

**Novel genome-wide methylation profile of
pancreatic ductal adenocarcinoma patients in Indian
population along with key modulated pathways and
its associated disease defining genes**

**Thesis submitted by
Akash Bararia
Doctor of Philosophy (Science)
in
Life Science and Biotechnology**

**Department of Life Science and Biotechnology
Jadavpur University
Kolkata- 700032
West Bengal, India
2024**

INDIAN STATISTICAL INSTITUTE

TELEPHONE: +91-33-2575-3240

-3241

Mobile: +91 9830780397 (P)

E. Mail: snilabja@gmail.com

F. Mail: snilabja@isical.ac.in



Dr. Nilabja Sikdar
Ramalingaswami Fellow
(comp),
Ramanujan Fellowship Awardee,
Human Genetics Unit, BSD
Indian Statistical Institute
203 Barrackpore Trunk Road
Kolkata – 700108, India

To Whom It May Concern

This is to certify that the thesis entitled “Novel genome-wide methylation profile of pancreatic ductal adenocarcinoma patients in Indian population along with key modulated pathways and its associated disease defining genes” submitted by Mr. Akash Bararia, registered in the Department of Life Sciences and Biotechnology, Jadavpur University, with Index No.: 24/22/Life Sc./27, registration dated 10th February, 2022 is his original work, based on the results of the experiments and investigations carried out by his under my supervision. He has presented his work in a hybrid mode Pre-thesis presentation seminar in front of Ph.D. Research Advisory Committee (RAC) on 13th June, 2024 and the Ph.D. Research Advisory Committee has recommended him to submit the thesis.

Further, it is also to certify that the above said work has not been previously submitted for the award of any degree, diploma, or fellowship in any Indian or foreign University. He has published 4 articles (1 research article and 3 review articles as first authors) in peer-reviewed journals including 1 book chapter and 2 letters to Editor. 3 more research paper are communicated as first author by him.

His thesis is free from substantial plagiarism.

I wish him all success in life.

Dated: 24/09/24

Nilabja Sikdar, Ph.D

NILABJA SIKDAR, Ph.D.
Scientist D/Ramalingaswami Fellow (comp)
Human Genetics Unit
Indian Statistical Institute
203, B. T. Road, Kol-108



Dr. Sankhadeep Dutta, Ph.D.

Senior Scientific Officer, Grade II

Department of Oncogene Regulation

Contact: +91-9051449396 (M); 91-33-2474-3922 (O);

Email: eordeep@gmail.com / sankhadeepdutta@cnci.ac.in

Chittaranjan National Cancer Institute

37 S P Mukherjee Road, Kolkata – 700026.

(An Autonomous Body under Ministry of Health and
Family Welfare, Government of India)

Phone: 2475 9313, 2476 5101 (Extn. 309)

Fax: 2475 7606; email cncinstkol@gmail.com

<http://www.cnci.ac.in>

September 23, 2024

CERTIFICATE FROM THE SUPERVISOR

This is to certify that the thesis entitled “**Novel genome-wide methylation profile of Pancreatic Ductal Adenocarcinoma patients in Indian population along with key modulated pathways and its associated disease defining genes**” submitted by **Mr. Akash Bararia**, registered in the Department of Life Sciences and Biotechnology, Jadavpur University (Index No.: 24/22/Life Sc./27) for the award of Ph. D. (Science) degree, is absolutely based upon his own work under my joint supervision along with Dr. Nilabja Sikdar.

This is also to certify that his thesis is free from substantial plagiarism. Neither this thesis nor any part of it has been submitted for either any degree/ diploma or any other award anywhere before.

(Dr. Sankhadeep Dutta)

Joint Supervisor

Dr. Sankhadeep Dutta, Ph.D.

Senior Scientific Officer, Grade-II

Dept. of Oncogene Regulation

CHITTARANJAN NATIONAL CANCER INSTITUTE

37, S. P. Mukherjee Road, Kolkata-700 026

ACKNOWLEDGEMENTS

On a sultry summer morning of April 2016, I walked through the campus of the Indian Statistical Institute for the first time as I embarked upon a new journey. It feels surreal that here I started with my internship and years later, penning down the concluding lines of my thesis. A multitude of emotions wash over me as I look back upon these years. Elation, disappointment, optimism, apprehension, anxiety and catharsis swirl around me like a kaleidoscope. Yes, it has been quite a journey, and by no means an easy one. It would be unforgiving of me if I did not reflect upon the people who have undertaken this journey with me, without whose constant efforts, perhaps I would not have got this far. So, here's to you.

I would like to begin by expressing my heartfelt respect and gratitude for my supervisor, Dr. Nilabja Sikdar. His relentless energy, passion and meticulousness for science were infectious. I was awed by his unwavering resolve and optimism, even in the face of seemingly insurmountable odds. He has the rare ability to explain complicated topics in a very interesting and lucid manner. Throughout these years, he has transformed my scientific thinking and perception. I shall always look back fondly upon our bi-weekly sessions of data presentation. Thank you, Sir for always being so caring and gentle, for believing in me and for never losing patience with me when things looked bleak. I would also like to thank Dr. Sankhadeep Dutta, my Co-supervisor from Chittaranjan National Cancer Institute for his constant support and faith during the entire tenure.

I would also like to thank the other faculty members of the Human Genetics Unit, for providing unconditional support and assistance whenever required. My heart breaks to think that our beloved Saurabh Sir is no longer with us, and I shall deeply miss his charisma, his humor and his witticism that was like a breath of fresh air. Prof. Bidyut Roy, Prof. Indranil Mukhopadhyay, and Dr. Shalini Datta were all very welcoming, kind and helpful right from the day of my joining in this department. National Institute of Biomedical Genomics (NIBMG) Core (Co-Teri) for 450K methylation array outsourcing facility for pancreatic cancer tissue samples. My gratitude also goes to my respected professors beyond the department with whom I was lucky enough to partake in scientific collaborations. I would like to thank Dr. Biswarup Basu, Prof. Priyadarshi Basu, Dr. Aniruddha Chatterjee, Dr. Nidhan K Biswas, Dr. Sumit Mukherjee (bioinformatics support) and Dr. Ankita Chatterjee for their valuable inputs and guidance. Thank you for providing me the collaborative opportunities to diversify my knowledge and experience.

I feel honored to have carried out my Ph.D. work in the esteemed Indian Statistical Institute, Kolkata. I extend my sincere gratitude to former Deans of studies Prof. Debashis Mukherjee. I would like to thank the department of Life Sciences and Biotechnology, Jadavpur University for deeming me worthy of registering as a Ph.D. scholar.

I could scarcely forget to mention my lab members who made Human Genetics Unit feel like a second home all these years. My PhD lab senior Gourab da, and Sayantan has been a pillar of strength with their vibrant attitude and ceaseless support, not only in academics but beyond. Words fail me to describe the profound impact they had in my life and I am utterly grateful to have them as a senior for the major part of my Ph.D. tenure. Pronoy Da, with his unparalleled dedication towards statistics and hard work has been a source of inspiration.

My gratitude also extends to my adjoining lab members of my biology department. Thank you, Dipjit, Debashish da, Joyita di, Sayani, Debakrito, Neelotpal, Madhurima, and Rahul da for your help and for making this journey so colorful. All the trips we went on, the birthday celebrations in the Students Room, the Holi celebrations and the tea sessions at the canteen will always remain evergreen. I am also greatly thankful to Tilak da, Tanmoy da, Raju da, Raj, Madhabi di, and Bela di, for their constant help whenever required.

I would like to acknowledge our clinical collaborators including Prof. Bitan Kr. Chattopadhyay, Dr. Paromita Roy, Dr. Shibajyoti Ghosh, Dr. Sumit Gulati, Dr. Supriyo Ghatak and Dr. Sudeep Banerjee, who enhanced my understanding of the clinical perspectives of the disease. All PGT students worked under my supervisor and provided clinical samples along with clinic pathological and demographic data from hospitals. My gratitude also goes to all the volunteers who willingly participated in the study, without whom this thesis would be impossible. Thank you for your selfless contribution to science.

I extend my gratitude to my fellow collaborators, who helped enrich my knowledge and experience. I thank Mr. Agnishwar Chakraborty for his support in machine learning algorithms, and Dr. Debabrata Ghosh Dastidar for giving me the opportunities to work with you over these years. You have been very patient and understanding of me, even when I took some time to deliver results. A special thanks to Dr. Sourav Banerjee, from Dundee University and Dr. Amlan Das from Royal Global University for their support. I would also like to thank Mr. Tarun Dutta (Lead AI Scientist at IBM, India), and Mr. Chandranshu Chakrabarti (Bread Financial Ltd) along with Dr. Pronoy Kanti Mondal (Michigan State University) for their immense support in statistics. I would also like to thank Prof. Indranil

Mukhopadhyay and late Prof. Saurabh Ghosh from Indian Statistical Institute for their extensive support in understanding the basics of bio-statistics. I would also like to thanks Nicolas Keestra, and Morgan Jones from Dunedin school of medicine, New Zealand for their support in bioinformatic analysis of WGBS data.

I extend my warm regards Sandip Ghosh, Souvik Saha and Pratiti Bhattacharjee for our discussions of lab and life in general. A special shout-out goes to Dr. Sourav Chakraborty of IPGIMER, with whom I have forged a close bond over cake, and coffee.

My family has been an immeasurable source of strength for me not only over these years but throughout my life. Their selfless hard work day in and day out, all for me is something I can never repay. They have always looked out for me without asking anything in return. Thank you for all the love and affection, you have showered over me. I am so very grateful to have had a loving home and kind faces to return to at the end of each day, with food still kept warm for me no matter how late it was. And finally, thank to my friends for being the rock against which I could willingly crash again and again.

Akash Bararia
Akash Bararia

Human Genetics Unit
Indian Statistical Institute
203 B.T. Road,
Kolkata -700108, India

Contents:

INTRODUCTION	1
1.0: Introduction.....	2
1.1: Epidemiology of Pancreatic Cancer in Indian Subcontinent	3
1.2: Pancreas Anatomy	5
1.3: Types of Pancreatic Cancer.....	7
1.3.1: Pancreatic Ductal Adenocarcinoma.....	7
1.4: Precancerous Forms of Pancreas	9
1.5: Clinical Presentation of Pancreatic Ductal Adenocarcinoma	11
1.6: TNM Staging of Pancreatic Ductal Adenocarcinoma	11
1.6.1: T Category.....	11
1.6.2: N Category	11
1.6.3: M Category.....	11
1.6.4: Tumor Grade	12
1.7: Etiological Factors.....	12
1.7.1: Modifiable Risk Factors.....	12
1.8: Lifestyle Risk Factors	12
1.8.1: Smoking.....	12
1.8.2: Alcohol	13
1.8.3: Obesity.....	13
1.8.4: Dietary Factors	13
1.9: Environmental Risk Factors	14
1.10: Occupational Risk Factors.....	14
1.11: Non-Modifiable Risk Factors	15
1.11.1: Gender	15
1.11.2: Age	16
1.11.3: Ethnicity.....	16
1.11.4: Diabetes Mellitus.....	16
1.12: Chronic Infections	16
1.13: ABO Blood Group.....	17
1.14: Chronic Pancreatitis.....	17
1.15: Genetic Risk Factors	17
1.16: Inherited Risk Factors	18
1.17: Familial Pancreatic Cancer.....	19
1.18: Hereditary Pancreatitis.....	19
1.19: Hereditary and Other Factors	20
1.20: Diagnostic Evaluation and Treatment Modalities for Pancreatic Ductal Adenocarcinoma	20

1.20.1: Diagnosis	20
1.20.2: Survival Rates of Pancreatic Cancer	21
1.20.3: Early and Late Onset of Pancreatic Cancer.....	21
1.20.4: Treatment Strategy of Pancreatic Cancer.....	22
1.21: DNA Methylation Inhibitors	24
1.21.1: Histone Deacetylase Inhibitors (HDACi).....	24
1.21.2: Histone Modification Inhibitors	24
1.21.3: EZH2 Inhibitors.....	25
1.21.4: AURKA Inhibitors	25
1.21.5: Bromodomain Inhibitors	25
1.21.6: miRNA Targeting.....	25
1.22: Epigenetics and Pancreatic Cancer Development	25
1.22.1: DNA Methylation	26
1.22.2: Identification of DNA Methylation.....	28
1.23: Cancer Biomarkers Diagnosis, Prognosis and Its Clinical Implications and Therapy.....	31
1.24: Identification of Pancreatic Cancer Biomarker Using Multi Omics Dataset, Machine Learning Technologies and Approaches.....	31
REVIEW OF LITERATURE	34
2.0: Pancreatic Cancer and Its Epigenetics Landscape	35
2.1: Hypermethylation in Pancreatic Cancer	36
2.2: Hypomethylation in Pancreatic Cancer	38
2.3: Hypermethylation Identified in Pancreatic Cancer by NGS Studies	39
2.4: Next Generation Epigenetic Study of Hypomethylation of Pancreatic Cancer	40
2.5: Common Hypo and Hypermethylated Genes Between Pancreatic Ductal Adeno Carcinoma and Its Precancerous Lesions.....	40
2.6: Therapeutic Aspects of Epigenetic Modulated Pathways in PDAC and Its Aassociated Precancerous Lesions.....	41
2.6.1: Different Types of Biomarkers in Pancreatic Cancer Diagnosis and Prognosis	42
2.7: Identified Methylated Biomarkers by Multiomics Methods in Pancreatic Cancer	46
OBJECTIVES.....	48
3.0: Aims and Objectives	49
MATERIALS AND METHOD.....	50
4.0: Study Design	51
4.1: Sample Collection and Ethical Clearance from Different Hospital and Respective Institute.....	51
4.2: Inclusion Criteria of Patient Samples.....	52
4.3: Exclusion Criteria of The Patient Samples.....	52

4.4: Description of Study Cohorts	52
4.5: DNA Extraction and Bisulfite Treatment	53
4.6: DNA Methylation Data Generation by 450K Bead Chip Illumina Array and Analysis	57
4.7: RNA Isolation from Patient Tissue Samples	58
4.8: cDNA Synthesis	58
4.9: Gene Expression Study by Quantitative Real-Time PCR Using Sybr Green Chemistry	59
4.10: Bisulfite Conversion as Preprocessing Step for Methylation Specific PCR	60
4.11: Validation of Differential Methylation Using Methylation Specific PCR	61
4.12: Study of Relationship Between Differential Methylation and Gene Expression	62
4.13: Validation of Differential Methylation Using Immunohistochemical (IHC) Analysis of Pancreatic Ductal Adenocarcinoma Tissue Samples	62
4.14: Statistical Analysis	62
4.15: Validation of Differential Methylation Using Functional Enrichment	63
4.16: Validation of the Target Genes Using Publicly Available Data Source	64
4.17: Multi-Omics Validation Using Differentially Methylated and Methylation Altered Differentially Expressed Genes	64
4.18: Identification of Crucial Methylation Altered Ion Transporter Signatures	65
4.19: Identification of Methylation-Regulated Differentially Expressed Genes	65
4.20: Whole Genome Bisulfite Sequencing Based Validation of Targeted Methylation-regulated Differentially Expressed Genes	65
4.21: Dimensional Reduction and Unsupervised Clustering	66
4.22: Supervised Algorithm Based Machine Learning Models	66
4.23: Prediction Model Constructed Using Adaptive LASSO-Cox Regression Model	67
4.24: Tissue of Origin of Methylation-regulated Differentially Expressed Genes	68
4.25: Functional Relevance of Methylation-regulated Differentially Expressed Genes	68
4.26: Pathway Analysis of Methylation-Driven Differentially Expressed Genes	69
4.27: Immune and Cell Ecosystem Profile of Methylation-regulated Differentially Expressed Genes	70
4.28: Immune Function Correlation Analysis	70
4.29: Identification of Candidate Small Molecules Drug Candidates Prediction	70
4.30: Prediction of Drugs Against Kcnj5	71
4.31: Identification of TFs that Regulate Differentially Methylated Genes	71
4.32: Analysing RNA-Seq Datasets to Identify Ion Channel Genes in Pancreatic Ductal Adenocarcinoma	71
4.33: Relevant Dataset's Based Validation of <i>KCNJ5</i> 's Expression in Pancreatic Cancer	72
4.34: Validation Of <i>KCNJ5</i> Gene Expression in HPA Database	72
4.35: Study Cohort for Validation of Identified Genes	72
4.36 Data Availability	73

RESULTS AND ANALYSIS.....	74
5.1: Results and Analysis.....	75
5.1.1: Characteristic of Patients Recruited in the Study.....	75
5.1.2: Differential Methylation in Pancreatic Cancer	75
5.1.3: Change of Methylation Status at the DMPs with PDAC Cancer Prognosis	78
5.1.4: Relation Between Promoter Methylation at DMPs with Gene Expression	80
5.1.6: Assessment of Methylation Along with Gene and Protein Expression of Npy and Faim2 Genes in PDAC	87
5.1.7: Cross-validation of Both <i>FAIM2</i> and <i>NPY</i> Hypermethylation by IHC Experiment in Another Validation Set.....	90
5.1.8: Association of <i>NPY</i> and <i>FAIM2</i> Expression with Clinicopathological Risk Factors Among PDAC Patients	91
5.2 Identification of Ion Channel Transporters Using Differentially Expressed Genes (DEGs) in PDAC	92
5.2.1: Clinical Relevance of Significant Differentially Methylated Genes in Ion Channel Pathway.....	93
5.2.2: <i>KCNJ5</i> Expression Validation and Kaplan-Meier Overall Survival Analysis.....	98
5.2.3: Correlation of Immune Cell With Differentially Methylated Genes.....	98
5.2.4: Cross -Validation of <i>KCNJ5</i> Expression in HPA Database	99
5.2.5: Functional Network Analysis of <i>KCNJ5</i>	100
5.2.6: Functional Enrichment of <i>KCNJ5</i> and Associated Coexpressed Genes.....	101
5.2.7: Identification of Differentially Expressed Ion Channel Gene for PDAC and Protein- Protein Interaction Enrichment Analysis.....	102
5.2.8: Revalidation of Methylation Status in 450K Dataset from Indian Cohort	103
5.2.9: Transcription Factors Regulate the Expression of <i>KCNJ5</i>	103
5.2.10: <i>KCNJ5</i> Expression in Various PanCa Cell Lines	103
5.2.11: Identification of Small Molecules and Drug Targeting Kcnj5 Protein.....	104
5.3: Methylation-regulated Differentially Expressed Genes Identificztion.....	105
5.3.1: Identification of Methylation-regulated Differentially Expressed Genes in PAAD.....	105
5.3.2: Clustering of Methylation-regulated Differentially Expressed Genes Across Global Pancreatic Cancer Databases.....	106
5.3.3: Implementation of Based Random Forest Classification Using Methylation-regulated Differentially Expressed Genes	108
5.3.3.1: Data Splitting	108
5.3.3.2: Feature Selection	108
5.3.3.3: Model Training.....	109
5.3.3.4: Cross-Validation.....	109
5.3.3.5: Performance Metrics Indicators.....	109
5.3.3.6: The Determination of the Cut-off	109
5.3.3.7: Concluding Model Assessment	109
5.3.4: Construction of Methylation-regulated Differentially Expressed Genes Based Nomogram	112

5.3.4.1: Regularization Strength:	113
5.3.4.2: Sparsity in Coefficients:.....	113
5.3.4.3: Model Interpretability:.....	113
5.3.5: Identification of Biological Process Involving the Methylation-regulated Differentially Expressed Genes.....	114
5.3.6: Targeted Functional Enrichment of <i>MOGAT2</i> , <i>MIA</i> , and <i>CD68</i>	116
5.3.7: Cross Validation of Methylation -regulated Differentially Expressed for 5 Genes	118
5.3.8: Validation of CpGs (MEDEGs) of Interest in Independent GEO Datasets Containing Methylation and Expression Data of Pancreatic Cancer.	120
5.3.9: Parallel Screening of Methylation-regulated Differentially Expressed of 5 Genes.....	122
5.4 Targeting of Methylation-regulated Differentially Expressed Genes	128
5.4.1: Targeting Concurrently Significant Methylation-regulated Differentially Expressed Genes in PAAD	128
5.4.2: Calibration and Clinical Application of Lasso-Cox Rregression.....	128
5.4.3: Implementation of Supervised Learning Algorithms Using Methylation-regulated Differentially Expressed Genes	131
5.4.4: Validation of 7 Signatures Using Global Obtained Datasets	132
5.4.5: <i>TFF1</i> and <i>S100P</i> Showcasing Transition from Pre-cancerous to Metastatic Stage	133
5.4.6: Evaluation the Methylation-regulated Differentially Expressed Genes Role in Translational Way.....	133
5.4.7: Immune Profile Deconvolution Concerning <i>CD36</i> as Top Clinically Relevant Signature.....	138
5.4.8: Multistage Evaluation of Methylation-regulated Differentially Expressed Genes.....	139
5.4.9: Functional Enrichment of Methylation-regulated Differentially Expressed Genes	140
5.4.10: Evaluating Prognostic Implications of <i>CD36</i> and <i>UGT1A1</i>	143
5.4.11: <i>CD36</i> , <i>TFF1</i> , <i>S100P</i> , <i>MUC13</i> and <i>UGT1A1</i> Expression in Various PanCa Cell Lines	144
5.4.12: Drug Prediction on <i>TFF1</i> , <i>MUC13</i> and <i>S100P</i>	145
5.4.13: Transcription Factor- mRNA Relationship for <i>TFF1</i> , <i>MUC13</i> and <i>S100P</i>	145
DISCUSSION	146
6.0: Discussion.....	147
LIMITATIONS.....	161
7.0: Limitations of The Study	162
BIBLIOGRAPHY	163
8.0 Bibliography.....	164
PUBLICATIONS.....	194
9.0: List of Publications.....	195
10.0: Conferences and Workshops attended.	196
APPENDIX.....	197

LIST OF FIGURES

Figure Number	Figure Legend	Page Number
Figure 1	Bar chart of the region-specific incidence age-standardized rate by sex for pancreatic cancer in 2022	3
Figure 2	Worldwide incidence rate of pancreatic cancer	4
Figure 3	Worldwide mortality of pancreatic cancer	4
Figure 4	Anatomical structure of pancreas	6
Figure 5	Structures in relation with pancreas	6
Figure 6	Histology of normal pancreas	6
Figure 7	Normal pancreatic histology (Magnification view)	7
Figure 8	Histopathology (HPE) picture of pancreatic ductal adenocarcinoma (Marked nuclear pleomorphism with 4x anisonucleosis)	8
Figure 9	HPE picture of pancreatic ductal adenocarcinoma (Atypical glands immediately adjacent to artery)	8
Figure 10	Stages of different PanIN	10
Figure 11	Different pre-cancerous forms of pancreatic cancer	10
Figure 12	Methylation of cytosine	26
Figures 13	DNA methylation bringing about repression of gene expression	28
Figure 14	A conceptual review of pancreatic cancer biomarkers from a variety of “omics” levels.	46
Figure 15	Proteogenomic analysis of pancreatic ductal adenocarcinoma	46
Figure 16	Flow chart showing the study design	51
Figure 17	Representative examples of H&E-stained slides showing the tumor cell percentage in PDAC patient samples.	52
Figure 18	Methodology of the bisulfite conversion	60
Figure 19	Differentially methylated positions (DMPs) identified in the pancreatic cancer samples of the TCGA cohort.	76
Figure 20	Representative images of 156 DMP's	77
Figure 21	Representative images of 156 DMP's	78
Figure 22	Changes of methylation with increasing severity of cancer phenotypes in the TCGA cohort.	79
Figure 23	Differences in expressions of genes with promoter hypomethylation and hypermethylation.	81
Figure 24	Agarose gel electrophoresis of methylation specific PCR products:	82
Figure 25	Validation of differential expressions of gDMs in validation cohort 3 (n = 26).	83
Figure 26	Distribution of average beta values at the promoters of 6 target genes between tumor and adjacent normal. Tumor is represented as the red curve against the black curve of adjacent normal.	84

