

Supplementary Material for
Development of a consensus molecular classifier for pancreatic ductal
adenocarcinoma

Pablo Villoslada-Blanco *et al.*

Supplementary Table 1	Page 2
Supplementary Table 2	Page 3
Supplementary Figure 1	Page 5
Supplementary Figure 2	Page 6
Supplementary Figure 3	Page 7
Supplementary Figure 4	Page 8

Table S1. Balanced accuracy of the different Machine Learning algorithms.

		NB	KNN	RF	LSVC	NLSVC	RLRC
TUMOR CLASSIFIERS	Collisson	0.9585	0.8442	0.8470	0.9234	0.9122	0.9744
	Moffitt	0.8429	0.8999	0.9046	0.8934	0.8938	0.9378
	Bailey	0.8039	0.7795	0.6186	0.7490	0.7920	0.9495
	Puleo	0.8313	0.6870	0.6098	0.8538	0.8568	0.9795
	Chan-Seng-Yue	0.8514	0.8221	0.7673	0.8889	0.8891	0.9842
STROMA CLASSIFIERS	Moffitt	0.9542	0.9359	0.9680	0.9679	0.9658	0.9886
	Maurer	0.8827	0.9140	0.9068	0.8973	0.9007	0.9110

KNN (K-Nearest Neighbour Classifier), LSVC (Linear Support Vector Classifier), NB (Naive Bayes Classifier), NLSVC (Non-Linear Support Vector Classifier), RF (Random Forest Classifier), RLRC (Regularized Logistic Regression Classifier).

Table S2. Genes belonging to the Classical and Non-Classical signatures present in other classification systems.

	Collisson_Cl assical	Moffitt_Cla ssical	Bailey_Prog enitor	Chan- Seng- Yue_Clas sical-A	Chan- Seng- Yue_Clas sical-B	Collisson _QM	Moffitt_B asal-like	Bailey_Squa mous	Chan- Seng- Yue_Ba sal-like- A	Chan- Seng- Yue_Ba sal-like- B	Collisson_Ex ocrine-like	Bailey_A DEX	Bailey_Immun ogenic	Pul eo	Consensus_cl assical	Consensus _non- classical
GATA6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
ADGR G7	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
APOB EC1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
BTNL3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
BTNL8	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0
GPA33	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
GUCY 2C	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
HMGC S2	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0
ITLN1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0
KRT20	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0
MUC2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
MUC1 7	0	0	1	1	0	0	0	0	0	0	0	0	0	1	1	0
MYO7 B	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0
PHGR 1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0
REG4	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	0
SPINK 4	0	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0
TFF1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0
TM4SF 20	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0
FAT2	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1
KRT5	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1
KRT6A	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	1
KRT13	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1
KRT14	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	1
RHCG	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1
TNNT3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1

		Collisson_Cl assical	Moffitt_Cla ssical	Bailey_Prog enitor	Chan- Seng- Yue_Clas sical-A	Chan- Seng- Yue_Clas sical-B	Collisson _QM	Moffitt_B asal-like	Bailey_Squa mous	Chan- Seng- Yue_Ba sal-like- A	Chan- Seng- Yue_Ba sal-like- B	Collisson_Ex ocrine-like	Bailey_A DEX	Bailey_Immun ogenic	Pul eo	Consensus_cl assical	Consensus _non- classical
Typical acinar genes	AMY2 A	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
	CEL	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
	CELA3 A	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
	CPA1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
	CPA2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
	CPB1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
	CTRB1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
	CTRB2	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
	CTRC REG1 B	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Typical immune genes	IGKV1- 12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Other genes present in Classical signature	PRAP1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	SLC13 A2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	B3GN T6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	FEZF1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Other genes present in Non- classical signature	DNER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	H19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	TENM 2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

