Intraductal Papillary Mucinous Neoplasm Cellular Plasticity linked with Repeat Element Dysregulation

SHORT TITLE: Intraductal Papillary Mucinous Neoplasm Cellular Plasticity

Amaya Pankaj[1,*], Michael J. Raabe[1,*], Yuhui Song[1], Bidish K. Patel[1], Katherine Xu[1], Joshua R. Kocher[1], Peter Richieri[1], Nicholas J. Caldwell[2], Peiyun Ni[7], Maria L Ganci[2], Linda T. Nieman[1], Mingjuan L. Zhang[2], Mari Mino-Kenudson[2], Vikram Deshpande[3], Nabeel Bardeesy[1], Carlos Fernandez-Del Castillo[4], Martin J. Aryee[5,6], David T. Ting[1,#], Yasmin G. Hernandez-Barco[7,#]

[1] Mass General Cancer Center, Harvard Medical School, Charlestown, MA, 02129, USA.
[2] Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, 02114, USA.
[3] Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA.
[4] Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA, 02114, USA.
[5] Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, 02215, USA.
[6] Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA.
[7] Department of Medicine, Division of Gastroenterology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, 02114, USA.

* Equal contribution

[#] Co-corresponding authors

GRANT SUPPORT

Dana Farber/Harvard Cancer Center: GI SPORE 5P50CA127003-12 Career Development Award (YH-B)

MGH Executive Committee on Research Center for Diversity and Inclusion Physician Scientist Award (YH-B)

National Institutes of Health grant U01CA228963 (DTT), NIH/NCI Diversity Supplement: R01 CA215498-02 (YH-B)

Massachusetts Life Sciences Center Research Infrastructure Program (DTT)

ABBREVIATIONS

IPMN – intraductal papillary mucinous neoplasm

CORRESPONDANCE

David T. Ting MD
Mass General Cancer Center
149 13th St. Room 6003
Charlestown, MA 02129
617-240-9402
dting1@mgh.harvard.edu

Yasmin G. Hernandez-Barco MD
Massachusetts General Hospital
15 Parkman St., Wang 5
Boston, MA 02114
yhernandez-barco@mgh.harvard.edu

DISCLOSURES: DTT received in kind research support and a speaking honorarium from NanoString Technologies, whose technology was used in this paper. DTT has received consulting fees from ROME Therapeutics, Sonata Therapeutics, Leica Biosystems Imaging, PanTher Therapeutics, and abrdn. DTT is a founder and has equity in ROME Therapeutics, PanTher Therapeutics and TellBio, Inc., which is not related to this work. DTT is on the advisory board with equity for ImproveBio, Inc. DTT has received honorariums from Astellas, AstraZeneca, Moderna, and Ikena Oncology that are not related to this work. DTT receives research support from ACD-Biotechne, AVA LifeScience GmbH, Incyte Pharmaceuticals, Sanofi, and Astellas which was not used in this work. DTT's interests were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict of interest policies.

MMK has served as a compensated consultant for AstraZeneca, Innate, BMS, Roche, Boehringer Ingelheim, AbbVie, Daiishi-Sankyo, and Sanofi, and received royalties from Elsevier, all of which are nor related to this work.

AUTHOR CONTRIBUTIONS

Conceptualization AP, DTT, YGH-B
Formal Analysis AP, MJR, DTT, YGH-B
Investigation by AP, MJR, YS, BKP, KX, JRK, NJ-C, SM, LTN, MLZ, MM-K, VD, MJA, DTT, YGH-B
Methodology by AP, MJR, YS, BKP, KX, NP, LTN, MM-K, VD, MJA, DTT, YGH-B
Resources by LTN, MM-K, VD, CF-D
Writing AP, MJR, DTT, YGH-B
Visualization by AP, MJR, YS, BKP, KX, MJA, DTT
Supervision by AP, LTN, MJA, DTT, YGH-B
Project Administration AP, MJA, DTT, YGH-B
Funding Acquisition DTT, YGH-B

DATA TRANSPARENCY STATEMENT: All spatial transcriptomic data will be uploaded to NCBI GEO or other appropriate public repository. Otherwise, all data will be provided upon request. All analysis codes will be provided upon request.

ABSTRACT (260 words)

Background and aims: Intraductal papillary mucinous neoplasms (IPMNs) are preneoplastic pancreatic lesions that present on a spectrum of histologic phenotypes with variable risk in progressing to invasive cancer. Aberrant repetitive element expression has been shown to be functionally linked to cell state changes in pancreatic cancer. This study utilized spatial transcriptomics with customized repeat element probes to better understand the relationship of histologic subtypes, repeat element dysregulation, and molecular profiles of different cell populations in IPMN.

Methods: A total of 63 lesions from 22 patients with resected IPMNs of different histologies and degrees of dysplasia were analyzed with whole transcriptome spatial analysis (GeoMx). Of these, 49 lesions from 18 patients were also processed for single-cell spatial molecular imaging (CosMx). Repeat element probes for LINE1, HSATII, HERVK, and HERVH were used for GeoMx and CosMx.

Results: Pancreaticobiliary IPMN was enriched for basal-like epithelium and infiltration of Treg cells. Intestinal IPMN was enriched for classical epithelium and macrophage infiltrates. Gastric IPMN was found to have equal basal-like and classical epithelium with a diverse immune infiltrate. Repeat RNAs were expressed at high levels across IPMN phenotypes and enriched in high-grade dysplasia. Single-cell transcriptional trajectory analysis reveals a phylogeny starting from Gastric IPMN to Intestinal and Pancreaticobiliary branches associated with higher-grade dysplasia and repeat RNA expression.

Conclusion: Spatial transcriptomics of IPMN identified transcriptional plasticity between gastric, pancreaticobiliary, and intestinal histologies and a molecular continuum between these cell

states. Elevated repeat element expression was linked with cells in transition between cell states and identify cellular progression to high-grade dysplasia.

Keywords: IPMN, Spatial transcriptomics, cell plasticity, heterogeneity, repeat element

WHAT YOU NEED TO KNOW

Background and Context

Intraductal papillary mucinous neoplasms (IPMNs) present with a wide range of morphologies with variable potential to progress to invasive cancer.

New Findings

This study utilized spatial transcriptomic analysis in IPMN that revealed that gastric, pancreaticobiliary, and intestinal morphologies are on a molecular continuum with aberrant repeat RNA expression as a marker of transdifferentiation.

Limitations

This study provides the basis of repeat RNAs as a marker of transdifferentiation in a cohort of 22 patients, which will need validation in a larger study.

Clinical Research Relevance

Transcriptional program changes in epithelial and mesenchymal cell states with repeat RNA sequences reflect cellular plasticity that are associated with IPMN progression. Application of RNA phenotype biomarkers for IPMN risk stratification merit further investigation.

Basic Research Relevance

Single-cell transcriptional trajectory analysis reveals IPMN histologies are on a continuum that begins with low-grade gastric morphology with different cell state changes towards a high-grade gastric/pancreaticobiliary or an intestinal phenotype. Understanding the genetic and epigenetic drivers of cell states can uncover potential strategies to intercept IPMN progression to cancer.

Lay summary (25-39 words): We have used spatial transcriptomic technology to create a high-resolution map of intraductal papillary mucinous neoplasms (IPMNs), a significant precursor to pancreatic cancer, which revealed distinct cell type composition that can be used to inform risk of developing invasive cancer.

INTRODUCTION

Intraductal Papillary Mucinous neoplasms (IPMN) and Pancreatic intraepithelial neoplasia (PanINs) are common precursors to deadly pancreatic ductal adenocarcinoma (PDAC) **(1, 2)**. Whereas PanINs are microscopic and elude detection with standard imaging methods, IPMNs are larger cystic tumors that are detected in 15% of the population, with sharply increased incidence at age >70 years. The likelihood of branch duct IPMNs developing worrisome features and progressing to invasive cancer varies significantly, ranging from 4.8% to 48.8% depending on size (3). Although there have been improvements in clinical and imaging criteria for risk stratification of patients with IPMN, a significant portion of patients still undergo surgical resection that would not have progressed to invasive cancer. Pancreatic surgery has greatly improved, but there is still the risk of major morbidity of 14.4% and mortality of 0.9% as reported recently from our institution (4). The combination of increasing detection of IPMN and the highly heterogenous outcomes of patients highlights the need to obtain a more comprehensive molecular understanding of these lesions. Conventional histological analysis of IPMN has described a spectrum of morphologies including gastric, pancreaticobiliary, and intestinal subtypes. Within these morphologies, analysis of high-grade and low-grade dysplasia (HGD and LGD) have been used to help provide prognostic information, but the ability to determine dysplasia from fine needle aspirate or biopsies remains a clinical challenge (5). Promising diagnostic biomarkers from cyst fluid, pancreatic fluid, and blood are actively being developed, but there continues to be interest in identifying novel biomarkers to apply to these biospecimens (5, 6).

Given the histological heterogeneity of IPMN, there has been interest in determining if there are differences in the microenvironment. There have been extensive studies in PanINs that describe activation of inflammatory pathways, but with defects in anti-tumoral immunity (7-20). Notable features include a striking absence of CD8+ cytotoxic lymphocytes (CTLs) and early infiltration of Foxp3+ T-regulatory (T-reg) cells, macrophages, and myeloid-derived suppressor cells (MDSCs) which are considered tumor-promoting immune cells. Studies in IPMN have been more limited, but a series of studies have described the evolution of the immune microenvironment showing infiltration of T and B cells in LGD IPMN with progression to increased Tregs and macrophages in HGD and invasive PDAC (21-24). In addition, the contribution of stromal fibroblasts and their different phenotypes with primarily myofibroblast and inflammatory cancer-associated fibroblasts (myCAFs and iCAFs) has been well described in PDAC (25-27) with some studies in IPMN that have shown the presence of CAFs associated with HGD and progression to invasive disease (28, 29).

To obtain a more comprehensive view of the cellular ecosystem of IPMN, recent studies have utilized spatial transcriptomic technologies (30-32). Eckhoff et al. (30) used the Nanostring GeoMx cancer transcriptome atlas of approximately 1800 genes on 11 IPMN specimens. They identified that the immune infiltrate of LGD IPMN is enriched for T-cells, while HGD IPMN had a predominance of macrophage infiltrates. Although epithelial and fibroblast AOIs were obtained, they were not the focus of this initial spatial transcriptomic study of IPMN. Sans et al. (31) utilized 13 snap-frozen IPMN samples and used the Visium high density spot assay of 8000 genes. This study had 12,685 spots of mixed cell populations and utilized a robust cell-type deconvolution method to digitally separate cell populations. Separating samples based on dysplasia, the authors identified gastric isthmus-pit (GIP) phenotype with high *NKX6-2* in LGD IPMN with associated plasma cell infiltrates, while elevated *CLDN4* and *CEACAM6* in HGD IPMN and invasive carcinoma. Loss of NKX6-2 was also associated with increased epithelial-mesenchymal transition (EMT) markers consistent with cell state changes toward invasive

cancer. Agostini et al. (32) utilized Visium in 18 IPMN samples from 14 patients and GeoMx whole transcriptome atlas (18,676 genes) in 57 IPMN samples from 40 patients. Interestingly, this work also found *NKX6-2* correlating with gastric isthmus markers, but they found enrichment in HGD IPMN, contrary to Sans et al. This work also identified *HOXB3* and *ZNF117* as markers for LGD. In summary, the IPMN field has applied spatial transcriptomics tools to refine the molecular phenotypes that evolve during IPMN progression that have revealed some consistent themes, but others including *NKX6-2* have been discordant.

