

Use of spatial transcriptomics to identify molecular features associated with African American heritage in pancreatic cancer

Yuesong Wu¹, Aoqi Xie¹, Daniel Sálas-Escabillas², Ian Loveless², Madison George², Kendyll Gartrelle², Julie Clark², David Kwon², Brian Theisen³, Ralph Francescone², Débora Barbosa Vendramini Costa², Howard Crawford², Yuehua Cui¹, Ling Huang², Nina Steele²

¹-Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, ²-Henry Ford Pancreatic Cancer Center, Department of Surgery, Henry Ford Health, Detroit, MI 48202, ³-Department of Pathology, Henry Ford Health, Detroit, MI 48202

Background

- Pancreatic ductal adenocarcinoma (PDAC) is a deadly malignancy with a 5-year survival rate of only 12% (SEER).
- Black or African American patients (BAA) have a 20% higher incidence of pancreatic cancer, and recent studies suggest that there are biological factors that may influence this disparity, in addition to complex socioeconomic factors.
- Pancreatic tumors are highly enriched with stroma and exhibit significant inter- and intra-patient morphological and molecular heterogeneity.
- To improve the care for all patients, we need to better understand the disease mechanisms by comprehensively mapping different spatial locations within individual patients' tumors and determining whether there are common patterns of pathogenesis across diverse patient populations.
- In this study, we utilized the 10X Visium spatial transcriptomics platform on eight treatment-naïve primary pancreatic tumors.

Visium spatial transcriptomics and bioinformatics workflow

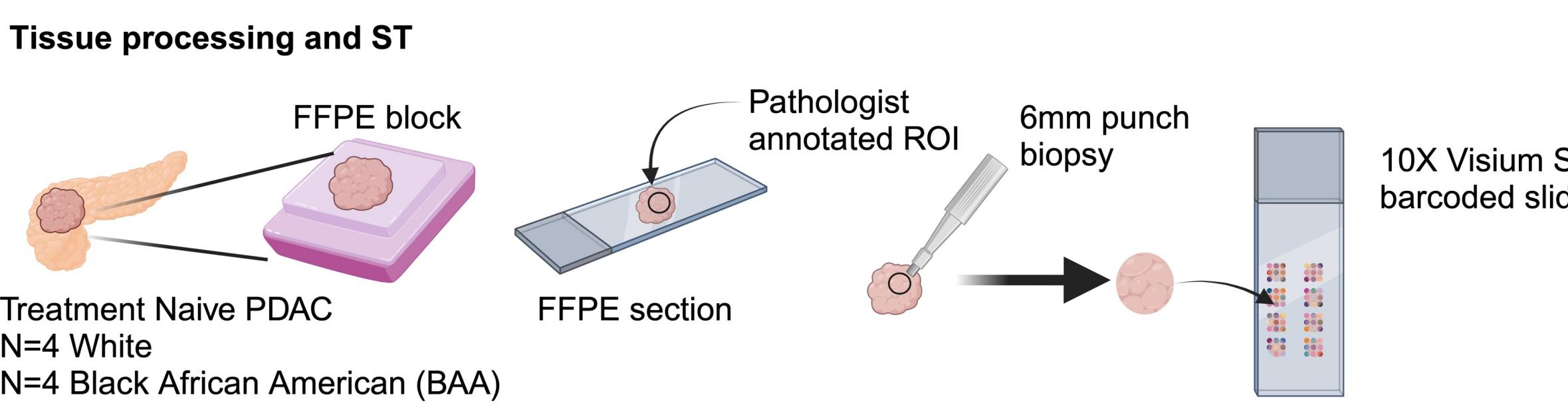


Figure 1. Workflow for spatial transcriptomics of human PDAC FFPE tissues.
Schematic diagrams describe procedures of tissue processing, spatial barcoding with 10X Visium platform and bioinformatic pipelines for spatial deconvolution and identification of differentially expressed genes.