Figure Number	Figure Legend	Page Number
Figure 27	Functional annotations of gDMs.	85
Figure 28	Differential survival of genes annotated in MCODE1(<i>MX2, OAS2, IRF4, PML, HLA-A</i>) and MCODE2 (<i>ADYC3, GALR1, NPY, OPRK1</i>), in the TCGA pancreatic cancer cohort.	86
Figure 29	Differential methylation and expressions of <i>NPY</i> and <i>FAIM2</i> in the TCGA cohort.	88
Figure 30	Differential methylation of <i>NPY</i> and <i>FAIM2</i> in CCLE cell lines along with Pan cancer expression.	89
Figure 31	Representative images of immunohistochemistry showing expressions of Faim2 and Npy proteins in PDAC and paired control tissues	90
Figure 32	Protein expression of <i>NPY</i> and <i>FAIM2</i> along with joint survival patterns.	92
Figure 33	Pathway associated landscape evaluation: differentially methylated 91 gene signature status evaluation through	93
Figure 34	Pan cancer profiling of 6 different methylated genes targeting all cancer subtypes (tumor vs normal) and respective significance.	94
Figure 35	Normal tissue expression profile using GTEx database.	94
Figure 36	Clinical relevance of significant differentially methylated genes. Differential expression (TCGA, tumor Vs GTEx, normal)	95
Figure 37	Clinical relevance of differentially methylated genes:	96
Figure 38	Correlation between differential methylation and expressions of ion transporter genes	96
Figure 39	Stage wise expression of ion transporter genes	97
Figure 40	Mutational status of 4 ion transporter genes	97
Figure 41	Mutational status of 3 ion transporter genes	98
Figure 42	Deconvolution immune profiling using LM22 cell types	99
Figure 43	Functional validation of <i>KCNJ5</i> gene.	100
Figure 44	Functional enrichment of <i>KCNJ5</i> :	101
Figure 45	Identification, functional enrichment and cluster analysis of protein-protein interaction network of differentially expressed ion channel genes in PDAC.	102
Figure 46	Methylation status of <i>KCNJ5, CACNB2, CLIC5, GABBR2, KCNA3, RASA3 and KCNA6</i> in Indian cohort of pancreatic cancer.	103
Figure 47	Gene expression of <i>KCNJ5</i> in pancreatic cancer cell lines	104
Figure 48	Identification of the methylation driven differentially expressed genes	105
Figure 49	Pattern understand and clustering of DNA methylation driven genes across global pancreatic cancer databases	107

Figure Number	Figure Legend	Page Number
Figure 50	Supervised Clustering to develop methylation driven predictive models	111
Figure 51	Development of a seven-CpG-sites-based prognostic prediction model of pancreatic cancer using adaptive Lasso regression.	113
Figure 52	Functional characterization of 27 methylation driven differentially expressed genes	115
Figure 53	Functional enrichment of <i>MOGAT2</i> , <i>MIA</i> and <i>CD68</i> :	117
Figure 54	Methylation markers (<i>MIA</i> , <i>MOGAT2</i> and <i>CD68</i>) evaluated using WGBS data which consisted of multiple tumors from same patients, which included normal pancreas, primary tumor, local peritoneal metastasis and distant metastasis samples.	118
Figure 55	Tissue of origin targeting for respective methylation driven differentially expressed genes	119
Figure 56	Independent database associated validation of methylation driven differentially expressed genes	121
Figure 57	Survival associated screening of methylation driven differentially expressed genes	122
Figure 58	Parallel screening of gene sets using alternative screening processes	123
Figure 59	Cross Validation of methylation driven differentially expressed genes and their CpG's considering their clinical significance:	124
Figure 60	Validation of 5 significant genes across varied platforms:	125
Figure 61	Immune profiling to associate immune dysregulation associated with methylation driven differentially expressed genes	126
Figure 62	Immune landscape evaluation using multiple cell type on sample pool:	127
Figure 63	Detailed workflow of the study to establish methylation driven differentially expressed genes as potential pancreatic cancer diagnostic biomarkers.	128
Figure 64	Clinical model development for identification of prognostic biomarkers	129
Figure 65	Clinical visualization of the 7 top gene set	130
Figure 66	Prediction model development using supervised learning algorithms	132
Figure 67	<i>CD36</i> and <i>UGT1A1</i> clinical relevance various cancer types	134
Figure 68	Overall survival and correlation analysis of nomogram identified gene sets	135
Figure 69	Correlation analysis of nomogram identified gene sets along with joint overall survival	135
Figure 70	Protein level expression of 4 gene sets	136

Figure Number	Figure Legend	Page Number
Figure 71	Pan-cancer profiling of nomogram identified gene sets	137
Figure 72	Mutational profiling of nomogram identified gene sets	137
Figure 73	Pathway analysis using seven genes of interest	138
Figure 74	Immune cell types dysregulation based on <i>CD36</i> expression	139
Figure 75	Stage specific targeted gene expression across TCGA pancreatic cancer patient pool	140
Figure 76	Functional enrichment of <i>CD36</i>	141
Figure 77	Functional enrichment of <i>TFF1</i>	141
Figure 78	Functional enrichment of <i>UGT1A1</i>	142
Figure 79	Functional enrichment of <i>S100P</i>	142
Figure 80	Survival and expression evaluation of <i>CD36</i> and <i>UGT1A1</i>	143
Figure 81	Targeting tissue specific expression using normal tissue datasets	144
Figure 82	Gene expression of methylation driven differentially expressed genes in 50 pancreatic cancer cell lines	145

LIST OF TABLES

Table Number	Table Legend	Page Number
Table 1	PanIN Classification	198
Table 2	TNM Classification	198
Table 3	Syndromes and Genes associated with pancreatic ductal adenocarcinoma	198
Table 4	Details of the PAAD samples in the TCGA cohort (n=185). The red marked samples were excluded from the analysis. Highlighted samples are used in discovery cohort	199
Table 5	Primer sequences	200
Table 6	Details of the PAAD samples in the TCGA cohort (n=185). The red marked samples were excluded from the analysis.	206
Table 7	Characteristics of the samples included in validation cohorts	213
Table 8	Details of the probes that were differentially methylated between pancreatic cancer and solid tissue normal samples in the TCGA cohort	215
Table 9	Details of 156 differentially methylated positions, identified in validation cohort 1 including PDAC tumors compared to adjacent normal tissue samples.	216
Table 10	Details of the probes that were differentially methylated both in the TCGA cohort and in our study (n=68).	226
Table 11	Details of the probes that were not differentially methylated commonly in the TCGA cohort and in our study (n=88).	229
Table 12	Details of the 91 genes with differential methylation (gDMs)	232
Table 13	Summary of the result obtained from PCMDB.	243
Table 14	Fold changes of the expressions of gDMs in the validation cohort.	246
Table 15	Comparison of methylation status and gene expression status of selected genes mapped to enriched pathway in pancreatic cancer.	247
Table 16	Functional analysis of gDMs with Metascape	248
Table 17	Top 4 enriched pathways observed from our dataset.	249
Table 18	Results of GO enrichment analysis on each MCODE network (Log (p-value) is mentioned with each annotation)	249
Table 19	Summary of the output obtained from the Pearson Correlation Test	250
Table 20	Summary of the output obtained from the Kendell Tau Correlation Test	255
Table 21	Validation of <i>KCNJ5</i> expression and methylation	260
Table 22	CMap derived <i>KCNJ5</i> associated small molecule targeting	260
Table 23	47 Probes	261
Table 24	Top 8 probes obtained using Nomogram	263
Table 25	450K based methylation data validation	263
Table 26	RNA seq based data validation	264
Table 27	MicroArray based expression data validation	265

Abbreviations

5HMC	–	5-hydroxymethylcytosine
5MC	–	5-methylcytosine
ACC	–	Acinar Cell Carcinoma
ACS	–	American Cancer Society
AI	–	Artificial Intelligence
AJCC	–	American Joint Committee on Cancer
AP	–	Acute Pancreatitis
APCGI	–	Australian Pancreatic Cancer Genome Initiative
AUC-ROC	–	Area Under the Receiver Operating Characteristic Curve
BET	–	Bromodomain Inhibitors
BP	–	Biological Process
BRCA	–	Breast Invasive Carcinoma
BSP	–	Bisulfite Sequencing PCR
CA-19.9	–	Carbohydrate Antigen 19-9,
CAFs	–	Cancer-associated fibroblasts
CCLE	–	Cancer Cell Line Encyclopedia
CGIs	–	CpG-islands
CIFC	–	Cancer Incidence in Five Continents
C-Index	–	Concordance Index
CMRI	–	Calcutta Medical Research Institute
COBRA	–	Combined Bisulfite Restriction Analysis
CP	–	Chronic Pancreatitis
CPTAC	–	Clinical Proteomic Tumor Analysis Consortium
CRC	–	Colorectal Cancer
CRG	–	Cancer Related Genes
DEGs	–	Differentially Expressed Genes
DEICG	–	Differentially Expressed Ion Channel Genes
DL	–	Deep Learning
DMG	–	Differentially Methylation Genes
DMPs	–	Differentially Methylation Position
DNMT	–	DNA methyltransferases
DNMTIS	–	DNMT Inhibitors

EMT	– Epithelial-Mesenchymal Transition
EPIC	– European Prospective Investigation into Cancer
ER	– Estrogen Receptor
EUS	– Endoscopic Ultrasound
FAIM2	– Fas Apoptotic Inhibitory Molecule 2
FNAB	– Fine-Needle Aspiration Biopsy
FPC	– Familial Pancreatic Cancer
GBC	– Gallbladder Cancer
gDM	– Genes with Differential Methylation
GO	– Gene Ontology
GSEA	– Gene Set Enrichment Analysis
GWA	– Genome Wide Analysis
HBV	– Hepatitis B virus
HCV	– Hepatitis C virus
HDACi	– Histone Deacetylase inhibitors
HGNC	– HUGO Gene Nomenclature Committee
HH	– Hedge Hog
HNPCC	– Hereditary Non-Polyposis Colorectal Cancer
HP	– Hereditary Pancreatitis
ICMR	– Indian Council of Medical Research
ICRD	– Indian Cancer Registry Database
IPA	– Ingenuity Pathway Analysis
IPGME&R	– Institute of Post Graduate Medical Education and Research
IPMN	– Intraductal Papillary Mucinous Neoplasms
IRB	– Institutional Review Board
ISI	– Indian Statistical Institute
KEGG	– Kyoto Encyclopedia of Genes and Genomes
KIRCCC	– Kidney Renal Clear Cell Carcinoma
KNN	– K-Nearest Neighbors
LAP	– Latency-Associated Peptide
LNM	– Lymph Node Metastasis
LUSC	– Lung Squamous Cell Carcinoma
MBD	– Methyl-CpG Binding Domain
MCN	– Mucinous Cystic Neoplasm

MDA	– Moderately-Differentiated Adenocarcinoma
MEDEG	– Methylation Regulated Differentially Expressed Genes
MeDIP	– SEQ– Methylated DNA Immunoprecipitation Sequencing
MLA	– Machine Learning Algorithm
ML	– Machine Learning
MOGAT2	– Monoacylglycerol O-Acyltransferase 2
MPC	– Metastatic Prostate Cancer
MPD	– Metastatic Pancreatic Duct
MSP	– Methylation-Specific PCR
MSRDA	– Methylation-Sensitive Representational Difference Analysis
MSRE	– Methylation Sensitive Restriction Endonucleases
MTHFR	– Methylene Tetra Hydro Folate Reductase
MYC	– Master Regulator of Cell Cycle
NCBI	– National Center for Biotechnology Information
NGS	– Next Generation Sequencing
NPY	– Neuropeptide Y
ORA	– Over-representation Analysis
OS	– Overall Survival
PanCa	– Pancreatic Cancer
PanIN	– Pancreatic Intraepithelial Neoplasia
PCA	– Principal Component Analysis
PCR	– Polymerase Chain Reaction
PDA	– Poorly-Differentiated Adenocarcinoma
PDAC	– Pancreatic Ductal Adenocarcinoma
PFAAs	– Plasma Free Amino Acids
P-MAPA	– Protein Aggregate Magnesium-Ammonium Phospholipoleate-Palmitoleate Anhydride
PR	– Progesterone Receptor
RF	– Random Forest
RR	– Relative Risks
RRBS	– Reduced Representation Bisulfite Sequencing
SAGE	– Serial Analysis of Gene Expression
SAM- S	– Adenosyl Methionine
SCLC	– Small Cell Lung Cancer

SCN	– Serous Cystic Neoplasm
SNPs	– Single Nucleotide Polymorphisms
TCGA	– The Cancer Genome Atlas
TGF β	– Transforming Growth Factor- β
TLR	– Toll-Like Receptors
TMBIM	– Transmembrane Bax Inhibitor Motif-Containing
TMC	– Tata Medical Center
TME	– Tumor Micro Environment
TSG	– Tumor suppressor genes
TSP1	– Thrombospondin 1
VIM	– Variable Importance Measures
WDA	– Well-Differentiated Adenocarcinoma
WGBS	– Whole-Genome Bisulfite Sequencing

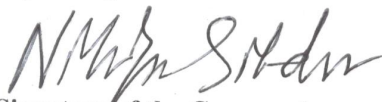
Abstract

Novel genome-wide methylation profile of Pancreatic Ductal Adenocarcinoma patients in Indian population along with key modulated pathways and its associated disease defining genes

Submitted by Akash Bararia (Index No.: 24/22/Life Sc./27)

Pancreatic ductal adenocarcinoma (PDAC) is a lethal cancer with a poor prognosis globally. Epigenetic changes, particularly DNA methylation, have a significant role in PDAC. Our study focused on the epigenetic landscape of PDAC in Indian patients and identified methylation markers like *NPY* and *FAIM2*, associated with poor prognosis. The study used DNA methylome analysis and compared findings with global datasets. The research highlighted the importance of ethnic factors in epigenetic research and identified potential diagnostic and therapeutic targets for PDAC in India. In our previous study, researchers found differences in genetic and epigenetic features of PDAC patients from India compared to worldwide patterns. Enrichment analysis revealed the significance of "regulation of ion transport" genes, with specific genes like *KCNJ5* showing variations in methylation and expression between Indian and Western populations. Several small molecule/ drugs targeting *Kcnj5* identified through CMap database based on their high enrichment score are Amiloride and Vernakalant hydrochloride which are FDA approved. The study also identified novel gene interactions in PDAC that could impact treatment strategies.

Understanding DNA methylation and its effects on tumor characteristics is crucial for developing new treatments, particularly in advanced PDAC cases. Targeting epigenetic changes in signalling pathway-associated genes may offer new therapeutic options. Integrating machine learning and artificial intelligence in precision oncology can enhance prognostication and treatment selection in PanCa patients. Machine learning methods, combined with bioinformatics, can help identify new prognostic and predictive biomarkers for PDAC. Research into epigenetic processes and gene expression dysregulation in PDAC is ongoing to improve diagnosis and treatment outcomes. Measuring Methylation-regulated Differentially Expressed Genes (MeDEGs) may provide valuable insights for enhancing prognostic tools and targeted therapies for PDAC. MeDEGs obtained by us includes genes (*MOGAT2*, *MIA*, *CD68*, *CD36*, *TFF1*, *S100P*, and *MUC13*) that are novel to our observation in PanCa but are previously reported in other cancer types., studies like these shed light on the complex interplay between genetics, epigenetics, and environmental factors in PDAC. By identifying novel methylation markers, researchers aim to improve the diagnosis, prognosis, and treatment of this deadly cancer. Collaboration across institutions and countries is essential for advancing our understanding of PDAC and developing targeted therapies for improved patient outcomes.


Signature of the Cosupervisor


Signature of the Cosupervisor

Dr. Sankhadeep Dutta, Ph.D.
Senior Scientific Officer, Grade-II
Dept. of Oncogene Regulation
CHITTARANJAN NATIONAL CANCER INSTITUTE
37, S. P. Mukherjee Road, Kolkata-700 026

Akash Bararia
Signature of the Candidate

NILABJA SIKDAR, Ph.D.
Scientist D/Ramalingaswami Fellow (comp)
Human Genetics Unit
Indian Statistical Institute
203, B. T. Road, Kol-108

Preface

Pancreatic ductal adenocarcinoma (PDAC) is a highly lethal malignancy with a dismal prognosis, ranking as the third leading cause of cancer-related deaths globally. The disease is characterized by late diagnosis, aggressive progression, and resistance to conventional therapies. Epigenetic alterations, particularly DNA methylation, play a pivotal role in the pathogenesis of PDAC. We conducted the first study which was published at *Cancer Cell International* (2022) (PMID: 36329447) delves into the epigenetic landscape of PDAC in the Indian population, identifying novel methylation markers that could potentially serve as diagnostic, prognostic, and therapeutic targets. The research utilized a multi-institutional approach, collecting samples from various hospitals in Kolkata, India. The study employed the Illumina 450 K bead chip for DNA methylome analysis, followed by validation using methylation-specific PCR and immunohistochemistry. The authors also compared their findings with global datasets from The Cancer Genome Atlas (TCGA) and the Clinical Proteomic Tumor Analysis Consortium (CPTAC). Our study identified two novel promoter hypermethylated genes, *NPY* (Neuropeptide Y) and *FAIM2* (Fas apoptotic inhibitory molecule 2), which were significantly associated with poor survival and advanced tumor stages in Indian PDAC patients, that was not previously associated with TCGA along with other Caucasian pancreatic cancer (PanCa) patient population from different datasets available throughout the globe. Our study proposed a unique genomic architecture that is totally different from other published data of the western world. These genes were found to be epigenetically regulated, with their hypermethylation leading to down-regulation of gene expression and also confirmed by immunohistochemical staining of clinical samples from another independent Indian patient cohort. The study also observed hypermethylation in other genes, including tumor suppressors, transcription regulators, ion transporters, and immune regulators, which are implicated in cancer progression and metastasis. The identification of *NPY* and *FAIM2* as novel epigenetic markers in the Indian context is a significant contribution to the field of personalized medicine. These markers could potentially be used to detect PDAC at an early stage, predict patient outcomes, and guide the development of targeted therapies. The study also highlights the importance of considering demographic and ethnic factors in epigenetic research, as evidenced by the differences in methylation patterns and gene expression between the Indian cohort and Western populations. Our study provides valuable insights into the epigenetic underpinnings of PDAC in the Indian population. The study identifies novel methylation signatures, paving the

way for epigenetically targeted therapies and biomarkers, highlighting the need for further research in PDAC diagnosis and treatment.

A recent study conducted by our lab using 450 K methylome data demonstrated that the epigenetic and genetic characteristics of a group of PDAC patients from India differ to some extent depending on the frequency of gene expression and the alternating driver behaviours, while being consistent worldwide. Utilizing GO/KEGG terms, canonical pathways, and hallmark gene sets, enrichment analysis of these 91 gDMs revealed that "regulation of ion transport" was the most statistically significant enriched term. The correlation between Metascape data from our previous study and IPA data from this investigation led to the discovery of 7 gDMs related with 'ion transport regulation': *RASA3*, *CLIC5*, *KCNA3*, *KCNA6*, *KCNJ5*, *CACNB2*, and *GABRR2*. An analysis revealed notable variations in the expression of *RASA3*, *KCNJ5*, and *CLIC5* between normal tissue samples in the GTEx cohort and cancer samples in the TCGA cohort. However, the expression of *CACNB2* was not statistically significant. The *KCNJ5* gene exhibited hypomethylation and overexpression in our first Indian cohort, whereas hypermethylation and under expression was observed in the TCGA cohort as well as few Caucasians PDAC patient sample datasets. Overall survival data from an Indian cohort revealed a notably lower survival rate in the high *KCNJ5* group compared to the low *KCNJ5* group. This observation is reinforced by the substantial upregulation of the protein in PanCa cell lines, SU86.86, YAPC, and HuP T4. However, the expression of its protein, Kir3.4, has not been shown in any cell lines thus far, which aligns with the identified *KCNJ5* expression patterns based on TCGA. Several small molecule/ drugs targeting Kcnj5 identified through CMap database based on their high enrichment score are Amiloride and Vernakalant hydrochloride which are FDA approved.

Several kinds of cancer display distinct methylation patterns that impact the development of tumors and the effectiveness of treatment, typically in combination with changes in gene expression. PanCa, in its advanced stages, is a highly destructive illness characterized by a low overall survival rate. This highlights the pressing requirement for innovative and effective treatment protocols. In addition to genetic modifications, epigenetic dysregulation of signalling pathway-associated enriched genes is considered a new therapeutic target. Hence, a comprehensive comprehension of Methylation-regulated Differentially Expressed Genes (MeDEGs) may facilitate the development of enhanced prognostic instruments, molecular subtyping, and targeted treatments. An essential aspect of clinical model development, such as Nomogram-based patient survival prediction, is the parsing of these complex information and their integration into appropriate models. The investigation of modified DNA methylation and

its resulting effects on tumor characteristics has shown potential as a research area for the development of biomarkers, namely for PanCa. Nevertheless, the therapeutic properties of MEDEGs as biomarkers have not been sufficiently investigated on a global scale. Within the present era of digitalization, the involvement of artificial intelligence (AI) and machine learning (ML) in the healthcare sector is growing, namely in the domain of precision oncology. The capacity to classify situations into high- or low-risk categories according to precise thresholds has assisted researchers in evaluating the influence of machine learning approaches. These applications include cancer survival, prognostication, clinical-histological interpretation, the significance of epigenetic characteristics, and gene-level expression dysregulation. This work is the first of its kind to assess global PanCa datasets in order to create prognostic and predictive biomarkers using machine learning methods followed by bioinformatics-based functional evaluation. The goal is to screen the overall landscape of PanCa in a statistically meaningful manner and considering the biologically functional processes that contribute to patient survival. In previous studies published by TCGA consortium, they have used unsupervised clustering to identify purity clusters by matching mRNA expression with DNA methylation profiles of 150 samples. We have shown a unique approach by comparing differentially methylated genes (DMGs) from 9 paired samples with differentially expressed genes (DEGs) from 179 tumor samples vs 7 neighbouring normal samples, regardless of purity-based clustering. This funnel enables the discovery of target genes that are new to the detection of PanCa, taking into account the existing published literature. To summarize, our investigation has identified new gene connections (*MOGAT2*, *CD68*, *MIA*, *CD36*, *TFF1*, *S100P*, and *MUC13*) in PDAC. These MEDEGs were observed to have significant overall survival along with significant change in expression status too. Protein level status alterations along with clinicopathological correlations were also reported to be significant. Immune cell dysfunction and biological processes concerned with PanCa were also noticed to be dysregulated. This nature of the genes makes them probable prognostic biomarkers and also suitable targets for drug targeting. Followed by this we obtained a very prominent data concerning again *TFF1* and *S100P* as we found them to be dysregulated across precancerous lesions to PanCa to metastasis. These genes are likely to be grouped together in the same module if they exhibit a progressive upregulation in pancreatitis, metastatic PanCa, PanCa, and normal pancreas. Hence this shows that our data obtained have significant potential to be used as diagnostic cell type specific biomarkers. This data obtained is very useful for future use by clinicians and basic science researchers and also can open the domain of mechanistic understanding of PanCa disease progression. The findings highlight the importance of multi-omics data integration in cancer research and the potential of personalized

medicine in improving patient outcomes. However, it is crucial to highlight the need of thorough confirmation before these results can be applied in clinical settings by use of validation using patient samples.



