.

Data analysis

UV-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.

Quality control (QC) metrics for each AOI were computed, including the receiver operating characteristic (ROC) curve and the associated area under the curve (AUC) metric of gene probes (N = 8,584) versus negative probes (N = 75)*(48)*. This was used to denote the signal-to-noise AUC (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 quality control criteria were used to filter AOIs: 1) >100,000 aligned deduplicated total counts and 2) snAUC >0.65. AOIs that did not meet these QC criteria were removed.

The CTA kit features multiple probes per gene (8,584 independent probes targeting 1,829 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*(49)*.

Analysis of filtered, normalized gene expression data was performed in the R language with Bioconductor*(50)*. Dimensionality reduction analysis was performed with principal components analysis (PCA). Differential expression (DE) analysis was conducted using the *limma* package*(51)*. Criteria for calling DE genes included absolute log2 fold change > 1.0 and adjusted p-value < 0.05. Heatmap plots were generated using the *pheatmap* R package with row and column clustering using the "ward.D2" method*(52)*.

Gene Set Enrichment Analysis (GSEA) was conducted using *fgsea(53–55)*. Publicly available gene sets were downloaded from MSigDB version 7.5.1*(55)*. External RNA-Seq analysis results from the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) proteogenomic characterization of pancreatic adenocarcinoma were obtained from Supplementary Table S3*(28)*. Genes were merged by official gene symbol. Genes with an absolute log2 fold change ≥ 1.0 and adjusted p-value < 0.05 were considered differentially expressed and included in gene sets. RNA-Seq analysis from Mao *et al*. comparing PDAC versus matched normal pancreas was curated into gene sets (log2 fold change ≥ 1.0, adjusted p-value < 0.01)*(29)*. Prognostic pathology information from the Human Protein Atlas (HPA) resource was downloaded from the HPA website (https://www.proteinatlas.org/download/pathology.tsv.zip)*(30)*. Genes with prognostic significance were merged by official gene symbol and included in gene sets.

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 ≥ 0.7 and q-value < 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 using Gephi using the ForceAtlas 2 layout algorithm*(58)*. Edge weights was set to the correlation coefficient raised to the fourth power. Community enrichment analysis was performed using the "enricher" function from the clusterProfiler package*(59)*.

List of Supplementary Materials

Fig S1 to S8

Tables S1 to S11 (including in Data File S1)

Data File S1 (Contains Tables S1 to S11)

Supplemental References *(60)*

References

1. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians* **72**, 7–33 (2022).

2. R. H. Hruban, A. Maitra, S. E. Kern, M. Goggins, Precursors to Pancreatic Cancer. *Gastroenterology Clinics of North America* **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. Ehehalt, C. Pilarsky, S. Post, H.-D. Saeger, R. Grützmann, Pathohistological Subtype Predicts Survival in Patients With Intraductal Papillary Mucinous Neoplasm (IPMN) of the Pancreas. *Annals of Surgery* **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. Ingkakul, R. P. Johnston, B. G. Turner, V. Androutsopoulos, 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: A Matched Control Study with Conventional Pancreatic Ductal Adenocarcinoma. *Annals of Surgery* **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. *Journal of the American College of Surgeons* **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. *Journal of the American College of Surgeons* **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. *Journal of Cancer Metastasis and Treatment* **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. *Journal of the American College of Surgeons* **225**, 481–487 (2017).

18. mAb Das-1 is specific for high-risk and malignant intraductal papillary mucinous neoplasm (IPMN) | Gut (available at https://gut.bmj.com/content/63/10/1626.1).

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. *The American Journal of Pathology* **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. *Clinical Cancer Research* **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* (2022), doi:10.1245/s10434-022-12157-0.

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. *Pathology - Research and Practice* **217**, 153312 (2021).