Identification of malignant tumor boundary in PDAC

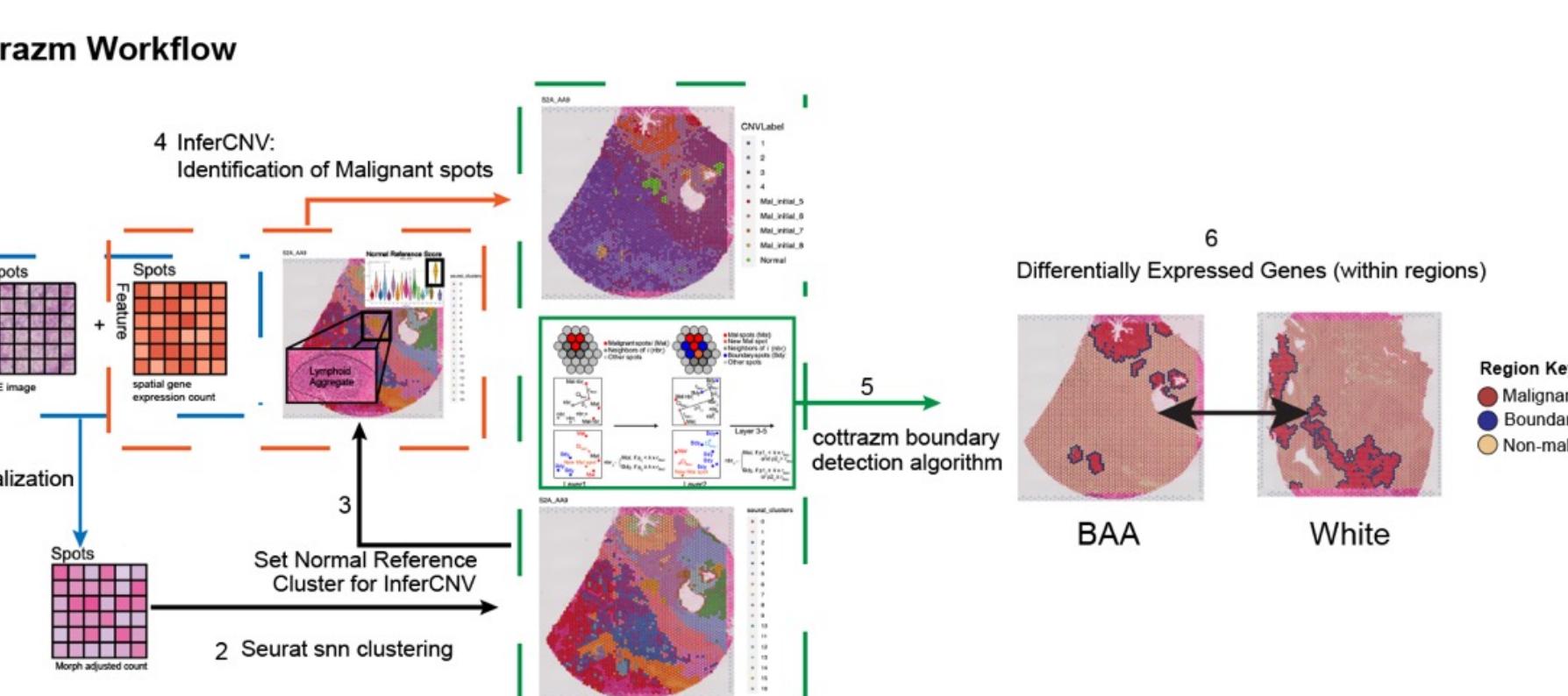


Figure 2. Data processing procedures in Cottrazm method. The stLearn package integrates spatial distances, tissue morphology and gene expression measurements from spatial transcriptomic data to normalize gene expression and obtain morphologically adjusted gene expression matrix. Normalized spots were clustered using the Seurat's shared nearest neighbor module. The inferCNV uses RNA expression to estimate copy number variations (CNV) in each spot and classifies tissues into malignant and non-malignant regions based on inferred CNVs. Normalized gene expressions of malignant and non-malignant regions from Black and White patients were compared for identification of differentially expressed genes (Wilcoxon test).

