

Supplementary information

Meta-analysis of transcriptomics data identifies a secretory gene classifier for pancreatic adenocarcinoma

Indu Khatri¹, Manoj K. Bhasin^{1,2*}

Affiliations:

¹Division of IMBIO, Department of Medicine, Beth Israel Lahey Health, Harvard Medical School, Boston MA 02215

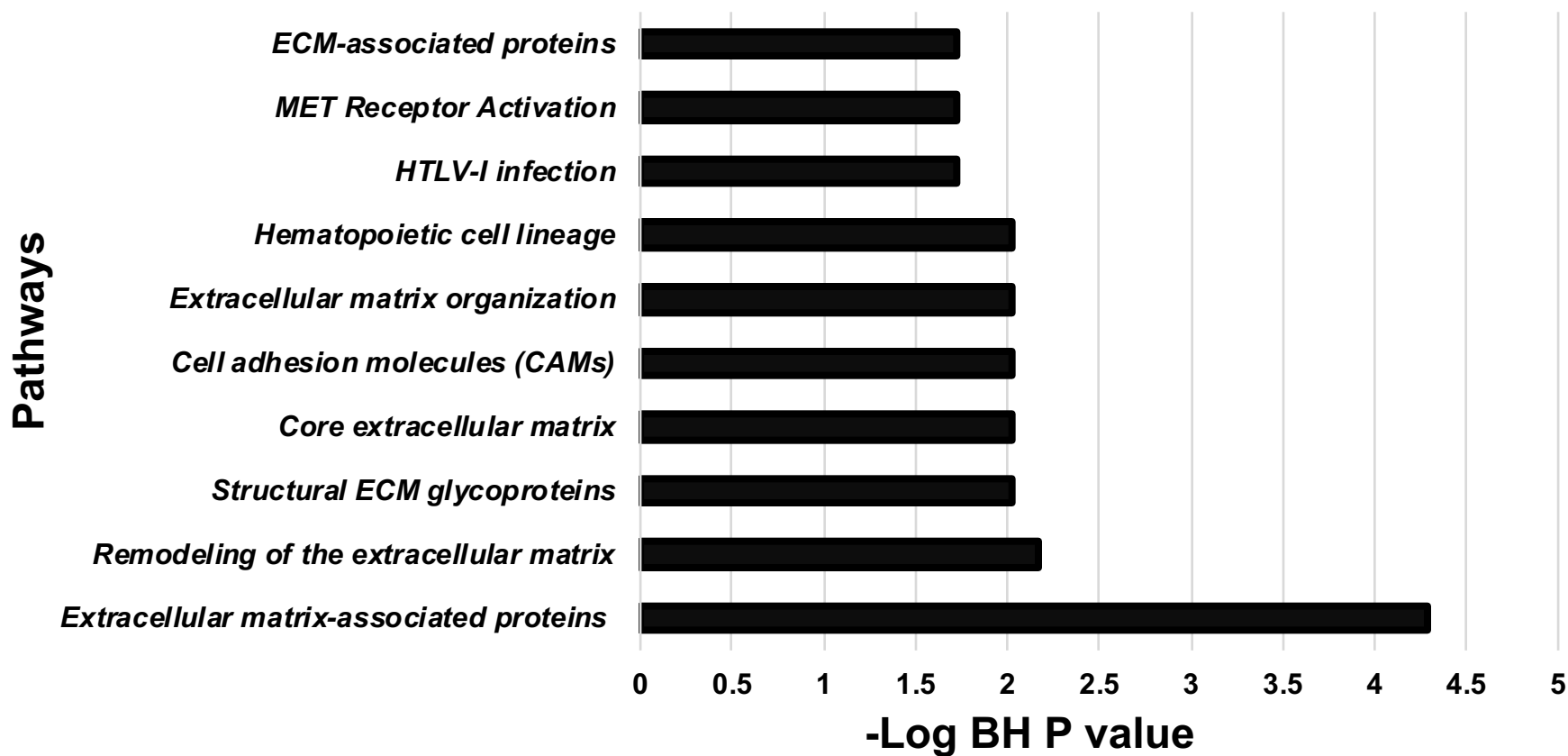
²Department of Pediatrics and Biomedical Informatics, Children Healthcare of Atlanta, Emory School of Medicine, Atlanta, GA 30322

***Corresponding Author**

Manoj K. Bhasin, PhD

E-mail: manoj.bhasin@emory.edu

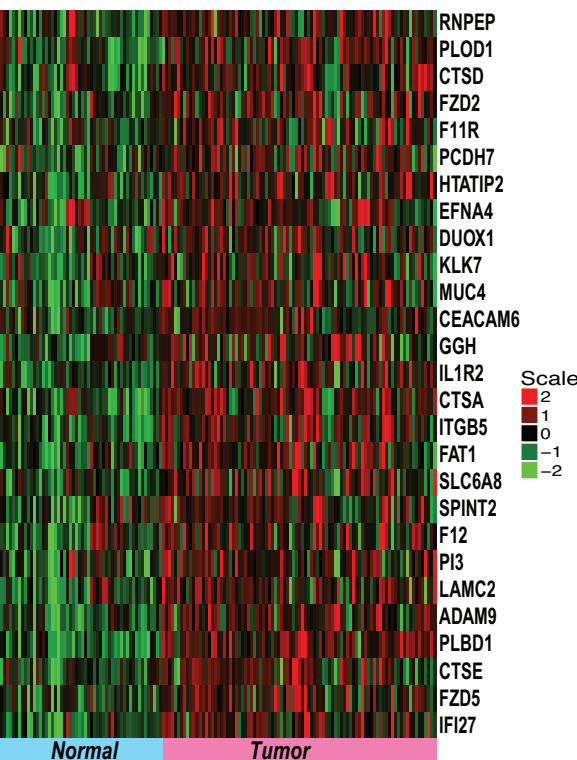
Supplementary Figure S1. Pathway enrichment analysis of the 74 PDAC-specific secretory genes.