INTRODUCTION



SEPTEMBER 16. 2024

PHD THESIS

Department of Life Sciences and Biotechnology, Jadavpur University

1.0: Introduction

Pancreatic cancer (PanCa) presents a significant challenge in oncology due to its aggressive nature, typically late-stage diagnosis, and poor prognosis. Its profound impact on public health and the healthcare system is evident through the high costs associated with specialized treatment and the even greater societal economic burden from workforce losses. Thus, it is crucial to thoroughly investigate the epidemiological characteristics of PanCa before implementing preventive measures and early screening strategies for high-risk populations (*Bararia et al. 2023*, and *Park et al. 2019*).

Pancreatic ductal adenocarcinoma (PDAC), the most common form of PanCa, has a five-year survival rate of approximately 12%, making it one of the cancers with the worst prognoses. Despite its relatively low incidence, PanCa accounts for a disproportionately large number of cancer-related deaths, with a mortality/incidence ratio of 94% (*Bararia et al. 2023*, and *O'Neill et al. 2021*). The epidemiology of PanCa involves a range of factors, including incidence, risk factors, and prognosis, which are significantly influenced by demographic and geographic disparities. Interestingly, according to GLOBOCAN 2022, there is a positive correlation between PanCa incidence rates and the Human Development Index, with higher rates observed in North America, Europe, and parts of Asia. Within these regions, incidence rates are further influenced by non-modifiable factors such as age, gender, and ethnicity, as well as modifiable factors like smoking, diet, and exercise (*Bray et al. 2024*, and *Sung et al. 2021*). PanCa is the 10th most common cancer in men and the 7th most common in women, but it is the 4th leading cause of cancer deaths in men and the 3rd leading cause in women; it accounts for about 3% of all cancers in the United States but is responsible for about 8% of all cancer-related deaths. Both the incidence and the death rates for PanCa are increasing. (*Luo et al. 2023*)

Globally, gastrointestinal cancers represent more than one-fourth of all cancer incidence and one-third of cancer-related mortality. Although there has been much progress in screening colorectal cancer (CRC), the prognosis of other gastrointestinal cancers tends to be poor. The highest burden of gastrointestinal cancers, including stomach, liver, oesophageal and gallbladder cancers (GBC), was observed in regions in East Asia. The increasing burden of gastrointestinal cancers in East Asian regions is related to population growth, ageing and the westernization of lifestyle habits in this region. Furthermore, the rising incidence of young-onset CRC is an emerging trend in East Asia (*Huang et al. 2023*).

1.1: Epidemiology of Pancreatic Cancer in Indian Subcontinent

The GLOBOCAN-2022 study, Cancer Incidence in Five Continents (CI5- XI), and the Indian Council of Medical Research (ICMR) publication on the Indian Cancer Registry Database (ICRD) are the sources of the incidence and fatality statistics (Figure 1) (Bray *et al.* 2024). The incidence is higher in the older population (over 50% in those aged 65 to 75). The Northeastern Indian regions have the highest incidence. Risk factors that cannot be changed include

genetic predispositions, age, race, smoking, excessive alcohol use, and non-vegetarian diets. There are currently no published case-control studies on risk variables pertaining to the Indian population. The rise in PanCa numbers is concerning because it primarily stems from lifestyle variables, which are changing quickly in India and other countries across the world. In 2023, there are approximately 64,050 new cases of PanCa, accounting for 3.3% of all new cancer cases (Siegel *et al.* 2023). Sadly, it is estimated that there will be 50,550 deaths from PanCa by the end of 2023, making up 8.3% of all cancer-related deaths (Figure 2 and 3) (Gaidhani *et al.* 2021, and Singh *et al.* 2022).

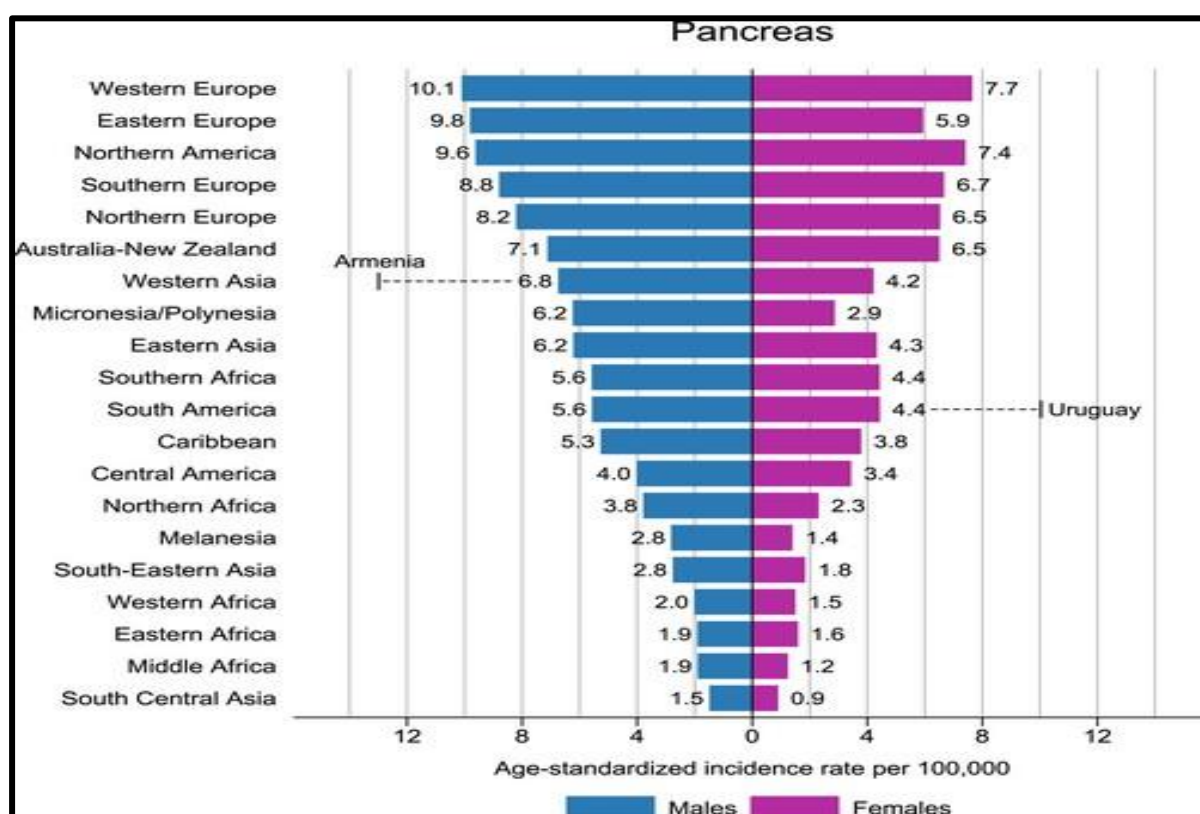


Figure 1: Bar chart of the region-specific incidence age-standardized rate by sex for pancreatic cancer in 2022 (Globocon 2022)

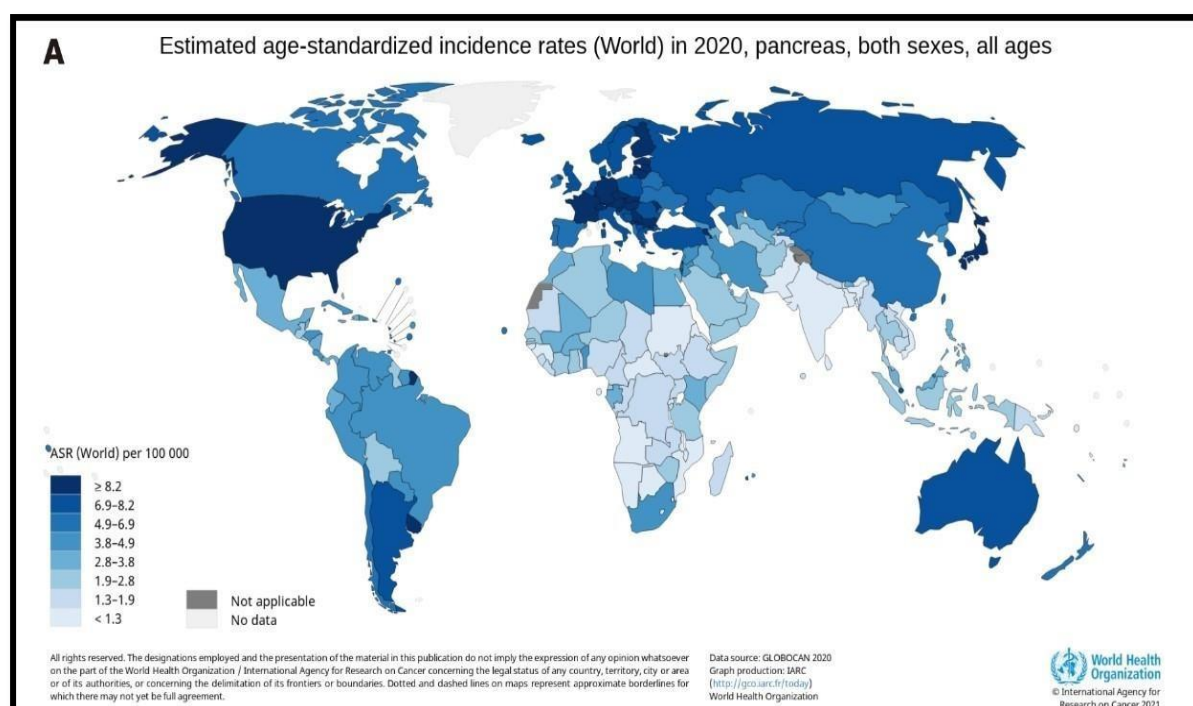


Figure 2: Worldwide incidence rate of pancreatic cancer (Globacon 2022)

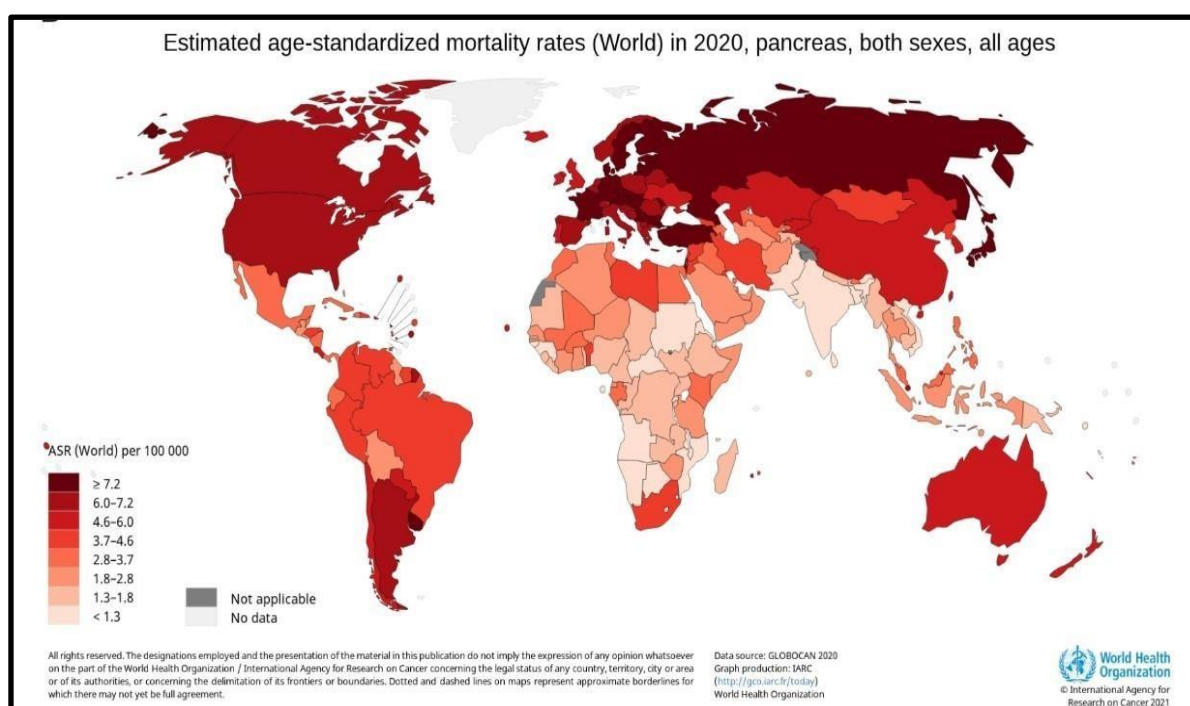


Figure 3: Worldwide mortality of pancreatic cancer (Globacon 2022)

Identifying predictive risk factors for PanCa is crucial for implementing cost-effective prevention and early screening strategies to reduce both the physiological stress on patients and the public health burden. Among non-modifiable factors, genetic predisposition is significant,

with individuals having certain hereditary inflammatory and neoplasm syndromes at higher risk. Extensive research and technological advancements are needed to better understand the genetic mechanisms underlying PanCa in predisposed populations. (*Bray et al. 2024*, and *Luo et al. 2023*)

PanCa is one of the tumors with the worst prognosis, and unlike other cancers, few advances have been made in recent years. The only curative option is surgery, but only 15-20% of patients are candidates, with a high risk of relapse. In advanced PanCa there are few first-line treatment options and no validated biomarkers for better treatment selection. (*Vázquez et al. 2024*)

1.2: Pancreas Anatomy

The pancreas is a glandular organ of about 15 cm (6 inches) long, with an internal hormonal role (endocrine) and an external digestive role (exocrine). It is located in the abdominal cavity behind the stomach with the head of the pancreas surrounded by the duodenum (*Karpińska et al. 2022*). Anatomically, the pancreas is divided into a head, which rests within the concavity of the duodenum, a body, lying behind the base of the stomach, and a tail, which ends abutting the spleen (Figure 4, and 5). The tissues with an endocrine role can be seen under staining as lightly-stained clusters of cells, called pancreatic islets (also called islets of Langerhans). Darker-staining cells form clusters called Acini which are arranged in lobe separated by a thin Fibrous barrier. The secretory cells of each acinus surround a small Intercalated Ducts (Figure 6). Because of their secretory function, these cells have many small granules of zymogens that are visible (*Fenge et al. 2016*). These intercalated ducts empty into bigger ducts located within the lobule, and ultimately into interlobular ducts. One layer of columnar epithelium lines the channels. Multiple layers of columnar cells may be observed as the diameter increases. The pancreas contains around 3 million compact cell clusters known as pancreatic islets. Included within these islets are four distinct cell types that play a role in the control of blood glucose levels. Every cell type produces a distinct hormone: α alpha cells release glucagon (which raises blood glucose levels), β beta cells release insulin (which lowers blood glucose levels), ϵ delta cells release somatostatin (which regulates and inhibits α and β cells), and PP cells, or γ (gamma) cells, release pancreatic polypeptide. (Figure 7). They act to control blood glucose through secreting glucagon to increase the levels of glucose, and insulin to decrease glucose levels (*Hezel et al. 2006*).

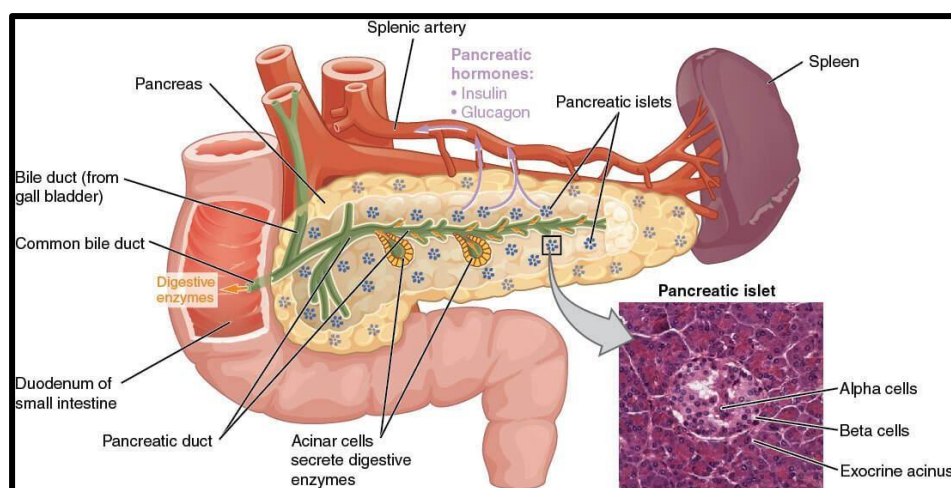


Figure 4: Anatomical structure of pancreas (<https://liverdrashok.com/pancreas/>)

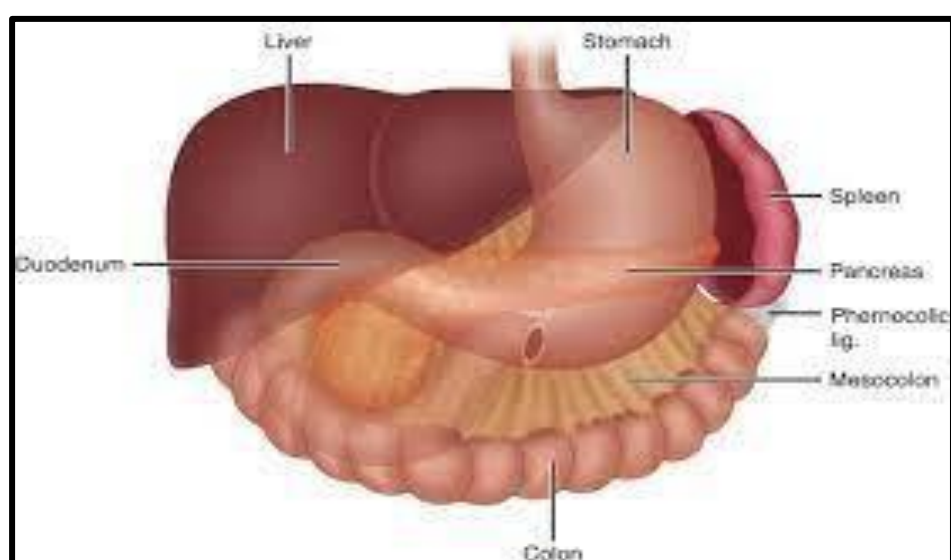


Figure 5: Structures in relation with pancreas (<https://liverdrashok.com/pancreas/>)

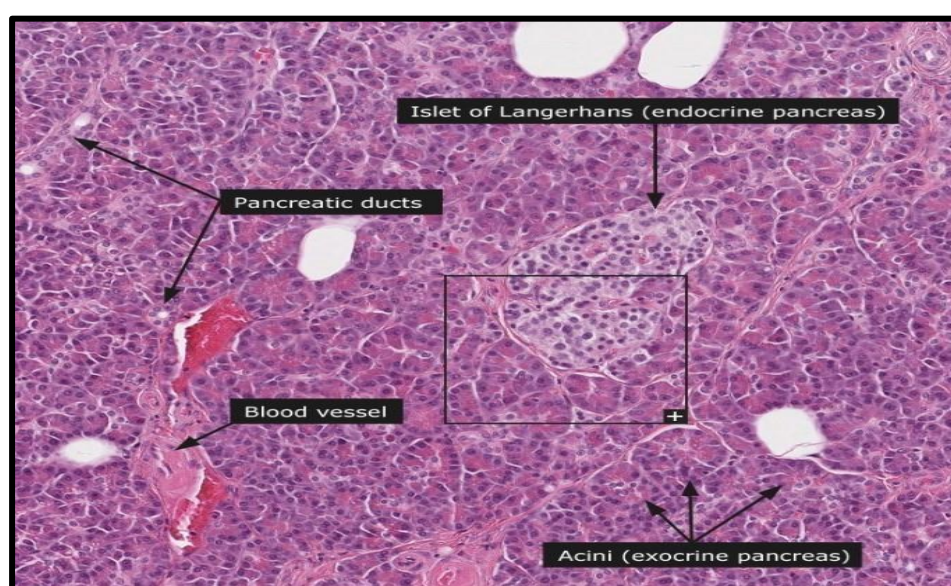


Figure 6: Histology of normal pancreas (Longnecker et al. 2021)



Figure 7: Normal pancreatic histology (Magnification view) (Longnecker et al. 2021)

1.3: Types of Pancreatic Cancer

More than 95% of PanCas are classified as exocrine tumors. These tumors start in the exocrine cells of the pancreas- PDAC, acinar cell carcinoma (ACC), pancreatic neuroendocrine tumors account for less than 5% of all pancreatic tumors. They may be benign or malignant and they tend to grow slower than exocrine tumors. They develop from the abnormal growth of endocrine (hormone- producing) cells in the pancreas called islet cells. This is why these tumors are sometimes referred to as islet cell tumors (Sikdar et al. 2018).

1.3.1: Pancreatic Ductal Adenocarcinoma

Pancreatic Ductal Adenocarcinoma presents as a poorly demarcated, firm white–yellow mass, with the surrounding non-malignant pancreas typically showing atrophy, fibrosis, and dilated ducts due to the obstructive effects of expanding carcinoma. Microscopically, these neoplasms vary from well-differentiated gland- forming carcinomas to poorly differentiated sarcomatoid carcinomas. Pancreatic ductal adenocarcinoma evolves from well-defined precursor lesions that, in the context of their genetic features, define the genetic progression model of pancreatic carcinogenesis. Early disease histology PDAC presents as a poorly demarcated, firm white–yellow mass, with the surrounding nonmalignant pancreas typically showing atrophy, fibrosis, and dilated ducts due to the obstructive effects of expanding carcinoma (Figure 8 and 9) (Bararia et al. 2023).