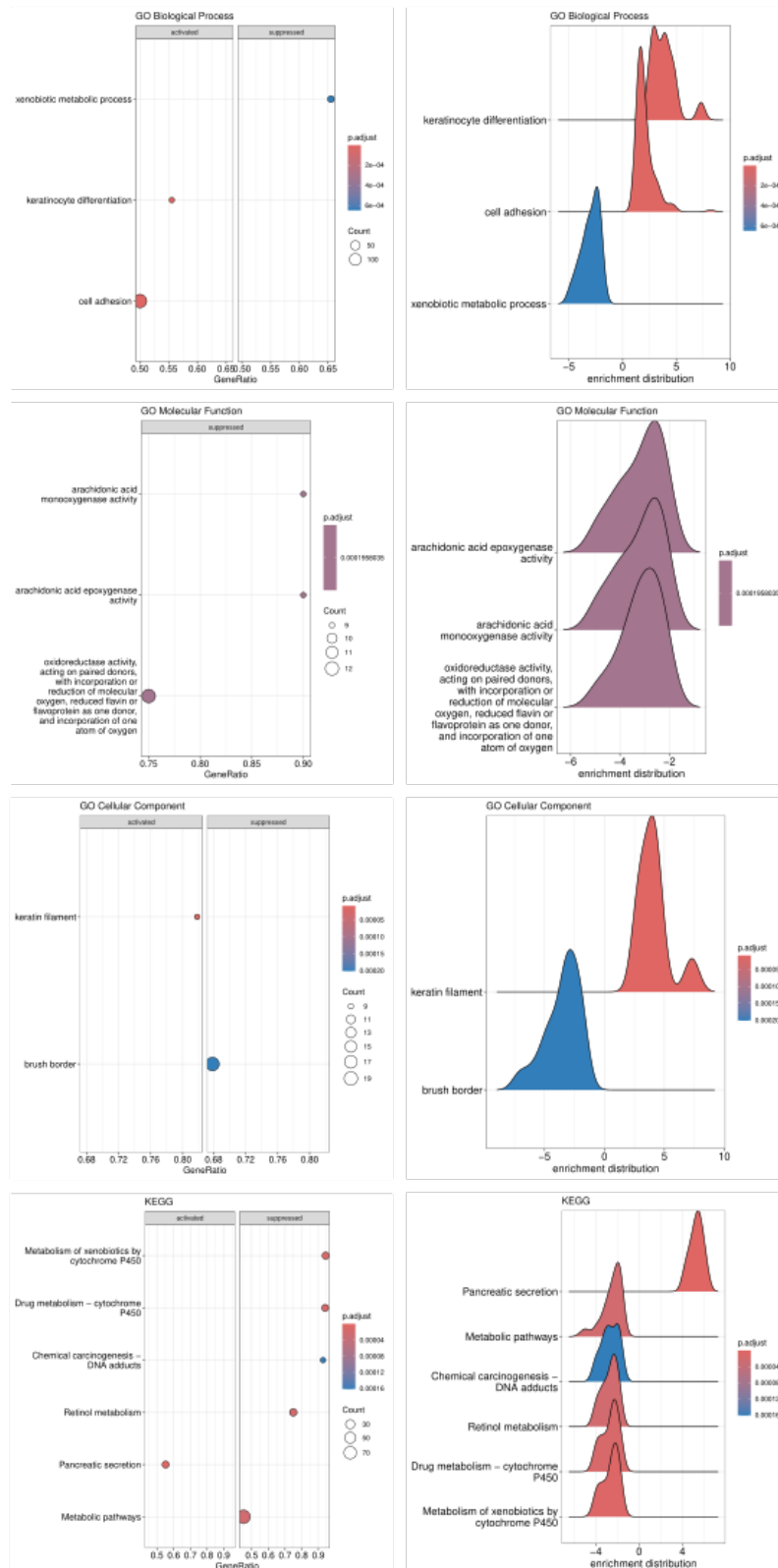


Fig S1. Gene Set Enrichment Analysis of tumor consensus classes. *GeneRatio* measures the proportion of genes in a particular gene set that are found within the ranked lists of genes, while *enrichment distribution* quantifies the degree to which a gene set is overrepresented at the top or bottom of a ranked list of genes. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