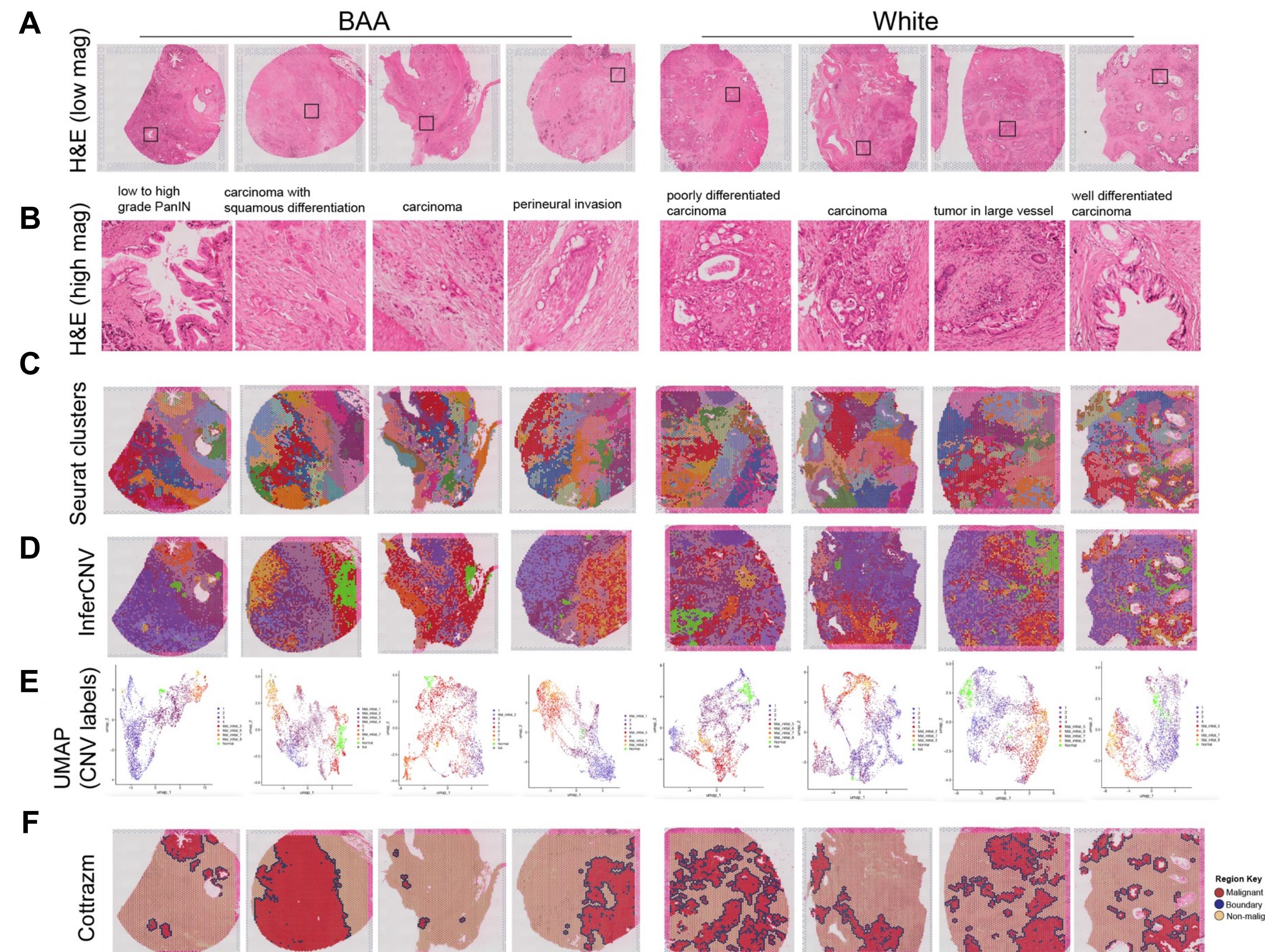


Figure 3. Tissue segmentation based on histological morphology and gene expression. (A) H&E morphology of tumor tissues used in our study. (B) Typical histological features of tissues used in our study. (C) Tissue segmentation determined by gene expression analyzed in Seurat. (D) Frequencies of gene copy number variation in tissues estimated by InferCNV. (E) UMAPs with InferCNV labels indicating putative malignant clusters. (F) Putative malignant and non-malignant regions determined by Cottrazm. Note the Cottrazm utilizes both spatial and transcriptomic information to determine malignancy.