Here, we have expanded upon these studies with GeoMx analysis of 63 IPMN samples from 22 patients combined with CosMx single-molecule imaging of 1,000 genes in 49 samples from 18 patients from the same cohort. Using both technologies we have characterized transcriptional signatures and trajectories that support the transdifferentiation of HGD intestinal and pancreaticobiliary IPMN morphologies from a common LGD gastric IPMN. In addition, we have incorporated customized probes designed to detect repeat RNA sequences, which we previously identified as being elevated in human PanIN and IPMN lesions (33, 34). These repeat RNA sequences have recently been demonstrated in our studies to play a functional role in driving critical cell state changes in both cancer cells and cancer-associated fibroblasts (CAFs) within human pancreatic ductal adenocarcinoma (PDAC) (35). We find that these repeat RNAs are dysregulated in HGD IPMN and track with transdifferentiation from LGD gastric IPMN, which presents a potential new IPMN biomarker of progression to invasive PDAC. By including these probes, we aim to further explore their contribution to tumor progression and the tumor microenvironment in preneoplastic lesions.

MATERIALS AND METHODS

Specimen acquisition and histological review

This discarded excess tissue study was approved by the Institutional Review Board at the Massachusetts General Hospital (IRB Protocol 2013P001854). Patients with Intraductal Papillary Mucinous Neoplasm (IPMN) with formalin-fixed paraffin-embedded (FFPE) tissue from surgical resections were used for this project. Serial sections from FFPE blocks were then made with the first section stained using hematoxylin and eosin (H&E) and then evaluated by pathologists (MM-K, NJC). We identified 63 distinct IPMN lesions from 18 patients who had undergone pancreatic resection, for tissue microarray (TMA) construction. Three TMAs were constructed of the 63 lesions. Each TMA consisted of 24-3 mm cores with one core each of normal pancreas and normal lung as control. The presence of the lesion in each TMA core on the section for immunohistochemistry was confirmed by pathologists (MM-K, NJC). Another 11 tissue sections from a cohort of four patients who had undergone resection for IPMN were obtained for full slide analysis to evaluate intratumoral heterogeneity. Each specimen was graded into LGD and HGD areas based on the degree of dysplasia by pathologists (VD, BKP). Additional slides were used for spatial transcriptomics and validation as below. Pertinent demographic and clinical data were summarized in Supplementary Tables 1 and 2.

Spatial transcriptomics data generation

GeoMx™ Digital Spatial Profiler (DSP)

We used the GeoMx™ Digital Spatial Profiler (DSP) platform (NanoString Technologies, Seattle, WA) to collect genetic information from transcribed mRNA. This involved staining the slide with the GeoMx™ Human Whole Transcriptome Atlas (a mix of unique oligo-labeled probes against 18,676 mRNAs). This was followed by staining with fluorescent morphology markers, PanCK (NanoString), CD45 (NanoString), and Alpha-smooth muscle Actin (clone 1A4; Novus Biologicals, cat. no. IC1420R conjugated with lyophilized Alexa Fluor 647 mix ab274049, Abcam) to visualize the epithelial, immune and fibroblast compartments respectively at different emission spectra. Post staining, multiple regions of interest (ROI) were marked out with H&E as

a reference. A total of 274 ROIs were marked out, following which the instrument accurately segmented the ROIs into distinct immune/epithelial/fibroblast “areas” of interest (AOI). This was followed by the collection of the barcodes (attached to oligos by ultraviolet light cleavable bonds) onto a 96-well collection plate. The sampled oligos were then pooled following the NanoString GeoMx-NGS Readout Library Prep User Manual (MAN-10117-03). Illumina’s i5 x i7 dual indexing system was used for amplification and ligation of the adapters. Quantitation of final library was done by qPCR using KAPA estimation (KAPA code: KK4854, Roche catalog: 07960298001) using Roche LightCycler® 480 program and eventually sequenced with a 1%PhiX spike-in on a NextSeq 1000/2000 using a 2 x 100 cycle MidOutput v3 flow cell following instructions on the Illumina NextSeq manual (15048776 v16). The generated FASTQ files were processed to generate target probe-specific count files and saved in DCC format.

CosMx™ Spatial Molecular Imager (SMI)

For this study, we utilized the CosMx™ Spatial Molecular Imager (SMI) from Bruker Technologies to perform a single-cell transcriptomic analysis on three IPMN tissue microarrays (TMAs). Sample preparation followed the standardized workflow detailed in the CosMx™ SMI Manual Slide Preparation for RNA Assays (MAN-10184-04), which includes deparaffinization, target retrieval, tissue permeabilization, and overnight hybridization with RNA-specific probes. We used the Human Universal Cell Characterization Panel (1000-plex, RNA) with an additional set of custom probes capturing retrotransposons including long interspersed nuclear element-1 (LINE-1) retrotransposon open reading frames (ORF1 and ORF2), HSATII, and two human endogenous retroviruses (HERV-K and HERV-H). For cellular segmentation, we employed DAPI Nuclear Stain, CD298/B2M, and PanCK/CD45 as markers. The processed slides were imaged on the CosMx SMI platform, which combines cyclic in situ hybridization and high-resolution fluorescence imaging to detect multiplexed RNA transcripts at single-cell resolution. Field of view (FOV) selection was performed to capture representative regions of interest within the

TMAs, including areas with diverse cellular compositions and stromal interactions with varying degrees of dysplasia, among the three histological subtypes. Data analysis was conducted using the AtoMx™ Spatial Informatics Platform (SIP), enabling spatial visualization and analysis of transcriptional heterogeneity and retrotransposon expression within the IPMN microenvironment.

GeoMx computational methods

Probes were collapsed to the geometric mean of each target after removing outliers detected through a Grubbs test conducted on a per-gene basis across areas of Interest (AOIs). AOIs that exhibited low read alignment quality, low total read counts, or insufficient segment area were excluded from the final analysis to ensure data reliability. The data was quantile normalized such that each AOI has the same 75th percentile of expression. For clustering, AOIs were grouped using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA), applied to all genes that displayed a standard deviation greater than 15. This clustering approach helped identify biologically meaningful patterns, and AOIs were subsequently classified by their predominant cellular composition into three key categories: Epithelial, Immune, and Fibroblast segments, which served as the basis for further downstream analyses.

Moffitt meta-scores and fibroblast meta-scores were calculated to evaluate the molecular characteristics of different subtypes. Specifically, the Moffitt meta-scores were determined by calculating the difference in mean z-score between the Basal-like and Classical subtypes, while fibroblast meta-scores were computed as the difference in mean z-score between inflammatory cancer-associated fibroblasts (iCAFs) and myofibroblastic cancer-associated fibroblasts (myCAFs).

For differential expression analysis, a Mann-Whitney U-test was applied to compare expression levels, with p-values being FDR-adjusted. Gene Set Enrichment Analysis (GSEA) was performed using the GSEA R package version 1.2. Additionally, immune cell subtype

composition within the ROIs was assessed using the Spatial Decon R package version 1.6.0, allowing for a detailed characterization of the immune microenvironment.

CosMx computational methods

For quality control, 32 out of 390 fields of view (FOVs) were excluded from the analysis due to significant signal loss, primarily caused by high tissue autofluorescence. Individual cells were removed from the dataset if they exhibited low transcript counts (fewer than 20 transcripts per cell) or fell below a predefined size threshold of less than 25 pixels. Cell classification was done using the Insitutype v2.0 algorithm in R. The algorithm was run in an unsupervised approach, using unbiased clustering of cells based on their gene expression profiles. Dimensionality reduction techniques were employed for further analysis. Principal component analysis (PCA) was performed using the irlba R package (version 2.3.5.1). Uniform manifold approximation and projection (UMAP) was then applied using the uwot R package (version 0.2.2). To quantify the activity of gene programs across cell populations, meta-scores were calculated as the mean z-score of gene expression within each defined program. Additionally, a repeat score was computed for each cell, defined as the sum of the expression levels of repetitive element-associated genes to provide a measure of repeat element activity. Pseudotime trajectory analysis was carried out using the Slingshot R package (version 2.4.0).

RESULTS

Characterization of the subtype-specific IPMN microenvironment with whole transcriptome digital spatial profiling

We utilized the Nanostring GeoMx platform with the Whole Transcriptome Atlas panel, comprising an oligonucleotide probe set targeting 18,676 genes, to investigate the spatial organization of compartment-specific gene expression programs between the histological

subtypes of IPMN. Our cohort included 52 formalin-fixed, paraffin-embedded (FFPE) clinical specimens from 18 treatment-naive, resected IPMN patients. The cohort was distributed across histological subtypes, including 24 Gastric, 14 Intestinal, and 14 Pancreaticobiliary cases. A multi-step approach was employed to capture specific areas of interest (AOIs) corresponding to epithelial, immune, and fibroblast compartments within the IPMN microenvironment. Oligonucleotide probes were hybridized to target mRNAs, and fluorescently labeled antibodies against Pan-CK (epithelial), CD45 (immune), and Alpha-SMA (fibroblast) were utilized to segment these compartments. Following segmentation, 274 regions of interest (ROIs) were identified across the 52 samples, with a median of 5 ROIs (range 6) per core (**Figure 1A**). A surgical pathologist with expertise in pancreatic pathology (MM-K) evaluated each ROI for histological grade and sub-type, enabling compartment-specific profiling of epithelial, immune, and stromal regions (**Figure 1B**). Normalized gene expression profiles from each AOI were subjected to unsupervised hierarchical clustering analysis of the 900 most variant genes, demonstrating a clear distinction between the epithelial and stromal compartments (**Figure 1C**). Principal component analysis (PCA) further revealed a distinct separation of the three histological subtypes—gastric, intestinal, and pancreaticobiliary—along the PC1 and PC2 axes (**Figure 1D**), reflecting the unique transcriptomic landscapes between each subtype and underscoring the inter- and intra-tumoral heterogeneity of preneoplastic epithelial cells.

The Pancreaticobiliary subtype – A Basal-like, Immunosuppressed cluster

Unsupervised hierarchical clustering of all Pancreaticobiliary (PB) epithelial AOIs did not reveal distinct clusters by grade (**Figure 2A**). To identify PB subtype-specific epithelial cell phenotypes, using Moffitt gene signatures for classical/epithelial and basal/mesenchymal phenotypes previously defined for pancreatic ductal adenocarcinoma, we calculated a meta score defining each epithelial AOI into Classical or Basal-like. This analysis classified 58% of epithelial AOIs in the PB subtype as Basal-like and 42% as Classical, revealing a predominantly basal-like

signature on a heatmap (**Figure 2B**). To investigate the immune microenvironment of our PB cohort, we proceeded with an immune deconvolution analysis to determine the immune cell composition of the PB sub-type. Tregs exhibited the highest relative enrichment, accounting for 24.1% of the total immune infiltrate, with T.CD4.naive and T.CD8.memory cells contributing 15.7% and 13.4%, respectively (**Figure 2C**). Other immune cell populations, including macrophages and NK cells, showed lower representation. We used a meta score derived from previously defined myCAF and iCAF gene signatures annotated in PDAC (25-27) to characterize the fibroblast niche. This analysis classified 62% of fibroblast AOIs in the PB subtype as iCAF and 38% as myCAF (**Figure 2D**). In summary, the PB subtype predominantly exhibits a Basal-like epithelial cell phenotype, and an immunosuppressive microenvironment enriched in Tregs and iCAF fibroblasts.

