Transcriptomic profiling of pancreatic adenocarcinoma for identification of differentially expressed genes

Akanksha Varshney, Deepak Tiwari, Shabi, Ashutosh Mani*
Department of Biotechnology Motilal Nehru National Institute of Technology Allahabad– INDIA
*Corresponding author: amani@mnnit.ac.in

Abstract

Pancreatic adenocarcinoma (PAAD) is among the most aggressive malignancies and is associated with an extremely poor prognosis due to late diagnosis and limited therapeutic options. High-throughput transcriptomic profiling provides an effective approach to identify molecular alterations underlying disease progression. In this study, RNA sequencing data from The Cancer Genome Atlas (TCGA) were analyzed to identify differentially expressed genes (DEGs) between pancreatic adenocarcinoma and normal pancreatic tissues. Differential expression analysis was performed using Bioconductor packages in R, followed by functional enrichment analysis using ShinyGO. Protein–protein interaction (PPI) networks were constructed using STRING and Cytoscape, and hub genes were identified using cytoHubba. Survival analysis of hub genes was carried out using GEPIA2. A total of 418 DEGs were identified, of which 410 were upregulated and 8 were downregulated. Enrichment analysis revealed significant involvement of immune-related biological processes and signaling pathways. Network analysis identified several hub genes, among which PTPN6 showed a significant association with patient survival. These findings highlight PTPN6 as a potential prognostic biomarker and provide insights into the molecular mechanisms of pancreatic adenocarcinoma.

Keywords: Pancreatic adenocarcinoma, TCGA, RNA-Seq, Differentially expressed genes, Network analysis, PTPN6

Received on: 12.11.2025
Accepted on: 22.12.2025
Published on 26.12.2025

Introduction

Pancreatic adenocarcinoma is one of the most lethal human malignancies, ranking among the leading causes of cancer-related mortality worldwide. The five-year survival rate remains below 10%, primarily due to late-stage diagnosis, aggressive tumor biology, and resistance to conventional therapies [1,2]. Current clinical management strategies rely mainly on surgery combined with chemotherapy or neoadjuvant therapy; however, these approaches offer only modest survival benefits [1].

Molecular profiling studies have revealed extensive genetic and transcriptomic heterogeneity in pancreatic cancer [3]. Several biomarkers, including CA19-9, KRAS mutations, MIC-1, and GPC1, have been proposed for diagnosis and prognosis, but their clinical utility is limited by insufficient sensitivity and specificity [4]. Advances in next-generation sequencing technologies and the availability of large-scale datasets such as TCGA have enabled systematic identification of differentially expressed genes and dysregulated pathways in pancreatic adenocarcinoma [3,5].

In the present study, we performed a comprehensive transcriptomic analysis of

TCGA-PAAD RNA-Seq data to identify DEGs between tumor and normal pancreatic tissues. Functional enrichment and network-based approaches were applied to uncover key biological processes and hub genes involved in disease progression, followed by survival analysis to assess their clinical relevance.

Materials and Methods

Data acquisition

RNA sequencing (RNA-Seq) data for pancreatic adenocarcinoma were obtained from The Cancer Genome Atlas (TCGA-PAAD) database using the TCGAbiolinks package in R [5]. HTSeq-Counts data corresponding to primary tumor (TP) and normal tissue (NT) samples were included in the analysis.

Differential gene expression analysis

Raw count data were normalized and filtered using Bioconductor workflows. Differential expression analysis was carried out using a generalized linear model likelihood ratio test (glmLRT). Genes with a false discovery rate (FDR) < 0.01 and an absolute log2 fold change ≥ 1.5 were considered significantly differentially expressed.

Functional enrichment analysis

Gene ontology (GO) enrichment analysis for biological process (BP), cellular component (CC), and molecular function (MF) categories was performed using the ShinyGO web tool [6]. Homo sapiens was selected as the reference organism, and a p-value cutoff of 0.05 was applied.

Protein–protein interaction network construction and hub gene identification

Differentially expressed genes were mapped to the STRING database to construct a protein–protein interaction (PPI) network. The network was visualized using Cytoscape software. Hub genes were identified using the cytoHubba plugin based on multiple topological algorithms, including Maximal Clique Centrality (MCC), Density of Maximum Neighborhood Component (DMNC), and Bottleneck methods [7].