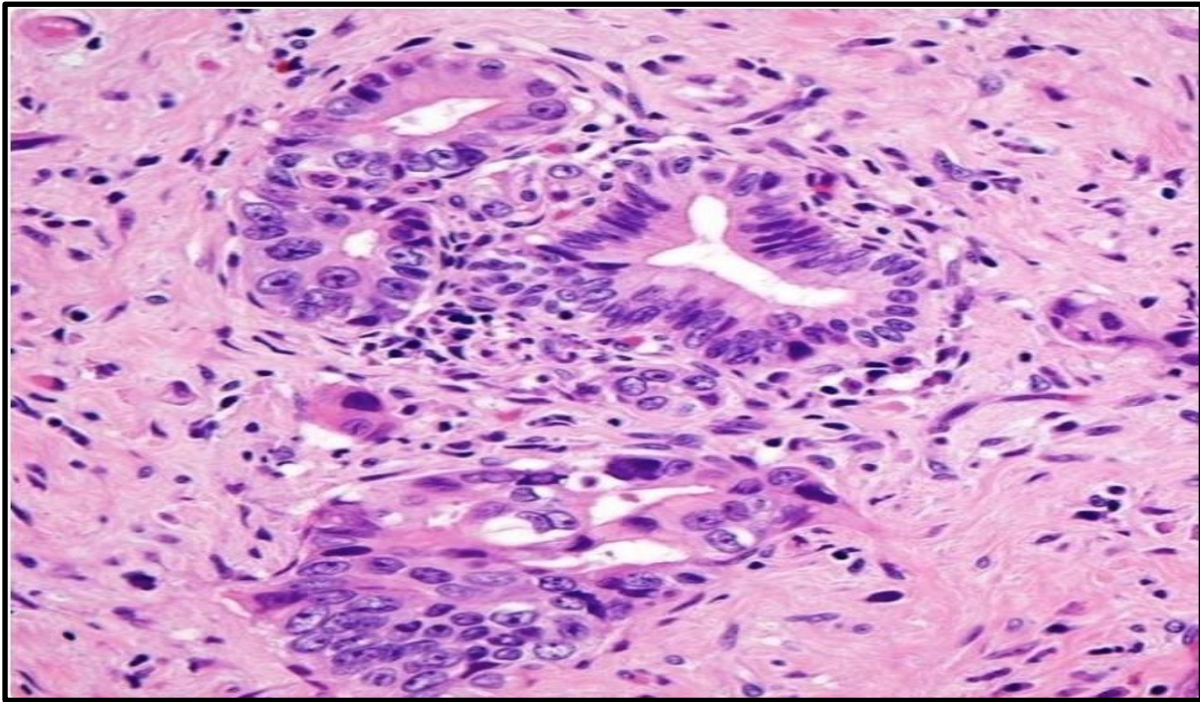


Figure 8: Histopathology (HPE) picture of pancreatic ductal adenocarcinoma (Marked nuclear pleomorphism with 4x anisonucleosis) (*Longnecker et al. 2021*)

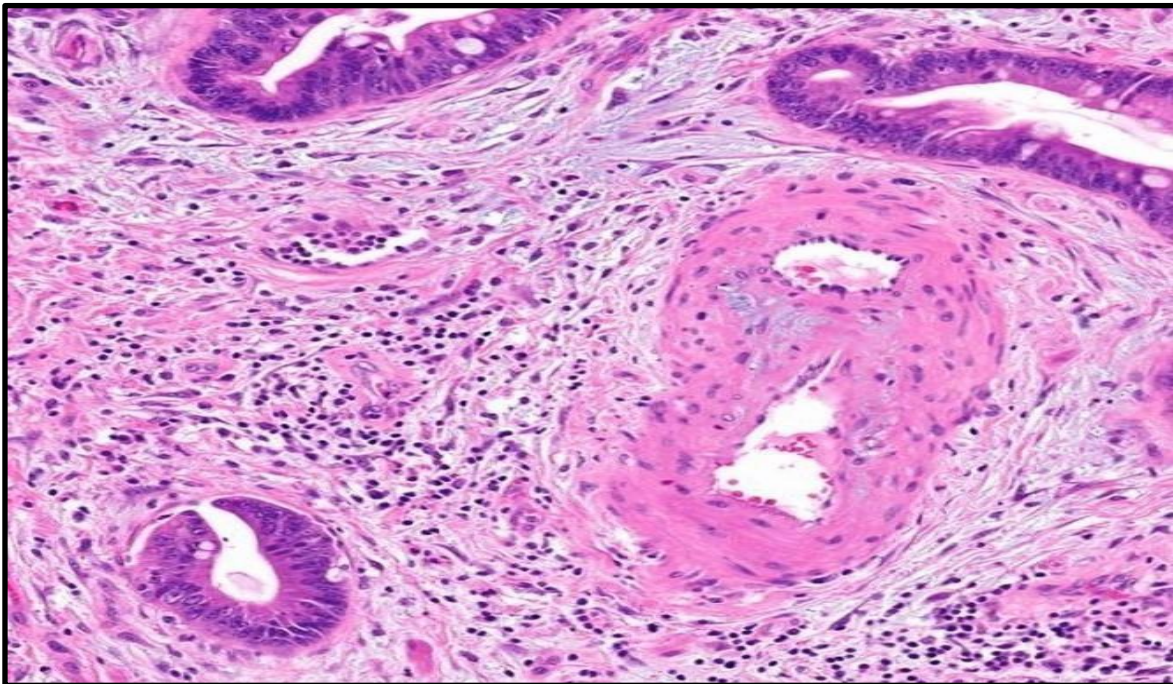


Figure 9: HPE picture of pancreatic ductal adenocarcinoma (Atypical glands immediately adjacent to artery) (*Longnecker et al. 2021*)

Under microscopic examination, infiltrating ductal adenocarcinoma of the pancreas is a malignant epithelial tumor that invades the pancreas and has glandular (ductal)

differentiation. It does not have a prevalent component of any other type of cancer. (*Schawkat et al. 2020*). The extent of gland development can range from fully developed glands, to partially developed glands, to localized intracellular mucin synthesis by poorly aligned cells infiltrating individually, to entire sheets of malignant cells. Many characteristics can be utilized to differentiate between benign reactive glands and infiltrating ductal adenocarcinoma, and the majority of these factors are related to the specific site of the tumor. If assessing the position of a duct, it is important to remember that non-cancerous glands, especially in cases of severe chronic pancreatitis, tend to develop predictable lobular units. The ducts are located in the middle of the lobular unit and are encircled by clusters of acini called grape-like structures. (*Hruban et al. 2014*). The glands violate the lobular architecture, and are strewn in the parenchyma without rhyme or reason. The neoplastic glands can be found adjacent to muscular arteries without intervening pancreatic arenchyma (*Warshaw et al. 1992*, and *Hruban et al. 2007*).

1.4: Precancerous Forms of Pancreas

Initial disease histology presents with various discrete forms of precursor lesions, with the most prevalent being microscopic Pancreatic Intraepithelial Neoplasia (PanIN), followed by larger cysts known as the Intraductal Papillary Mucinous Neoplasm (IPMN) and Mucinous Cystic Neoplasm (MCN) (*Patra et al. 2017*). The Non-invasive PanIN lesions were formerly classified into 3 grades according to the extent of cytological and architectural atypia - PanIN 1A (flat lesion) and PanIN 1B (Micro papillary type) show low grade dysplasia; PanIN 2 exhibits additional loss of polarity, nuclear crowding, cell enlargement, and hyperchromasia with frequent papillary formation; PanIN 3 is advanced lesions with severe nuclear atypia, luminal necrosis, and manifest epithelial cell budding into the ductal lumen (*Bararia et al. 2020*). A documented increase in incidences of PanINs in patients with PDAC initially suggested their biologic relationship. The high grade PanIN is ultimately transform into frank PDAC with evidences of areas of invasion beyond the basement membrane (Figure 10) (*Ciernikova et al. 2020*). Although PanINs are the most characterized precursors, other non-invasive lesions are also thought to precede PDAC formation. These are IPMNs and MCNs named as they produce mucins (*Nasca et al. 2020*). PanINs are columnar, mucinous epithelial structure characterized by progressive architectural disorder. Pancreatic ductal neoplasms are distinguished by their bigger size and their involvement of the main ductal pancreatic duct or its larger ductal branches. Metabolic carcinomas do not originate in the main pancreatic ducts. Instead, they are characterized by their ovarian type stroma, which exhibits varying

degrees of epithelial dysplasia and localized foci of invasion. Morphologically unrelated to the ductal system, MCN often manifests in females. Serous cystic neoplasms (SCNs) are managed as non-malignant benign growths. It is well recognized that superficial cystic neoplasms (SCN) seldom exhibit a correlation with the main pancreatic duct (MPD), which aids in distinguishing them from other cystic lesions, such as IPMN. Very uncommon instances in which the suprachiasmatic nucleus is linked to the metastatic prostate cancer (MPC) (Figure 11) (Park et al. 2023, and Jonckheere et al. 2010)

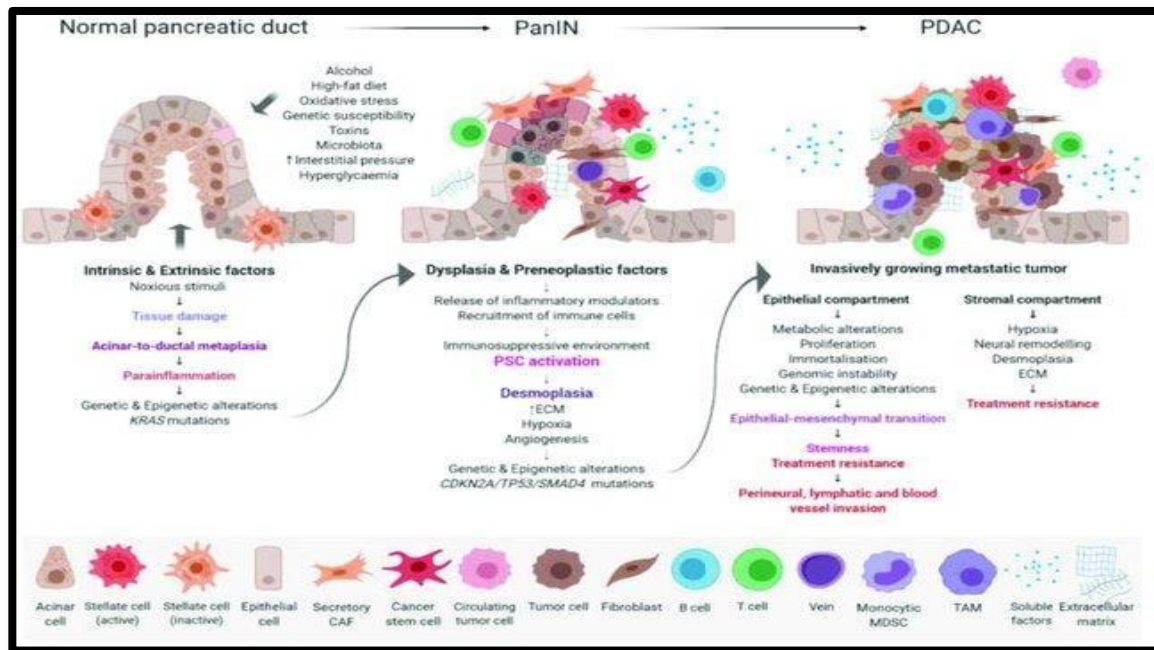


Figure 10: Stages of different pancreatic intraepithelial neoplasia (Ciernikova et al. 2020)

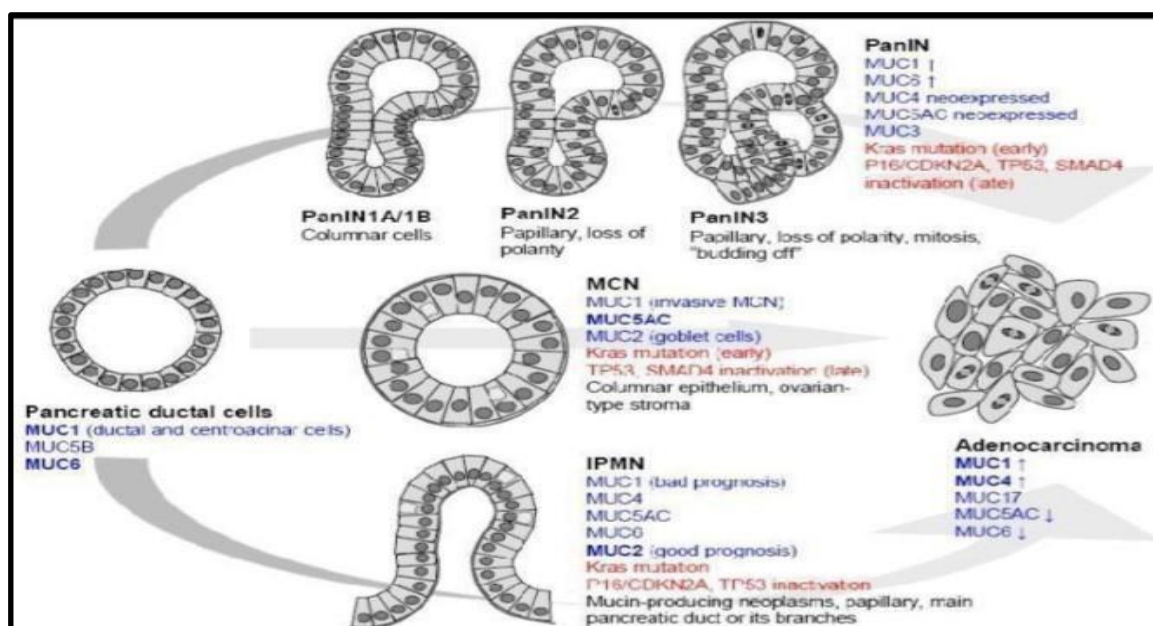


Figure 11: Different precancerous forms of pancreatic cancer (Jonckheere et al. 2010)

1.5: Clinical Presentation of Pancreatic Ductal Adenocarcinoma

Pain and weight loss are the most typical PDAC symptoms, and jaundice is the most typical clinical marker. To be more precise, one study listed the incidence of signs and symptoms in PDAC patients in the following order, descending: Anorexia (lack of appetite) 85%, weight loss 85%, abdominal pain 79%, and dark urine 70% were the results of the survey (*Wood et al. 2022*).

There may be associated weight loss, anorexia, pain or cholangitis. Duodenal ampullary cancers may present with anemia secondary to gastrointestinal bleeding. PanCa's may also present with paraneoplastic syndromes 59%, Jaundice 56%, Nausea 51%, Back pain 49%, Diarrhea 44% and Vomiting 33% (*Maitra et al. 2008*).

1.6: TNM Staging of Pancreatic Ductal Adenocarcinoma

A approach for categorizing malignancies is called the TNM Classification. It helps with prognostic cancer staging and is mainly employed in solid tumors. A common classification scheme facilitates improved information exchange and cross-population research while also enhancing provider-provider communication. The method is predicated on evaluating the metastatic distant, local lymph nodes, and tumor (*Rosen et al. 2023*).

1.6.1: T Category

Tx Primary tumor cannot be assessed T0 No evidence of primary tumor Tis Carcinoma in situ

T1 Tumor ≤ 2 cm in greatest dimension

T2 Tumor > 2 cm and ≤ 4 cm in greatest dimension T3 Tumor > 4 cm in greatest dimension

T4 Tumor involves celiac axis, superior mesenteric artery, and/or common hepatic artery, regardless of size

1.6.2: N Category

Nx Regional lymph nodes cannot be assessed N0 No regional lymph node metastases

N1 Metastasis in one to three regional lymph nodes N2 Metastasis in four or more regional lymph nodes

1.6.3: M Category

M0 No distant metastasis

M1 Distant metastasis (Table 2)

1.6.4: Tumor Grade

Based on the degree of tumor differentiation, pancreatic adenocarcinoma is divided into four grades. These are; GX: Grade cannot be assessed, G1: Well differentiated adenocarcinoma (WDA), G2: Moderately differentiated adenocarcinoma (MDA), G3: Poorly differentiated adenocarcinoma (PDA). Patients with poorly differentiated tumors have worse prognosis than those with well differentiated tumors (*Xi et al. 2018*).

1.7: Etiological Factors

Although exact causes of PDAC and PAC still remain unclear, several environmental, lifestyle and genetic risk factors have been identified by different studies. But there are no extensive epidemiological study on PAC. Risk factors for PanCa can be classified in two groups: i. Modifiable risk factors, ii. Non modifiable risk factors (*Midha et al. 2016*).

1.7.1: Modifiable Risk Factors

The modifiable risk factors that pose risk of PanCa are smoking, alcohol, obesity, dietary factors and occupational exposure.

1.8: Lifestyle Risk Factors

1.8.1: Smoking

Tobacco smoking is one of the major risk factors for PanCa. The International Agency for Research on Cancer (IARC) has stated that tobacco smoking is causally associated with PanCa. It is found that the risk is nearly two times higher in smokers compared to non-smokers. Recently a meta-analysis based on 82 studies related to PanCa reported that the Relative Risk (RR) of PanCa was 1.74 for current smokers while 1.2 for former smokers (*Iodice et al. 2008*; and *Rastogi et al. 2008*). Most epidemiologic studies reported strong associations between cigarette smoking and PanCa with RR ranging 1.5 - 3.0 for current smokers compared to non smokers (*Rawla et al. 2019*). Although the association between smoking and PanCa is not as strong as the smoking-lung cancer relationship, but it is found that even after smoking cessation the risk persist till 10 years. The European Prospective Investigation into Cancer (EPIC) study in 2012 showed that PanCa risk elevated for every 5 cigarettes smoked per day (*Vrieling et al. 2010*).

In addition to that, it is also found passive smoking increases the risk of PanCa by 50%. In developing countries like India, where tobacco use is high in population, it remains a major concern for PanCa risks. Interestingly, risk of PanCa associated with tobacco use increases more with addition of confounding factor such as alcohol consumption (*Hassan et al. 2007*).

1.8.2: Alcohol

Most studies have reported either “no” or only weak association between alcohol consumption and risk of PanCa. Several reports suggest (*Gupta et al. 2010*), risk of PanCa is significantly increased by higher consumption of alcohol (more than three drinks per day), whereas in case of low-to-moderate alcohol consumption “no” association was found (*Wang et al. 2016*). A case-control study with large number of patients in 2010 showed intake of 60 ml/day or more of liquor (spirits) increase the risk of PanCa but found “no” association with beer, wine or other alcoholic beverages (*Michaud et al. 2010*). Alcohol drinking in conjunction with tobacco smoking significantly increases the risk of PanCa at a much higher rate than consuming alcohol alone. Similarly, a recent study revealed that heavy alcohol consumption was linked to a substantial rise in PanCa risk among current smokers (age-adjusted odds ratio (OR) = 4.04, 95% CI: 1.58 - 10.37), but not among non-smokers (age-adjusted OR = 2.01, 95% CI: 0.50 - 8.18). (*Rahman et al. 2015*). In addition, it is also reported that low-to-moderate alcohol consumption is associated with increased PanCa risk among current smokers. Thus, it is much challenging to identify alcohol as a strong independent risk factor for PanCa.

1.8.3: Obesity

Several studies have shown that high caloric intake, obesity or both are the risk factors for PanCa as well as many other cancers. It is still uncertain, but there is an indirect link between obesity and inflammatory response. Few studies reported obesity is associated with PanCa incidence as well as mortality. Interestingly, a study by *Li et al.* reported BMI ≥ 30 kg/m² in early adulthood associated with higher risk of PanCa (*Li et al. 2009*). Furthermore, it was also found that obesity at an older age was associated with lower overall survival in PanCa. Based on study of American Cancer Society (ACS) it was found that RR of PanCa among obese group was higher (RR=2.08) compared healthy group (*Calle et al. 2003*). In addition, physical inactivity mostly results in accumulation of fat overweight and increased risk of PanCa.

1.8.4: Dietary Factors

There are reports that suggest several dietary factors are associated with PanCa. Several foods impose higher risk while few is reported as protective. Intake of red meat, processed meats, fried foods, cholesterol, and foods that contain nitrosamines may increase risk of PanCa. This indicates carcinogens in meat and preservatives such as nitrite or N-nitroso compounds which widely used for processed meats maybe involved in PanCa. A meta-analysis (*Rawla et al.*

2019) result based on 11 case-control studies showed intake of red meat increased PanCa risk by about 48% (95% CI: 1.25 - 1.76). However, intake of fruits and vegetables that are enriched in citrus and antioxidants, showed protective nature, by decreasing PanCa risk by 29% (95% CI: 0.59 - 0.84) and 38% (95% CI: 0.54 - 0.73) respectively. Another meta-analysis (Larsson *et al.* 2012) also reported an association between PanCa incidence and high intake of red (120 g/day) or processed meat (50 g/day) with risks (RR = 1.13 and RR = 1.19) respectively. However, few studies failed to validate such findings and stated association found only in men. Interestingly, a couple of study reported that consumption of nut frequently lowers the risk of PanCa in women. In addition, it was found in a large UK cohort study in 2016 that PanCa mortality was lesser in vegetarians (about 50% lower) as well as low or non-meat eaters (about 30-45% lower) compared to regular meat eaters (Appleby *et al.* 2016).

1.9: Environmental Risk Factors

Among environmental risk factors diabetes, obesity, smoking, alcohol, increased oil fat consumption, coffee intake, decreased physical activity are associated with increased risk of PDAC. The proportion of PDAC that may be attributable to occupational exposures have been estimated to be 12% (Ojajärvi *et al.* 2000). Chemical substances that cause cancer (called carcinogens) include asbestos, pesticides, dyes, and petrochemicals. Exposure to them may be linked to PanCa. Among the environmentally acquired occupational risk factors, chlorinated hydrocarbons,

organochlorines, ionizing radiation, airborne particles, nitrosamines have been associated with increased risk of PDAC. The strongest associations were seen between smoking and diabetes mellitus and smoking and family history of PanCa. A person who possesses both of these risk factors is far more likely to develop PDAC since the combined effect of smoking and either risk factor appears to be cumulative.

.1.10: Occupational Risk Factors

The proportions of PDAC that may be attributable to occupational exposures have been estimated to be 12% (Rawla *et al.* 2019). Among occupational risk factors, chlorinated hydrocarbons, organochlorines, ionizing radiation, airborne particles, nitrosamines have been associated with increased risk of PDAC.

A meta-analysis of occupational exposures in PanCa stated higher risk with nickel exposure. Carcinogenic role of nickel is thought to be due to increasing DNA methylation and

inhibiting DNA repair (*Ojajarvi et al. 2000*). However, in occupational environment, nickel mostly associated with polychlorinated biphenyls, which could be reason for increased PanCa risk. In addition, PanCa risk has been reported by few studies due to exposure of cadmium and arsenic. Cadmium, which is a non-essential metal accumulates in the human pancreas thus increases the risk and mortality of PanCa. It causes trans-differentiation of pancreatic cells, DNA repair inhibition and induces or regulates several oncogenes or tumor-suppressor genes (TSGs), which are found to express in PanCa. Similarly, exposure to arsenic also found to be associated with cancer risk, but a very little study is reported for PanCa risk. A recent study showed a link between exposures to milk powder contaminated with arsenic in childhood and almost two- fold increase in PanCa mortality (*Yorifuji et al. 2011*).

Chlorinated hydrocarbon exposure is one of the most established occupational risk factors for PanCa. In industries, several workers are exposed to chlorinated hydrocarbons, such as, solvents, pesticides and plastics. Compounds like trichloroethylene, polychlorinated biphenyl methylene chloride, vinyl chloride, and tetra-chloro-ethylene are reported to increase risk of PanCa. Organochlorines, such as DDT, which are used worldwide for agricultural and pest control, is found to be associated with increase in PanCa risk. It was found in a case– control study that among chemical manufacturing workers in US, individuals, ever exposed to DDT, was associated with a 4.8-fold risk, and workers with a mean exposure of 47 month had a 7.4-fold risk of PanCa compared to those who were never exposed (*Sikdar et al. 2018*). Interestingly, in a study from Spain, it was found concentrations of DDT were significantly higher in *KRAS* mutated PanCa cases than controls ($p < 0.01$), but similar among *KRAS* wild type cases and controls (*Porta et al. 2009*).