The Intestinal sub-type – A classical, inflammatory cluster

Unsupervised hierarchical clustering of all Intestinal epithelial AOIs did not reveal distinct clusters by grade (**Figure 3A**). To identify intestinal subtype-specific epithelial cell phenotypes, using Moffitt gene signatures for classical/epithelial and basal/mesenchymal phenotypes previously defined for pancreatic ductal adenocarcinoma, we calculated a meta score defining each epithelial AOI into Classical or Basal-like. This analysis identifies that 67% of the intestinal AOIs are classical, and 33% are Basal-like revealing a predominantly classical phenotype on a heatmap (**Figure 3B**). To investigate the immune microenvironment of our Intestinal cohort, we proceeded with an immune deconvolution analysis to determine the immune cell composition of the sub-type. Macrophages exhibited the highest relative enrichment, accounting for 44.5 % of the total immune infiltrate, with Mast cells, CD4 and CD8 memory cells representing 10.8% and 9.6% respectively (**Figure 3C**). Other immune cell populations, including B memory and CD8 naive, showed lower representation. We used a meta score derived from myCAF and iCAF gene signatures annotated in PDAC to characterize the fibroblast niche. This analysis classified

61% of fibroblast AOIs in the PB subtype as iCAF and 39% as myCAF (**Figure 3D**). In summary, the Intestinal subtype predominantly exhibits a Classical epithelial cell phenotype and an inflammatory microenvironment enriched in Macrophages and iCAF fibroblasts.

The Gastric sub-type – A heterogenous transitionary cluster

Unsupervised hierarchical clustering of all Gastric epithelial AOIs did not reveal distinct clusters by grade (**Figure 4A**). To identify gastric subtype-specific epithelial cell phenotypes, using Moffitt gene signatures for classical/epithelial and basal/mesenchymal phenotypes previously defined for pancreatic ductal adenocarcinoma, we calculated a meta score defining each epithelial AOI into Classical or Basal-like. This analysis identifies that 51% of AOIs are classical, and 49% of them are Basal-like with a heterogenous expression of both cell states on the heatmap (**Figure 4B**). To investigate the immune microenvironment of our Gastric cohort, we proceeded with an immune deconvolution analysis to determine the immune cell composition of the sub-type. The Gastric IPMN subtype exhibits a diverse immune profile, highlighted by mast cell enrichment (15.1%), significant monocyte (C) enrichment (12.5%), and prominent T.CD4.Memory (14.1%) and T.CD8.Memory (11.4%) cells (**Figure 4C**). We used a meta score derived from myCAF and iCAF gene signatures annotated in PDAC to characterize the fibroblast niche. This analysis classified 63% of fibroblast AOIs in the PB subtype as iCAF and 37% of them as myCAF (**Figure 4D**). In summary, the Gastric subtype predominantly exhibits a heterogenous epithelial cell phenotype and microenvironment enriched with diverse immune cells and iCAF fibroblasts.

The IPMN Transcriptomic Spectrum

To compare the transcriptomic range between IPMN subtypes, we identified the top 200 highly expressed genes in the Pancreaticobiliary, Gastric, and Intestinal subtypes. Shared genes including the custom repeat genes were plotted against the median expression across the subtypes. The Gastric subtype aligned closely with the median, the Intestinal subtype showed

overexpression, and the Pancreaticobiliary subtype exhibited an under-expression pattern (**Figure 5A**). Using the Moffitt meta score, we then observed a phenotypic spectrum with the Intestinal subtype being predominantly classical, the Pancreaticobiliary subtype being primarily basal-like, and the Gastric subtype heterogeneous with near-equal representation of both phenotypes (**Figure 5B**). We then did a comparison of Low-Grade Dysplasia (LGD) and High-Grade Dysplasia (HGD), which did not reveal clear stratification in terms of classical versus basal-like phenotypes (**Figure 5C**). On comparing the Immune signatures between the subtypes, Intestinal has a macrophage-enriched response, PB has a T-reg enriched response whereas Gastric seems to have a diverse immune microenvironment (**Figure 5D**). Notably, analysis of *NKX6-2* in our cohort revealed significant enrichment in Gastric IPMN compared to Intestinal and Pancreaticobiliary subtypes as well as in LGD compared to HGD (Mean 58.36 vs 35.77, p-value = 0.021; (**Figure 5E**). Differential expression between LGD versus HGD identifies pathways such as MYC targets, oxidative phosphorylation, WNT Beta-catenin, etc. enriched in HGD versus Protein secretion, and Pancreas Beta cells, etc. enriched in LGD (**Supplementary Figure 1**). Immune cell analysis of LGD and HGD revealed significant enrichment of macrophages in HGD (**Supplementary Figure 2**). To confirm the transitional and heterogeneous nature of Gastric IPMN, we analyzed 11 additional FFPE whole resection specimens from 4 patients that were all identified as Gastric using the GeoMx Whole Transcriptomic Atlas panel. The Moffitt meta-score analysis of this cohort revealed significant intra- and inter-patient heterogeneity, with 5 out of 11 resections classified as basal-like and 6 out of 11 resections classified as classical, aligning with our earlier findings. (**Figure 5F**). These results suggest that IPMN represents a dynamic spectrum of cell phenotypes spanning between these histological subtypes.

Retrotransposons in the IPMN Spectrum

To investigate the heterogeneous landscape of IPMN, we analyzed the expression of retrotransposons, including long interspersed nuclear element 1 (LINE-1) retrotransposon open reading frame 1 (ORF1) and open reading frame 2 (ORF2), HSATII, and two human endogenous retroviruses (HERV; HERV-K and HERV-H), across the spectrum of IPMN subtypes. Repetitive elements emerged as some of the most highly expressed genes in the Intestinal, Gastric, and Pancreaticobiliary subtypes. While no significant differential expression of repetitive elements was observed between the histological subtypes, we found that their expression levels were significantly higher than *KRT7* in IPMN (**Figure 6A**). To determine whether these repetitive elements contribute to the observed cell-type differences in IPMN, we stratified epithelial AOIs into Classical/Epithelial, Transitional, and Basal-like phenotypes using the Moffitt meta-score. Interestingly, repetitive elements were significantly overexpressed in the Classical/Epithelial and Transitional phenotypes compared to the Basal-like phenotype (**Figure 6B**). Repeat expression was higher in IPMN with HGD compared to LGD (**Figure 6C**). Additionally, fibroblasts were categorized into iCAF, Transitional, and myCAF groups using their meta scores, and we compared repetitive element expression between these groups. We observed significantly higher expression of repetitive elements in the iCAF group compared to the myCAF group (**Figure 6D**). These findings suggest that repetitive elements mark the epithelial, transitional, and iCAF-predominant cell states in IPMN.

Single-cell Transcriptomic Analysis of the IPMN Spectrum

To comprehensively characterize cell phenotype variations between different histological subtypes of IPMN at the single-cell level, we employed the NanoString CosMx Spatial Molecular Imager (SMI) platform. This approach utilized a 1,000-plex RNA panel with custom probes targeting long interspersed nuclear element-1 (LINE-1) retrotransposon open reading frames (ORF1 and ORF2), HSATII, and two human endogenous retroviruses (HERV-K and HERV-H), enabling the in-situ visualization of individual RNA molecules (**Figure 7A**). Spatially resolved

single-cell transcriptomic data were collected from 49 IPMN resections across 18 patients, encompassing 247,944 cells imaged from 390 fields of view (FOVs) (**Figure 7A, 7B**). We identified six major clusters (**Figure 7C**) - Cluster 2 is predominantly Gastric and benign ductal or acinar cells, Clusters 1, 4, and 6 are primarily Intestinal, and Clusters 3 and 5 were enriched for Pancreaticobiliary (**Figure 7D).** On interposing the information regarding the grade of disease, Cluster 1 and 2 are comprised predominantly of LGD, and Clusters 3, 4, 5, and 6 are predominantly composed of HGD and invasive (**Figure 7E).** Leveraging the Moffitt gene set for defining Classical/Epithelial versus Basal-like/Mesenchymal phenotypes, we observe that Clusters 2 demonstrated a heterogeneous, mixed phenotype; Clusters 1, 4 and 6 are predominantly Classical; and Clusters 3 and 5 exhibited a basal-like phenotype (**Figure 7F**), consistent with our previous findings. Repeat element expression analysis revealed higher levels in Cluster 2 and 4 compared to the other clusters, aligning with previous observations (**Figure 7G**). To explore the dynamics of cell phenotypes and subtypes, we performed Pseudotime trajectory analysis. Three primary lineages were observed (**Figure 7H**). The starting point of the trajectory was predominantly formed by Low-Grade Gastric IPMN, which trifurcated into three branches: one leading to High-Grade Gastric and Pancreaticobiliary IPMN, the other to High-Grade Intestinal IPMN, and another branch that loops in (**Figure 7I**). Interestingly, the transition from Low-Grade Gastric IPMN to High-Grade Gastric and Pancreaticobiliary IPMN was evident within one branch, the other branch transitioned into Intestinal IPMN, while the third looped branch may indicate the transitional continuum these cells are in, emphasizing the pivotal role of the Gastric subtype as a transitional state (**Figure 7J).** When stratified by the Moffitt meta score, the Pancreaticobiliary branch displayed a clear enrichment of Basal-like phenotypes, whereas the Intestinal branch retained Classical phenotypes (**Figure 7K)** Repeat expression analysis showed a trend of higher enrichment in the transitional cluster compared to the other clusters (**Figure 7L**) This suggests that repeat element expression may be a marker of transition or plasticity within IPMN. In summary, these findings support the hypothesis that

Gastric IPMN represents a transitional or plastic state, and this state appears to transition to distinct transcriptomic subtypes, including Pancreaticobiliary and Intestinal IPMN, through differential cell states and phenotypic transitions.

DISCUSSION

The clinical dilemma of managing patients with IPMN has led to a variety of studies focused on identifying biomarkers of IPMN progression to invasive PDAC. This current work using spatial transcriptomic technologies has validated findings from other studies and contributes new analyses involving repeat RNA characterization linked with changes in morphology and dysplasia. The combination of focal whole transcriptome and single cell trajectory analysis has provided support of a molecular continuum of IPMN starting from low grade gastric morphology with progression to a pancreaticobiliary or intestinal subtype with increasing dysplasia.

The current work has been consistent with other studies of the IPMN immune microenvironment with a diverse immune microenvironment in gastric IPMN including plasma cells as seen by Sans et al. (31) and infiltration of macrophages with HGD consistent with Eckhoff et al. (30). Notably, our analysis of CAFs reveals no major differences in myCAF and iCAF states between the different IPMN morphologies. However, we do see repeat RNA expression enriched in iCAFs (**Figure 6**) consistent with our recent findings demonstrating a functional effect of repeat RNAs in activating innate immune signaling in CAFs that drives a shift from myCAF to iCAF states in PDAC (35). This recent work demonstrated the ability of repeat RNAs to be transmitted from PDAC cells to CAFs analogous to a “viral-like” infection, which suggests that repeat RNA delivery to the microenvironment occurs in preneoplastic IPMN lesions.

