An Integrative Bioinformatics Approach to Biomarker Discovery in Pancreatic Cancer

Thomas I. Blair III[1], Dr. Duaa Mohammad Alawad[2], Dr. Chindo Hicks[2]

Loyola University of New Orleans[1], Bioinformatics and Computational Medicine Program[2], Department of Genetics, LSU School of Medicine[2]

Introduction

•Pancreatic ductal adenocarcinoma (PDAC), is among the most lethal cancers with a 5-year survival rate below 10%.

•The American Cancer Society projected 66,440 new Pancreatic cancer cases and 51,750 related deaths in the U.S. in 2024.

•Clinical management of PDAC is challenged by poor response to chemotherapy and the lack of early diagnostic or prognostic biomarkers.

•This study uses integrative bioinformatics to identify diagnostic and prognostic biomarkers in PDAC from RNA-Seq and mutation data.

Objective

To discover **clinically actionable diagnostic and prognostic biomarkers** and potential **therapeutic targets** for pancreatic cancer by:

•Identifying differentially expressed genes (DEGs) between tumor vs. control, and dead vs. alive groups.

•Integrating gene expression with somatic mutation data to identify functionally relevant mutated genes.

•Performing functional enrichment and pathway analysis to identify biologically meaningful signatures.

Materials and Methods

TCGA Pancreatic Cancer Cohort (PDAC)

•RNA-Seq expression data (Tumor vs. Control; Dead vs. Alive).

•Matched somatic mutation data from whole-exome sequencing.

Table 1: Data distribution

Total	Control	Tumor		
		N= 556		
639	83	Alive	Dead	Unknown
		261	160	135

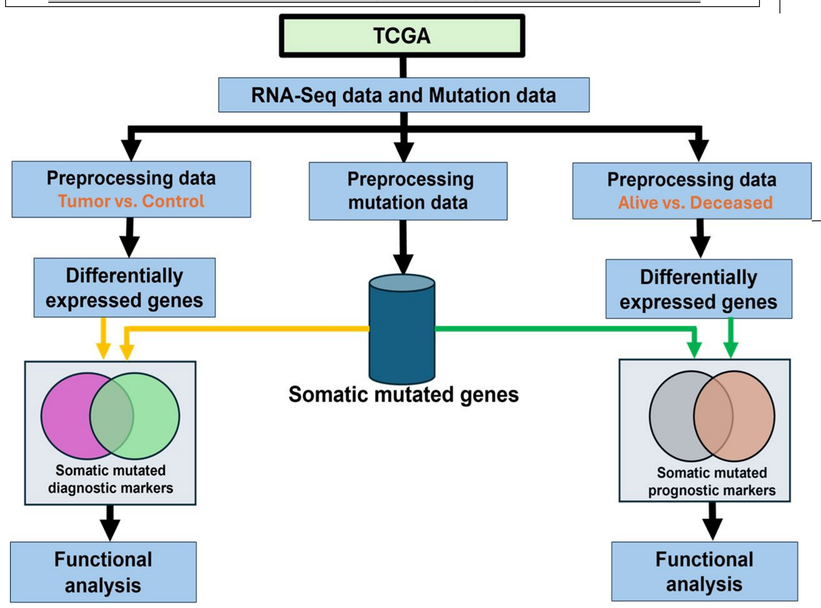

Figure 1: Overall study design and execution workflow.

Results

Tumor Vs Normal

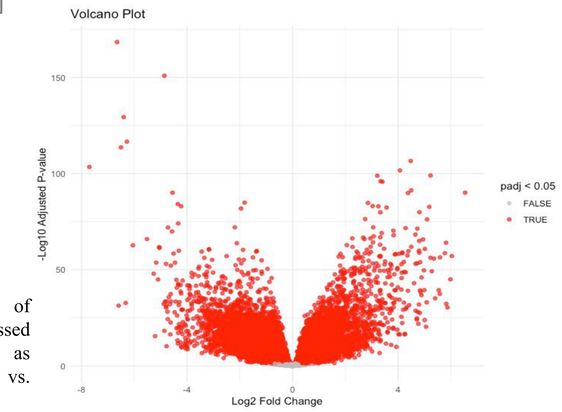

Figure 2: Volcano plot of significantly differentially expressed somatically mutated genes as measured by the log2FC (Tumor vs. Normal).