Single cell RNAseq human atlas of PDAC (N=61 patients) for reference-based ST deconvolution

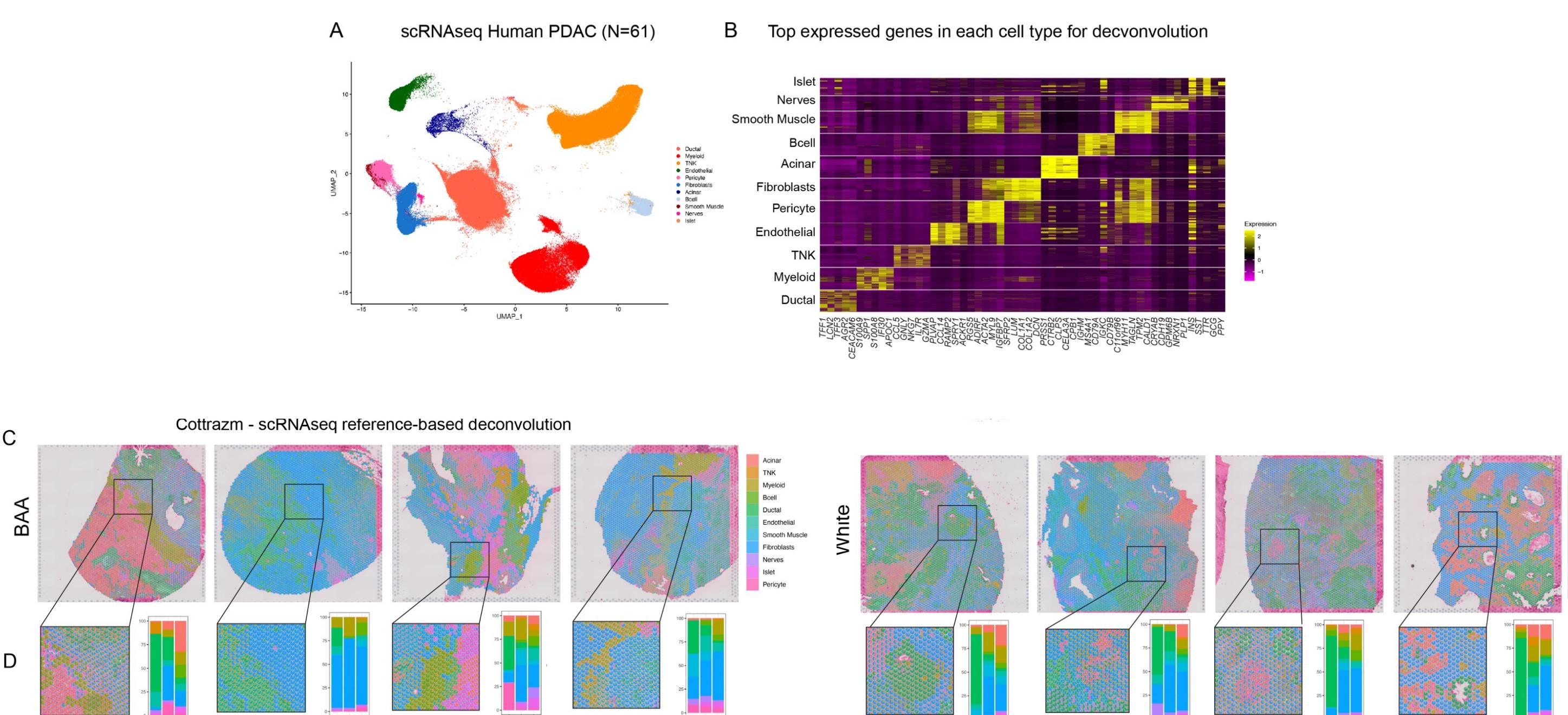


Figure 4. Reference based spatial deconvolution using Cottrazm. (A) Single cell transcriptomes of normal and cancerous pancreatic tissues from 61 subjects. This database was used as reference and to derive cell identity markers as shown in (B). (C) Spatial distributions of multiple cell types in pancreatic tumors from Black and White patients. Color codes for cell types are shown in the labels on the right. (D) Enlarged view of regions showing spatial localizations of different types of cells. Stacked bar charts depict the percentages of different cell types in the malignant (Mal), border (Bdy) and non-malignant (nMal) regions within individual patient's tumor tissues.