Repeat RNA expression in the context of PDAC cells was shown to be associated with EMT (35), which we found in IPMN to be more enriched in classical epithelial states. On the surface, this would appear to indicate a discordance of repeat RNA expression in PDAC compared to

IPMN, but instead, this supports the model that repeat RNAs are linked with the transitional state between cell phenotypes and not a specific cell state. In this study, we have found that repeat RNA expression was at higher levels in HGD compared to LGD (**Figure 6**). This indicates that dysplasia and associated repeat dysregulation can occur in either epithelial or mesenchymal/basal-like states. This relationship of morphology, EMT cell state, and repeat expression became more evident with single-cell spatial analysis revealing dysregulated repeats associated with higher classical epithelial transcriptional profiles in intestinal IPMN subtype and also basal-like transcriptional enrichment in pancreaticobiliary IPMN subtype (**Figure 7**). Altogether, these data support repeat RNA expression as a marker of cellular transdifferentiation and dysplasia that can serve as a risk stratification biomarker independent of morphology. These repeat RNAs can be used to complement other IPMN risk stratification markers that have been identified, including the Das-1 antibody (36-38), mutation-based assessment (39), cyst fluid CEA (40), and other multimodality approaches (41). We evaluated the previously reported IPMN dysplasia gene *NKX6-2* in our cohort and identified significant enrichment in LGD compared to HGD. This is consistent with Sans et al. (31) and contrary to the study by Agostini et al. (32), but the consensus would support *NKX6-2* as an important marker of IPMN development that merits further study. It is possible that in some cellular contexts, increased NKX6-2 leads to worsening dysplasia, and in other cell states, loss of NKX6-2 leads to a different cellular path of progression. In summary, the collective spatial transcriptomics studies have refined our understanding of IPMN progression to PDAC that has opened a new era of precision oncology to identify biomarkers for risk stratification and implications for therapeutic cancer interception.

ACKNOWLEDGEMENTS

We are grateful to Danielle Bestoso and Angelique N. Gilbert for administrative support.

Funding support from:

INCLUSION AND DIVERSITY STATEMENT

All patient samples included in this study were obtained on a discarded excess tissue IRB approved protocol that utilized archived tissue biopsied or resected as part of clinical standard of care. Although not specified, we included samples to capture the diversity of gender and ethnic backgrounds of the patient population served by the Massachusetts General Hospital. We worked to ensure sex and diversity of patient derived cell lines used in this manuscript. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location.

REFERENCES

1. Farrell JJ, Fernandez-del Castillo C. Pancreatic cystic neoplasms: management and unanswered questions. Gastroenterology. 2013;144(6):1303-15.
2. Rech AJ, Vonderheide RH. Dynamic interplay of oncogenes and T cells induces PD-L1 in the tumor microenvironment. Cancer Discov. 2013;3(12):1330-2.
3. Ohtsuka T, Fernandez-Del Castillo C, Furukawa T, Hijioka S, Jang JY, Lennon AM, et al. International evidence-based Kyoto guidelines for the management of intraductal papillary mucinous neoplasm of the pancreas. Pancreatology. 2024;24(2):255-70.
4. Roldan J, Harrison JM, Qadan M, Bolm L, Baba T, Brugge WR, et al. "Evolving Trends in Pancreatic Cystic Tumors: A 3-Decade Single-Center Experience With 1290 Resections". Ann Surg. 2023;277(3):491-7.
5. Gardner TB, Park WG, Allen PJ. Diagnosis and Management of Pancreatic Cysts. Gastroenterology. 2024;167(3):454-68.
6. Wood LD, Canto MI, Jaffee EM, Simeone DM. Pancreatic Cancer: Pathogenesis, Screening, Diagnosis, and Treatment. Gastroenterology. 2022;163(2):386-402.e1.
7. Bayne LJ, Beatty GL, Jhala N, Clark CE, Rhim AD, Stanger BZ, et al. Tumor-derived granulocyte-macrophage colony-stimulating factor regulates myeloid inflammation and T cell immunity in pancreatic cancer. Cancer Cell. 2012;21(6):822-35.
8. Beatty GL, Winograd R, Evans RA, Long KB, Luque SL, Lee JW, et al. Exclusion of T Cells From Pancreatic Carcinomas in Mice Is Regulated by Ly6C(low) F4/80(+) Extratumoral Macrophages. Gastroenterology. 2015;149(1):201-10.
9. Chang JH, Jiang Y, Pillarisetty VG. Role of immune cells in pancreatic cancer from bench to clinical application: An updated review. Medicine (Baltimore). 2016;95(49):e5541.
10. Clark CE, Hingorani SR, Mick R, Combs C, Tuveson DA, Vonderheide RH. Dynamics of the immune reaction to pancreatic cancer from inception to invasion. Cancer Res. 2007;67(19):9518-27.
11. Erkan M, Hausmann S, Michalski CW, Fingerle AA, Dobritz M, Kleeff J, et al. The role of stroma in pancreatic cancer: diagnostic and therapeutic implications. Nat Rev Gastroenterol Hepatol. 2012;9(8):454-67.

12. Jang JE, Hajdu CH, Liot C, Miller G, Dustin ML, Bar-Sagi D. Crosstalk between Regulatory T Cells and Tumor-Associated Dendritic Cells Negates Anti-tumor Immunity in Pancreatic Cancer. Cell Rep. 2017;20(3):558-71.
13. Marks DL, Olson RL, Fernandez-Zapico ME. Epigenetic control of the tumor microenvironment. Epigenomics. 2016;8(12):1671-87.
14. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 2015;47(10):1168-78.
15. Patra KC, Kato Y, Mizukami Y, Widholz S, Boukhali M, Revenco I, et al. Mutant GNAS drives pancreatic tumourigenesis by inducing PKA-mediated SIK suppression and reprogramming lipid metabolism. Nat Cell Biol. 2018;20(7):811-22.
16. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell. 2015;160(1-2):48-61.
17. Timosenko E, Hadjinicolaou AV, Cerundolo V. Modulation of cancer-specific immune responses by amino acid degrading enzymes. Immunotherapy. 2017;9(1):83-97.
18. Zhang Y, Yan W, Mathew E, Bednar F, Wan S, Collins MA, et al. CD4+ T lymphocyte ablation prevents pancreatic carcinogenesis in mice. Cancer Immunol Res. 2014;2(5):423-35.
19. Zhang Y, Zoltan M, Riquelme E, Xu H, Sahin I, Castro-Pando S, et al. Immune Cell Production of Interleukin 17 Induces Stem Cell Features of Pancreatic Intraepithelial Neoplasia Cells. Gastroenterology. 2018;155(1):210-23 e3.
20. Zhu Y, Herndon JM, Sojka DK, Kim KW, Knolhoff BL, Zuo C, et al. Tissue-Resident Macrophages in Pancreatic Ductal Adenocarcinoma Originate from Embryonic Hematopoiesis and Promote Tumor Progression. Immunity. 2017;47(3):597.
21. Roth S, Zamzow K, Gaida MM, Heikenwälder M, Tjaden C, Hinz U, et al. Evolution of the immune landscape during progression of pancreatic intraductal papillary mucinous neoplasms to invasive cancer. EBioMedicine. 2020;54:102714.
22. Hernandez S, Parra ER, Uraoka N, Tang X, Shen Y, Qiao W, et al. Diminished Immune Surveillance during Histologic Progression of Intraductal Papillary Mucinous Neoplasms Offers a Therapeutic Opportunity for Cancer Interception. Clin Cancer Res. 2022;28(9):1938-47.
23. Mo S, Zou L, Hu Y, Chang X, Chen J. Expression of PD-L1 and VISTA in Intraductal Papillary Mucinous Neoplasm With Associated Invasive Carcinoma of the Pancreas. Mod Pathol. 2023;36(9):100223.
24. Bernard V, Semaan A, Huang J, San Lucas FA, Mulu FC, Stephens BM, et al. Single-Cell Transcriptomics of Pancreatic Cancer Precursors Demonstrates Epithelial and Microenvironmental Heterogeneity as an Early Event in Neoplastic Progression. Clin Cancer Res. 2019;25(7):2194-205.
25. Biffi G, Oni TE, Spielman B, Hao Y, Elyada E, Park Y, et al. IL1-Induced JAK/STAT Signaling Is Antagonized by TGFbeta to Shape CAF Heterogeneity in Pancreatic Ductal Adenocarcinoma. Cancer Discov. 2019;9(2):282-301.
26. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, et al. Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. Cancer Discov. 2019;9(8):1102-23.
27. Ohlund D, Elyada E, Tuveson D. Fibroblast heterogeneity in the cancer wound. J Exp Med. 2014;211(8):1503-23.
28. Kakizaki Y, Makino N, Tozawa T, Honda T, Matsuda A, Ikeda Y, et al. Stromal Fibrosis and Expression of Matricellular Proteins Correlate With Histological Grade of Intraductal Papillary Mucinous Neoplasm of the Pancreas. Pancreas. 2016;45(8):1145-52.
29. Shindo K, Aishima S, Ohuchida K, Fujino M, Mizuuchi Y, Hattori M, et al. Podoplanin expression in the cyst wall correlates with the progression of intraductal papillary mucinous neoplasm. Virchows Arch. 2014;465(3):265-73.

30. Eckhoff AM, Fletcher AA, Landa K, Iyer M, Nussbaum DP, Shi C, et al. Multidimensional Immunophenotyping of Intraductal Papillary Mucinous Neoplasms Reveals Novel T Cell and Macrophage Signature. Ann Surg Oncol. 2022;29(12):7781-8.
31. Sans M, Makino Y, Min J, Rajapakshe KI, Yip-Schneider M, Schmidt CM, et al. Spatial Transcriptomics of Intraductal Papillary Mucinous Neoplasms of the Pancreas Identifies NKX6-2 as a Driver of Gastric Differentiation and Indolent Biological Potential. Cancer Discov. 2023;13(8):1844-61.
32. Agostini A, Piro G, Inzani F, Quero G, Esposito A, Caggiano A, et al. Identification of spatially-resolved markers of malignant transformation in Intraductal Papillary Mucinous Neoplasms. Nat Commun. 2024;15(1):2764.
33. Franses JW, Basar O, Kadayifci A, Yuksel O, Choz M, Kulkarni AS, et al. Improved Detection of Circulating Epithelial Cells in Patients with Intraductal Papillary Mucinous Neoplasms. Oncologist. 2017.
34. Ting DT, Lipson D, Paul S, Brannigan BW, Akhavanfard S, Coffman EJ, et al. Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science. 2011;331(6017):593-6.
35. You E, Danaher P, Lu C, Sun S, Zou L, Phillips IE, et al. Disruption of cellular plasticity by repeat RNAs in human pancreatic cancer. Cell. 2024.
36. Das KK, Brown JW, Fernandez Del-Castillo C, Huynh T, Mills JC, Matsuda Y, et al. mAb Das-1 identifies pancreatic ductal adenocarcinoma and high-grade pancreatic intraepithelial neoplasia with high accuracy. Hum Pathol. 2021;111:36-44.
37. Das KK, Geng X, Brown JW, Morales-Oyarvide V, Huynh T, Pergolini I, et al. Cross Validation of the Monoclonal Antibody Das-1 in Identification of High-Risk Mucinous Pancreatic Cystic Lesions. Gastroenterology. 2019;157(3):720-30 e2.
38. Das KK, Xiao H, Geng X, Fernandez-Del-Castillo C, Morales-Oyarvide V, Daglilar E, et al. mAb Das-1 is specific for high-risk and malignant intraductal papillary mucinous neoplasm (IPMN). Gut. 2014;63(10):1626-34.
39. Springer S, Wang Y, Dal Molin M, Masica DL, Jiao Y, Kinde I, et al. A combination of molecular markers and clinical features improve the classification of pancreatic cysts. Gastroenterology. 2015;149(6):1501-10.
40. Brugge WR, Lewandrowski K, Lee-Lewandrowski E, Centeno BA, Szydlo T, Regan S, et al. Diagnosis of pancreatic cystic neoplasms: a report of the cooperative pancreatic cyst study. Gastroenterology. 2004;126(5):1330-6.
41. Springer S, Masica DL, Dal Molin M, Douville C, Thoburn CJ, Afsari B, et al. A multimodality test to guide the management of patients with a pancreatic cyst. Sci Transl Med. 2019;11(501).