Table 2. List of the top 10 most highly significantly differentially expressed somatically mutated genes (tumor versus control),

Gene	ChromosomePosition	log2FoldChange	padj	Mutation_Count
KRAS	12p12.1	0.485752	1.08E-12	401
TP53	17p13.1	0.344114	0.000653	352
SMAD4	18q21.2	-0.85207	3.49E-11	107
CDKN2A	9p21.3	1.649951	4.25E-09	105
TTN	2q31.2	-2.1241	5.47E-20	93
MUC16	19p13.2	3.597126	2.42E-30	55
OBSCN	1q42.13	-0.79275	7.35E-09	34
KMT2D	12q13.12	-0.44719	1.73E-05	34
RYR1	19q13.2	-0.47291	0.011233	34
SYNE1	6q25.2	-1.59638	4.65E-13	30

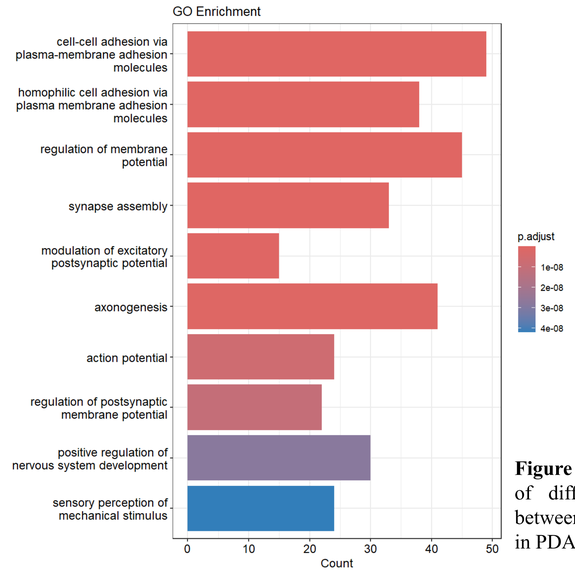

Figure 4: GO enrichment analysis of differentially expressed genes between normal and tumor samples in PDAC.

Dead Vs. Aive

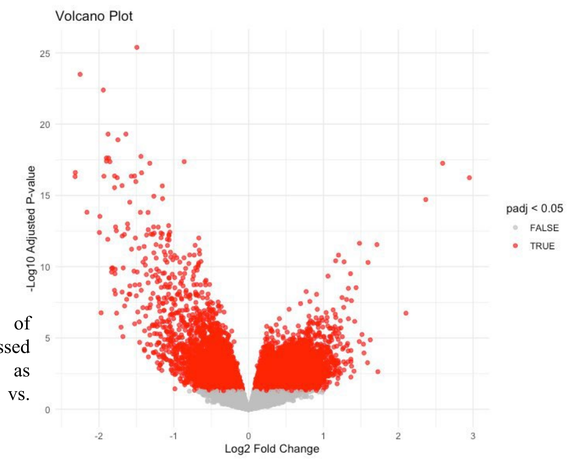

Figure 3: Volcano plot of significantly differentially expressed somatically mutated genes as measured by the log2FC (Dead vs. Alive).

Table 3. List of the top 10 most highly significantly differentially expressed somatically mutated genes (Dead versus Alive),

Gene	Chromosome Position	log2FoldChange	padj	Total_Mutations
KRAS	12p12.1	0.138476	0.025643	106
TP53	17p13.1	-0.20133	0.036571	100
TTN	2q31.2	0.773293	0.000612	30
MUC16	19p13.2	1.024793	0.000537	16
RNF213	17q22	0.265439	0.020433	10
LRP1B	2q22.1-q22.2	0.578667	0.01844	9
FAT2	5q33.1	0.73216	0.000855	8
CACNA1B	9q34.3	-1.00805	1.56E-05	8
OBSCN	1q42.13	0.519068	6.11E-05	8
SCN5A	3p22.2	1.162279	3.84E-11	8

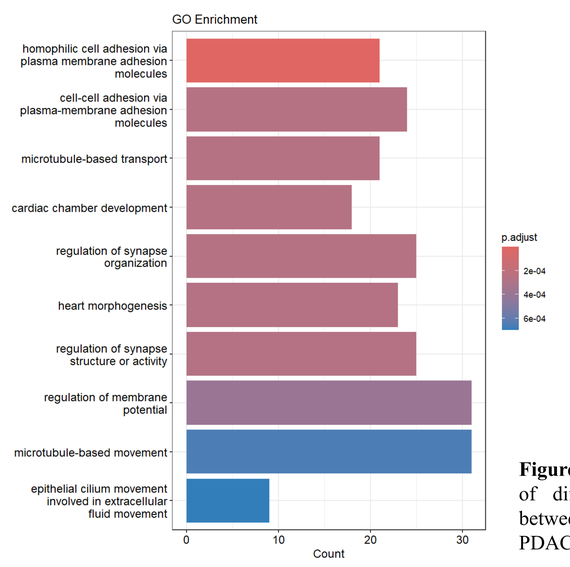

Figure 5: GO enrichment analysis of differentially expressed genes between Dead and Alive samples in PDAC.

Conclusion

•Integrative analysis of gene expression and somatic mutation data revealed key diagnostic and prognostic biomarkers in PDAC, including KRAS and TP53.

•Functional enrichment identified critical pathways involved in tumor progression and patient survival.

•This approach supports precision oncology and lays the foundation for developing predictive models to prioritize high-risk patients.

•Future work will focus on validating candidate biomarkers experimentally and applying machine learning models to predict high-risk patients for clinical prioritization.

This research project was supported by Award Number: DBI-2349224 through the National Science Foundation (NSF), Research Experiences for Undergraduates (REU) Program