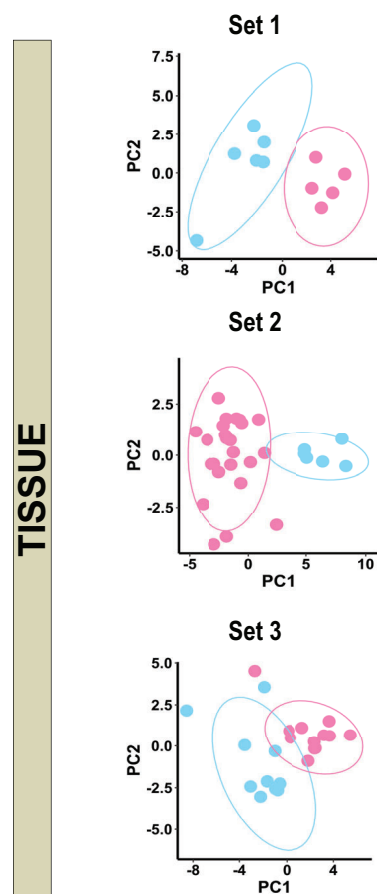
Supplementary Figure S2: Upregulated secretory genes in training datasets.

A) Heatmap of 27 upregulated secretory genes in PDAC for two of the three tissues and one of the two blood datasets. **B)** PCA plots for each training datasets using 27 upregulated secretory genes.

A

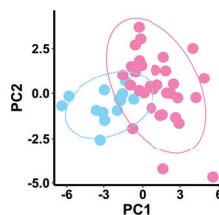


B

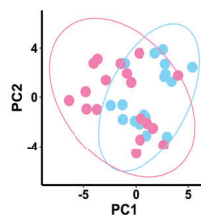


BLOOD

Set 4

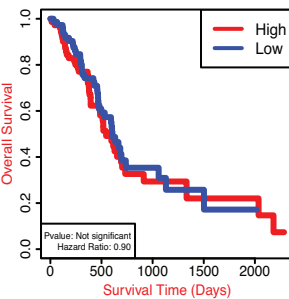


Set 5

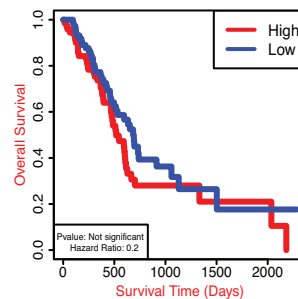


Supplementary Figure S3: Survival curve of 9-gene-based PDAC classifier and combined genes.

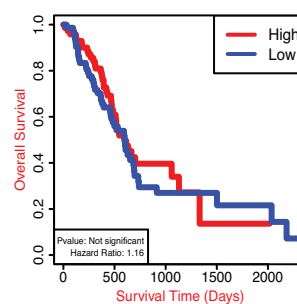
CTSA



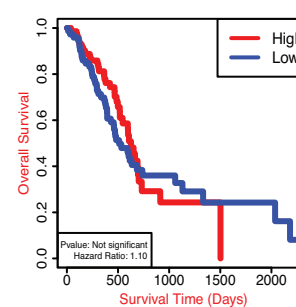
CTSD



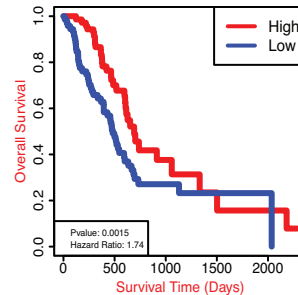
GGH



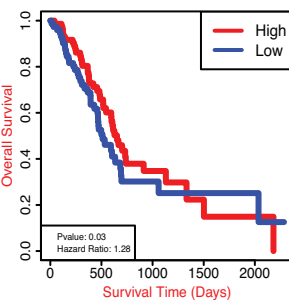
EFNA4



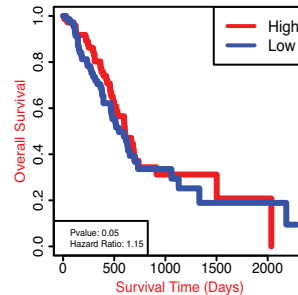
HTATIP2



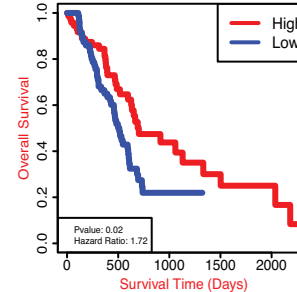
IFI27



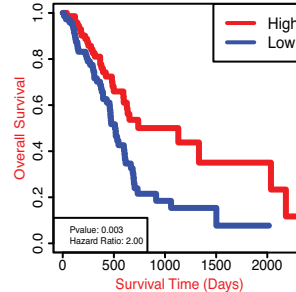
IL1R2



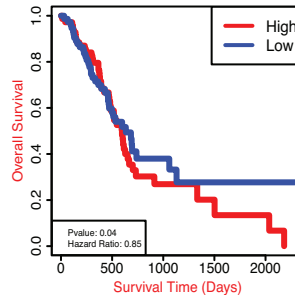
ITGB5



PLBD1

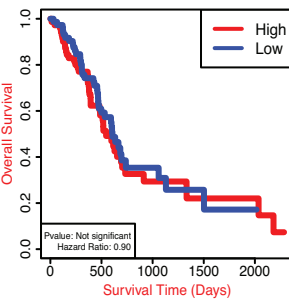


Combined Survival

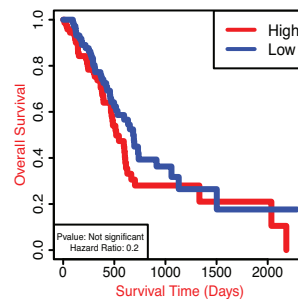


Supplementary Figure S3: Survival curve of 9-gene-based PDAC classifier and combined genes.