28. L. Cao, C. Huang, D. Cui Zhou, Y. Hu, T. M. Lih, S. R. Savage, K. Krug, D. J. Clark, M. Schnaubelt, L. Chen, F. da Veiga Leprevost, R. V. Eguez, 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, M. Amin, E. An, C. Ayad, T. Bauer, C. Birger, M. J. Birrer, S. M. Boca, W. Bocik, M. Borucki, S. Cai, S. A. Carr, S. Cerda, H. Chen, S. Chen, D. Chesla, A. M. Chinnaiyan, A. Colaprico, S. Cottingham, M. Derejska, S. M. Dhanasekaran, M. J. Domagalski, B. J. Druker, E. Duffy, M. A. Dyer, N. J. Edwards, M. J. Ellis, J. Eschbacher, A. Francis, J. Francis, S. Gabriel, N. Gabrovski, J. Gardner, G. Getz, M. A. Gillette, C. A. Goldthwaite, P. Grady, S. Guo, P. Hariharan, T. Hiltke, B. Hindenach, K. A. Hoadley, J. Huang, C. D. Jones, K. A. Ketchum, C. R. Kinsinger, J. M. Koziak, K. Kusnierz, T. Liu, J. Long, D. Mallery, S. Mareedu, R. Matteotti, N. Maunganidze, P. B. McGarvey, P. Minoo, O. V. Paklina, A. G. Paulovich, S. H. Payne, O. Potapova, B. Pruetz, L. Qi, N. Roche, K. D. Rodland, D. C. Rohrer, E. E. Schadt, A. V. Shabunin, T. Shelton, Y. Shutack, S. Singh, M. Smith, R. D. Smith, L. J. Sokoll, J. Suh, R. R. Thangudu, S. X. Tsang, K. S. Um, D. R. Valley, N. Vatanian, W. Wang, G. D. Wilson, M. Wiznerowicz, Z. Zhang, G. Zhao, Proteogenomic 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 Feilitzen, 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. *Annals of Surgery* **268**, 340–347 (2018).

36. G. Gemenetzis, 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. *Annals of Surgery* **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 Research* **49**, D325–D334 (2021).

39. A. C. Yopp, P. J. Allen, Prognosis of invasive intraductal papillary mucinous neoplasms of the pancreas. *World Journal of Gastrointestinal Surgery* **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. M. Mino-Kenudson, C. F. Castillo, Y. Baba, N. P. Valsangkar, A. S. Liss, M. Hsu, C. Correa-Gallego, T. Ingkakul, R. P. Johnston, B. G. Turner, V. Androutsopoulos, 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).

42. 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. *European Journal of Internal Medicine* **26**, 652–657 (2015).

43. 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 (Basel)* **12**, 2383 (2020).

44. 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. *Clinical Cancer Research* **20**, 5787–5795 (2014).

45. 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 (Oxford)* **20**, 597–604 (2018).

46. M. T. Tun, R. K. Pai, S. Kwok, A. Dong, A. Gupta, B. C. Visser, J. A. Norton, G. A. Poultsides, 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).

47. 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. *Annals of Surgery* **273**, e273 (2021).

48. 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).

49. 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).

50. 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).

51. 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 Research* **43**, e47 (2015).

52. Kolde, Raivo, pheatmap: Pretty Heatmaps. (2019) (available at https://CRAN.R-project.org/package=pheatmap).

53. 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. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005).

54. G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, M. N. Artyomov, A. Sergushichev, Fast gene set enrichment analysis, 060012 (2021).

55. A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, J. P. Mesirov, Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

56. J. Storey, A. Bass, A. Dabney, qvalue: Q-value estimation for false discovery rate control. (2022) (available at 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, in *International AAAI Conference on Weblogs and Social Media.*, (2009).

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. *The Innovation* **2**, 100141 (2021).

60. J. Larsson, eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses Version 6.1.1 (2021) (available at https://CRAN.R-project.org/package=eulerr).

Acknowledgements: The authors thank Hariharan K. Iyer for critical review of statistical and bioinformatics methodology. The authors also thank the Duke University BioRepository & 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:
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
Cancer Center Support grant P-30 (P30-CA014236): Duke Cancer Institute

Author contributions:

Conceptualization: PJA

Methodology: MKI, CS, AME, AF, PJA
Investigation: MKI, CS, AME, AF, DPN, PJA
Visualization: MKI, DPN, AME, PJA
Funding acquisition: AME, PJA
Project administration: AF, PJA
Supervision: PJA
Writing – original draft: MKI
Writing – review & editing: MKI, DPN, CS, AME, PJA

Competing interests: Authors declare that they have no competing interests.

Data and materials availability:

All data are available in the main text or the supplementary materials.

Code to reproduce the analysis and figures is provided at:

https://github.com/mkiyer/spatial_rna_ipmn

Figures: Figures have been embedded within the results.