FIGURE LEGENDS

Figure 1 GeoMx Spatial Transcriptomic Profiling identifies sub-type specific gene expression in IPMN

(A) A low-power GeoMX DSP generated image of 3 IPMN TMA slides with multiple Regions of Interest (ROI) selected on each core with their corresponding Hematoxylin and Eosin (HE) scan and a high-power magnification of one ROI with fluorescent staining PanCK (green), CD45 (magenta), and alpha-SMA (blue) depicting the Epithelium, Immune, and Fibroblast segments respectively.

(B) Heatmap depicting unsupervised hierarchical clustering of normalized gene expression from areas of interest (AOIs) shows distinct separation between epithelial (green) and stromal compartments (magenta and blue).
(C) Principal component analysis (PCA) of gene expression highlights distinct transcriptomic landscapes of the Gastric (green), Intestinal (blue), and Pancreaticobiliary (red) subtypes along PC1 and PC2.

Figure 2 The Pancreaticobiliary Subtype - a Basal-Like, Immunosuppressed Cluster

(A) Heatmap depicting the unsupervised hierarchical clustering of Pancreaticobiliary (PB) epithelial AOIs using the 200 most variably expressed genes interposed with the grade of the disease does not reveal clustering by grade.
(B) Heatmap depicting the distribution of Basal-like versus Classical AOIs. Classification of PB epithelial AOIs using the Moffitt metascore for each AOI identifies 58% as Basal-like and 42% as Classical phenotypes.
(C) Bar graph representing Immune deconvolution analysis using the Immune AOIs shows enrichment of Tregs (24.1%), followed by T.CD4.naive (15.7%) and T.CD8.memory (13.4%) cells.
(D) Fibroblast profiling using a Metascore for iCAF versus myCAF gene signatures identifies 62% of fibroblast AOIs as iCAF and 38% as myCAF.

Figure 3 The Intestinal subtype – a classical, inflammatory Cluster

(A) Heatmap depicting the unsupervised hierarchical clustering of Intestinal epithelial AOIs using the most variably expressed genes interposed with the grade of the disease does not reveal distinct stratification by grade.
(B) Heatmap depicting the distribution of Basal-like versus Classical AOIs. Classification of Intestinal epithelial AOIs using the Moffitt metascore for each AOI identifies 33% as Basal-like and 67% as Classical phenotypes.
(C) Bar graph representing Immune deconvolution analysis reveals macrophages as the dominant immune cell type (44.5%), followed by mast cells (8.98%) and T.CD4.memory cells (10.8%)
(D) Fibroblast profiling using a metascore for iCAF versus myCAF gene signatures identifies 61% of fibroblast AOIs as iCAF and 39% as myCAF.

Figure 4 The Gastric subtype – a heterogenous, transitional cluster

(A) Heatmap depicting the unsupervised hierarchical clustering of Gastric epithelial AOIs using the most variably expressed genes interposed with the grade of the disease does not reveal distinct stratification by grade.
(B) Heatmap depicting the distribution of Basal-like versus Classical AOIs. Classification of Gastric epithelial AOIs using the Moffitt metascore for each AOI identifies 49% as Basal-like and 51% as Classical phenotypes.

(C) Bar graph representing Immune deconvolution reveals mast cells (15.1%) and monocytes (12.5%) as prominent populations, along with T.CD4.memory (14.1%) and T.CD8.memory (11.4%) cells
(D) Fibroblast profiling using a Metascore for iCAF versus myCAF gene signatures identifies 63% of fibroblast AOIs as iCAF and 37% as myCAF.

Figure 5 – IPMN is on the spectrum of cellular phenotypes

(A) Plot depicting the gene expression (y-axis) of the shared genes (x-axis) among Pancreaticobiliary, Gastric, and Intestinal subtypes highlights that Gastric (green) aligns closely with the median (black), Intestinal (blue) shows overexpression, and Pancreaticobiliary (red) exhibits a trend of under-expression of genes.
(B) Moffitt metascore analysis highlights a phenotypic spectrum, with Intestinal (blue) predominantly Classical = -1, Pancreaticobiliary (red) primarily Basal-like = 1, and Gastric (gastric) heterogeneous = 0
(C) Comparison of Low-Grade Dysplasia (LGD) and High-Grade Dysplasia (HGD) using the Moffitt Metascore shows no stratification by phenotype.
(D) Heatmap depicting the supervised clustering of Classical versus Basal-like phenotypes using Moffit classification in an additional cohort of 11 IPMN whole resections of the Gastric subtype from 4 patients identifies 45.5% of them as Basal-like and 54.5 % being classical

Figure 6 Retrotransposon Expression in IPMN Phenotypes

(A) Heatmap of repetitive element expression (LINE-1 ORF1/ORF2, HSATII, HERV-K, and HERV-H) across Pancreaticobiliary (red), Gastric (green), and Intestinal (blue) subtypes. Repetitive elements are expressed at levels significantly higher than KRT7 but show no statistically significant differential expression between the histological subtypes
(B) Violin plots of repetitive element expression across epithelial phenotypes are binned as Classical (-2 to -0.5), Transitional (-0.5 to 0.5), and Basal-like (0.5 to 2) based on their Moffitt metascore. Repetitive elements show a trend of overexpression in Classical and Transitional phenotypes compared to Basal-like phenotypes
(C) Violin plots of repetitive element expression across epithelial phenotypes comparing LGD versus HGD show enrichment of LINE1_ORF, LINE1_ORF2, and HSATII in HGD.
(D) Violin plots of repetitive element expression across fibroblast phenotypes classified as iCAF (-2 to -0.25), Transitional (-0.25 to 0.25), and myCAF (0.25 to 2) based on their Fibroblast metascore. iCAF fibroblasts exhibit significantly higher repetitive element expression compared to myCAF fibroblasts

Figure 7 Single-Cell Spatial Transcriptomic Analysis of IPMN Subtypes

(A) Spatial distribution of single cells annotated by histological subtype (Gastric, Intestinal, Pancreaticobiliary) across all fields of view (FOVs).
(B) Representative image (top) of FOV segmentation highlighting IPMN (green) and stromal (blue) regions. The corresponding segmented FOV (bottom) shows the spatial resolution of individual cell compartments.
(C) UMAP visualization of single-cell transcriptomic data identifying six unique clusters, distinguishing epithelial (IPMN), stromal, benign acinar epithelium, and benign ductal epithelium compartments.
(D) UMAP visualization of single cells color-coded by histological subtype, showing clustering of Gastric (green), Intestinal (blue), and Pancreaticobiliary (red) subtypes.
(E) (E) UMAP clustering showing cells annotated by grade, including Low-Grade Dysplasia (LGD), High-Grade Dysplasia (HGD), HGD with a nearby invasive focus, invasive lesions, benign acinar epithelium, and benign ductal epithelium.
(F) UMAP overlay of Moffitt metascore visualizing Classical to Basal-like phenotypic transitions with Clusters 1, 4, and 6 being predominantly classical and Clusters 3 and 5 being Basal-like
(G) UMAP overlay of repeat element expression (Repeat Score) clusters 2 and 4 having the highest repeat expression
(H) Pseudo-time trajectory analysis plotted in PCA space shows differentiation trajectories diverging into three distinct lineages.
(I) PCA plot with cells annotated by histological subtype, visualizing subtype-specific distributions with Gastric (green), Intestinal (blue), and Pancreaticobiliary (red) along the three lineages.
(J) PCA plot with cells annotated by grade, highlighting the transition from LGD (green) to HGD (orange), HGD with a nearby invasive focus, and invasive (purple) disease states.
(K) PCA overlay of Moffitt metascore, illustrating the phenotypic spectrum from Classical (blue) to Basal-like (yellow) across the pseudo-time trajectory.
(L) PCA overlay of repeat element expression (Repeat Score), showing the transition between transitional and specific cell states.

Figure 1

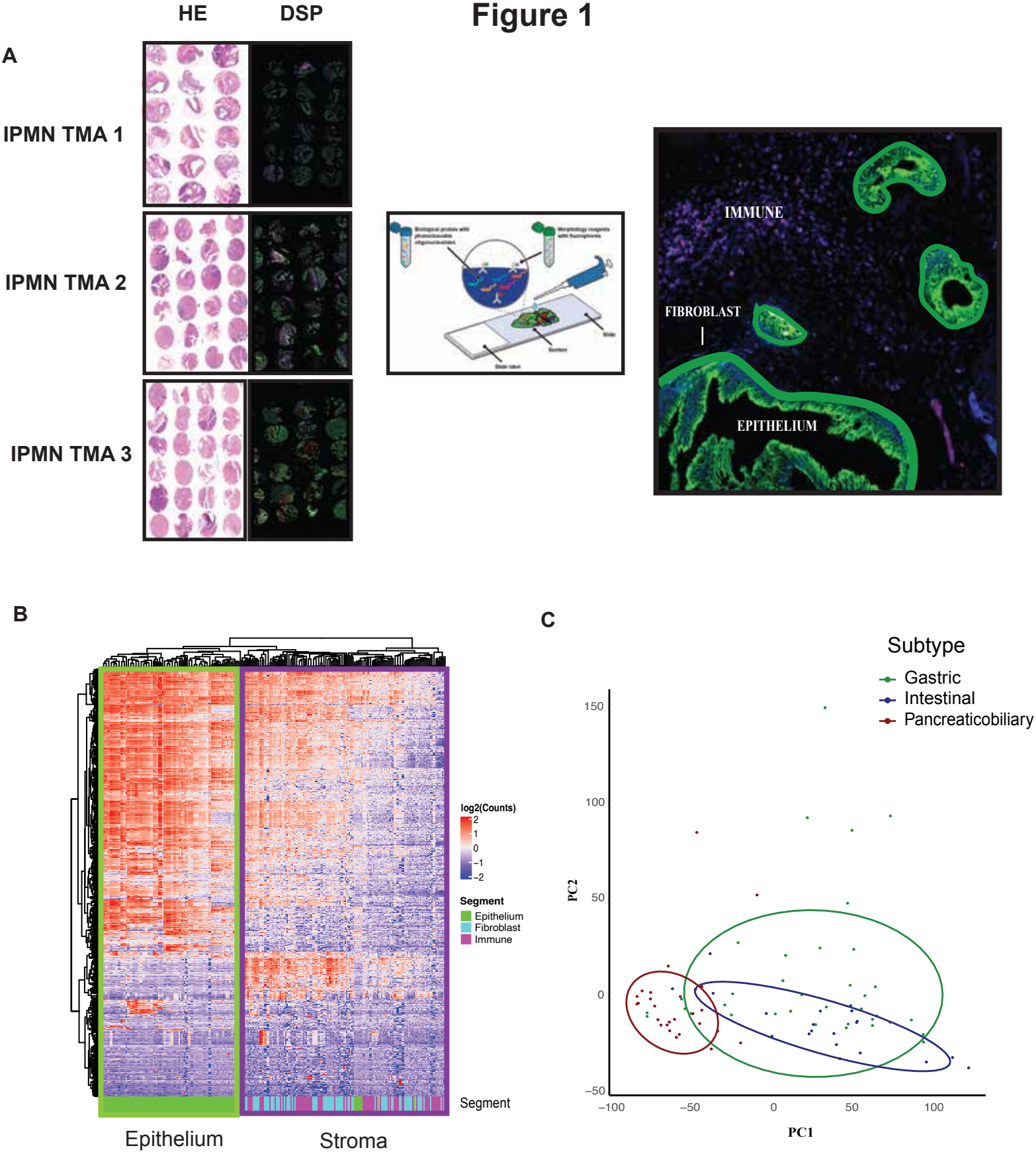
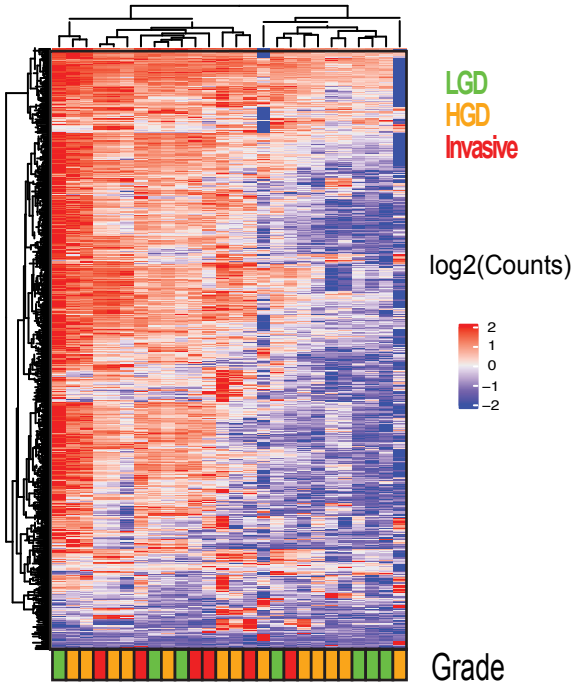