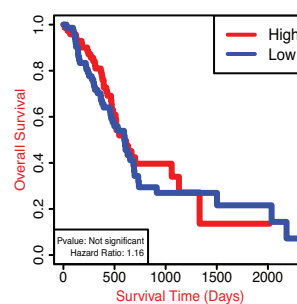
CTSA



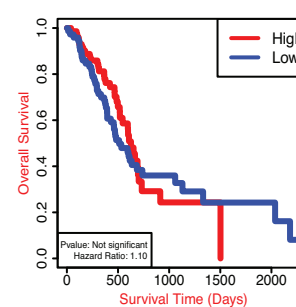
CTSD



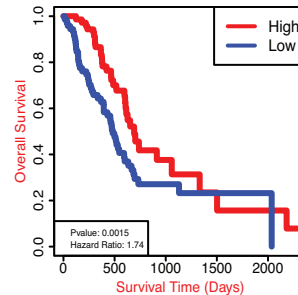
GGH



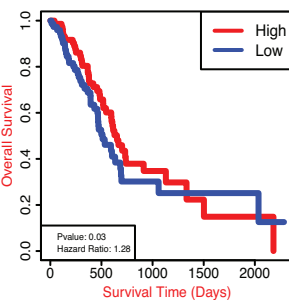
EFNA4



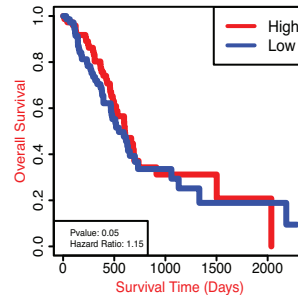
HTATIP2



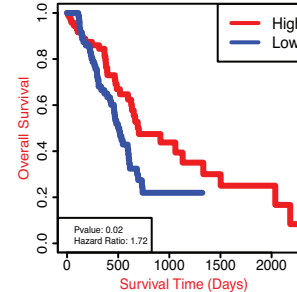
IFI27



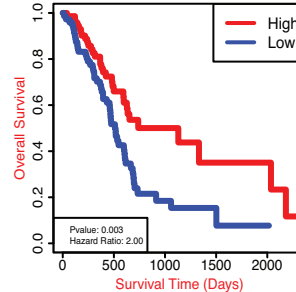
IL1R2



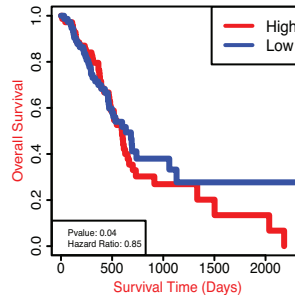
ITGB5



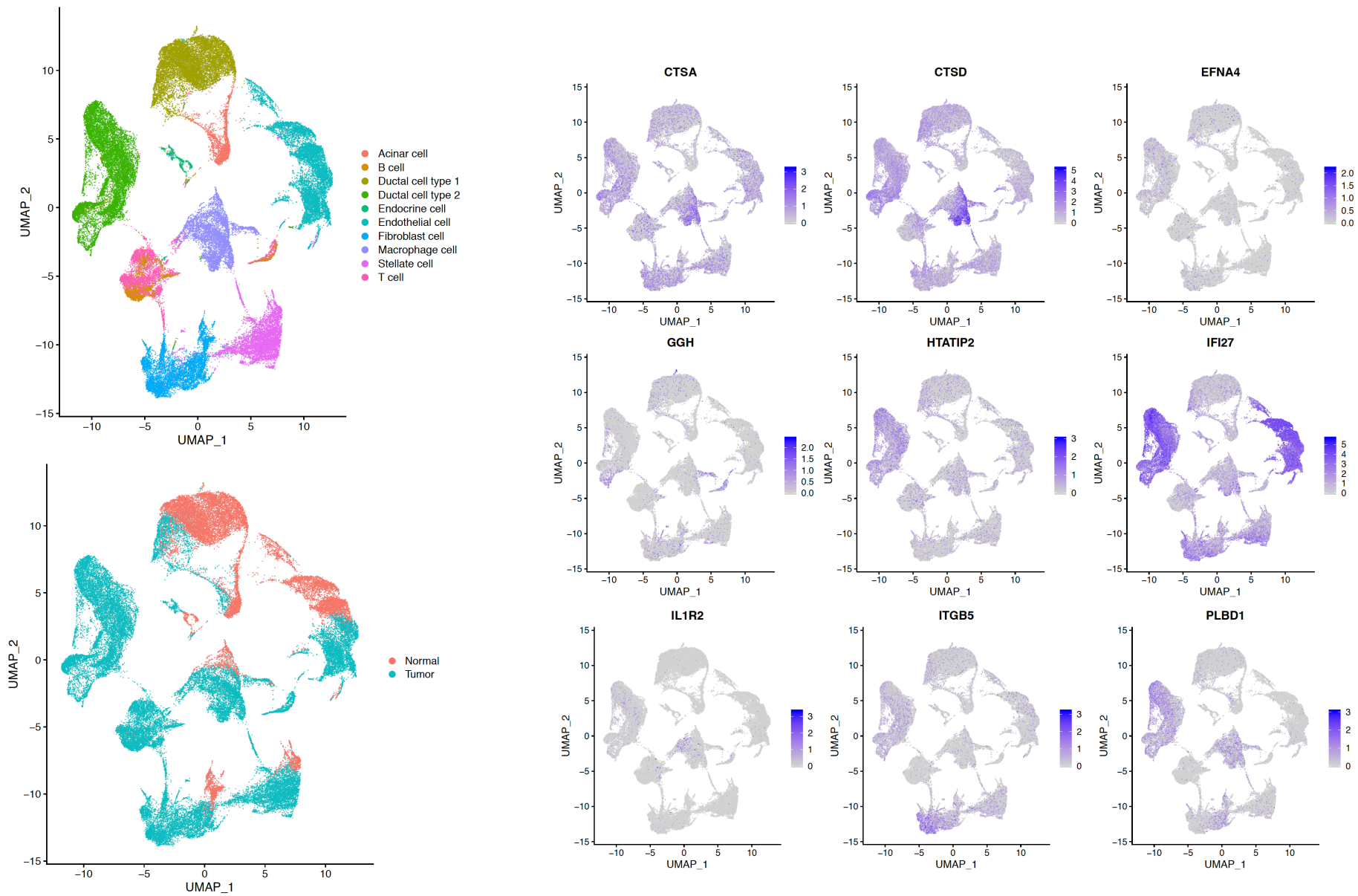
PLBD1



Combined Survival

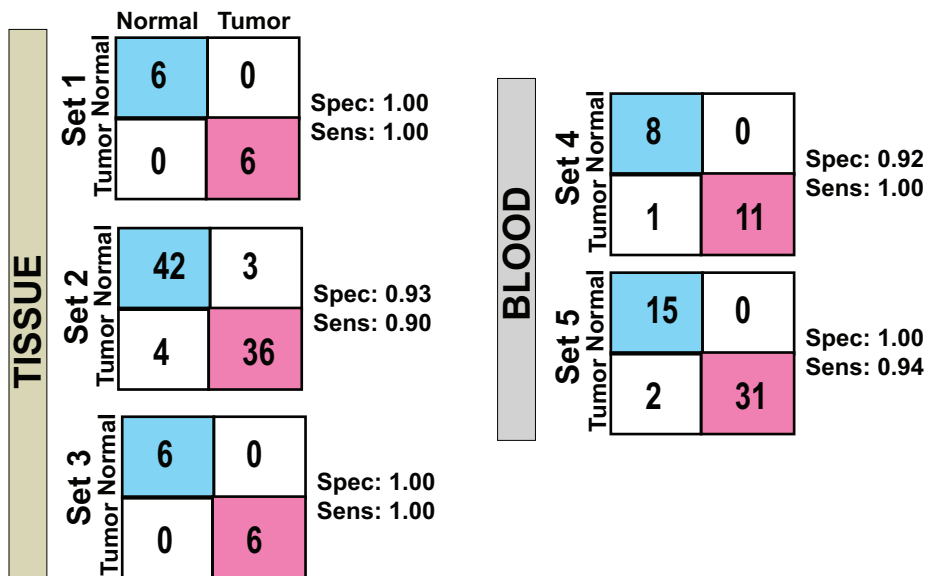


Supplementary Figure S4: Expression of 9-gene markers in different pancreas cell-types in both healthy and tumor states. The expression of these genes is high in tumor state (CTSA, CTSD, EFNA4, GGH, HTATIP2, IFI27 and ITGB5) or they are not expressed at all in healthy