Tables

Table 1: Patient characteristics

Patients	12
Median Age: years (IQR)	70 (63-77)
Female Sex	5 (42%)
Race	
Caucasian	8 (67%)
Asian	3 (25%)
Black	0 (0%)
Unidentified	1 (8%)
Operation	
Pancreaticoduodenectomy	9 (75%)
Distal Pancreatectomy	2 (17%)
Other	1 (8%)
Median Lesion Size	5.2cm
Invasive Carcinoma Within Lesion	4 (33%)
Cancer Stage	IA (2), IIA (2)
Median Follow up (months)	27
Cancer Recurrence	2 (50%)
Death	3 (25%)
Anatomic Subtype	
Main Duct	5 (42%)
Branch Duct	3 (25%)
Mixed	4 (33%)
Predominant Histologic Subtype	
Intestinal	6 (50%)
Pancreatobiliary	6 (50%)
Gastric Foveolar	0 (0%)

SUPPLEMENTARY MATERIALS

Spatial Transcriptome Profiling of Intraductal Papillary Mucinous Neoplasms Establishes a Molecular Framework for Risk Stratification

Authors: Matthew K. Iyer, MD, PhD[1], Chanjuan Shi, MD, PhD[2], Austin M. Eckhoff, MD[1], Ashley Fletcher[1], Daniel P. Nussbaum[1], Peter J. Allen, MD[1*]

Affiliations:

[1]Department of Surgery, Duke University; Durham, North Carolina, USA.

[2]Department of Pathology, Duke University; Durham, North Carolina, USA.

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

Contents:

Fig. S1 to S8

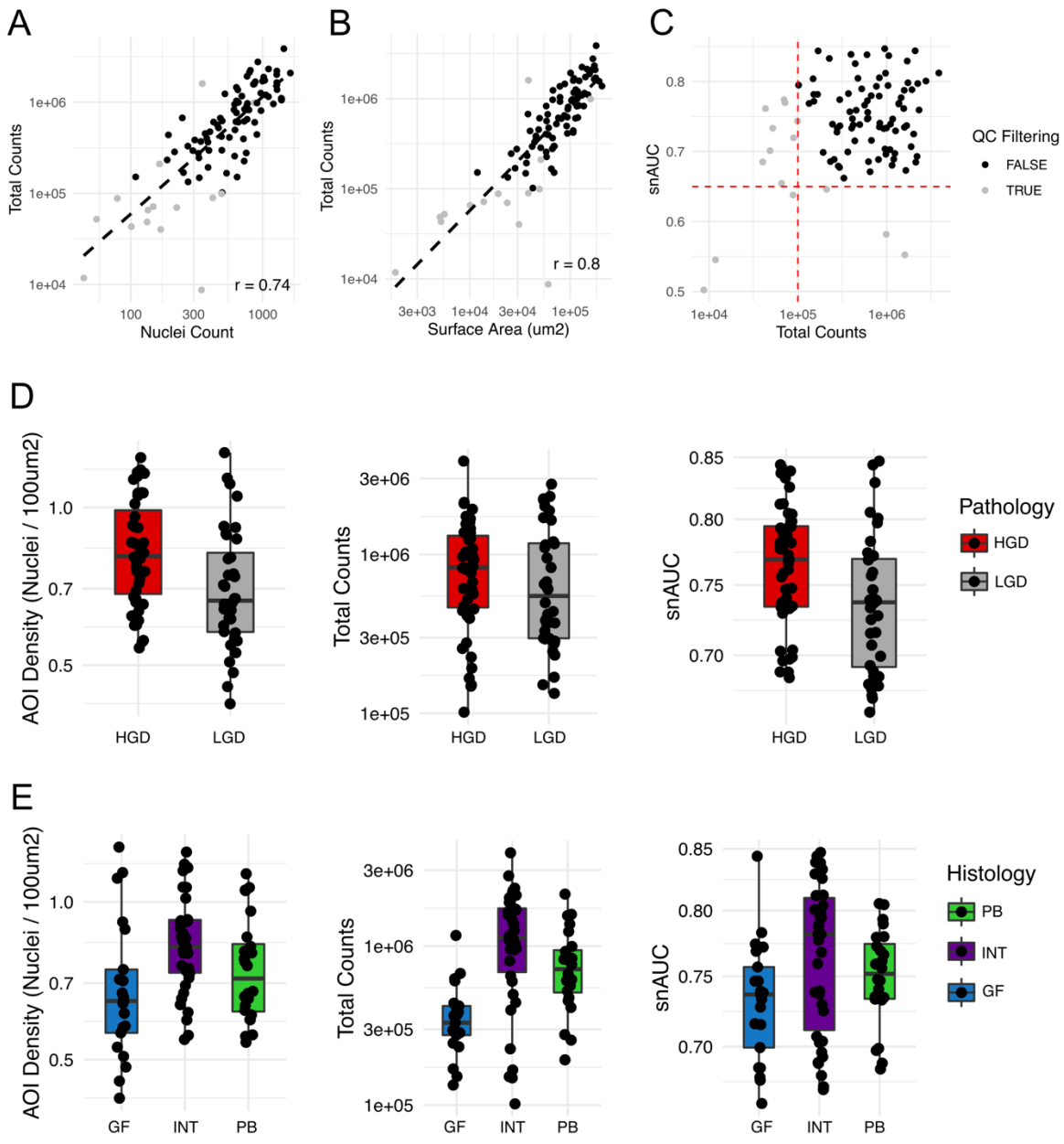

Fig. S1: QC analysis of epithelial AOIs. (**A-B**) Scatter plots of AOI QC parameters showing filtered AOIs (grey) and retained AOIs (black) based on nuclei count (**A**) and surface area (**B**) relative to total deduplicated sequencing counts. (**C**) Scatter plot depicting total deduplicated sequencing counts versus snAUC with dotted red lines indicating filtering cutoffs, and filtered and retailed AOIs indicated by grey and black dots, respectively. (**D**) Box plots by grade of dysplasia (red = HGD, grey = LGD) comparing the distribution of (*left*) AOI density (Nuclei / Surface Area), (*middle*) total deduplicated sequencing counts, and (*right*) snAUC. (**E**) Box plots by epithelial subtype (blue = GF, purple = INT, green = PB) comparing the distribution of (*left*) AOI density (Nuclei / Surface Area), (*middle*) total deduplicated sequencing counts, and (*right*) snAUC.