A
HE
DSP
IPMN TMA 1
IPMN TMA 2
IPMN TMA 3
IMMUNE
FIBROBLAST
EPITHELIUM
B
log2(Counts)
2
1
0
−1
−2
Segment
Epithelium
Fibroblast
Immune
Segment
Epithelium
Stroma
C
Subtype
Gastric
Intestinal
Pancreaticobiliary
150
100
50
0
−50
PC2
−100
−50
0
50
100
PC1

Figure 2

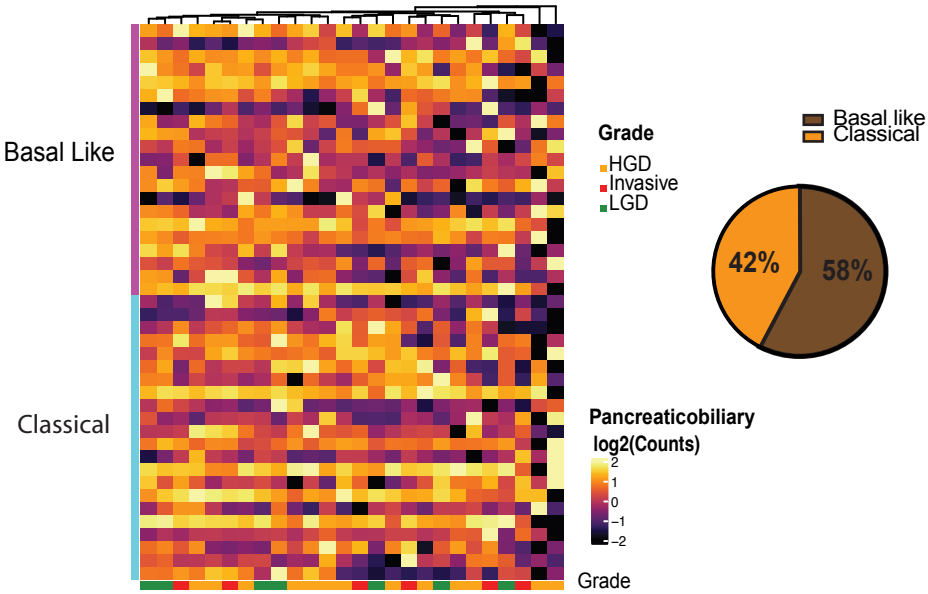
A

PB - Epithelial



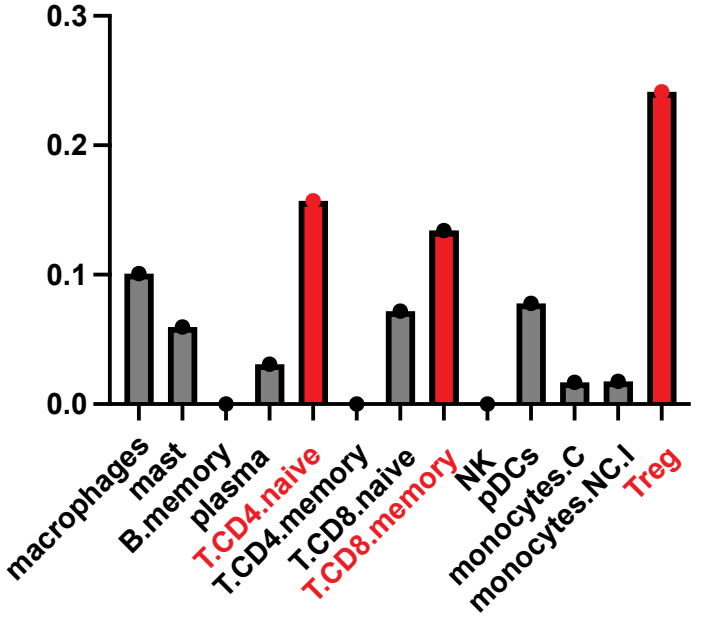
B

PB - Basal-like versus Classical



C

PB - Immune deconvolution analysis



D

PB - iCAF versus myCAF

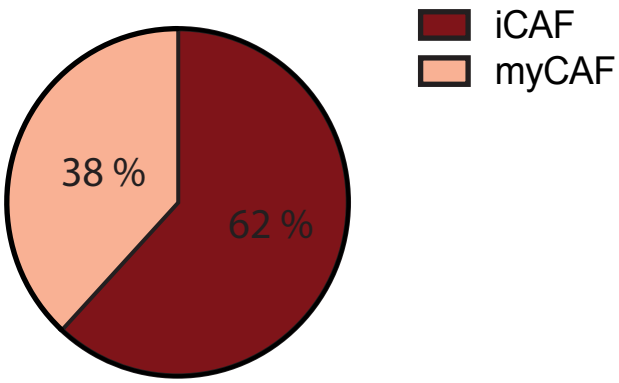
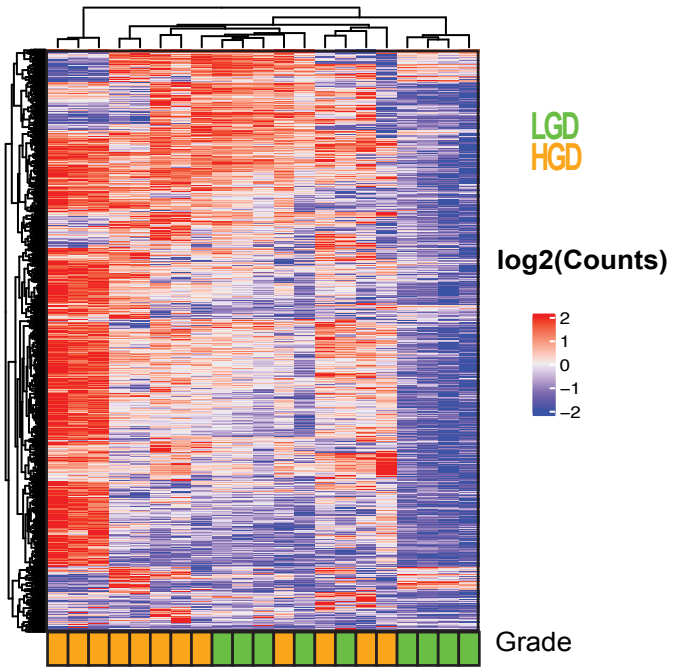