state (IL1R2 and PLBD1). This is also consistent with protein expression of the genes as measured by antibody staining experiments by Human protein atlas.

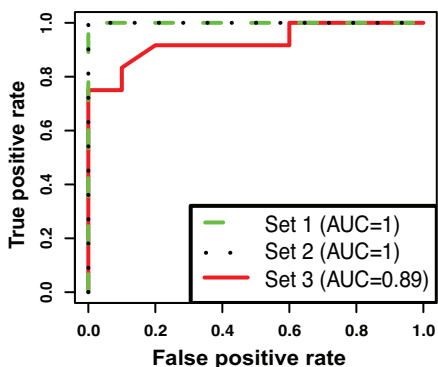


Supplementary Figure S5: Performance of 9-gene PDAC classifier on test sets using leave one out cross-validation (LOOCV). A) Diagnostic performance of the 9-gene PDAC classifier on the five training sets. Sensitivity (Sens) and Specificity (Spec) are indicated for each dataset. **B)** AUC plot for 9-gene PDAC classifier on the three tissue training datasets. **C)** AUC plot for 9-gene PDAC classifier on the two blood training datasets.

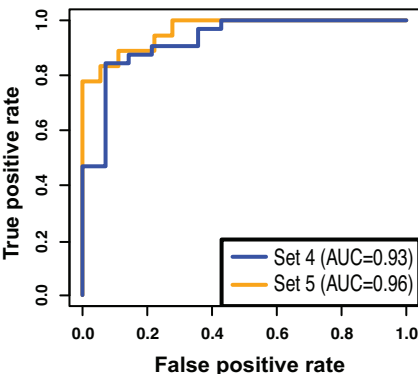
A



B

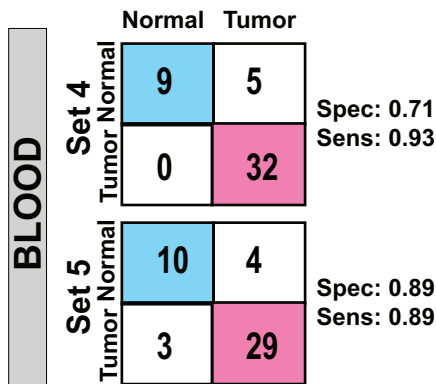
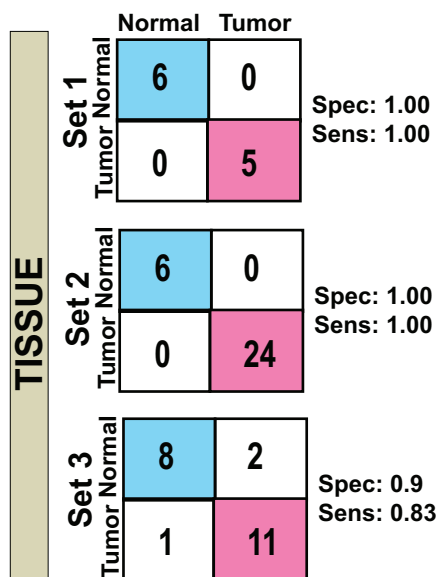