Classical and Basal spot scoring of PDAC spatial data

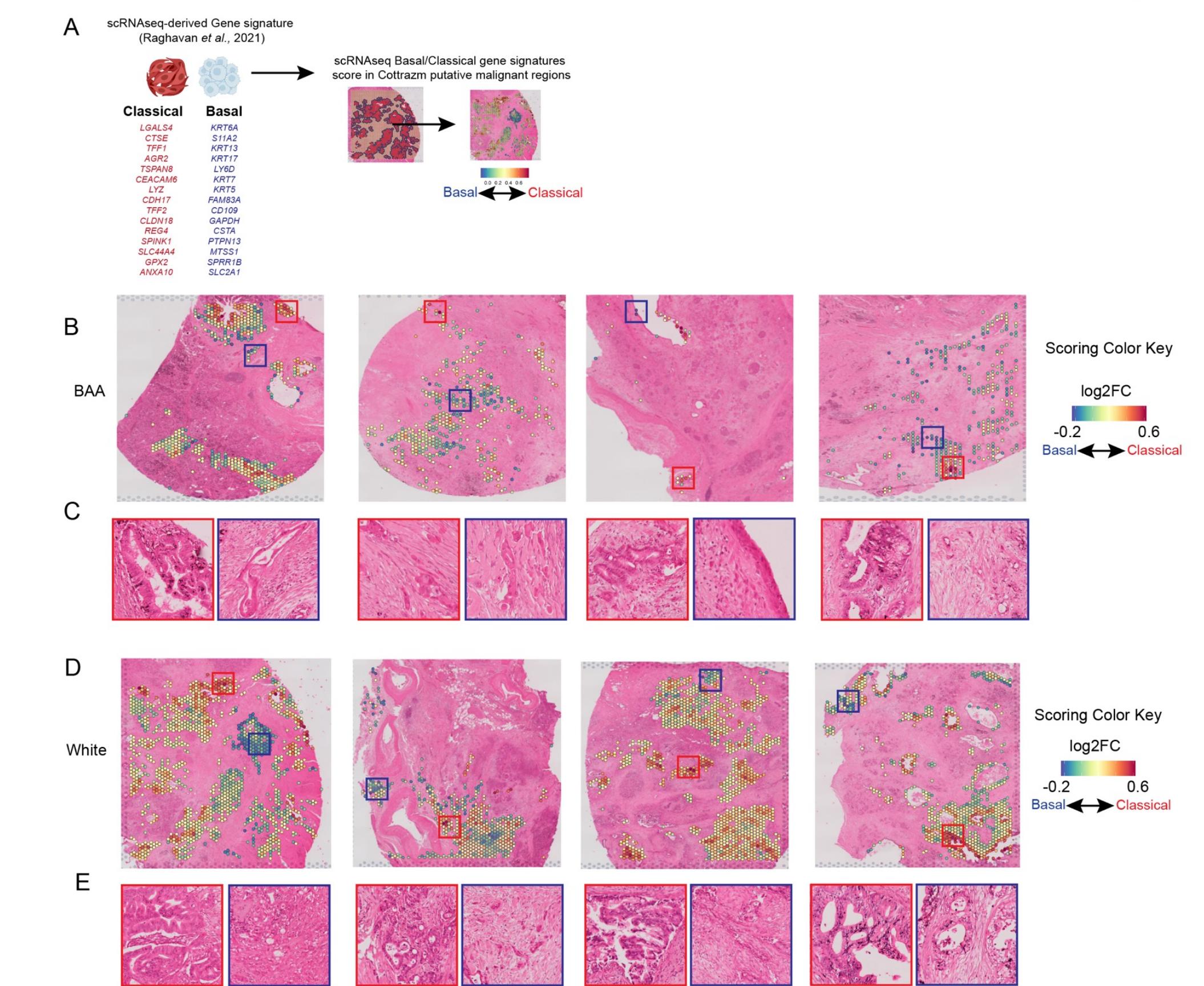


Figure 5. Tumor subtype spot scoring and histology in putative malignant regions. (A) Schematic of classical vs basal scoring of ST data. The scRNAseq classical and basal gene signatures from Raghavan et al were applied to putative malignant spots in Black (B-D) and White (D-E). Classical is indicated in Red and basal subtype in Blue. Log2 Fold change normalized values were used to plot classical versus basal scores within malignant spots.