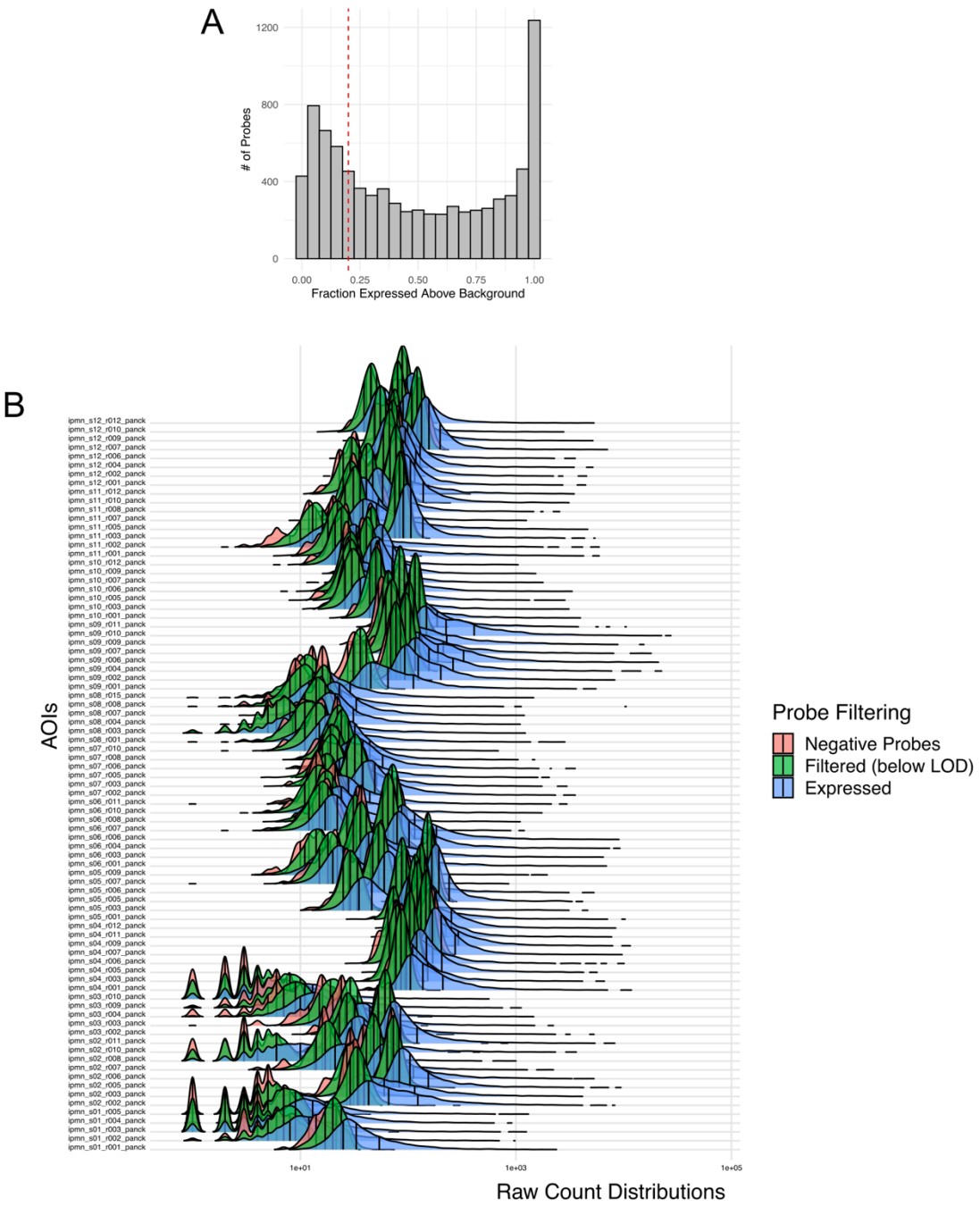

Fig. S2: Probe QC analysis. (**A**) Histogram depicting the fraction of probes with counts above the limit of detection across all AOIs. Dotted red line indicates filtering cutoff. (**B**) Ridge plot (layered density plot) showing raw (unnormalized) count distributions for all AOIs. Probes groups include background (red), filtered (green), and expressed (blue).