Figure 3

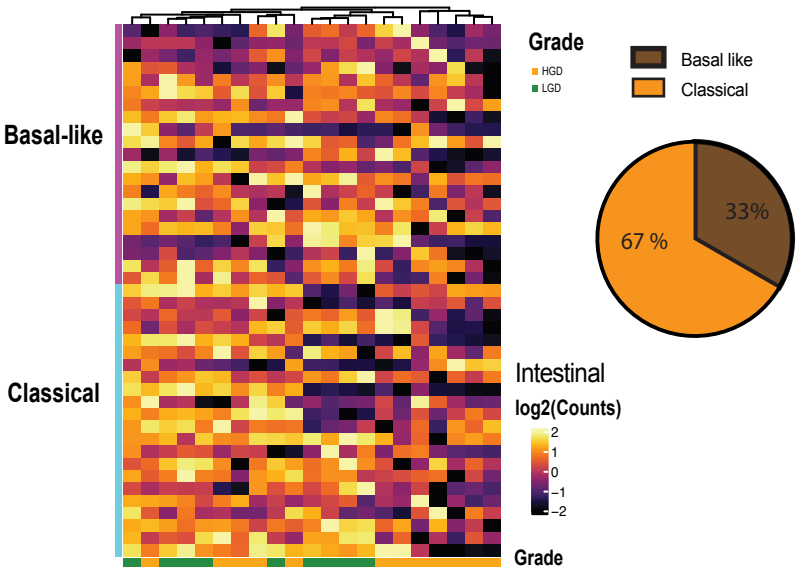
A

Intestinal - Epithelial



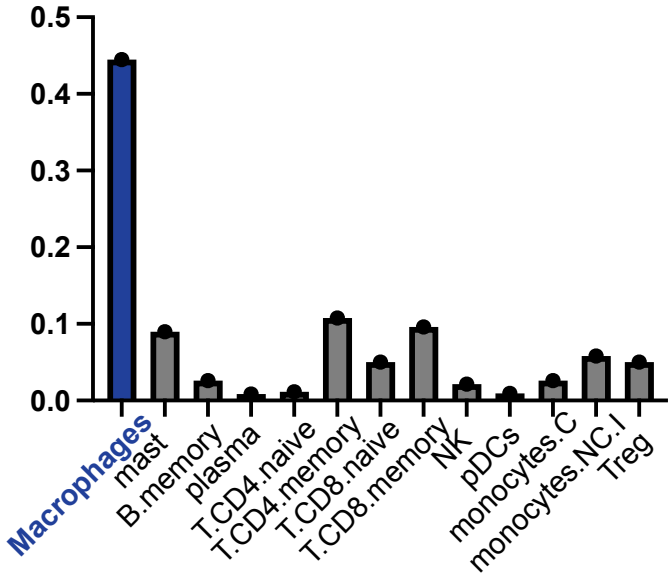
B

Intestinal - Basal-like versus Classical



C

Intestinal - Immune deconvolution analysis



D

Intestinal - iCAF versus myCAF

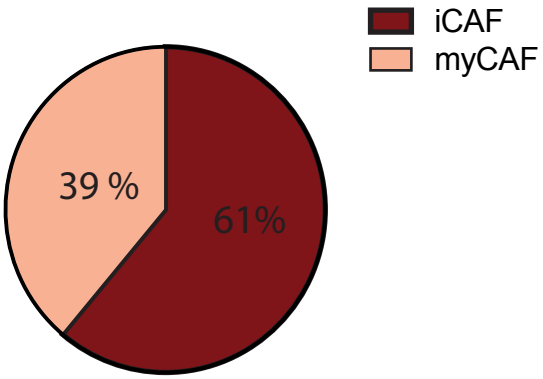
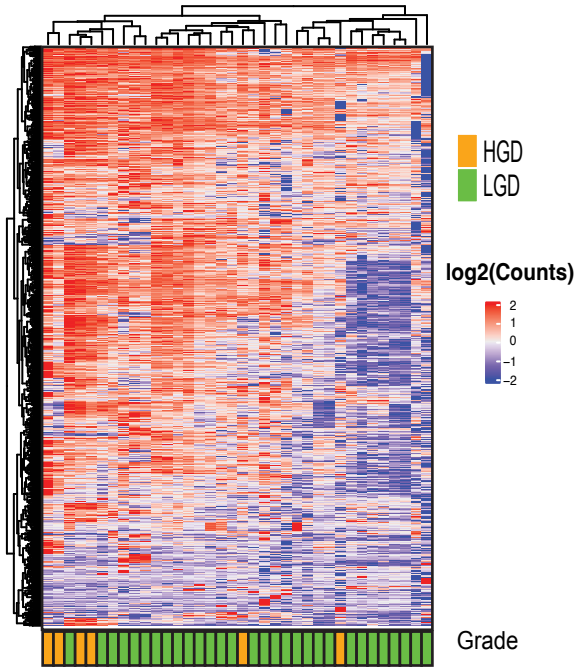
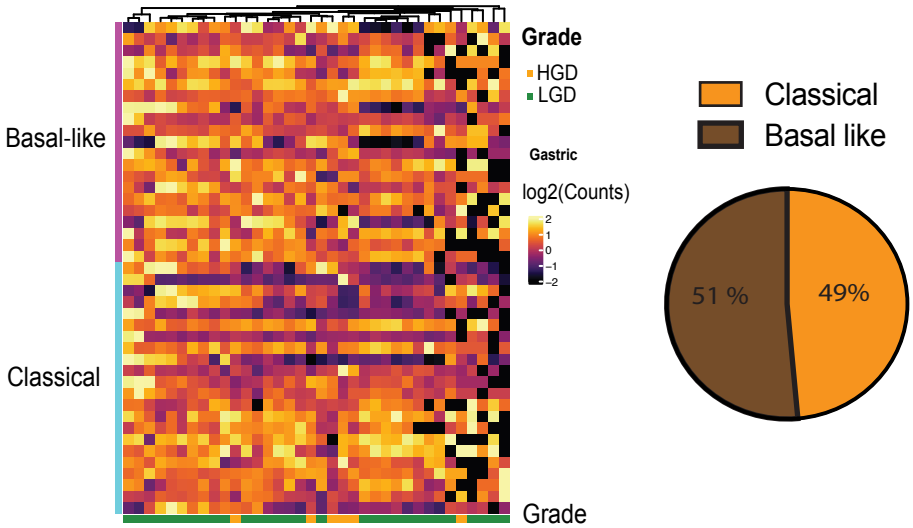