1.11: Non-Modifiable Risk Factors

The non-modifiable risk factors that are associated with risk of PanCa are gender, age, ethnicity, diabetes mellitus, chronic infections, non-O blood group, chronic pancreatitis, and genetic factors. Currently, screening and manipulating non-modifiable factors are not the most cost-effective ways to significantly reduce PanCa incidence. However, modifiable factors should be prioritized in public health initiatives. Comprehensive smoking cessation and outreach programs integrated into primary care have shown effective results in increasing cessation rates and cost-effectiveness

1.11.1: Gender

Worldwide, incidence of PanCa is 5.5 per 100,000 for men and 4.0 per 100,000 for women

(Bray *et al.* 2018). This suggests for male prevalence of this cancer. This is possibly due to the reasons that men are more exposed to previously mentioned occupational exposure as well as lifestyles factors such as heavy smoking and alcohol consumption (Hu *et al.* 2021).

1.11.2: Age

Based on SEER Cancer Statistics as well as all previous reports, it is well documented that PanCa is predominant in older age people (>50 years). The highest peak age for occurring PanCa is between 60 and 80 years (Siegel *et al.* 2018). The possible explanation may be: it takes many years for a pancreatic lesion or inflammatory condition to eventually switch into a malignant neoplasm. Although this explanation needs to be validated by conclusive evidence.

1.11.3: Ethnicity

Worldwide variable incidence rates of PanCa are observed in different races and geographical locations. For example, PanCa incidence rates are much higher in African-Americans than Caucasians (McGuigan *et al.* 2018). On the other hand, PanCa incidence is the lowest in Asian-Americans and Pacific Islanders (Rawla *et al.* 2019). One possible reason for such variability could be due to genetic factors and modifiable risk factors such as diet, alcohol, smoking and vitamin D insufficiency (Gad *et al.* 2020).

1.11.4: Diabetes Mellitus

Both the type I and II diabetes are associated with increased risk of PanCa and reported in many studies. Study from Italy estimated that diabetes is linked to almost 9.7% of PanCa (Rosato *et al.* 2015). It is also reported that diabetes mellitus increases 1.8-fold risk of developing PanCa, particularly in Asians and Hispanic men compared to Whites and Blacks (Li *et al.* 2011). In a case-control study it was reported RR of PanCa for diabetic people compared to non-diabetics was found 2.1 (95% CI 1.6-2.8) (Li *et al.* 2011). Additionally, another case control study reported incidence of diabetes was almost 3 times higher in the cases (16%) than controls (6%) (OR = 2.52, 95% CI of 1.04-6.11) (Li *et al.* 2011). One interesting study observed that risk of PanCa decreases with duration of diabetes, while 30% risk increases when diabetes persists for more than 2 decades after diagnosis (Bosetti *et al.* 2014). In addition, they also concluded that diabetes plays a role in non-familial fashion.

1.12: Chronic Infections

Helicobacter pylori (*H. pylori*) colonization is also associated with greater risk of PanCa. Report from a meta-analysis of 7 studies showed an increased risk of PanCa in people infected

with *H. pylori*, specifically the CagA-positive strain (Risch *et al.* 2010). Possible explanation is that *H. pylori* colonization enhances the pancreatic carcinogenic effect of N-nitrosamines in pancreas due to smoking or dietary sources. In contrast, several studies failed to observe any correlation between *H. pylori* infection and the risk of PanCa. Further studies needed to conclude *H. pylori* as an independent risk factor for PanCa. In addition, few studies also reported association between PanCa with chronic infections such as hepatitis B and C virus (HBV and HCV) (Ben *et al.* 2011).

1.13: ABO Blood Group

Interestingly, it was found Non-‘O’ blood types account for 17% of all new PanCa, as was reported by a study based on almost one million person and years of follow up. However, the exact mechanism for this association is unknown. Different studies showed different blood groups are associated with PanCa risk. For example, a study in the UK reported blood group ‘A’ individuals has more risk of developing PanCa, while, study in Italy reported an increased risk of PanCa among ‘B’ blood group individuals (Zhang *et al.* 2014). Whereas, a study in the USA found increased risk of PanCa for individuals who had blood types A, B and AB compared with O blood group (Wolpin *et al.* 2009).

1.14: Chronic Pancreatitis

Chronic Pancreatitis (CP) is an independent risk factor for PanCa. It is found that PanCa risk increases 16-fold in patients with CP (Gandhi *et al.* 2022). However, only 3%-4% of PanCa cases are found to be attributed due to CP. Association of CP with PanCa are strong in patients with hereditary pancreatitis, which arises mainly during or soon after childhood (Rawla *et al.* 2019).

1.15: Genetic Risk Factors

Recent studies have demonstrated that PanCa is definitely hereditary, and that the chance of developing the illness is much increased in families where PanCa has been present. A small proportion of PanCa cases are caused by specific genetic alterations, while the majority are produced by random mutations. PanCa is primarily brought on by acquired and hereditary gene alterations. The risk of familial PanCa increases exponentially with the number of first-degree Relatives (Yousef *et al.* 2024). Chromosomal anomalies have currently been discovered in loci 13q22.1, 15q14,17,18, 6p25.3, 12p11.21, 7q36.2,19, 21q21.3, 5p13.1, 21q22.3, 22q13.32, and 10q26.120 by Chinese and Japanese scientists. In the largest genome-wide collection in Europe, 7p12, 1p36.33, 8q21.11, 17q12, and 18q21.32 were identified as new

PanCa susceptibility chromosomal target deletions. The most prevalent point mutations are found in *K-RAS*, *CDKN2A (P16)*, *TP53*, *SMAD4*, *BRCA2*, *BRCA1*, *STK11*, *PRSS1*, *MMR24*, and *MMR25*. PanCa stem cells reportedly display epigenetic changes. mostly emerged via the process of chromatin mutation. controlling proteins and the epithelial mesenchymal transition (EMT), but these changes do not include changes to the genetic makeup. only the chemistry and structure of chromatin Since changes are occurring, the overall state of the cell's phenotypic is eventually impacted by these changes. Given these ideas, a number of researchers are beginning to investigate whether severing the epigenetic regulatory system facilitates the creation of novel treatments for pancreatitis. (Zhao et al. 2023)

1.16: Inherited Risk Factors

Of the lifestyle risk factors, current cigarette smoking had the strongest connection with PDAC. A meta-analysis of 12 case-control studies, comprising 6507 PanCa cases and 12 890 control patients (absolute rates not available), yielded an odds ratio (OR) of 1.74 (95% CI, 1.61–1.87) for the link between current smoking and PanCa (Iodice et al. 2008). A moderate correlation between alcohol consumption and PDAC was found in a meta-analysis of 19 prospective studies using data from 4 211 129 people when intake surpassed 30 g daily, or around three drinks per day (relative risk, 1.22 [95% CI, 1.03–1.45]; absolute rates not reported). A 13-fold increased risk for PDAC was found in a pooled analysis of 14 prospective cohort studies with 862 664 people who had chronic pancreatitis (relative risk, 13.3 [95% CI, 6.1–28.9]) (Genkinger et al. 2009). Obesity, defined as a body mass index in the top 20% of the quintile (the fifth quintile), was also associated with an increased risk of PDAC (incidence, 14.1 vs. 5.7 per 100 000 person-years; adjusted relative risk, 1.54 [95% CI, 1.04–2.29]) (Johansen et al. 2010). This was found in a Norwegian study involving 940 060 participants. Sugar-filled drinks, processed meat, and saturated fat have been associated with an increased risk of diabetes, pancreatitis, and obesity. PDAC is becoming more common in people under 30 (Sung et al. 2019). From 1995 to 2014, the annual incidence of PDAC increased with younger age, consistent with an increase in younger patients. The percentages for those 45 to 49 years old were 0.77% (95% CI, 0.57%–0.98%), those 30 to 34 years old were 2.47% (95% CI, 1.77%–3.18%), and those 25 to 29 years old were 4.34% (95% CI, 3.19%–5.50%). (Park et al. 2021)

1.17: Familial Pancreatic Cancer

Familial PanCa (FPC) is defined as a family with at least one pair of first-degree relatives (parent- child or sibling pair) with PanCa without an identifiable syndrome in the family. Germline mutations of different DNA repair genes have been identified in familial PanCa (Hruban *et al.* 1999). The underlying mechanism is that the buildup of damaged DNA can induce several states, including unregulated cell proliferation, apoptosis, and senescence, which can ultimately lead to the development of malignant cells. The PanCa gene family includes *PALLD*, *BRCA1*, *BRCA2*, *PALB2*, *FANCC*, *FANCG*, *ATM*, *CDKN2A*, and *CHEK2*. (Bartsch *et al.* 2012) The protein palladin is encoded by the *PALLD* gene in humans. Palladin is a constituent of actin-containing microfilaments that regulate the morphology, attachment, and contraction of cells. Significant linkage to chromosome 4q32-34 was observed in a linkage analysis of a large PanCa pedigree characterized by autosomal dominant inheritance of PanCa with early onset and strong penetrance. (Geile *et al.* 2006). After few years an oncogenic germline mutation (Pro239Ser) was identified in the affected members of the family. Thus, it was suggested *PALLD* gene is a major susceptible for PDAC (Eberle *et al.* 2002).

1.18: Hereditary Pancreatitis

Within the entirety of inherited cancer predisposition syndromes, Hereditary Pancreatitis (HP) stands out as the only syndrome in which PanCa is the exclusive risk factor for cancer. Hunter syndrome is an uncommon variant of cerebral palsy, characterized by its autosomal dominant inheritance pattern and strong penetrance (>80%). It usually manifests before the age of 30, although its occurrence is influenced by ancestral origin within the family. Genetic mutations in the *PRSS1* gene result in an autosomal dominant manifestation of the disease, while mutations in the *SPINK1* gene result in an autosomal recessive mode of inheritance. The *PRSS1* gene encodes the precursor for the synthesis of an enzyme known as cationic trypsinogen. Cationic trypsinogen is synthesised in the pancreas and facilitates the process of food digestion. Too early activation of trypsinogen results in the development of acute pancreatitis (AP). (Whitcomb *et al.* 1996). The R117H germline mutation is a commonly seen germline mutation in *PRSS1*. Consequent to this mutation, the enzyme becomes unable of being enzymatically degraded, even in the absence of its calcium binding. Nevertheless, this mutation eradicates the trypsin hydrolysis site. A small number of uncommon mutations were also detected in cationic trypsinogen. This list provides the documented mutations of cationic trypsinogen: N29I, A16V, D22G, K23R, A121T, and R122C. R116C is an uncommon

genetic mutation that can lead to protein misfolding and propose another mechanism of activation. (Kereszturi *et al.* 2009). Patients with HP have recurrent acute pancreatitis (AP), which can often evolve into CP. Their lifetime risk for PDAC is 35-fold (or more) by ages 70-75 (Lowenfels *et al.* 1997).

1.19: Hereditary and Other Factors

In addition to senior age, other recognized host variables associated with many distal common bile duct malignancies include inflammatory bowel disease, sclerosing cholangitis, choledochal cysts, and intrahepatic or common bile duct stones. Furthermore, there is a documented correlation between bile duct tumors and specific geographic regions, identified by clusters in certain areas of the United States. Hereditary polyposis syndromes such as Hereditary non-polyposis colorectal cancer (HNPCC), Peutz-Jeghers syndrome, familial adenomatous polyposis, and Gardner's syndrome are associated with an increased incidence of duodenal and ampullary malignancies (Table 3) (Virnig *et al.* 2009).

1.20: Diagnostic Evaluation and Treatment Modalities for Pancreatic Ductal Adenocarcinoma

1.20.1: Diagnosis

The signs and symptoms of PanCa's are more or less similar. Signs and symptoms of PanCa often don't appear until the disease is advanced. The commonly associated symptoms are; abdominal pain that radiates to back, loss of appetite or unintended weight loss, yellowing of skin and the whites of eyes (jaundice), light-colored stools, dark- coloured urine, itchy skin, new diagnosis of diabetes or existing diabetes that's becoming more difficult to control, blood clots, fatigue, nausea, vomiting, fever, pain in abdomen etc. A tumor marker called CA19.9 is the only marker which is often checked for suspected individuals. However, elevated CA19.9 is also associated with many cancers as well as other inflammatory conditions (Lee *et al.* 2020, and Kim *et al.* 2020).

Likewise, symptoms diagnosis of PDAC is also very similar. Once cancer is suspected, it is diagnosed and staged by a high resolution, thin cut, triple-phase CT scan or PET scan. Usually, an endoscopic ultrasound (EUS) tests are done for imaging pancreas and associate organs inside the abdomen (Lee *et al.* 2014). Most frequently the tissue is collected during EUS by passing special tools through the endoscope for biopsy. Sometimes, a sample of tissue is collected from the suspected area by inserting a needle through the skin into the pancreas. Simultaneously, other blood-based tests, such as jaundice, kidney function, hemoglobin and blood clotting parameters are also done (Iwasaki *et al.* 2021).

1.20.2: Survival Rates of Pancreatic Cancer

The five-year survival rate for PanCa has increased to 13%, up for the third year in a row according to the American Cancer Society's Cancer Facts & Figures 2024. Ten years ago, the five-year survival rate stood at a just 6%, therefore the rise indicates significant advancement as more patients attain extended lifespans. The primary drivers of the trend in survival rates include patients diagnosed with localized disease and improved therapy of individuals diagnosed at earlier stages.

According to the report, an estimated 66,440 Americans will be diagnosed with PanCa in 2024, and 51,750 are predicted to die from the disease this year (*Qasim et al. 2024*). PanCa continues to be the third- leading cause of cancer-related death in men and women combined in the United States and is on track to become the second-leading cause before 2030.

These statistics reflect a rise in both deaths and the number of cases, and experts note an increase in the incidence of PanCa in younger women. The cause is unknown but potential factors include increases in obesity and shifts in the racial/ethnic demographics within the U.S. (*Herremans et al. 2021*)

The survival rate of PanCa varies depending upon the stage and type of cancer as stated by Max healthcare centre, India. Localised PanCa: For patients with localized, resectable tumors, the 5-year survival rate is around 20% (*Puckett et al. 2022*). Early detection and surgical removal of the tumor offer the best chance of long-term survival. Locally Advanced PanCa: When the cancer has spread to nearby tissues but is still resectable, the 5-year survival rate drops to approximately 12% (*McGuigan et al. 2018*). Metastatic PanCa: Unfortunately, most patients are diagnosed with metastatic disease, where the cancer has spread to distant organs. In such cases, the 5-year survival rate is only about 3%. Pancreatic Neuroendocrine tumors (PanNETs): Survival rates for PanNETs tend to be more favorable. The 5-year survival rate for localised PanNETs is around 80%, emphasizing the importance of early detection. (*Williams et al. 2023*)

1.20.3: Early and Late Onset of Pancreatic Cancer

The study by *Ramai et al.*, identified 136,100 patients with PanCa, with 1181 (8.7%) being early-onset and 134,919 (91.3%) late-onset. Both groups showed an increase in incidence, with early-onset PanCa showing a 3.69% annual change and late-onset PanCa a 6.25% annual change (*Ramai et al. 2021*). Early-onset PanCa was more frequently located in the head of the pancreas, while late- onset PAC showed a rising incidence in the body and tail. Most cancers were poorly differentiated, and the majority presented at Stage IV. Surgical resection rates

were higher in early-onset PanCa (19%) compared to late-onset PanCa (13%). Median survival times were 7 months for early-onset and 6 months for late-onset PanCa (*Leonhardt et al. 2022*). Cox regression analysis showed that late-onset PanCa had a higher risk of mortality compared to early-onset PanCa. Both early- and late-onset PanCa are increasing in incidence, with poor overall prognosis. Tumor anatomy trends differ by age group, with early-onset PanCa more common in the head of the pancreas and late-onset PanCa showing a rise in the body and tail. Survival outcomes are similar between age groups, with worse prognosis associated with cancer in the tail of the pancreas. (*Ramai et al. 2021*)

1.20.4: Treatment Strategy of Pancreatic Cancer

Treatment of PDAC depends on the stage and location of the tumor as well as overall health. The primary goal of treatment is to eliminate the cancer, when possible. Secondly, when that isn't an option, the focus may be on improving quality of life and restricting the cancer from growing or causing more harm (*Anderson et al. 2021*). In regular practice, treatment of PDAC include surgery, radiation, chemotherapy or a combination of these. The surgical procedure is known as pancreatoduodenectomy. When the cancer is in advanced stage then only treatment option remain is palliative care (symptomatic treatment) to keep the patient as comfortable as possible for as long as possible (*Wei et al. 2021*).

The development of targeted therapies in PanCa is increasingly feasible due to tumor-agnostic treatments, such as *PARP* inhibitors in patients with *BRCA1*, *BRCA2* or *PALB2* alterations or immunotherapies in patients with high microsatellite instability/tumor mutational burden (*Ragupathi et al. 2023*). In addition, other therapeutic molecules have been developed for patients with *KRAS* G12C mutation or fusions in *NTRK* or *NRG1*. Consequently, there has been a growing interest in biomarkers that may help guide targeted therapy in PanCa (*Vázquez et al. 2024*). *KRAS* is a critical target for therapeutic evaluation in PDAC and proof-of-principle approaches have validated targeting in *KRAS* G12C; ongoing trials are evaluating allele-specific inhibitors and pan-*KRAS* inhibitors against the more common allele variants G12D, G12V and G12R (*Waters et al. 2018*). Immune-checkpoint inhibitor blockade and other immune therapies have not had utility for most patients with PDAC; however, select rare individuals beyond mismatch repair deficiency have seen major benefit, leading to optimism that these results can be expanded to a broader patient population (*Bian et al. 2021*).

In Asia, S-1 adjuvant-based chemotherapy is primarily adapted to treat PDAC and its dose intensity highly influences survival potential (*Kumar et al. 2023*). PanCa patients with S-1

dose intensity of compared to its lower dose, which had a survival time of 20.2 months, $\geq 62.5\%$ demonstrated favorable in terms of survival, with a median length of 53.3 months. According to this research, S-1 adjuvant chemotherapy, which keeps the dose intensity over 60%, is a useful treatment against PanCa for extending postoperative survival. (Kobayashi *et al.* 2022).

Immune check-points inhibitors like antibodies PD-1/PD-L1 and *CTLA-4* are applied for immunotherapy against cancers (Araki *et al.* 2022). However, due to restricted mutation burden and weaker cytotoxic T cell infiltration, PDAC is less susceptible to immunotherapy. Using immunological therapy in conjunction with molecules that target toll-like receptors (TLRs) can be a successful cancer treatment strategy. (Fávaro *et al.* 2022). Erlotinib, an immunomodulatory medication associated with gemcitabine, functions by obstructing the tyrosine kinase of an epidermal-growth factor receptor. Compounds like protein aggregate magnesium-ammonium phospholinate-palmitoleate anhydride (P-MAPA) are emerging potential candidates against PanCa involving the *TLR4* signalling pathway (Fávaro *et al.* 2022). P-MAPA and gemcitabine-coupled therapy efficiently reduced the progression of neoplastic lesions and histopathological improvement by up to 80% in PanCa over only 40% repair using P-MAP alone. P-MAPA and P-MAPA/gemcitabine therapies enhance *Tlr4* protein contents that improve interferon signalling pathways via abnormal cell proliferation suppression (Fávaro *et al.* 2022).

In PDAC, thrombospondin-1 (*TSP-1*) associated with cancer-associated fibroblasts (CAFs) induces downregulation of *Smad4* expression and enhances malignant potential via activation transforming growth factor- β (*TGF- β*) signals (Matsumura *et al.* 2022). *SMAD4* involves *TGF- β* superfamily cytokines signal transduction by target genes transcriptional activation and plays a vital role in pancreatic tumor suppressor. Deletion of the *SMAD4* gene (*DPC4*) activates faster cell invasiveness and metastatic capability. In vivo, *TSP-1* acts as an essential activator of *TGF- β* . In PDAC cells, *DPC4* knockdown triggered *TSP-1* overexpression along with *TGF- β* signal initiation. Also, the overexpression of *TSP-1* caused *SMAD4* expression downregulation and improved the proliferation of cells. LSKL peptide (*TSP-1* inhibitor) treatment prevents *TSP-1* binding to a latency-associated peptide (LAP) and *TGF- β* signal activation that results in anticancer impacts via attenuation of cell proliferation, migration, and chemoresistance in PDAC cells. However, LSKL peptides as *TSP-1* inhibitors can potentially be therapeutics in PDAC patients (Sinha *et al.* 2020, and Deng *et al.* 2024).

The eco-evolutionary perspective offers a new framework for understanding PDAC, emphasizing the importance of the TME and the need for dynamic treatment strategies.

Epigenetic therapy is a novel treatment strategy that targets the epigenetic modifications of cancer cells, aiming to alter their gene expression and behavior without directly changing the DNA sequence (Versemann *et al.* 2022). This approach has shown promise in various cancers, including PDAC, due to its dynamic nature and ability to modulate the tumor microenvironment (TME). Epigenetic therapy represents a promising approach for treating PDAC by targeting the dynamic epigenetic landscape of cancer cells. By modulating DNA methylation, histone modifications, and miRNA expression, these therapies aim to alter the fitness of PDAC cells and overcome the limitations of current treatments (Elrakaybi *et al.* 2022).

1.21: DNA Methylation Inhibitors

These drugs target DNA methyltransferases (DNMTs) to reduce DNA methylation which include 5-azacytidine and decitabine. Some DNMT inhibitors have FDA approval. Trials have shown partial responses and increased sensitivity to standard chemotherapy. For instance, a phase I trial of azacitidine in combination with nab-paclitaxel showed a partial response in PDAC. These drugs can decrease tumor growth and increase drug sensitivity by altering DNA methylation patterns (Kagan *et al.* 2023).

1.21.1: Histone Deacetylase Inhibitors (HDACi)

HDACi have shown low efficacy and high toxicity in PDAC due to the dense stroma and CAFs within the TME. Combining HDACi with other treatments like chemotherapy, immunotherapy, and small molecule inhibitors has shown more promising results. Valproic acid, a class-I HDACi, was found to induce apoptosis in PDAC when used with etoposide (Sim *et al.* 2022).

1.21.2: Histone Modification Inhibitors

Inhibition of G9a (lysine-preferring mammalian histone methyltransferase), a histone methyltransferase, has been effective in increasing sensitivity to gemcitabine and reducing tumor metastasis and growth. G9a inhibition was effective both as a monotherapy and in combination with other chemotherapy drugs. When a G9a inhibitor was used with gemcitabine, it increased sensitivity to gemcitabine in PDAC cell lines and animal models, and reduced tumor metastasis and growth. Additionally, other studies demonstrated that dual targeting of G9a and Checkpoint Kinase 1 (*CHK1*) with prexasertib (LY2606368) and the G9a HMT inhibitor (BRD4770) decreased cell growth in PDAC (Urrutia *et al.* 2020).

1.21.3: EZH2 Inhibitors

EZH2, which is overexpressed in PDAC, can be targeted to decrease drug resistance and cell proliferation by reversing H3K27 methylation. Inhibiting EZH2 reversed H3K27 methylation, which reduced cell growth and reduced resistance to gemcitabine and doxorubicin-stimulated apoptosis. Additional treatments that focus on GSK126, one of the components of *EZH2*, has been shown to prevent the migration of cancer cells in a variety of cancer cell lines (*Jones et al. 2018*).