C

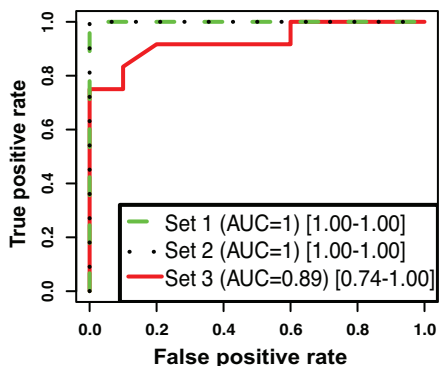


Supplementary Figure S5: Performance of 9-gene PDAC classifier on training sets using leave one out cross-validation (LOOCV). A) Diagnostic performance of the 9-gene PDAC classifier on the five training sets. Sensitivity (Sens) and Specificity (Spec) are indicated for each dataset. **B)** AUC plot for 9-gene PDAC classifier on the three tissue training datasets. **C)** AUC plot for 9-gene PDAC classifier on the two blood training datasets.

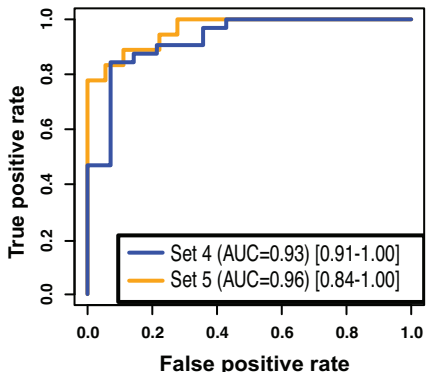
A



B

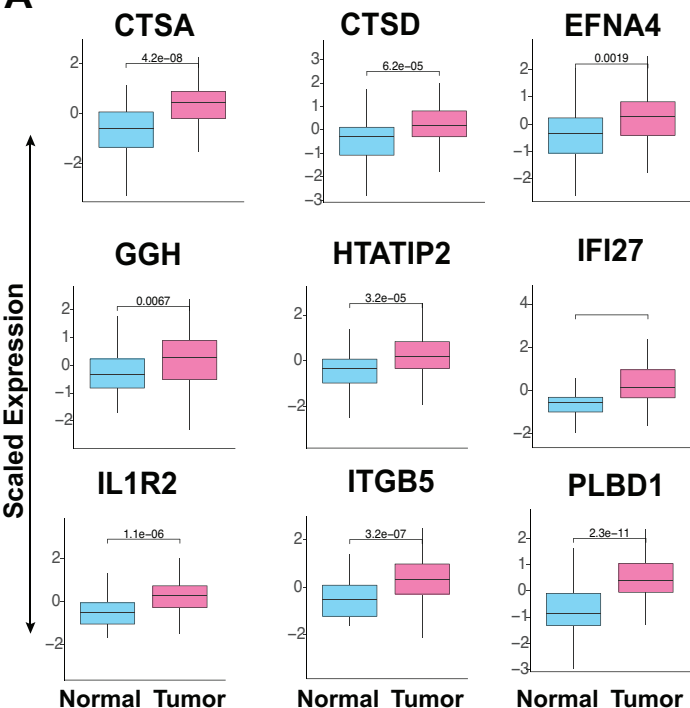


C

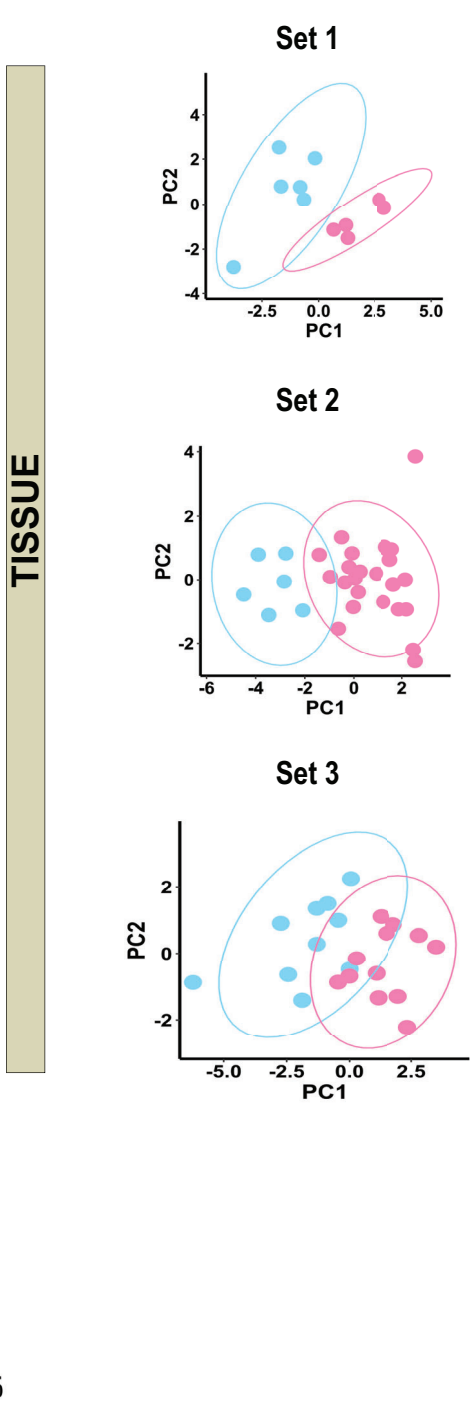


Supplementary Figure S6: The metrics for training datasets using the 9-biomarker panel genes.
A) Boxplot of the averaged expression of the genes across all the five training datasets. **B)** PCA plots for each training datasets using the 9-biomarker panel genes.