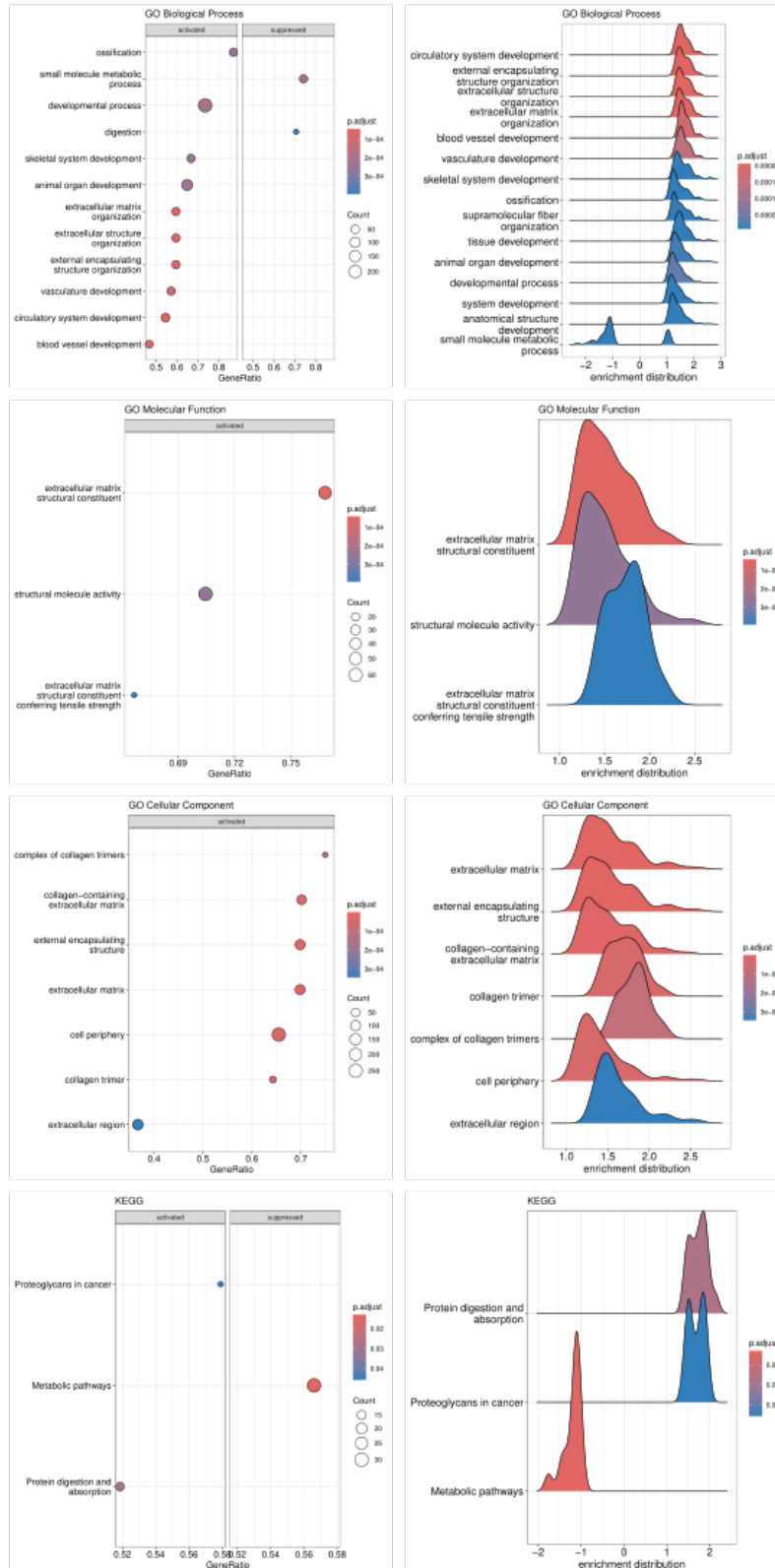


Fig S2. Gene Set Enrichment Analysis of stroma consensus classes. Gene Ratio measures the proportion of genes in a specific gene set found within the ranked lists of genes, while enrichment distribution quantifies how much a gene set is overrepresented at the top or bottom of a ranked list. GO refers to Gene Ontology; KEGG stands for Kyoto Encyclopedia of Genes and Genomes.