1.21.4: AURKA Inhibitors

Inhibition of Aurora kinase A (*AURKA*) can interfere with mitotic progression and reduce PDAC cell growth (*Zhang et al. 2022*).

1.21.5: Bromodomain Inhibitors

These inhibitors have shown increased survival in genetically engineered mouse models (GEMMs), decreased tumor growth, and reduced CAFs in the TME. This inhibitor, when combined with gemcitabine, has reduced tumor volume in PDAC mouse models (*Menezes et al. 2022*).

1.21.6: miRNA Targeting

miR-23b and miR-205 Overexpression of these miRNAs has been shown to increase sensitivity to radiotherapy and reduce chemotherapy resistance, respectively. miR-34a Introduction of miR-34a nanocomplexes has increased apoptosis and stabilized cell growth (*Chauhan et al. 2020*).

1.22: Epigenetics and Pancreatic Cancer Development

Epigenetic modifications, such as abnormal methylation of DNA, histone deacetylation, chromatin remodeling, impact gene expression and play a crucial role in the development of cancer. The process of DNA methylation is of paramount importance in determining illness outcome. Modified methylation status at the CpG islands located at the gene promoters restricts the availability of transcription factors and has a substantial impact on gene expression (*Chatterjee et al. 2018*). Extensive documentation has shown that hypermethylation at CpG islands is associated with decreased expression of TSG, while hypomethylation is associated with enhanced expression of oncogenes in diverse types of cancer. Epigenomics has emerged as a highly promising approach for diagnosing PDAC and has correctly identified certain pathways that can be focused on for treatment (*Lomberg et al.*

2019). Past researches have tried to identify altered methylation of cell-free DNA in PDAC as the potential blood-based diagnostic and prognostic biomarkers. However, limited validation studies across patient population and applicability of such biomarkers were seen till date. Thus, it is important to study the differential methylation marks associated with PDAC development (Vrba *et al.* 2022, and Henriksen *et al.* 2021).

Past studies had shown that both genetic and epigenetic alterations contributed to PDAC initiation and progression (Lomberk *et al.* 2019). PDAC diagnosis based on gene-mutation has only brought partial success in early diagnosis and overall patients' survival improvement (Lomberk *et al.* 2019, and Bararia *et al.* 2020). Distinct epigenetic markers have been observed for specific subtypes of PanCa's and for specific stages of PDAC (Biankin *et al.* 2012, and Mishra *et al.* 2017). Also, studies including PanCa samples from different populations have identified partially similar methylation marks. The inter-population epigenetic variation has been reported and this may be explained by varied ethnicity, demographic, and occupational factors. Globally few studies have been done on investigating the changing epigenetic landscapes through progressive stages of PDAC and on different ethnicity. (Chatterjee *et al.* 2022)

1.22.1: DNA Methylation

In the process of DNA methylation, a methyl group is added at the 5'-carbon position of cytosine nucleotide mostly in the context of CpG dinucleotide (Figure 12) (Nones *et al.* 2014 and Saini *et al.* 2013.).

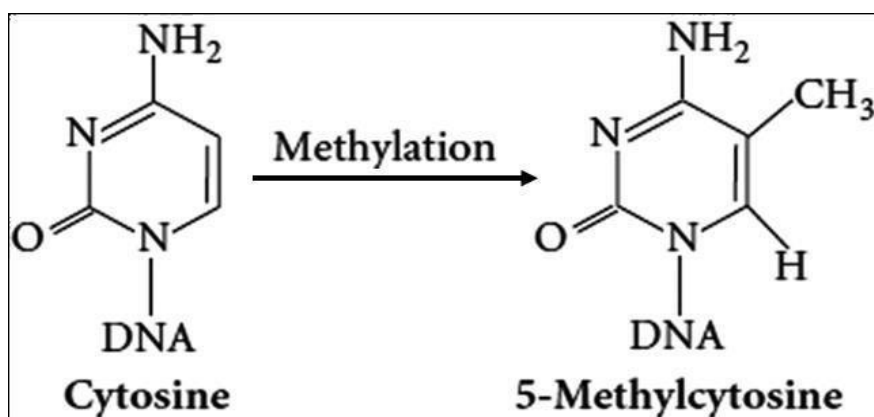
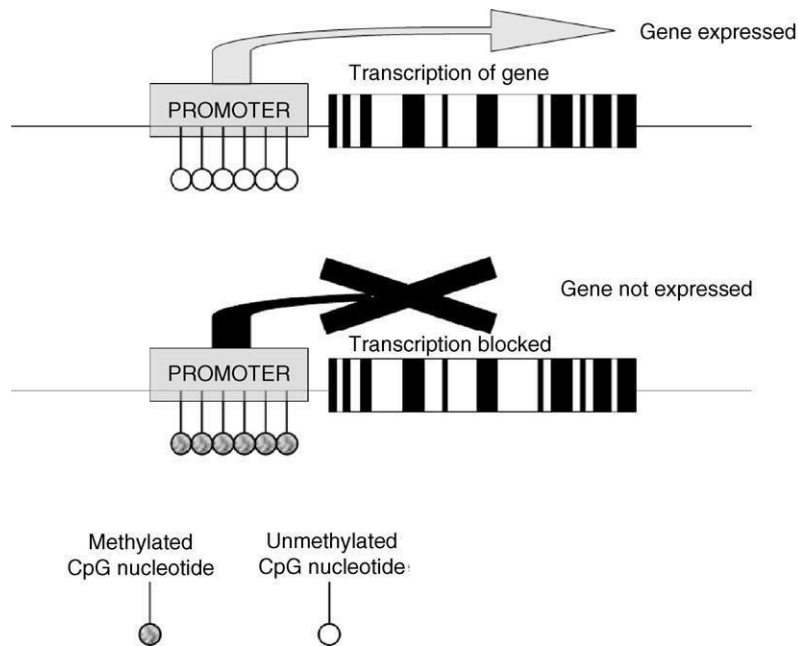


Figure 12: Methylation of cytosine (Saini *et al.* 2013).

Cytosine base (left) in DNA is usually methylated to 5-methylcytosine (right) in context of CpG dinucleotides (Figure 12). Frequency of the CpG dinucleotides is generally depleted across the human genome as the 5-methylcytosine (5mC) frequently tends to deaminate into

thymine. However, clusters of CpG dinucleotides are found in the human genome and are termed as CpG Islands (CGI). The CGIs are described as 200-4000 bp long regions in the genome with >50% GC composition, and observed/expected ratio of CpG dinucleotides greater than 0.6. Around 50% of the promoters in the human genome are associated with CGIs (Illingworth *et al.* 2009; and Nones *et al.* 2014). CGIs constitute for about 1% of the genome and ~15% of the total CpG present in the genome, and a total of about 45,000 CGIs have been identified so far (Antequera *et al.* 1993; and Robertson *et al.* 2000). Based on the CpG content, the promoters are divided into 3 groups viz; Low CpG containing promoters, High CpG containing promoters and the intermediate CpG containing promoters. High CpG promoters are usually associated with housekeeping genes and are mostly found to be unmethylated whereas the low CpG containing promoters are mostly reported to be hypermethylated (Weber *et al.* 2007; and Bararia *et al.* 2020). The enzymes that carry out the methylation of DNA are known as DNMTs (Shaw *et al.* 2006). Three major DNMTs have been identified that play a role in maintaining the epigenetic marks of the cell. DNMT1, the first discovered methyltransferase, is identified and reported as the maintenance methyltransferase which adds methyl group to the newly synthesized DNA strand following replication by recognizing the hemi-methylated CpG sites in the double stranded DNA. DNMT1 maintains the existing methylation marks in the DNA during replication (Bestor *et al.* 1988). DNMT1 is found abundantly in cells and show 10-40 folds higher affinity for hemi-methylated sequences (Pradhan *et al.* 1997, and Robertson *et al.* 1999). The DNMT3 family (consists of DNMT3A and DNMT3B) of enzymes were discovered and characterized almost after a decade of the discovery of DNMT1 (Okano *et al.* 1998, and Robertson *et al.* 2000). DNMT3A and DNMT3B, together known as the de-novo methyltransferases have equal preferences for unmethylated and hemimethylated DNA sequences and are identified as important players for establishing new methylation marks on CpG sites of the genome (Robertson *et al.* 2000). Cytosine methylation maps of human somatic cells have revealed that exons, introns and intergenic regions that have low CpG density are usually methylated and CGIs are unmethylated (Lister *et al.* 2009; and Laurent *et al.* 2010). The CpGs associated with the repeat regions in the genome are also found to be methylated in somatic cells (Lister *et al.* 2009). Methylation of CpG in the promoter regions are usually correlated with transcriptional repression as the methylation marks prevent transcription factors from binding to its cognate binding sites (Figure 13) (Shaw *et al.* 2006).



Figures 13: DNA methylation bringing about repression of gene expression (*Shaw et al. 2006*)

1.22.2: Identification of DNA Methylation

In last few years, several methods have been developed to identify the methylation pattern in DNA. Initial experiments relied on digestion of DNA with methylation sensitive restriction endonucleases (MSREs). The digested DNA samples were then resolved following southern blot to identify DNA methylation (*Issa et al. 1994*). Gradually digestion of genomic DNA with MSREs was followed by polymerase chain reaction (PCR) to identify methylated cytosines in genomic DNA (*Storz et al. 2020*). Digestion of DNA with restriction enzymes followed by bisulfite treatment of DNA was then identified and used to study methylation patterns in DNA. Following bisulfite modification of DNA, cytosines are converted to uracil whereas the methylated cytosine moieties remain unchanged following the reaction (*Clark et al. 1994*). Combination of the bisulfite modification of DNA, followed by amplification of the DNA and subsequently digestion of the modified and amplified DNA with restriction endonucleases was developed into a new methodology known as Combined Bisulfite Restriction Analysis (COBRA) for identifying DNA methylation (*Xiong et al. 1997*). Following bisulfite modification, the modified DNA was amplified using a PCR reaction, with specific primer sets mostly devoid of CpG dinucleotides, in a process called the Bisulfite sequencing PCR (BSP) (*Clark et al. 1994*). The BSP product was then used for Sanger sequencing either directly or following cloning (*Frommer et al. 1992*; and *Clark et al. 2006*). Direct sequencing of BSP product, although a time efficient method, is technically

challenging and has low sensitivity (*Leontiou et al. 2015*). In case of cloning and sequencing of the bisulfite modified DNA, the PCR products are first cloned into a suitable vector followed by transformation in a suitable *E.coli* strains. Following transformation, the region of interest is sequenced from the obtained clones. The methylation status of each CpG site is obtained as an average of the methylation status of the resulting clones, which then forms the DNA methyl haplotypes (*Frommer et al. 1992*, and *Clark et al. 2006*). Moreover, Methylation specific PCR (MSP) was identified as an effective way of studying the methylation status of CpGs in regions of interest. Two sets of primers, one designed for the methylated sequence and the other for the unmethylated sequence were used separately (*Herman et al. 1996*). Real time-based methods like Methyl Light, that uses probes, have also been deployed to study methylation patterns in regions of interest of DNA samples (*Eads et al. 2000*). Methylation detection using dyes like SYBR Green soon modified MSP reactions into quantitative MSP and were found to be cost effective than probe-based techniques. Methylation sensitive melting curve analysis were also implemented to study CpG methylation that identifies methylation pattern following bisulfite reaction using the differences in melting curve of cytosine and uracil bases (*Worm et al. 2001*). Pyrosequencing method, based on the detection of a pyrophosphate group has been widely used to study methylation patterns following bisulfite modification of DNA in a quantitative manner (*Colella et al. 2003*). Modern methods have also employed digital PCR platforms to study DNA methylation (*Liesbeth Van Wesenbeeck et al. 2018*). Gradually the focus of methylation studies has shifted towards identifying genome wide DNA methylation changes. Genome wide DNA methylation techniques usually employ hybridization methods or sequencing platforms. Usually, treatment of genomic DNA by sodium bisulfite or enrichment of methylated DNA is preceded before the genome wide methylation study. Enrichment based methods for identifying genome wide DNA methylation usually employ MSRE digestion or methyl-specific antibodies. This is then followed by size selection or capturing of CpG-rich regions by probe hybridization. Genome wide microarrays were initially developed to study cancer associated DNA methylation changes following probe hybridization. Illumina Infinium Golden Gate was first developed to study methylation status of 1536 CpG sites across human genome of which 1505 CpG sites were from over 807 cancer related genes identified previously (*Bibikova et al. 2006*, and *Fan et al. 2006*). In order to study genome-wide DNA methylation, Illumina later introduced the Infinium HumanMethylation27 Bead Chip array or the 27 K array where the methylation status of a total of 27,578 CpGs across the human genome could be profiled (*Bibikova et al. 2009*). Later the target CpG sites of the

array were increased to almost 450K in the Infinium Human Methylation 450 Bead Chip or the 450k array. The array included 485,577 CpGs sites of which 482,421 sites were CpG dinucleotides and 3091 non-CpG sites and 65 random Single Nucleotide Polymorphisms (SNPs) in the human genome (*Bibikova et al. 2011*). Illumina has recently introduced the Infinium Methylation EPIC Bead Chip array where methylation status of more than 850,000 CpG sites across the human genome could be profiled. Apart from most (90%) of the previously included CpG sites of the 450K array, it contains CpG from enhancer regions described in FANTOM5 and ENCODE projects (*Pidsley et al. 2016*). Analysis of the array data consists primarily of image processing and normalization steps. Usually, image processing requires the Illumina Beadscan software and normalization is done in Illumina Genome-Studio software (*Wang et al. 2013*). A number of R/Bioconductor packages are available for analyzing Infinium Methylation data (*Gentleman et al. 2004*, and *Aryee et al. 2014*). Many software packages have also been developed to define differentially methylated regions in the genome (*Butcher et al. 2015*). Sequencing of bisulfite modified DNA samples using Next generation sequencing platforms are also available to study genome wide methylation (*Lister et al. 2009*). Methylation of a particular CpG is obtained first by mapping the sequences to the reference genome and finally calculating the number of cytosine and thymine bases in the mapped reads for the particular CpG in question (*Bock et al. 2012*). For genome wide methylation study, whole genome bisulfite sequencing (WGBS) is available, that allows comprehensive sequencing of approximately 28 million CpGs in the human genome, but the method is not cost effective and requires high genomic coverage for methylation analysis (*Lister et al. 2009*). Moreover, reduced representation bisulfite sequencing (RRBS) is an alternative and efficient way to study genome wide methylation, as only CpG-rich target regions are selected and sequenced (*Meissner et al. 2005*). On the other hand, in the enrichment-based methods the methylated DNA can be precipitated with the help of antibodies, following which high throughput sequencing of the DNA is done. Methylated DNA immunoprecipitation sequencing (MeDIP-seq) and Methyl-CpG binding domain (MBD) based capture and sequencing (MBD-seq) are commonly used to study the epigenome (*Down et al. 2008*). Use of methyl sensitive restriction enzyme to capture methylated DNA which are then sequenced in high throughput experiments of MCRBC-seq are also used in genome wide methylation study (*Nagarajan et al. 2013*). The sequencing data are then analyzed using different bioinformatic tools.

1.23: Cancer Biomarkers Diagnosis, Prognosis and Its Clinical Implications and Therapy

A biomarker refers to any biological marker found in blood, body fluids, or tissues that signals the presence of normal or abnormal biological processes, conditions, or diseases. When tailored to the field of oncology, a cancer biomarker specifically identifies characteristics of cancer, ideally with a high degree of accuracy and reliability, reported as their sensitivity and specificity. The use of cancer biomarkers extends beyond merely determining the type of cancer a patient suffers from. Indeed, once a diagnosis is established, tumor markers can provide valuable insights into the likely progression of the disease, including the chances of recurrence and the expected outcomes of treatment. Cancer biomarkers play a crucial role in outlining the prognosis of a disease independently of any treatment (known as prognostic biomarkers) or in predicting how a cancer will respond to a specific treatment, which helps anticipate treatment outcomes (referred to as predictive biomarkers) (*Passaro et al. 2024*)

1.24: Identification of Pancreatic Cancer Biomarker Using Multi Omics Dataset, Machine Learning Technologies and Approaches

An effective early screening system for PanCa, essential for monitoring and reducing its incidence, requires advancements in detecting predictive biomarkers through non-invasive measures and a comprehensive risk assessment model incorporating patient history, symptoms, genetic susceptibility, etc. Currently, identifying predictive biomarkers for early PanCa detection remains challenging (*Kato et al. 2020*). However, research into accessible and accurate biomarkers and innovative diagnostic tools is ongoing. Advances in liquid biopsy, for instance, enable the detection of cellular and sub-cellular components non-invasively, including circulating tumor cells, circulating tumor DNA, and exosomes (*Bararia et al. 2023*). This technology plays a pivotal role in understanding early-stage PanCa. Patterns in patients' medical history require extensive analysis, often facilitated by machine learning algorithm (MLA) and artificial intelligence (AI) models. These models can integrate information from various biomarkers, imaging data, and medical records, enhancing early diagnosis and evaluation of PanCa (*Kourou et al. 2014*). In an ideal scenario, several technologies would be integrated to facilitate the diagnosis of diseases and to provide a comprehensive understanding of human phenotypes and symptoms. Nevertheless, the processing of multi-omics data presents novel informatics and interpretation difficulties. Particularly, there is a need for innovative analytical and statistical techniques to merge different data sets, together with defined quality control measures (*Albaradei et al. 2021*). Additionally, the field must address challenges in the interpretation of molecular events and, accordingly, their actionability

and whether they can guide therapeutics and clinical care. Choosing the right MLA techniques is essential for creating realistic risk prediction models. Deep learning techniques in particular, which operate on big and noisy sequential datasets, are particularly useful (*Tran et al. 2021*).

Advances in omics technologies — such as genomics, transcriptomics, proteomics and metabolomics — have begun to enable personalized medicine at an extraordinarily detailed molecular level. Individually, these technologies have contributed medical advances that have begun to enter clinical practice. However, each technology individually cannot capture the entire biological complexity of most human diseases. Integration of multiple technologies has emerged as an approach to provide a more comprehensive view of biology and disease (*Karczewski et al. 2018*). Multi-omics methodologies are innovative frameworks that combine several omics datasets produced from the same patients in order to gain a deeper understanding of the molecular and clinical characteristics of malignancies. A diverse array of developing omics and multi-view clustering algorithms now offer unparalleled possibilities to further categorize cancers into subtypes, enhance the prediction of survival and therapeutic outcomes for these subtypes, and comprehend fundamental pathophysiological processes across various molecular scales. A multi-omics study is a scientific inquiry that utilizes data analysis to examine many high-dimensional datasets at different levels and scales in order to uncover the intricate nature of cells and their associated surroundings. Such type of study can provide novel frameworks to untangle biological phenomena or models to test certain hypotheses using various datasets. The recently developed high-throughput technologies in genomes and transcriptomics, increased effort in large-scale research collaboration, and progress of computational algorithms have led to a paradigm change in cancer research towards multi-omics approaches (*Heo et al. 2021*).

Data integration may be categorized into two primary groups: (a) integrated analysis of data that includes information on a single type of variation, such as combining expression data from various studies (single-platform integration), and (b) integrated analysis of data that includes information on different types of variation, such as integrated analysis of expression and methylation data (*Thingholm et al. 2016*). Machine learning techniques can be employed to decipher intricate patterns presented in biological data. Unsupervised and supervised model learning are the two main categories of machine learning. An unsupervised machine learning approach is commonly employed for subgroup identification in cases when the labels for the data are not known. With supervised machine learning, the objective is to establish the correlation between a collection of input features and their matching labels. In

addition to the precision of this correlation, feature selection is a vital procedure in supervised machine learning and offers valuable understanding of the fundamental biology that underlies the data labels, which is essential for the translational interpretation. The random forest (RF) is a supervised machine learning technique that is extensively used for classification issues. It is renowned for its exceptional predicted accuracy and its capability to effectively handle high-dimensional and mixed-type data. Feature selection and variable importance measures (VIMs) are implicit techniques within Random Forest (RF) that are evaluated using the Gini impurity criteria index. (Zhou *et al.* 2018, and Toth *et al.* 2019). This index provides a measure of the prediction power of features in classification problems based on the principle of impurity reduction (Strobl *et al.* 2007). VIMs play an essential part in the identification of biologically relevant biomarkers, but can become unreliable in high-dimensional and mixed-type data (Parmar *et al.* 2015). In cancer classification, this is often the case when gene expression (real-valued continuous), copy number alterations (integer) and clinical characteristics (categorical/binary) are integrated. As such, features of high cardinality are more likely to be favored and binary features that contain information that could predict a data label are ignored. (Quist *et al.* 2021)



REVIEW OF LITERATURE



SEPTEMBER 16, 2024

PHD THESIS

Department of Life Sciences and Biotechnology, Jadavpur University

2.0: Pancreatic Cancer and Its Epigenetics Landscape

Epigenetic modifications give rise to characteristic carcinogenic signaling pathways that arise from the dysregulation of transcription, therefore establishing a distinctive feature of PanCa. 5-Hydroxymethylcytosine (5-hmC) is a genomic modification of DNA at regulatory areas that leads to the production of 5-methylcytosine (5-mC) residue. This modification has been extensively investigated in PanCa. A genome analysis was conducted on the cell lines of short-passaged PanCa, focusing on the 5-hmC occupied genomic regions. Consequently, unexpected patterns of modification were seen in neoplastic tissues of patients with primary malignancy (Neureiter *et al.* 2014). It was observed that near the open chromatin regions, the 5-hmC was very much enhanced and thus tends to show upregulation of the allied transliteration (Guler *et al.* 2020). The transcripts encompass several crucial oncogenic signaling pathways implicated in pancreatic neoplasia, including *KRAS*, master regulator of cell cycle (*MYC*), *BRD4*, and *VEGFA*. *BRD4* is often significantly upregulated in natural settings. The functional method involved the introduction of 5-hmC at promoter *BRD4*, which was associated with an increase in transcript expression particularly in samples from primary patients. (Fujikura *et al.* 2021). Furthermore, it was observed that the proliferation of PanCa is significantly suppressed by the *BRD4* blockade in the in-vivo trials. To summarize, it can be stated that partisan enhancement and 5-hmC reallocation are significant regulatory mechanisms in human PanCa and oncogenic enhancers (Lomberk *et al.* 2019).