A



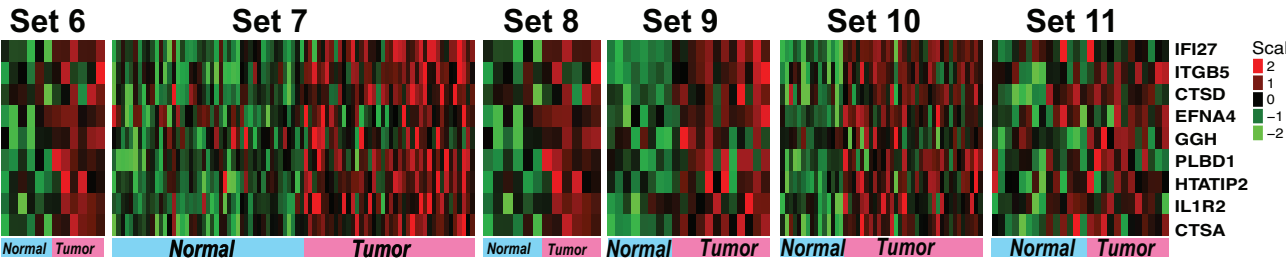
B



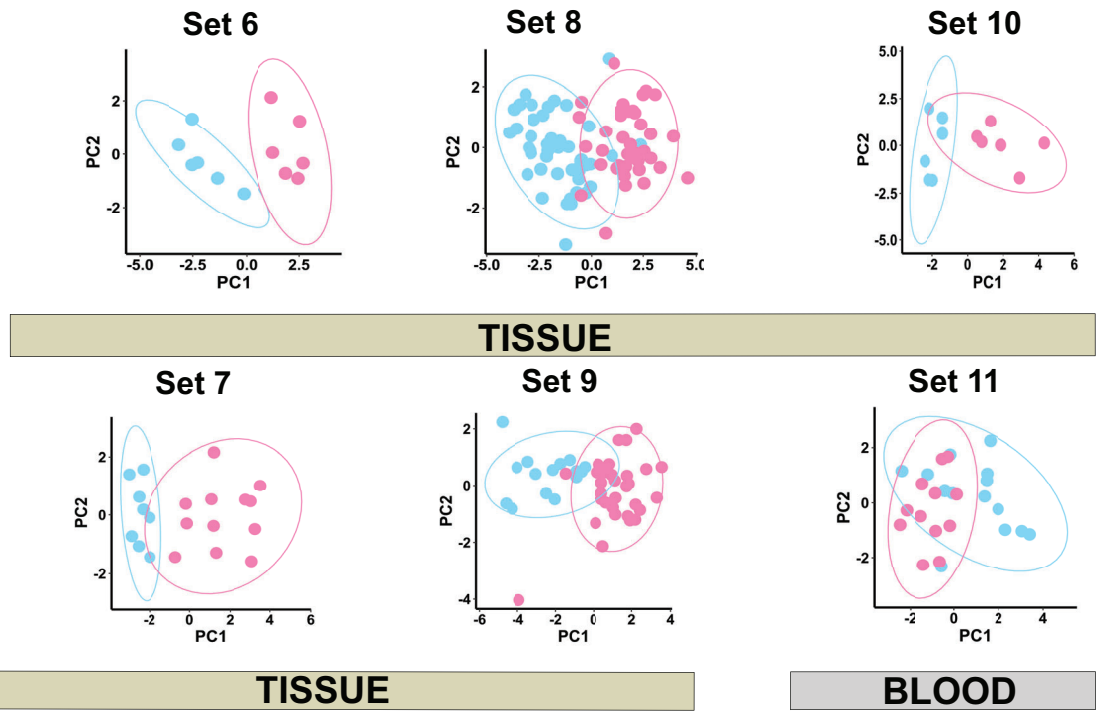
BLOOD

Supplementary Figure S7: The assessment metrics for testing datasets using the 9-biomarker panel genes. A) Heatmap of the 9 PDAC-upregulated marker genes. B) PCA plots in six independent testing datasets..