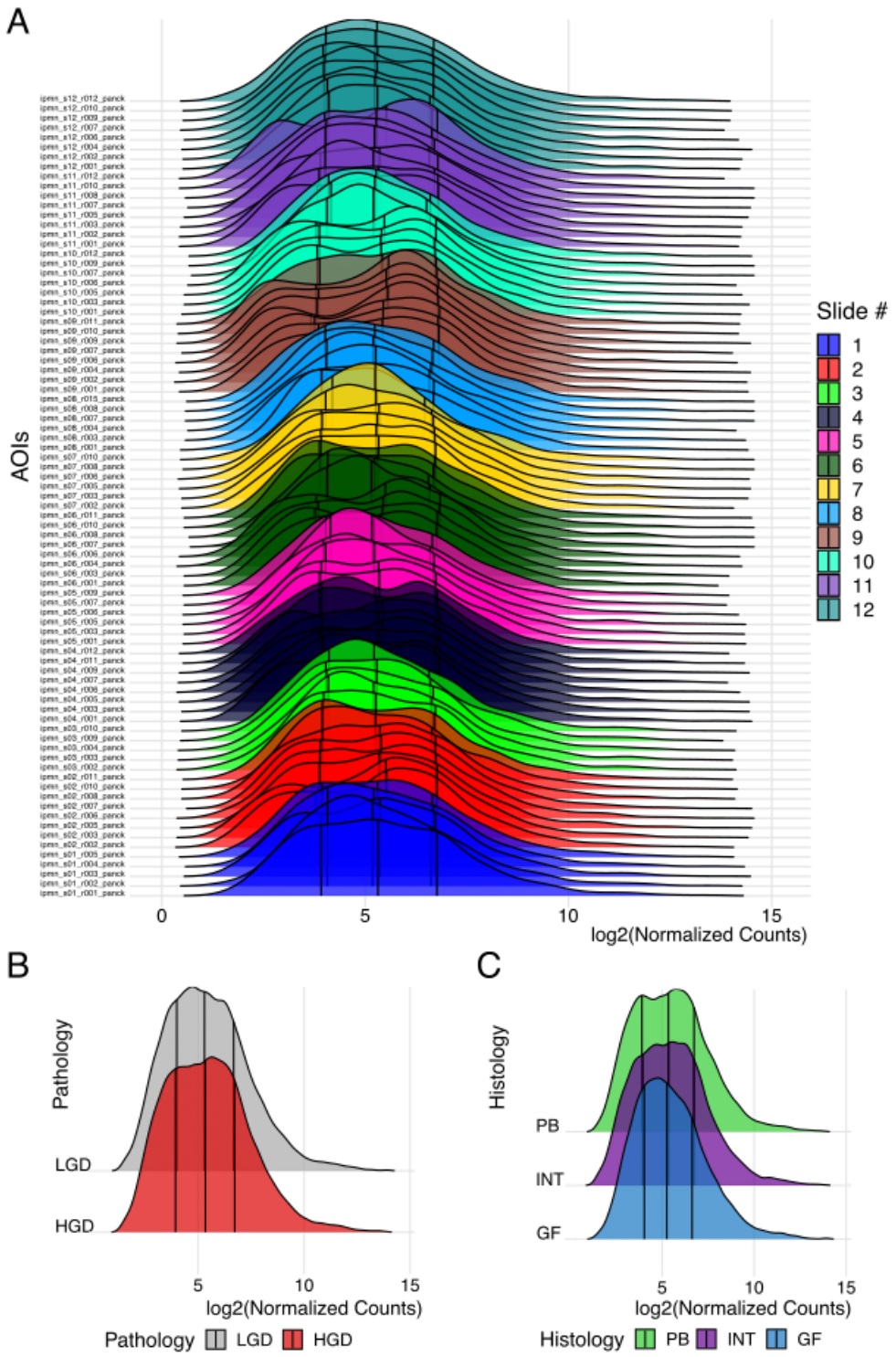

Fig. S3: Ridge plots of filtered log-normalized count distributions (x axis), (**A**) Individual AOIs colored by slide identifier. (**B**) Aggregated AOIs dichotomized by LGD or HGD. (**C**) Aggregated AOIs separated by PB, INT, and GF subtypes.

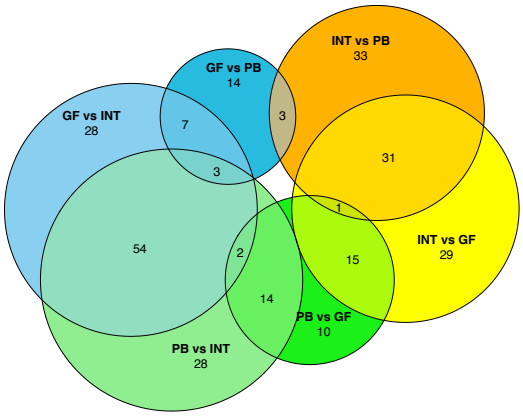

Fig. S4: Area-proportional Euler diagram *(60)* showing the intersection of DE gene sets from pairwise comparisons of PB, INT, and GF AOIs.

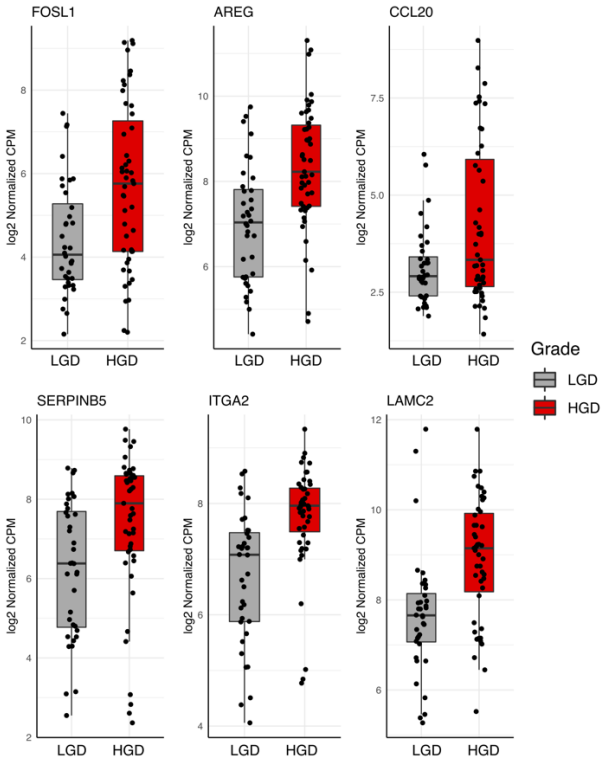

Fig. S5: Boxplots of differentially expressed genes in HGD versus LGD IPMN that were also represented within the CPTAC PDAC RNA-Seq dataset and associated with unfavorable prognosis in PDAC (Human Protein Atlas).