Aside from genetic alterations, another specific field of study centers on epigenetic regulation, which is crucial in the formation, persistence, and advancement of diseases. Epigenetic modifications associated with cancer progression include the suppression of Tumor suppressor genes (TSG) by DNA methylation, deactivation of DNA repair genes by histone acetylation, and control of cell proliferation, differentiation, and apoptosis by miRNA (miR) (Jones *et al.* 2002). The precise meaning of epigenetics is modification in gene expression without concomitant changes in DNA bases. Progression of PDAC can be attributed to the primary epigenetic modifications, namely DNA methylation. The process of DNA methylation is intricately linked to CpG islands. Some 60% of human genes are linked to CpG islands. (Jones *et al.* 2012). However, it is clear that there are normal tissue specific patterns of CpG island methylation and some CpG islands are prone to progressive methylation during aging and development of certain diseases such as PanCa (Rodríguez *et al.* 2011). The process of DNA methylation involves the covalent addition of a methyl group to the 5 positions of cytosine. This modification is sustained by a collection of genes known as DNA methyl-transferases (DNMTs) (Issa *et al.* 2004). DNA methylation is categorized

into two groups that are hypermethylation and hypomethylation (Kanwal *et al.* 2011). DNA hypermethylation of other genes except TSG has also been strongly linked with cancer initiation and progression (Esteller *et al.* 2011). DNA hypomethylation was observed in cancer cells in many genomic locations such as repetitive sequences, retrotransposons and introns resulting in genomic instability. In the repetitive sequences, a higher rate of chromosomal rearrangements and a higher probability of translocations were observed at retrotransposons. In the last two decades, the field of cancer epigenetics has attracted more interest among basic and clinical researchers, especially after the introduction of specific methods utilized for DNA methylation analysis, including MSP, array methodology, and next-generation sequencing (NGS) after bisulphite treatment, methylC-seq, ChIP-chip and ChIP-seq (Laird *et al.* 2003). It has been observed that PanIN, IPMN, and MCN have a higher frequency of developing into invasive PDAC by several genetic and epigenetic changes in the tissue environment (Lüttges *et al.* 2000). Observation of aberrant methylation in PanCa patient samples leads to further investigation of methylation patterns in precursor lesions which finally leads to detection of genes commonly methylated in PDAC patients (Distler *et al.* 2014).

2.1: Hypermethylation in Pancreatic Cancer

Pancreatic ductal adenocarcinoma progression via abrogation of various tumor suppressor pathways has been relatively well documented in the progression of PanCa. Nowadays, it is apparent that epigenetic alteration such as main methylation and the transcriptional changes of affected genes plays an important role for progression of various human cancers from its precancerous state to PDAC. Two forms of frequently observed methylation changes in cancer are global hypomethylation in repetitive position and gene specific hypermethylation of CpG islands (Ehrlich *et al.* 2013, and Sharma *et al.* 2010). As it has been mentioned above, DNA hypermethylation has been a major driving force in the progression of PDAC (Wilentz *et al.* 1998). CpG islands are usually targeted as the regions with maximal hypermethylation due to the presence of high amount of CpG dinucleotide in the promoter region of the DNA. Anomaly methylation of promoter CpG islands has been associated with gene silencing along with TSG inactivation. The TSG and CpG islands are unmethylated in healthy situations and act as potential targets for DNA methylation in PDAC patients. Ueki *et al.* emphasized on aberrant CpG is- land methylation. At least one out of 12 cancer related genes (CRG) were observed in 60% cases of carcinomas (Ueki *et al.* 2001). Promoter methylations were observed in genes like *p16*, *RAR β* , *CACNA1G*, *TIMP-3*, *THBS1*, *hMLH1* and *DBC1* with

frequency varying from ~5 to 20%. Loss of expression is the outcome of the above epigenetic changes. Higher rate of methylation was observed in 4 CpG islands namely *MINT1*, -2, -31, and -32 (11% –73%). Previous observations also stated that *hMLH1* methylation outside the promoter region inactivates transcription. During normal situation *hMLH1* stays methylated but in some cases the rate increases to inactivate its expression. In case of PDAC there were several CRGs that are hypermethylated in their promoter regions like *SPRAC*, *p16*, *RASSF1A*, *3-OST-2*, *Cyclin D2*, *SOCS-1*, *RAR-b*, *APC*, *RELN*, *TFPI-2* and *ppENK*, which play an important role for the progression of disease. DNA hypomethylation also occurs at the 5' region of certain genes in PanCa. For example, *S100A4* was associated with hypomethylation at specific CpG site (Rosty *et al.* 2002). It has been exhibited that *RASSF2A*, one of TSGs, is inactivated by promoter hypermethylation. The inactivation of *RASSF1A* may rule out the necessity of *KRAS* activation to alter Ras signalling in PanCa (Kuroki *et al.* 2004). The methylation status of *RASSF2A* was screened by Zhao *et al.* in 8 human PanCa cell lines and 41 paraffin embedded tissues of PDAC samples where they found 100% hypermethylation in PanCa cell line and 22% hypermethylation in paraffin embedded tissues (Zhao *et al.* 2012). Omura *et al.* used genome-wide profiling and concluded that aberrant methylation in PanCa cells targeted towards the promoter along with the CpG islands. They identified hypermethylation in *MDFI*, *has-miR-9-1*, *ZNF415*, *CNTNAP2* and *ELOVL4* (96%, 89%, 86%, 82% and 66%, respectively) of PanCa (Omura *et al.* 2008). In a study to characterize DNA methylation patterns, Tan *et al.* showed the methylation pattern of 807 genes and identified 23 candidate genes with hypermethylation on CpG islands in PanCa. According to them, the genes of tyrosine kinase family like *KIT*, *FLT1*, *FLT3*, *NTRK3* and *EPHA5* were also hypermethylated and as a result they contributed to the resistance of certain tyrosine kinase drugs in PanCa due to under expression. Some candidate genes that caused PDAC through hypermethylation were *LIG3*, *MEST*, *SFN*, *VAV2* and *Lif* (Tan *et al.* 2009). Syren *et al.* analyzed PDAC sample tissues and cell lines and found that PanCa displayed hypermethylation of promoter genes namely *BNIP3*, *ELOVL4*, *ZNF415*, *MDFI*, *CNTNAP2*, *SPARC/osteonectin*, *ppENK*, *DBC1*, *FLT1* and *EYA* (Syren *et al.* 2017). Matsubayashi *et al.* analyzed pancreatic juice collected from 155 individuals with pancreatic disease and identified 118 methylated genes out of 187 genes. The pancreatic juice samples of PanCa patient contain more methylated genes as compared to juice of patients without PanCa (Matsubayashi *et al.* 2006). Sato *et al.* found that aberrant hypermethylation of *TFPI-2* gene's promoter CpG island is associated with its loss of expression. Hence, they concluded that the epigenetic inactivation of *TFPI-2* gene is a contributing factor in PDAC aggressive (Sato *et*

al. 2005). The aberrant hypermethylation in promoter of certain cancer related genes can cause carcinogenesis by up-regulating or down-regulating the cell cycle genes, DNA repair genes and other signalling pathway. In PanCa several CRG has been reported to be hypermethylated which becomes the lead cause of carcinogenesis. Cancer related genes like *p16*, *RASSF*, *3-OST-2*, *Cyclin-D1*, *SOC-1*, *RAR- β* , *DBC1* and *APC* have become inactive after promoter hypermethylation. Another TSG, *RASSF1A* was reported in PanCa which after promoter hypermethylation becomes silent. Promoter hypermethylation can be detected using MSP as previously mentioned (*Liggett et al.* 2010).

This is a very advanced and sensitive technique with the ability to detect very small number of cancer cells in peripheral blood. Another diagnostic marker cell free DNA promoter hypermethylation for PDAC has the ability to differentiate between malignant and benign lesions. *Xu et al.* found that the hypermethylation of the control region of *DUSP6* is the reason behind its undoing nature in human PanCa cells (*Xu et al.* 2005). *Sato et al.* demonstrated that DNA methylation influences *CXCR4* expression in human PanCa. Again, they used MSP, combined with bisulfite restriction analysis, and bisulfite sequencing, and found that the 5 CpG islands of the *CXCR4* gene to be unmethylated in normal pancreas, whereas promoter hypermethylation was detected in 45% (9 of 20) of PanCa cell lines and in 46% (46 of 100) of primary PDAC (*Sato et al.* 2005). *Yi et al.* observed that the *BNC1* hypermethylation have a sensitivity of 79% and a specificity of 89% in PanCa. Several hypermethylated genes, such as Cadherin (*CDH13*), *HIC1*, *HIC 2*, *HOXA9* and *IGFBP5* are correlated with various types of cancers (*Yi et al.* 2013). *Hagihara et al.* used methylation-sensitive-representational difference analysis (MS-RDA) to isolate 111 DNA fragments derived from CpG islands and found that 35 of them were in the 5 regions of known genes (*Hagihara et al.* 2004). They also analyzed MSP of the 5 CpG islands in 7 PanCa cell lines and two pancreatic ductal epithelial cell lines and found that 13 genes (*RASGRF2*, *ADAM23*, *NEF3*, *NKX2-8*, *HAND1*, *EGR4*, *PRG2*, *FBN2*, *CDH2*, *TLL1*, *NPTX1*, *NTSR1* and *THBD*) were silenced by methylation of their 5 CpG islands in promoter region. MSP of 24 primary PanCa showed that all these genes, except for *THBD*, were methylated in at least one cancer including PanCa. Some of those genes were suggested to be potentially involved in PanCa development and progression (*Hagihara et al.* 2004).

2.2: Hypomethylation in Pancreatic Cancer

DNA hypomethylation was the initial epigenetic abnormality recognized in human tumors. Repeated DNA elements were adjacently connected with hypomethylation of DNA in cancer

biology, irrespective of that malignancy associated DNA hypomethylation has still received very minute recognition. Hypomethylation in highly repeated DNA sequences was only considered as reason for cancer. However, a hypomethylation transcription regulatory region in cancer has been recognized as an important epigenetic abnormality after the discovery of high-resolution genome wide studies. In PDAC, aberrant loss of methylation is detected at genomic level as well as sequence specific level. The loss of methylation leads to many genetic alterations and instability. It was discovered that the deficiency in folate and vitamin B 12 may contribute to hypomethylation by decreasing the level of methyl group donor S- adenosyl methionine (SAM) which resulted in decreased biosynthesis of thymidine from uracil (*Van et al. 1998*). The cancer development is also caused by misplacement of uracil into thymidine which not only imbalance the nucleotide pool but also increase DNA strand breaks. Methylene tetra hydro folate reductase (*MTHFR*) gene is very important for the PDAC prospective because defective *MTHFR* gene presents in PanCa that causes more DNA hypomethylation, resulting in increased chromosomal losses. Unlike hypermethylation, hypomethylation causes the overexpression of protein and genes (*Omura et al. 2009*, and *Rosty et al. 2002*). *Sato et al.* reported 19 out of 32 genes had methylation and 7 of them (*CLDN4*, *LCN2*, *TFF2*, *MSLN*, *S100A4*, *SFN/14-3-3 σ*, and *PSCA*) were hypomethylated in PanCa cell line as well as primary PanCa (*Sato et al. 2003*). The analysis of methylation with oligonucleotide array identified two genes including *S100P* and *MASP* which are aberrantly hypomethylated. *Rosty et al.* analyzed National Center for Biotechnology Information (NCBI) Serial Analysis of Gene Expression (SAGE) database and found that *S100A4* was expressed in 95% PanCa cell line, and 11 out of 12 PanCa cell lines had hypomethylation at specific CpG site within the first intron of *S100A4* (100%) gene (*Rosty et al. 2002*). *Tan et al.* identified 35 candidate genes with hypomethylation among 807 genes in PanCa (*Tan et al. 2009*). In several studies conducted by multiple research groups, it has been found that *CCND1* and *CCND3* genes were over expressed due to hypomethylation in PanCa (*Gansauge et al. 1997*). Mucins like *MUC2* and *MUC5AC* are not expressed in normal pancreas but the promoter methylation causes mucin expression in PanCa cells. About 86% –100% invasive adenocarcinoma expressed *MUC5AC* (*Kim et al. 2005*). *Zhu et al.* detected hypomethylation in *MUC4* gene at 116 micro dissected foci in 57 patients with PDAC (*Zhu et al. 2010*).

2.3: Hypermethylation Identified in Pancreatic Cancer by NGS Studies

The identification of genome wide pattern of DNA methylation is necessary to detect epigenetic dysregulation in PDAC. The high- density array was used by *Nones et al.* to capture the

methylation profile of 167 untreated resected PDAC (Nones *et al.* 2014). They compared PDAC samples to 29 adjacent untransformed pancreatic tissues and identified 3522 aberrantly methylated genes. The samples were collected as part of Australian Pancreatic Cancer Genome Initiative (APCGI). As documented in a review by Iguchi *et al.* hypermethylation in the genes *ROBO1*, *ROBO3*, *SLIT3* and *SLIT2* play an important role in PDAC (Nones *et al.* 2014). The epigenetic perturbation of axon guidance pathways was analyzed by using bisulfite amplicon deep sequencing and qRT-PCR and suppression of SLIT-ROBO signalling was observed (Blockus *et al.* 2016).

2.4: Next Generation Epigenetic Study of Hypomethylation of Pancreatic Cancer

To identify methylation pattern of genes in PDAC, Vincent *et al.* performed methylated CpG island amplification followed by CpG island microarray where they analyzed 27,800 CpG islands covering 21 MB of the human genome (Vincent *et al.* 2011). The frozen primary PanCa tissues and normal pancreatic tissues were obtained from patients at John Hopkins Hospital and then they analyzed number of genes that are silenced by DNA methylation. The reported genes were TSG genes (*STK11* and *WT1*), stem cell pluripotency genes (*BM11*, *BMP3*, and *FOXD3*), cell adhesion genes (*CDH2*, *CDH4*, *PCDH1* and *PCDH10*) and also genes involved in WNT signalling (*WNT5A*, *WNT7A* and *WNT9A*). By using MCAM (microarray hybridization) assay, they identified many candidate hypomethylated genes in PDAC (Vincent *et al.* 2011). Genes responsible for chromatin modification including *SETD8*, *PRMT1*, *CTR9*, *KDM6A* and chromatin assembly protein *HIRIP3* were over expressed in PanCa samples as compared to normal ductal epithelium. The Genome wide analysis (GWA) done by Vincent *et al.* also reported 3 oncogenes (*JUNB*, *MYB* and *FOS*) to be hypomethylated in PanCa (Vincent *et al.* 2011).

2.5: Common Hypo and Hypermethylated Genes Between Pancreatic Ductal Adeno Carcinoma and Its Precancerous Lesions

There are several methylated genes that were observed in both the precancerous lesions as well as in PDAC. These common methylation patterns observed are the early epigenetic drivers in the progression of precancerous lesions towards PDAC. These genes would have bigger impact on the development and the progression of PDAC from its precancerous stages. Previous studies had reported two genes, *ppENK* and *p16*, which are found aberrantly methylated in both PanIN and PDAC. As documented in a review by Omura and Goggins, partial methylation of *CDKN1C* promoter CpG island was observed in both IPMN and PanCa cell lines as a

reduction in expression (Omura *et al.* 2009). According to a review by Iguchi *et al.* *WNK2* was silenced during progression of PanIN to PDAC hence contributing in both the cases (Iguchi *et al.* 2016). It was also observed that *MUC4* was overexpressed in both precursor lesions and in PDAC. The *APC* and *p16* genes are vital in the cell cycle regulation as *APC* gene is an extremely important component in the *APC/ β-catenin* pathway. The *p16* also plays a vital role and DNA methylation of *p16* leads to silencing of it in PanIN and PDAC, hence highlighting the carcinogenic progression from precancerous lesions to PDAC (Iguchi *et al.* 2016). Peng *et al.* found that *APC* and *p16* genes were the most frequently methylated and the frequency of DNA methylation of these two genes were increased from PanIN to PDAC. Another observation reported was that the down-regulation of *BRCA1* gene may be due to DNA methylation and it was observed across CP to PDAC (Peng *et al.* 2006). Delpu *et al.* in a review pointed out that 8 genes were differentially methylated and the methylation pattern varies amongst normal epithelia and PanIN lesions (Delpu *et al.* 2011). It indicates that in case of PDAC initiation CpG island hypermethylation is an early event hence justifying the involvement of methylation pattern in progression from PanIN to PDAC. The genes reported in this study are *CDH3*, *reprim*, *CLDN5*, *LHX1*, *NPTX2*, *SARP2*, *SPARC* and *ST14*. Another subset containing *p16*, *ppENK*, *p14*, *p15*, *p17*, *APC*, *hMLH1*, E-Cadherin and *MGMT* were observed in IPMN showing similar results as in PanIN. These aberrant DNA hypermethylation highlights the fact that methylation acts as a driving force from precursor lesion progression to PDAC (Delpu *et al.* 2011). *WNK2*, a cytoplasmic serine/threonine kinase, was hypermethylated in both chronic pancreatitis and PDAC. *WNK2* silencing supports PDAC progression. Systematic hypomethylation of *MUC4* may play a key role in carcinogenesis in PDAC (Iguchi *et al.* 2016).

2.6: Therapeutic Aspects of Epigenetic Modulated Pathways in PDAC and Its Associated Precancerous Lesions

PDAC is a deadly illness with few therapy options. According to new research, PDAC includes numerous layers of epigenetic alterations. Because the change is possibly reversible, it is a possible therapeutic target. Epigenetic changes can potentially affect the tumor microenvironment, modulating and enhancing treatment. Because epigenetic changes occur early in the disease, epigenetic markers can also be employed as diagnostic screening tools. Immunotherapy is being used more frequently to treat solid organ tumors, however there is no benefit for PDAC because most patients do not respond to these new treatments (Zhang *et al.* 2022). Because epigenetic processes regulate underlying immune cell activities, resulting in

an anti- tumor response, combining immunotherapy and epigenetic therapy may improve patient outcomes even more. PDACs are currently classified as three to five subtypes according on the system used (*Moffitt et al. 2015*). Using transcriptomic profiling, two primary molecular subtypes of PDAC were discovered: classical and basal. The traditional kind has a better prognosis and clinical significance. Basal subtypes have altered the methylation of effectors and inhibitors of the WNT signalling pathway by analyzing the epigenomic landscape. Classical tumors are hypomethylated, resulting in upregulation of the cholesterol transporter *NCP1L1* (*Koni et al. 2020*). Furthermore, basal tumors were discovered to contain dysregulation of multiple genes related with established oncogenic signalling networks, including the *MYC*, erythroblastic oncogene B/*EGFR*, and transforming growth factor (*TGF*) signalling pathways. Chronic pancreatitis is a well-known risk factor for PDAC, which is consistent with the previously documented general link between tumor and inflammation (*Del et al. 2021*). Early stage PanCa caused by inflammation is linked to epigenetic alterations. Damage to the pancreatic epithelium during a pancreatitis episode results in long-lasting transcriptional and epigenetic remodeling that creates epithelial memory that protects against strokes in the future. Reader proteins, have lately been identified as prospective therapeutic targets, in particular the chromatin adaptors of the bromo and extra C-terminal (BET) family, after directly engaging with the histone tails with acetylated lysine residues, these proteins with bromodomains can bind transcription factors to the DNA, boosting the acetylation-induced transcriptional activation. BET proteins use the epigenetic landscape in this way to support the growth of PDAC cells (*Garcia et al. 2018*).

2.6.1: Different Types of Biomarkers in Pancreatic Cancer Diagnosis and Prognosis

A biomarker is defined as ‘any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease’. Currently carbohydrate antigen 19-9 (CA19-9) is regarded as the best serological biomarker available so far in the diagnosis of PanCa, however the majority of studies endorsing the use of CA19-9 as a complementary test in the diagnosis of PC acknowledge it is not specific or sensitive enough to be used for screening (*Gui et al. 2014*). A number of other biomarkers have been proposed and these will be reviewed here. Variation exists in the biomarker domain, with studies utilizing serum, biliary fluid, pancreatic juice, urine, faeces and pancreatic cystic fluid for analysis of potential agents to determine their worth as a malignancy biomarker, however these methods of assessment vary in their invasiveness, sensitivity and specificity (*O'Neill et al. 2021*).

It has been demonstrated that PDAC cells induced DNA methylation of the *SOCS1* gene, a cytokine suppressor and cancer promoting growth factor, to boost tumor cell proliferation in vitro. Clinical evidences demonstrating a higher 3-mo overall survival in patients missing *SOCS1* methylation end credence to this. In PDAC, lysine demethylase 3A (*KDM3A*) is an effective epigenetic regulator of immunotherapy responses. This enzyme controls the *EGFR* expressions (Li et al. 2022). Tumors produced by cancer cells deficient in *KDM3A* have infiltrating immune cells that are responsive to immunotherapy. To distinguish between PDAC and cancer precursor phase, methylation -specific electrophoresis was used to determine the methylation status of the *MUC1*, *MUC2*, and *MUC4* genes in pancreatic fluids. Additionally, the methylation status of the mucin genes was examined using machine learning, and it was discovered that *MUC1* and *MUC4* hypomethylation levels were significantly correlated with poor prognosis. Through the suppression of Hedgehog (Hh) signalling, improved gemcitabine delivery was shown in preclinical investigations (Yokoyama et al. 2020).

Based on the transcription factor GATA binding factor 6 (*GATA6*)'s function as a regulator of the traditional PDAC subtype identity, the method to induce subtype switching in PDAC has been further investigated. A basal state is provided in PDAC by *GATA6* depletion. As a regulator of *GATA6* transcription in PDAC, the histone methyltransferase zeste homologue 2 (*EZH2*) enhancer prevents the decreased *EZH2-GATA6* and induced gene signatures present in traditional PDAC subtypes. Therefore, a potential target for PDAC treatment in the future is the *EZH2-GATA6* axis (Kloesch et al. 2022). Using genome-wide epigenetic mapping of the alterations 5- methylcytosine and 5hmC, this epigenetic dysregulation was demonstrated. Due to decreased production of the enzyme 5-methylcytosine hydroxylase *TET2*, these transcriptional subtypes exhibit a higher loss of 5hmC. In addition, reduction of *SMAD4* expression revealed decreased 5hmC and *GATA6*, resulting in a more squamous-like tumor. Blocking DNA methylation by utilizing the DNA methyltransferase (*DNMT*) inhibitor 5-azacytidine slows the growth of typical PDAC tumor. In contrast, utilizing the same medication or *DNMT* knockdown via small interfering RNA boosted hyaluronic acid synthesis, ultimately increasing the advancement of PDAC (Kohi et al. 2016).

Additionally, RNA modification of N6-methyladenosine (*m6A*) is a unique strategy for dynamic and reversible epigenetic control that has been discovered by researchers. By triggering the Wnt signalling cascade and changing Wnt 1. Inhibitory factor 1 (*WIF-1*), *m6A* accelerates the course of PanCa. Demethylase, *m6A* rubber, and the alkylation repair protein 5 (*ALKBH5*) homolog are increased in gemcitabine-treated sensitized PDAC cells. In vitro and in vivo development and invasiveness are accelerated when PanCa cells lack *ALKBH5* (Ding

et al. 2020). As a result, *ALKBH5* might be a brand-new target for PanCa treatment. *Ding et al.* found that the natural flavonoid fisetin regulates H3K36 methylation to promote *RFXAP* and *KDM4A* expression and interferes with HR, leading to DNA damage and PDAC S-phase arrest (*Ding et al.* 2020). Therefore, this tactic may constitute a cutting-edge therapeutic method for treating PanCa. DNMT inhibitors (DNMTis) are undertaking In PanCa cells, HDAC is, which includes *SAHA* and *CUDC-101*, can downregulate apoptotic inhibitory proteins including survivin and *XIAP*. Additionally, these HDACs can make PanCa more radiosensitive and make it cytotoxic (*Funaki et al.* 1996).

AR-42, which is another potent HDACi against PanCa cells, can inhibit cell proliferation via inducing cell cycle arrest at G2 phase., *Bian et al.* defined a novel technique for PDAC classification and management based on sensitivity to JQ1 treatment (*Bian et al.* 2017). For instance, it was discovered by researchers that during PanCa, H3K4me3 of the *BCL2L1*, *CFLAR*, and *MCL-1* gene promoters upregulates the production of the anti-apoptotic proteins *Bcl-x*, *FLIP*, and *MCL-1*, as well as the *BAK1*, *BAX*, and *BCL2L11* gene promoters of Bak and Bax. Proapoptotic proteins like the Bim protein, for example, have their expression downregulated. These six apoptosis-controlling genes are all essential for PanCa growth and development (*Abulwerdi et al.* 2014).