Survival analysis

The prognostic significance of hub genes was evaluated using the GEPIA2 web server, which integrates TCGA gene expression and clinical data [8]. Kaplan–Meier survival curves were generated, and log-rank tests were performed. Genes with log-rank p-values < 0.05 were considered significantly associated with overall survival.

Results

Identification of differentially expressed genes

Comparative analysis between pancreatic adenocarcinoma and normal pancreatic tissues identified a total of 418 DEGs. Among these, 410 genes were significantly upregulated and 8 genes were downregulated in tumor samples, indicating widespread transcriptional reprogramming in pancreatic adenocarcinoma.

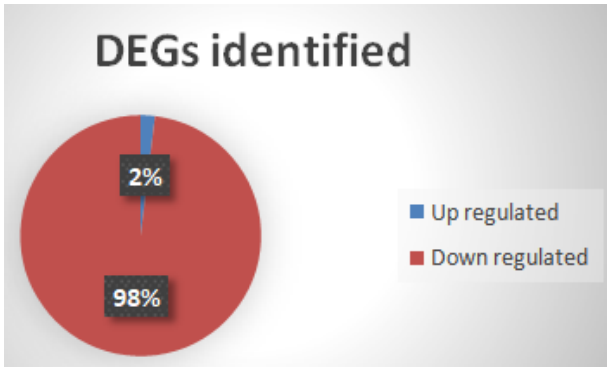

Figure 1: Number of Upregulated and Downregulated DEGs

ENSG	logFC	PValue	FDR
ENSG00000000938	2.1311277	4.62E-08	8.38E-06
ENSG00000005187	1.8290009	0.000101	0.004275
ENSG00000005844	2.4743562	2.08E-07	2.72E-05
ENSG00000007129	2.0409784	3.23E-07	3.96E-05
ENSG00000007264	2.1085381	1.54E-08	3.43E-06
ENSG00000007312	3.1875588	1.29E-06	0.000123

Table 1: DEGs in PAAD data extracted from TCGA

Gene ontology enrichment analysis

GO biological process enrichment analysis revealed that DEGs were predominantly associated with immune system processes, immune response, cell activation, and leukocyte activation, suggesting an important role of immune modulation in pancreatic cancer progression [6]. Cellular component analysis showed enrichment in intrinsic and integral components of the plasma membrane and cell surface proteins. Molecular function analysis indicated significant enrichment of signaling receptor activity and molecular transducer activity, highlighting altered signal transduction mechanisms in tumor tissues.

Protein–protein interaction network and hub genes

The PPI network constructed using STRING and Cytoscape revealed several highly interconnected nodes. Topological analysis using cytoHubba identified key hub genes, including PLEK, CD48, BTK, FCGR2B, and PTPN6. These genes are primarily involved in immune signaling and cellular communication pathways, consistent with enrichment analysis results [7].