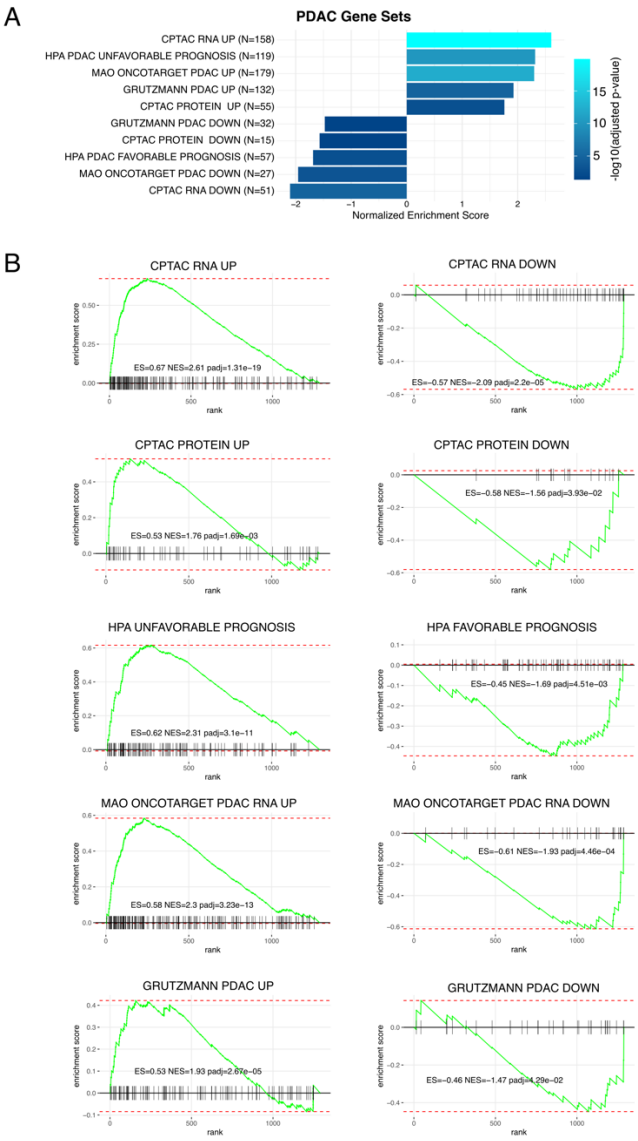

Fig. S6: GSEA analysis of AOIs with HGD versus LGD based on curated external gene sets. **(A)** Summary bar plot showing normalized enrichment score of each gene set. Bar colors reflect adjusted p-value. All adjusted p-values were <0.05. **(B)** GSEA plots of each individual analysis depicted in **(A)**.

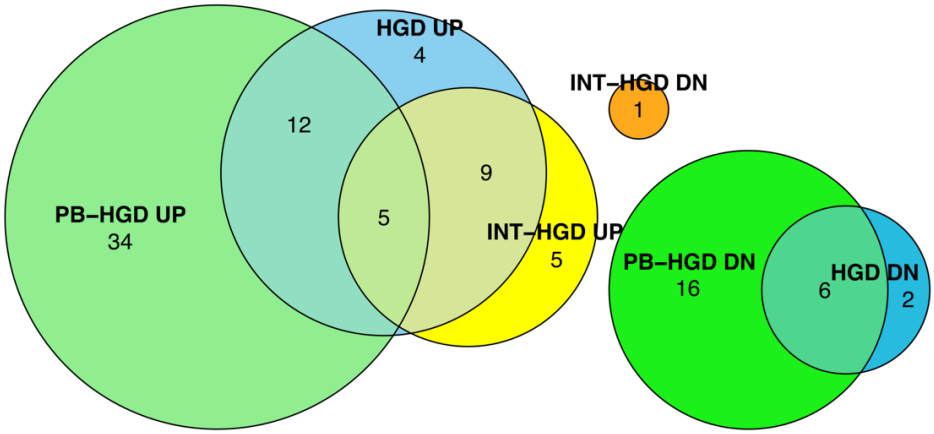

Fig. S7: Area-proportional Euler diagram showing the intersection of DE gene sets from multiple comparisons: HGD versus LGD, PB-HGD versus PB-LGD, and INT-HGD versus INT-LGD.