Figure 4

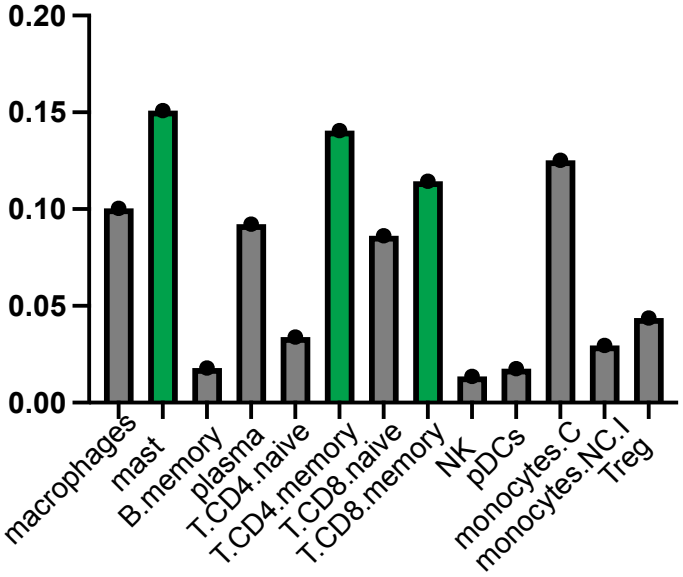
A Gastric - Epithelial



B Gastric - Basal-like versus Classical



C Gastric - Immune deconvolution analysis



D Gastric - iCAF versus myCAF

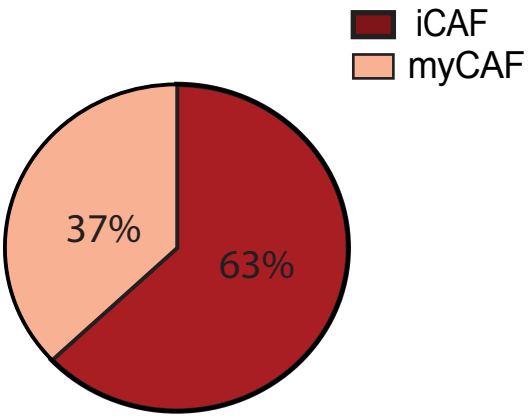


Figure 5

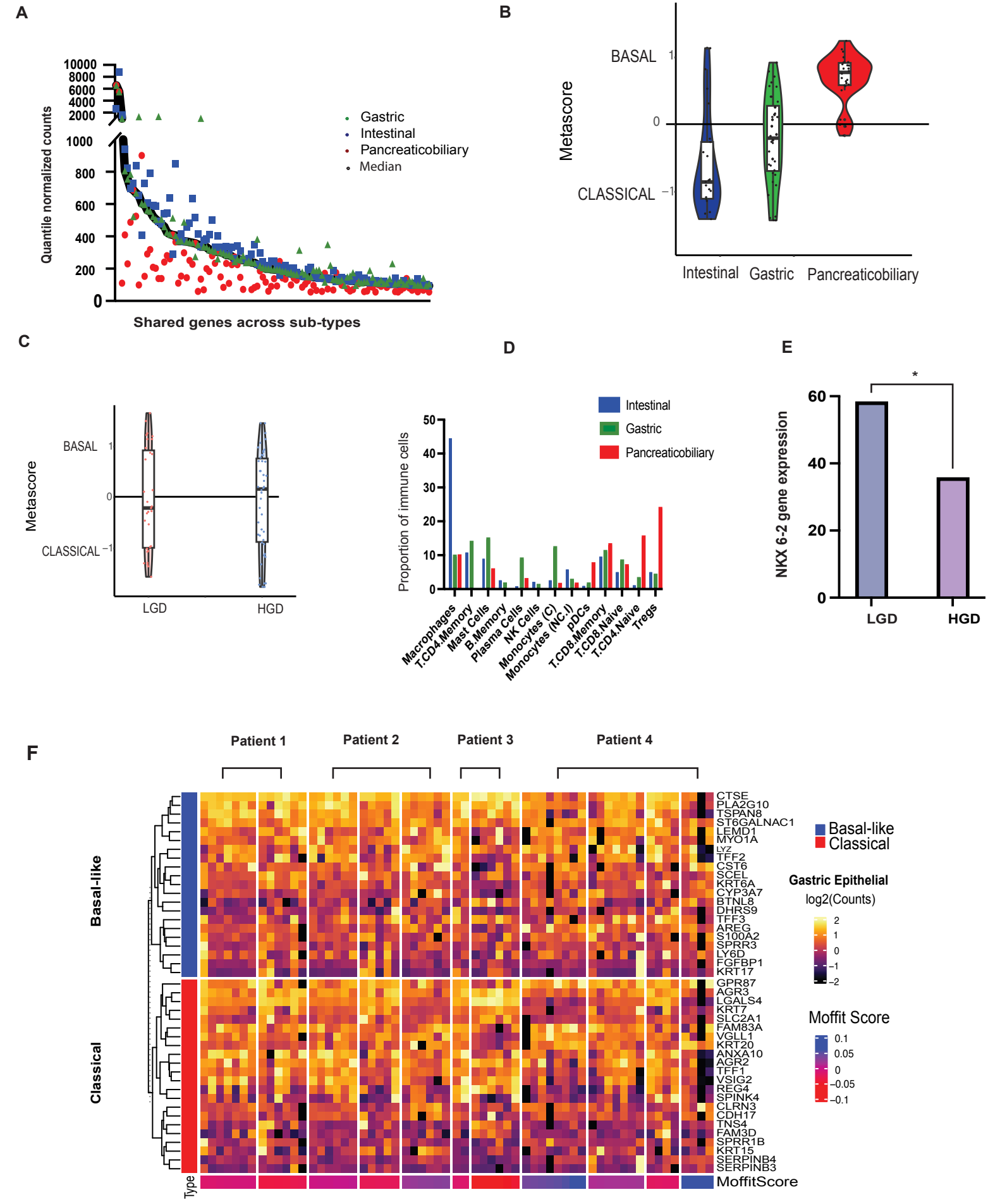
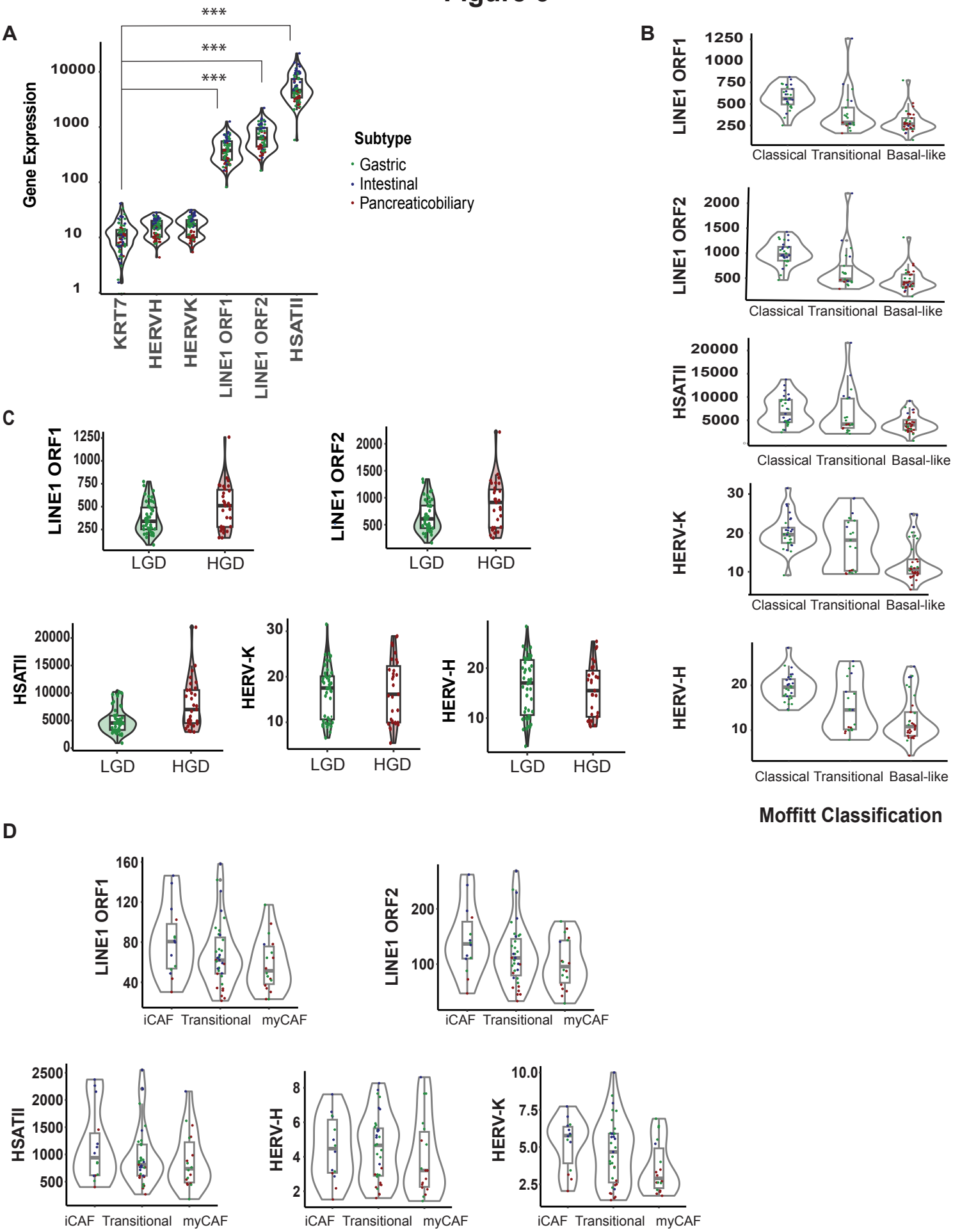
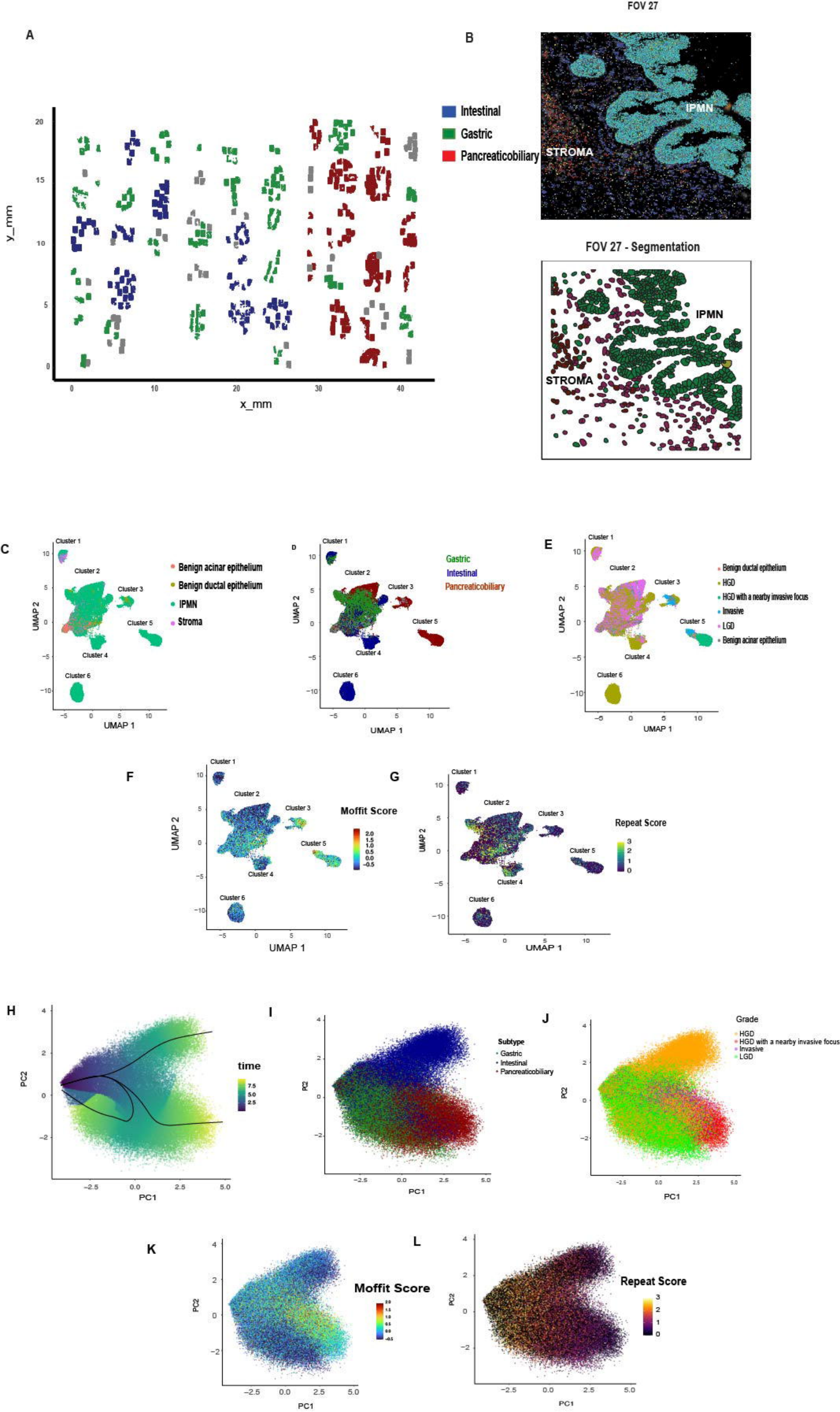
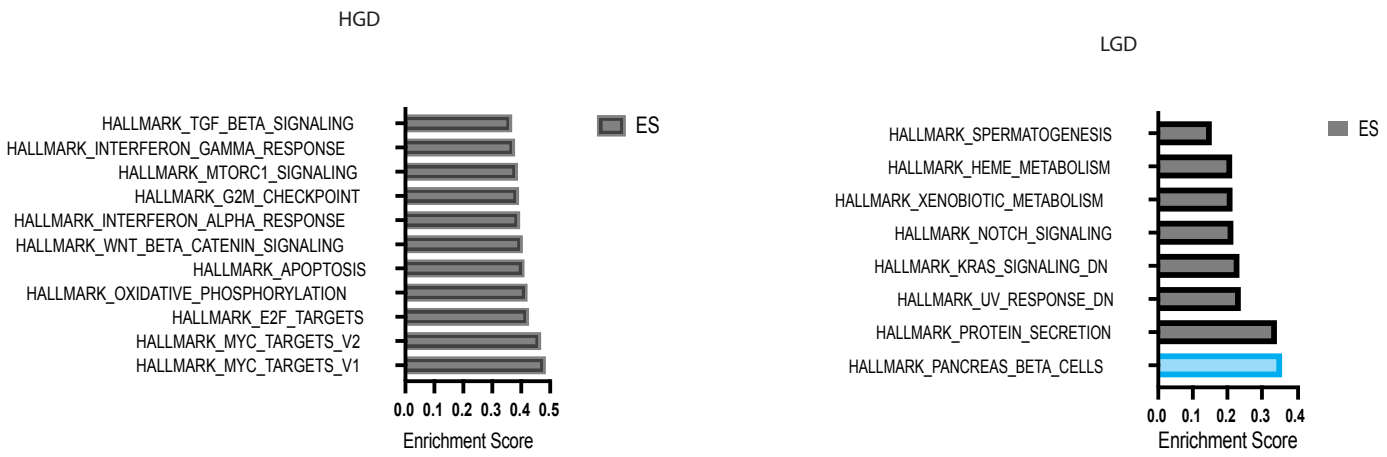
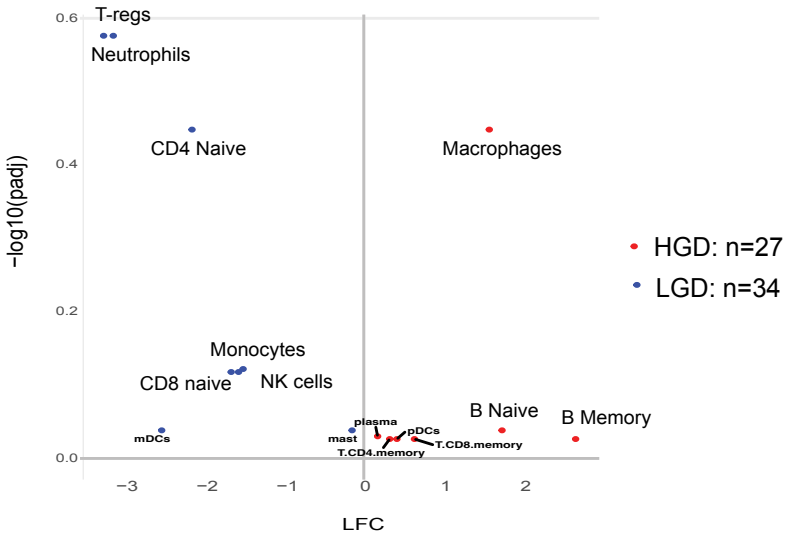

A
Quantile normalized counts
Shared genes across sub-types
Gastric
Intestinal
Pancreaticobiliary
Median
B
Metascore
BASAL
CLASSICAL
Intestinal
Gastric
Pancreaticobiliary
C
Metascore
BASAL
CLASSICAL
LGD
HGD
D
Proportion of immune cells
Intestinal
Gastric
Pancreaticobiliary
Macrophages
T.CD4.Memory
Mast Cells
B.Memory
Plasma Cells
NK Cells
Monocytes (C)
Monocytes (NC.I)
pDCs
T.CD8.Memory
T.CD8.Naive
T.CD4.Naive
Tregs
E
NKX 6-2 gene expression
LGD
HGD
F
Patient 1
Patient 2
Patient 3
Patient 4
Basal-like
Classical
CTSE
PLA2G10
TSPAN8
ST6GALNAC1
LEMD1
MYO1A
LYZ
TFF2
CST6
SCEL
KRT6A
CYP3A7
BTNL8
DHRS9
TFF3
AREG
S100A2
SPRR3
LY6D
FGFBP1
KRT17
GPR87
AGR3
LGALS4
KRT7
SLC2A1
FAM83A
VGLL1
KRT20
ANXA10
AGR2
TFF1
VSIG2
REG4
SPINK4
CLRN3
CDH17
TNS4
FAM3D
SPRR1B
KRT15
SERPINB4
SERPINB3
Type
MoffitScore
Basal-like
Classical
Gastric Epithelial
log2(Counts)
Moffit Score

Figure 6


A
Gene Expression
10000
1000
100
10
1
***
***
***
KRT7
HERVH
HERVK
LINE1 ORF1
LINE1 ORF2
HSATII
Subtype
Gastric
Intestinal
Pancreaticobiliary
B
LINE1 ORF1
LINE1 ORF2
HSATII
HERV-K
HERV-H
Classical Transitional Basal-like
Moffitt Classification
C
LINE1 ORF1
LINE1 ORF2
HSATII
HERV-K
HERV-H
LGD
HGD
D
LINE1 ORF1
LINE1 ORF2
HSATII
HERV-H
HERV-K
iCAF Transitional myCAF


FOV 27
A
B
Intestinal
Gastric
Pancreaticobiliary
IPMN
STROMA
FOV 27 - Segmentation
y_mm
x_mm
C
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Benign acinar epithelium
Benign ductal epithelium
IPMN
Stroma
UMAP 1
UMAP 2
D
Gastric
Intestinal
Pancreaticobiliary
E
Benign ductal epithelium
HGD
HGD with a nearby invasive focus
Invasive
LGD
Benign acinar epithelium
F
Moffit Score
G
Repeat Score
H
time
PC1
PC2
I
Subtype
Gastric
Intestinal
Pancreaticobiliary
J
Grade
HGD
HGD with a nearby invasive focus
Invasive
LGD
K
Moffit Score
L
Repeat Score

Supplementary Figure 1



Supplementary figure 2