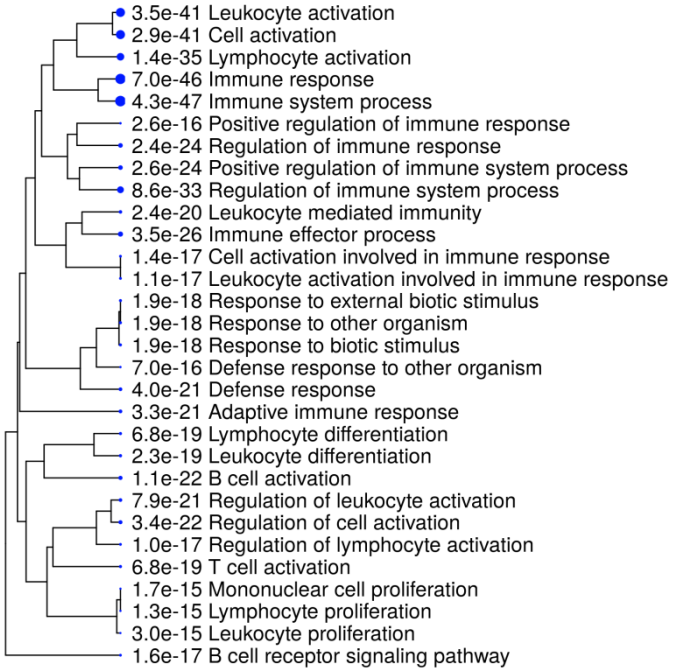

Figure 2: Hierarchial clustering summarizing GO molecular function terms

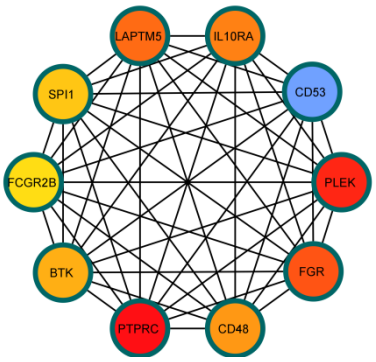

Figure 3: Interaction network of top 10 genes

Survival analysis of hub genes

Survival analysis using GEPIA2 demonstrated that among the identified hub genes, PTPN6 expression was significantly associated with overall survival in pancreatic adenocarcinoma patients (log-rank p <

0.05). Other hub genes did not show statistically significant survival associations. These findings suggest that PTPN6 may have prognostic relevance in pancreatic adenocarcinoma [8].

Discussion

The present study employed an integrative bioinformatics approach to analyze transcriptomic alterations in pancreatic adenocarcinoma using TCGA data. Consistent with previous reports, immune-related pathways and genes were prominently enriched among the DEGs, underscoring the critical role of the tumor microenvironment and immune regulation in pancreatic cancer progression [2,3].

Network-based analysis enabled the identification of hub genes with high topological significance, which are more likely to play central roles in disease-associated molecular networks [7]. Among these, PTPN6 emerged as a key gene associated with patient survival. PTPN6 encodes a protein tyrosine phosphatase involved in immune receptor signaling and has been implicated in the regulation of inflammatory responses and cancer-related pathways [9,10]. Dysregulation of PTPN6 may contribute to tumor immune evasion and disease progression in pancreatic adenocarcinoma.

Although this study provides valuable insights, it is limited by its reliance on publicly available datasets and in silico analyses. Experimental validation using clinical samples and functional assays is required to confirm the biological role of PTPN6 and other identified genes.

Conclusion

In conclusion, this study identified 418 differentially expressed genes between pancreatic adenocarcinoma and normal pancreatic tissues. Functional enrichment and network analyses highlighted immune-related pathways and identified PTPN6 as a key hub gene with prognostic significance. These findings provide a foundation for future experimental studies and support the potential of PTPN6 as a biomarker and therapeutic target in pancreatic adenocarcinoma.

Acknowledgements

The authors acknowledge the TCGA Research Network for generating and providing access to the genomic and clinical datasets used in this study.

References:

1. Vareedayah AA, Alkaade S, Taylor JR. Pancreatic adenocarcinoma. Missouri Medicine. 2018;115:230–235.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin. 2017;67:7–30.
3. Aguirre AJ, Hruban RH, et al. Integrated genomic characterization of pancreatic ductal adenocarcinoma. Cancer Cell. 2017;32:185–203.e13.
4. Ahmad S, Hussain I, Sarvepalli D. Tumor biomarkers and diagnosis of pancreatic adenocarcinoma. J Gastrointest Oncol. 2018;9:934–944.
5. Nicolle R, Raffenne J, Paradis V, et al. Prognostic biomarkers in pancreatic cancer: avoiding errata when using the TCGA dataset. Cancers (Basel). 2019;11:126.
6. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020;36:2628–2629.
7. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8(Suppl 4):S11.
8. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 2019;47:W556–W560.
9. Liu Y, Cen C, Peng S, et al. Identification of differentially expressed genes in pancreatic ductal adenocarcinoma and normal pancreatic tissues. Mol Med Rep. 2021;23:1–12.
10. Tonks NK. Protein tyrosine phosphatases: from genes, to function, to disease. Nat Rev Mol Cell Biol. 2006;7:833–846.