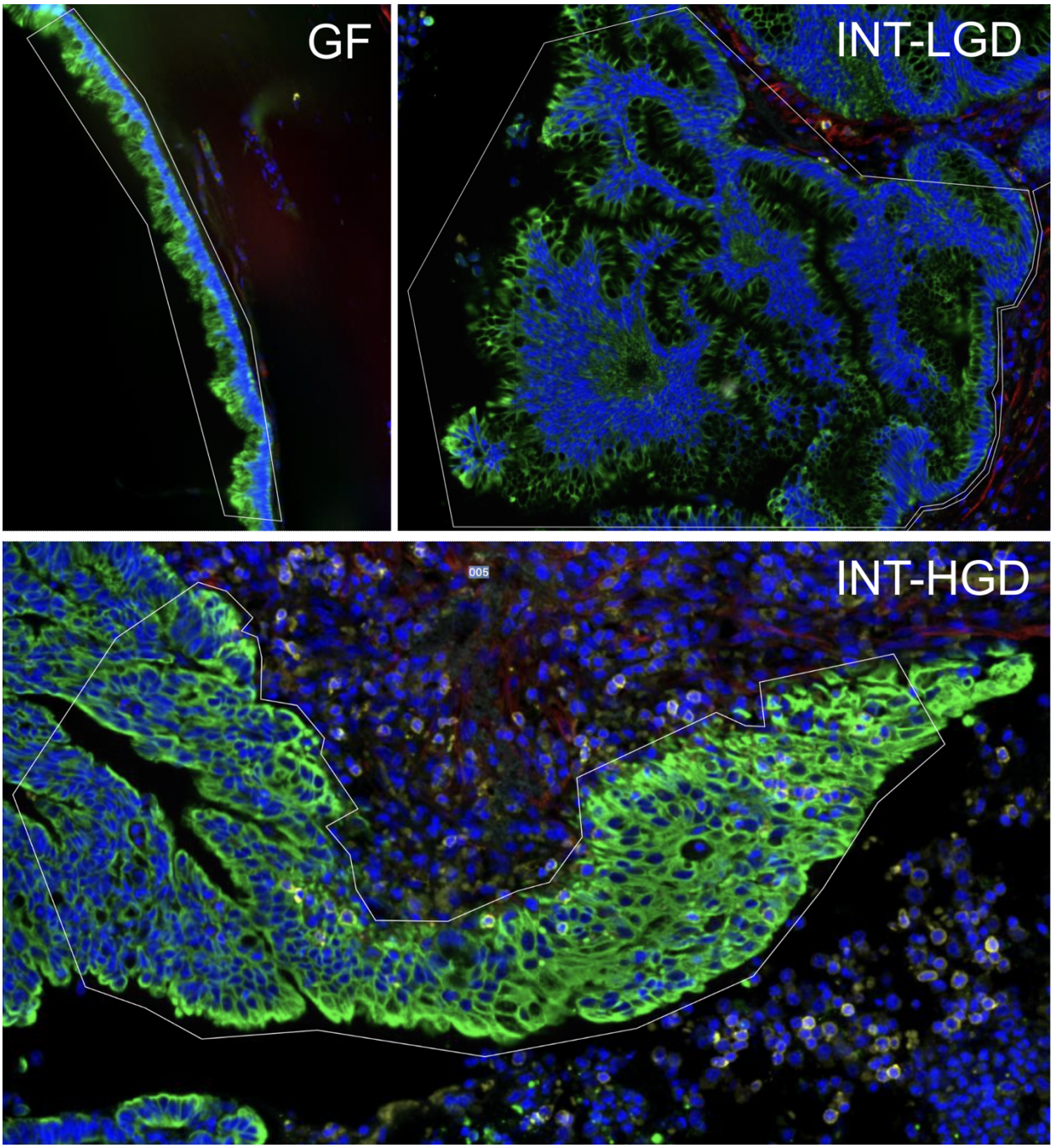

Fig. S8: Example AOIs from slide 2 harboring GF, INT-LGD, and INT-HGD subtypes.