Gene expression differences in Black and White patients

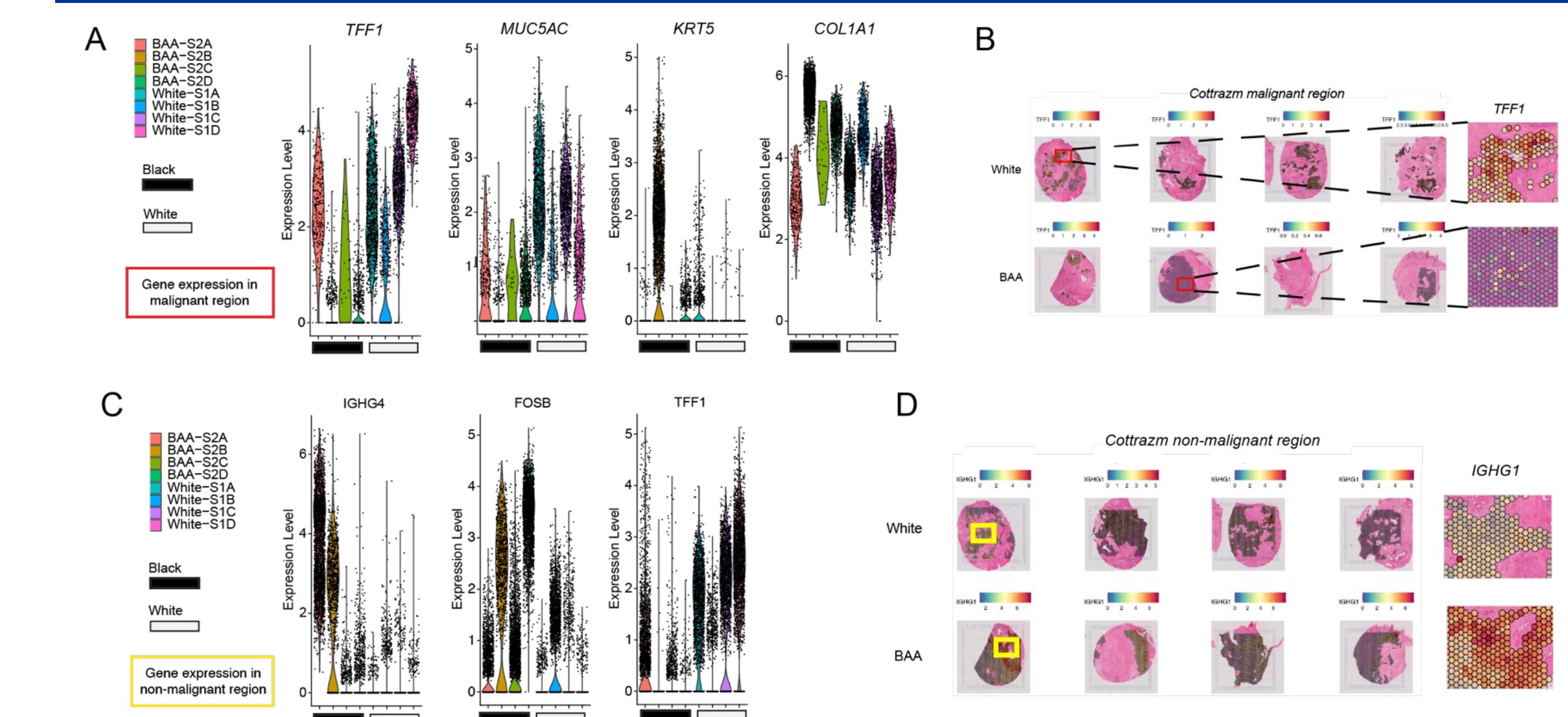
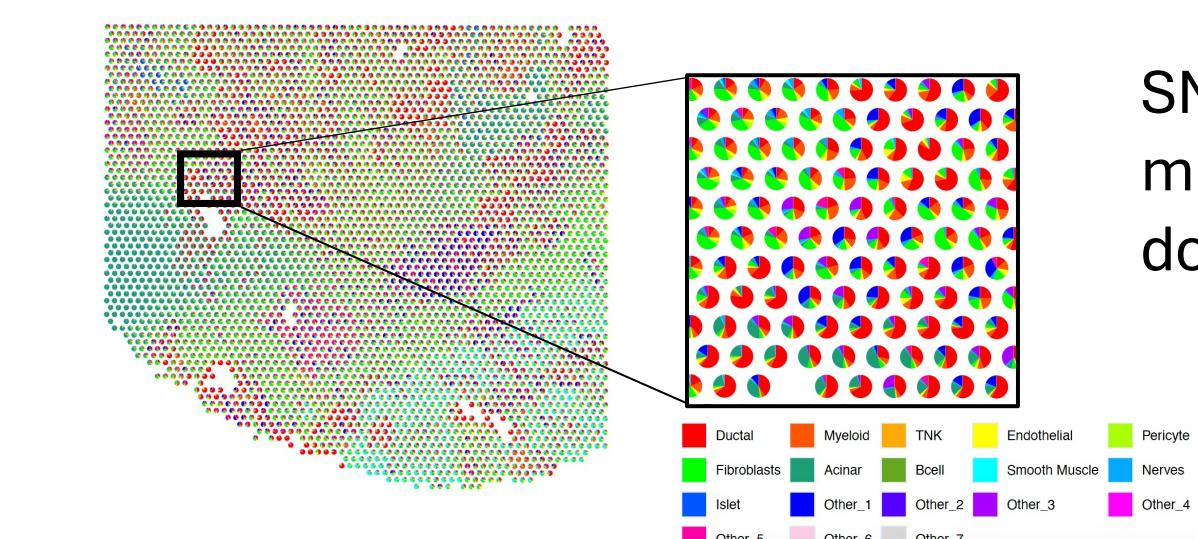


Figure 6. Genes differentially expressed in pancreatic tumor tissues from Black and White patients. (A) Examples of genes enriched in the malignant regions of Black (*KRT5*, *COL1A1*) and White (*TFF1*, *MUC5AC*) patients' tumors. Violin plots depict target gene expressions in spots of individual patients. Labels on the left indicate individual patients (color squares) or groups (black and white bars) in the violin plots. (B) Expression of *TFF1* in malignant regions of tumor tissues from Black and White patients. (C) and (D) show genes enriched in non-malignant regions in Black patient's tumors. Labels and layouts are similar to (A) and (B). (D) Expression of *IGHG1* in non-malignant regions of tumor tissues from Black and White patients.

Future Directions/Ongoing work



SMART-reference free deconvolution using marker-gene-assisted topic models
doi: <https://doi.org/10.1101/2023.06.20.545793>