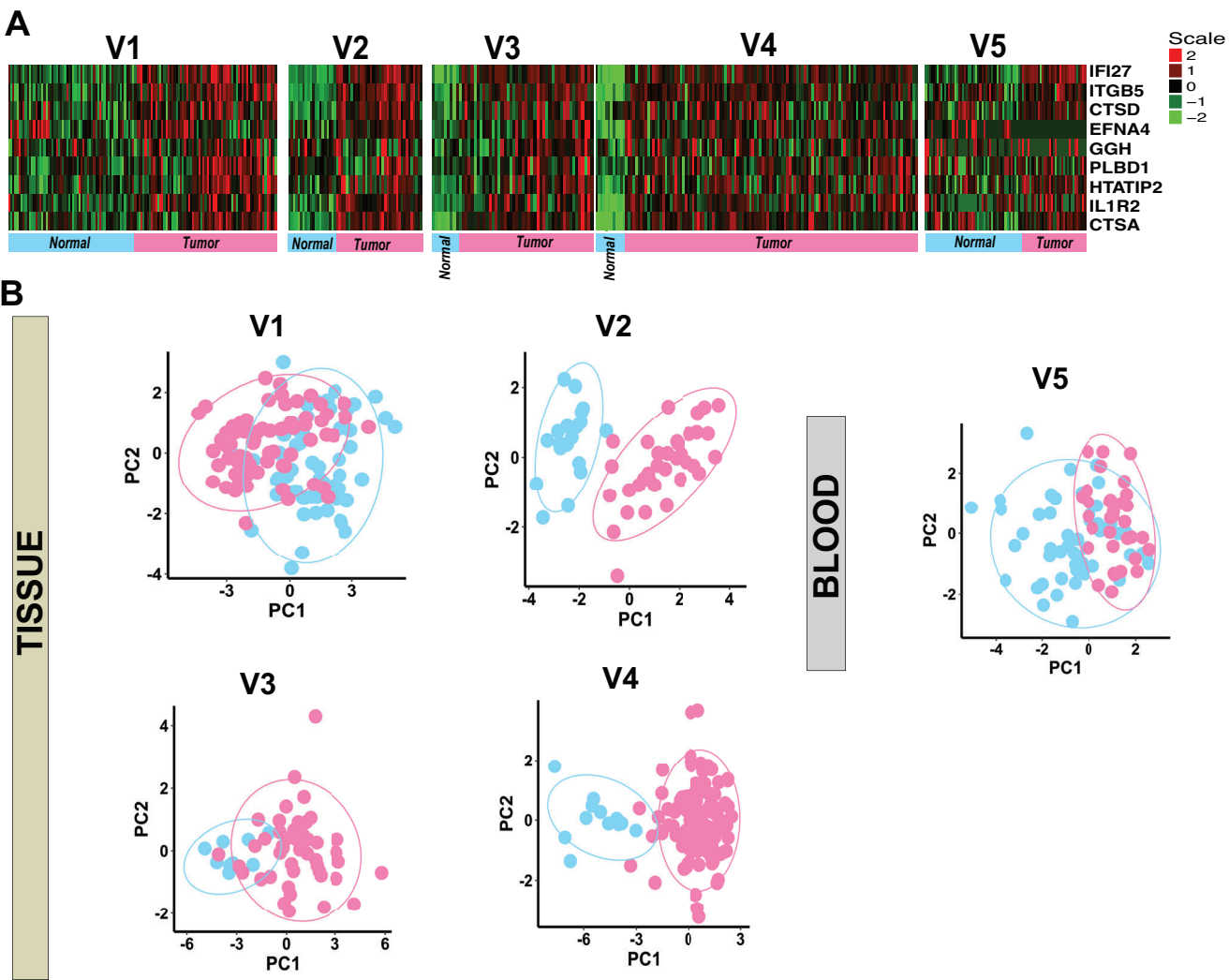
A



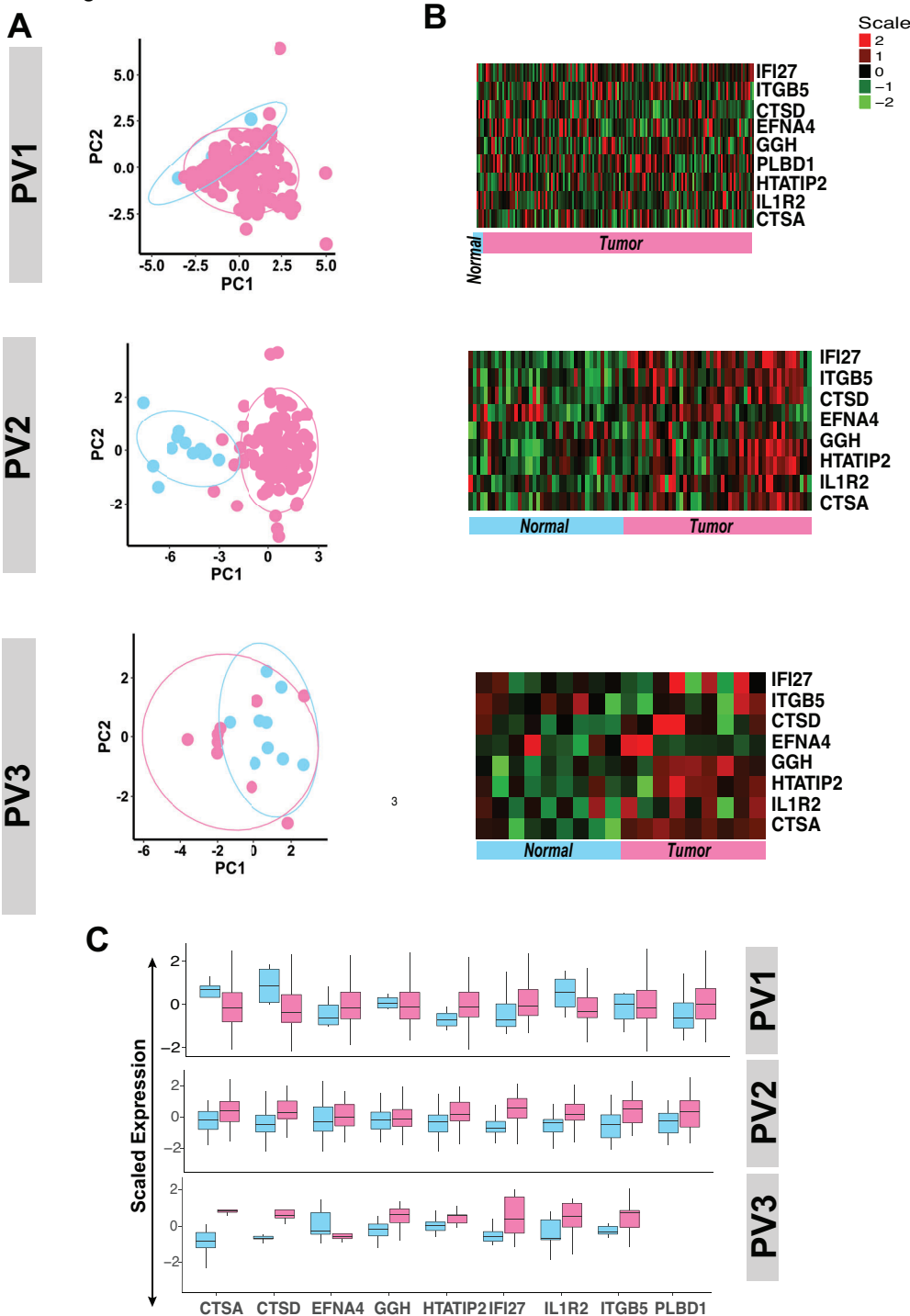
B



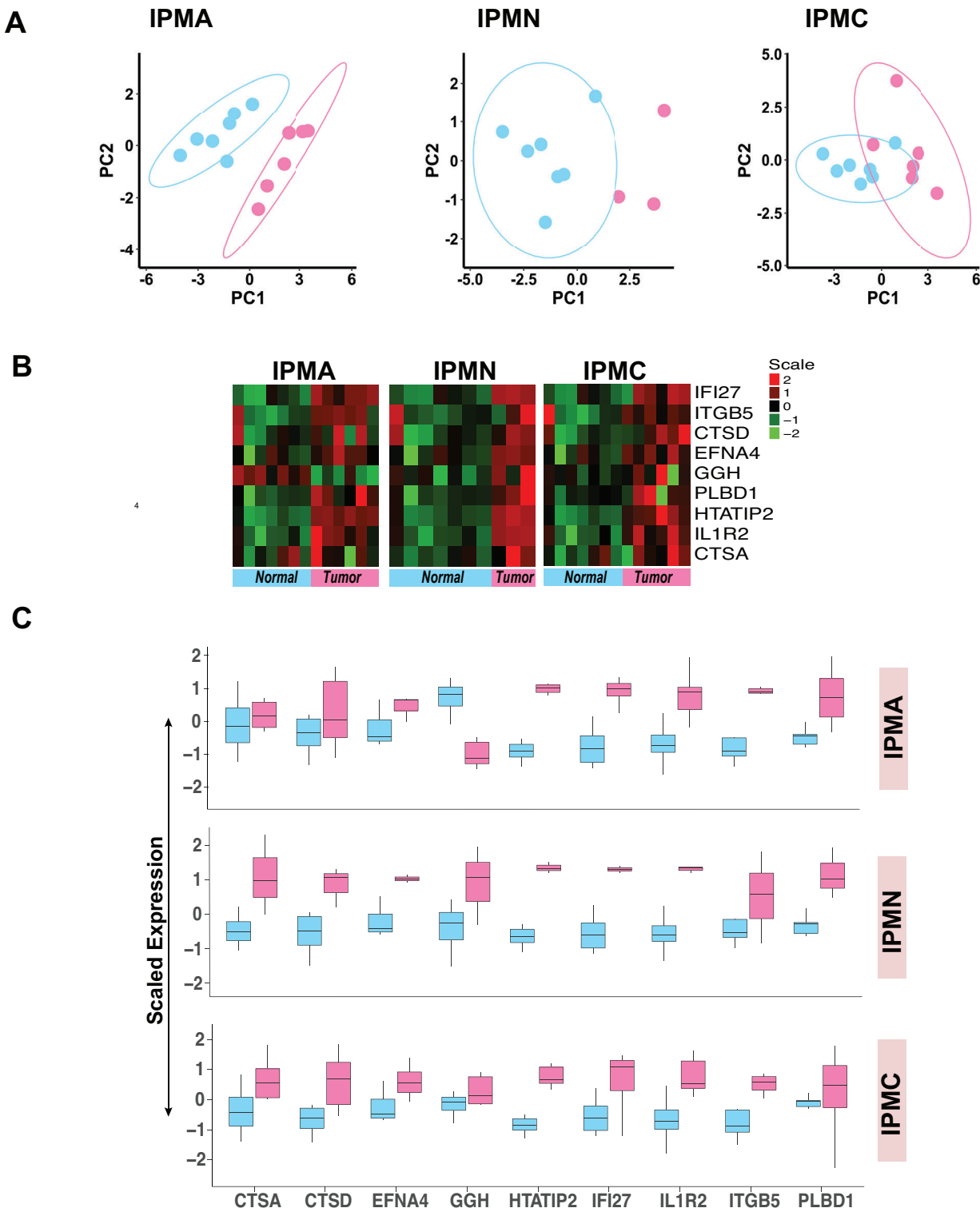
Supplementary Figure S8: The assessment metrics for validation datasets using the 9-biomarker panel genes. Heatmaps **(A)** and PCA plots **(B)** based on biomarker panel genes in validation sets.



Supplementary Figure S9: The assessment metrics for PV1-3 dataset using the 9-biomarker panel genes. A) PCA plots of three different prospective validation datasets. B) Heatmaps of the 9-marker genes panel. C) Boxplots of the expression of the genes.



Supplementary Figure S10: The assessment metrics for PV4 dataset using the 9-biomarker panel genes.
A) PCA plots for precursor lesions in three stages IPMA, IPMN and IPMC. **B)** Heatmaps of the 9-marker genes panel. **C)** Boxplots of the expression of the genes in precursor lesions.



Supplementary Figure S11: Comparative performance of 9-gene-based PDAC classifier with different previously established biomarkers. AUC plot for 9-gene-based PDAC classifier across the training and validation datasets. The measures of performances e.g. accuracy, sensitivity, specificity and AUC are mentioned in **Supplementary table 4**.