Upstream lncRNAs such as *BLACAT1*, *SNHG17*, and *CASC15* can decrease *P15* and *P21* expression and cause G0/G1 checkpoint deficit. DZNep (3-deazaneplanocin A), a powerful pharmacologic inhibitor of S-adenosylhomocysteine hydrolase, modifies chromatin accessibility via inhibiting histone methyltransferases such as *EZH2*. It results in a large decrease in H3K27me3 (a primary repressive histone mark) levels, as well as a significant decrease in cell proliferation and migration in CRC. Another study found that DZNep significantly modulates miR-663a and miR-4787-5p expression and consecutively suppresses *TGFb1*-induced EMT signalling in PanCa (*Anfossi et al.* 2014). *UNC1999*, an *EZH2*-specific inhibitor, not only lowers the abnormal H3K27 methylation that characterizes PanCa cells, but it also slows cancer cell proliferation in three model systems. As a result, diosgenin, garcinol, *FBW7* and curcumin analogue CDF have also been identified as potential agents targeting *EZH2* to prevent the development of PanCa (*Karandish et al.* 2016).

Clinicians may find it helpful that cell-free markers can distinguish PanCa from autoimmune pancreatitis, chronic pancreatitis, and other precancerous tumors. It is possible to diagnose PanCa at an early stage only if these indicators distinguish the non-carcinogenic illness with precision stage from cancer to the many stages of a tumor. Six genes—*PRKC*, *IKZF1*, *CD1D*, *KCNK12*, *CLEC11A*, and *NDRG4*—were found to have clinical potential and to be useful as

cfDNA diagnostic biomarkers for PanCa in an investigation of an 87 gene panel. A remarkable shift in methylation status was reported by *Liggett et al.* Using cfDNA DNA methylation profiling, a total of 14 genes—*CCND2*, *CDKN2B*, *DAPK1*, *MUC2*, *MYOD1*, promoter A of *MLH1*, *CDKN1C*, *ESR1*, *MGMT* *PGK1*, the proximal region of the *PGR* promoter, *RARB*, *RB1*, and *SYK*—were able to differentiate CP from PanCa with 91% sensitivity and specificity (*Liggett et al. 2010*). Six genes were shown to be differently methylated between CP and PanCa in a study employing pancreatic juice samples taken from PanCa and other related precancerous lesions. The six genes, *FOXE1*, *p16*, *ppENK*, *TFPI2*, *NPTX2*, and *CyclinD2*, were able to effectively discriminate between CP and PanCa because CP's methylation levels were lower than PanCa's (*Fukushima et al. 2002*).

In a parallel investigation, *Su et al.* used blood as a sample source and found that miR- 877, miR-3201, miR-890, miR-602, and miR-16-2-3p were effective in differentiating between PDACs that were high-risk and low-risk (*Su et al. 2018*). Research on urine proteome discrepancies between PDAC, CP, and healthy persons was done by *Weeks et al 2008*. They observed 101 proteins in CP and PanCa that had substantial alterations. It was discovered that S100A9 was overexpressed in CP but downregulated in PanCa. In PDAC samples, dysregulated levels of the proteins annexin A2, Cd59, and gelsolin were also discovered. Based on quantitative protein expression, an isotope-coded affinity tag tended to effectively differentiate between PanCa and CP (*Weeks et al. 2008*).

The usage of *BNC1* and *ADAMTS1* as novel serum-based biomarkers for early PanCa detection was discovered by *Yi et al.* in an intriguing finding (*Yi et al. 2016*). The MethDet56 test was created by *Melnikov et al.* They screened 56 gene promoters from PDAC plasma samples, and five of them—*CCND2*, *SOCS1*, *THBS1*, *PLAU*, and *VHL*—were found to have significantly aberrant DNA methylation, with corresponding sensitivity and specificity of 76% and 59% (*Melkinov et al. 2009*). The index of plasma-free amino acids (PFAAs) is becoming more and more well-liked as an appropriate method for PanCa diagnosis. According to a study by *Fukutake et al.* comparing the PFAAs index between PanCa and healthy controls, PFAAs is a strong biomarker that can differentiate PDAC (*Fukutake et al. 2015*). *Wu et al.* reported that the serum of PanCa patients included different amounts of *GPC1*, *PFAA*, *OPNT + TIMP-1*, *CPA4*, *MUC5AC*, and *C4BPA* (*Wu et al. 2022*) (Figure 14).

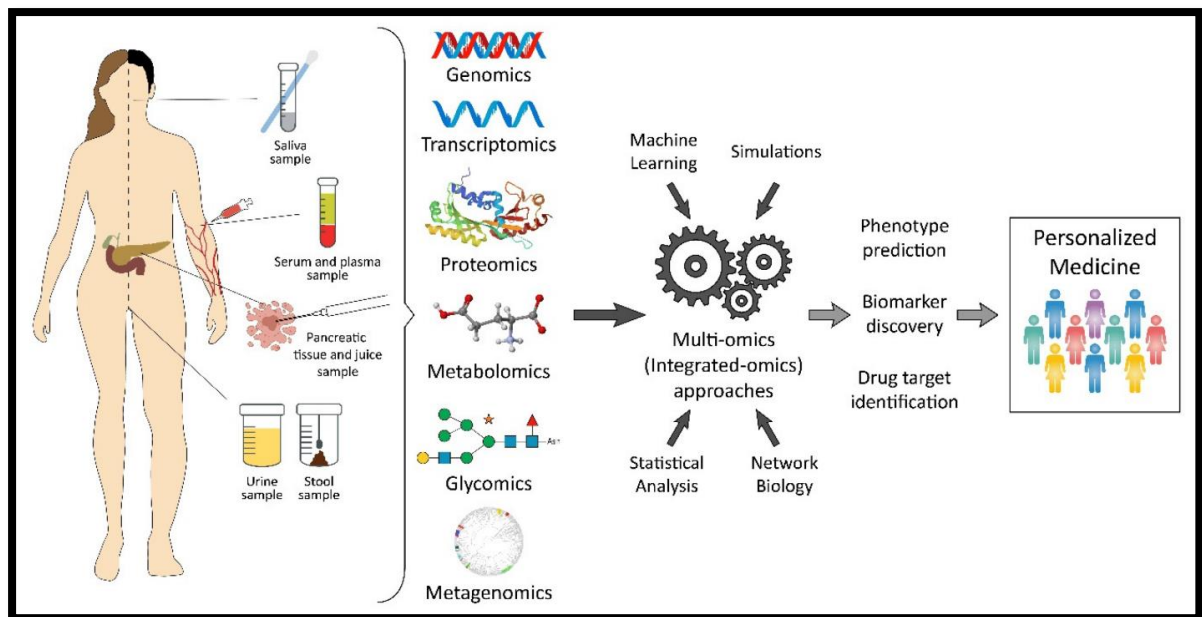


Figure 14: A Conceptual review of pancreatic cancer biomarkers from a variety of “Omics” Levels. (Turanil et al. 2021)

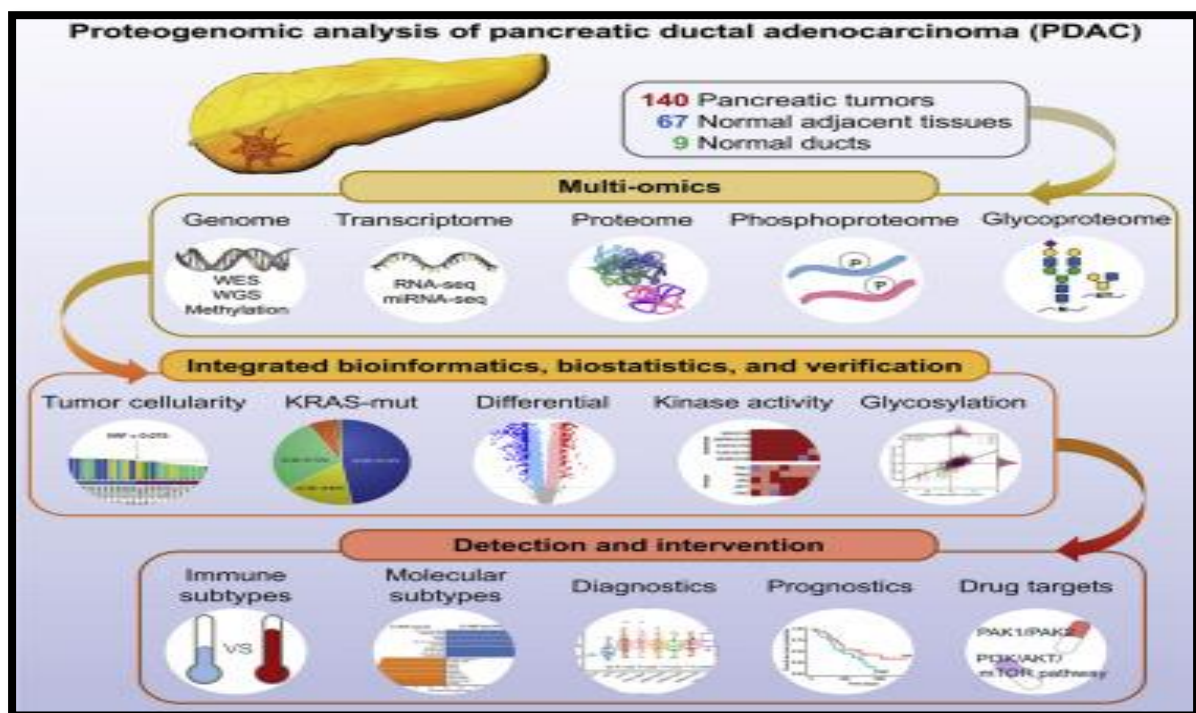


Figure 15: Proteogenomic analysis of pancreatic ductal adenocarcinoma (Cao et al. 2021)

2.7: Identified Methylated Biomarkers by Multiomics Methods in Pancreatic Cancer

By utilizing integrative methods to combine omics data from several layers, one can gain a deeper comprehension of the biology behind complex diseases like cancer. Finding more efficient treatment options is aided by the identification of biomarkers linked to the

development or prognosis of cancer. The following genes, which were identified as triple-evidence genes, were also found to be shared by the four cancer types in our study: *MCM4*, *GNG11*, *LGR4*, *LPCAT1* for Lung Squamous Cell Carcinoma (LUSC), *EGFR* for Kidney Renal Clear Cell Carcinoma (KIRCCC), and *ARHGEF10* for Breast Invasive Carcinoma (BRCA); *CBX2*, *CBX8* genes for BRCA, LUSC, and COAD; *EGFR* gene for KIRCCC and *ARHGEF10* gene for BRCA. A statistical integrative clustering analysis (iCluster+) was carried out in a different work by *Mo et al.* employing TCGA colorectal cancer exome sequencing, DNA copy number, promoter methylation, and mRNA expression data (*Cao et al. 2021*). The *PTDSS1*, *MCM4*, and *PRKDC* genes were shown to be molecular drivers associated with the same subgroup in this investigation to identify cancer subgroups. *MCM4*, *CDKN3*, and *EGFR* were found to be hub genes in a PPI network that *Qi et al.* built using aberrantly methylation and differentially expressed genes for breast cancer (*Qi et al. 2019*). Using a neural network-based approach to analyze gene expression and copy-number variations, a different study on breast cancer revealed *CDKN3* as one of the subtype-specific genes associated with the LumA subtype. Using gene expression, methylation, and miRNA data of Glioblastoma (GBM), *Fiorentino et al.* established a methodology that combines omics-specific similarity networks in a single network and validated the *SEC61G* gene as a predictive biomarker (*Fiorentino et al. 2021*). By utilizing somatic mutations, copy number variations, methylation, mRNA and miRNA expression data, *Dimitrakopoulos et al.*, established a network-based integration strategy that allowed for the identification of the known *EGFR* gene for lung cancer (*Dimitrakopoulos et al. 2018*). *Sheng et al.*'s study found circRNAs, miRNAs, and mRNAs that were differently expressed in breast cancer and created a regulatory network (*Sheng et al. 2017*). They next built a PPI network and used the MCODE method to investigate the important genes involved in the regulatory network; as a result of this study, *LPCAT1*, *CBX2*, and *EGFR* were discovered as possible hub genes. In their proposal, *Shi et al.* used gene networks, expression data, and mutation data to find driver genes. They also identified *EGFR* and *PRKDC* as putative driver genes for GBM (Figure 15) (*Shi et al. 2017*).



OBJECTIVES



SEPTEMBER 16, 2024

PHD THESIS

Department of Life Sciences and Biotechnology, Jadavpur University

3.0: Aims and Objectives

- 3.1:** To identify differential methylation sites by genome wide DNA methylation profile from pancreatic ductal adenocarcinoma samples along with their adjacent normal using 450K DNA methylome array in discovery patient cohort.
- 3.2:** To identify dysregulated pathways and key signalling networks targets involved in regards to the obtained methylome profile in our study.
- 3.3:** To validate the Differentially Methylated Positions (DMP) along with its associated genes via Methylation specific PCR, gene expression profiling and Immunohistochemistry methods in an independent patient cohort.
- 3.4:** To identify the novel methylome biomarker in pancreatic cancer by multiomics using available public datasets



MATERIALS AND METHOD



SEPTEMBER 16, 2024

PHD THESIS

Department of Life Sciences and Biotechnology, Jadavpur University

4.0: Study Design

In this study, we aimed to identify differential methylation marks associated with PanCa in Indian population, using global methylome analysis. The detailed study design is below in Figure 16

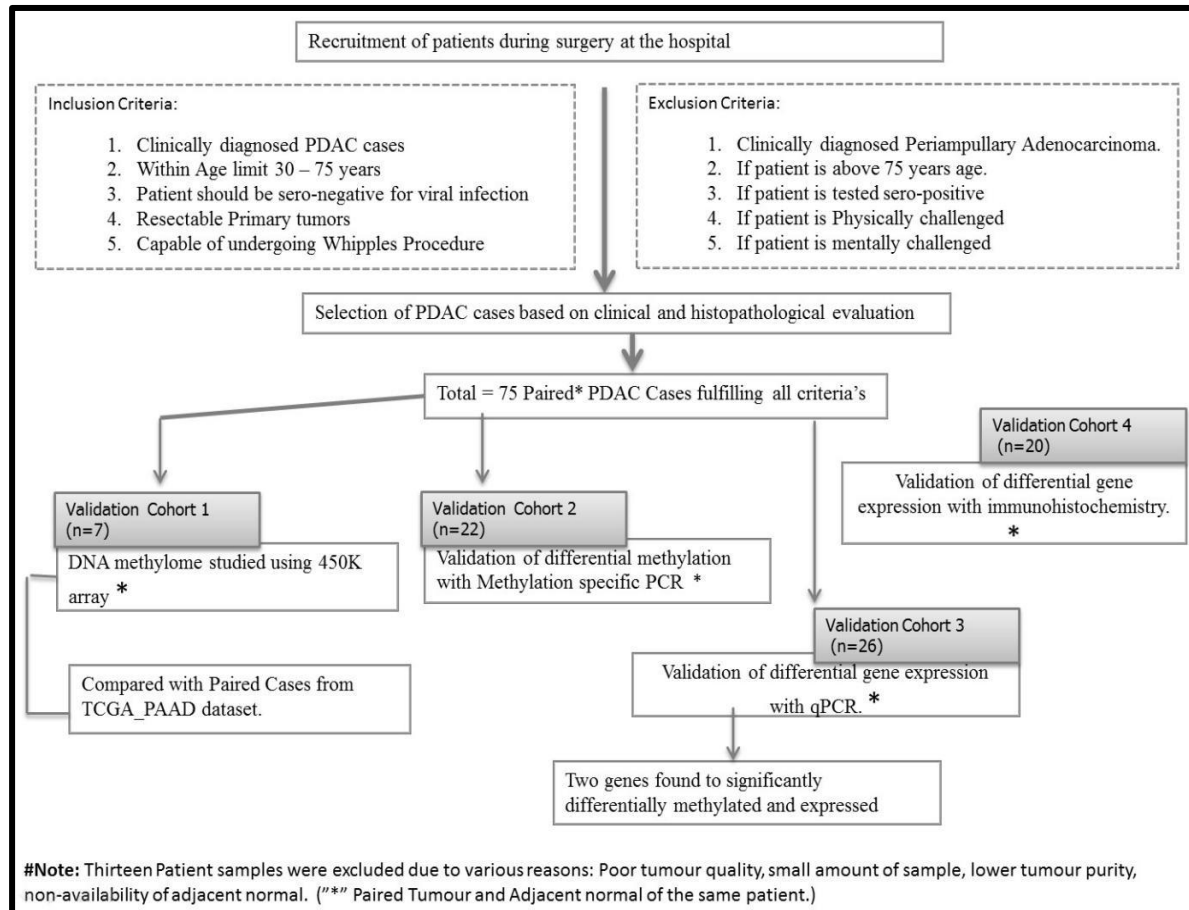


Figure 16: Flow chart showing the study design.

4.1: Sample Collection and Ethical Clearance from Different Hospital and Respective Institute

The study was approved by the Institutional Review Board (IRB) of Indian Statistical Institute (ISI), Kolkata and all involved hospitals namely Tata Medical Center (TMC), Calcutta Medical Research Institute (CMRI), Institute of Post Graduate Medical Education & Research (IPGME&R) and SSKM Hospital, Kolkata, and Medical College and Hospital, Kolkata, India. All clinical samples were collected only after taking consent from patient and/or family post debriefing about the study. Our study included samples of PDAC patients, collected from Government and private Hospitals, in the city of Kolkata during the time frame of 2013–2022. A systemic 2–5 year follow up was done for overall survival (OS).

tumor and adjacent normal tissue samples were collected during Whipple's surgery and stored in RNA later solution (Sigma-Aldrich Co. LLC). Histopathological examination and TNM staging were done according to the 8th edition guidelines by American Joint Committee on Cancer (AJCC), by all the associated clinician. We had investigated the tumor purity through histopathological evaluation, and it was found to be >80% in all samples (10× imagery by using Leica DM 1000, camera: Leica EC3). We have generated a score for each sample regarding tumor cell percentage by randomly taking 8 images for each slide (Figure 17). The inclusion and exclusion criteria of the patients were considered prior to specimen collection and consenting.

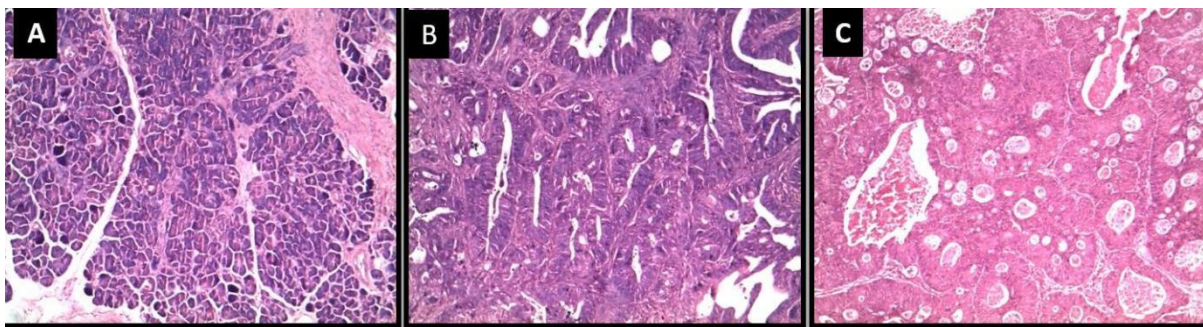


Figure 17. Representative examples of H&E-stained slides showing the tumor cell percentage in PDAC patient samples. Tumor purity is indicated as stated below: (A) 90% (B) 87% (C) 85%. (*Chatterjee et al. 2022*)

4.2: Inclusion Criteria of Patient Samples

The inclusion criteria started with collecting samples that were only clinically diagnosed PDAC samples within the age-limit 30–75 years and should positively be sero-negative and also capable of undergoing whipple's procedure.

4.3: Exclusion Criteria of The Patient Samples

The exclusion criteria were designed to target PDAC subtype, which is a specific subtype of PanCa. Hence periampullary adenocarcinomas were excluded and also PDAC samples with sero -positivity were excluded. Some other exclusion criteria's included were age greater than 75 years and if the patient were physically and mentally handicapped.

4.4: Description of Study Cohorts

(a) The TCGA PanCa cohort. We have downloaded DNA methylome data (level 3) and bulk transcriptome data on PanCa samples (n=185) submitted in TCGA database from GDC data portal (<https://portal.gdc.cancer.gov/>). Among these 185 patients, for 9 patients methylome

data on solid tissue normal samples was also obtained. Three cancer samples lacked information on tumor stage and were excluded from the study. The final 450 K methylation Infinium array data on DNA methylome of PanCa included 182 tumors. The phenotypes of the samples are listed in table 4

1. (b)The Validation Cohort from Indian Population in our study consists of a total 75 clinically and histopathologically confirmed PDAC patient samples (Paired solid tumor and normal samples). A validation cohort (named as validation cohort1) including PanCa patients from Indian population (n=7), on whom DNA methylome was studied 7 paired primary tumor and adjacent normal samples were collected along with all clinicopathological details.
2. We have further identified key differentially methylated genes and validated differential methylation of the genes in a second independent patient cohort (named as validation cohort2; n=22), using MSP.
3. Finally, we have validated inverse relationship between promoter methylation and gene expression of key genes in a third independent patient cohort (named as validation cohort3; n=26).
4. Validation cohort 4 (n=20): We have selected two genes that were differentially methylated and expressed in pancreatic cancer samples. Differential expressions of the genes at protein level were validated using immunohistochemistry from paraffin block tissue samples.

4.5: DNA Extraction and Bisulfite Treatment

DNA isolation from tissue samples (tumor and adjacent normal) was done using the DNeasy Blood and Tissue Kits (QIAGEN Inc., Germany). The method of DNA isolation was as follows: First a small piece of tissue approximately weighing 25 mg, stored in RNA Later, was taken and minced into several small pieces. The minced tissue pieces were then placed in a glass homogenizer and approximately 180 μ l ATL Buffer was added for homogenizing the tissue pieces. This homogenized mixture was added to a clean 1.5 ml micro centrifuge tube (MCT). Around 20 μ l of Proteinase K was then added, vigorously vortexed and kept at 56 °C water bath for incubation until the solution turned clear. Vigorous vertexing was done at regular intervals during the incubation. The approximate duration for this process of tissue lysis was 2-3 hours. After incubation, 200 μ l AL Buffer was added to the solution and vortexed. Following AL buffer addition, 200 μ l of ethanol was also added. The solution was then placed onto the spin column and centrifuged at 8,000 rpm for 1 minute. Following centrifugation, the

flow through was discarded and the column was washed with 500 µl of AW1 Buffer (Wash Buffer I) and centrifuged at 8,000 rpm for 1 minute. The flow through was then discarded and 500 µl of AW2 Buffer (Wash Buffer II) was added and centrifuged at 13,000 rpm for 3 minutes. The flow through was again discarded and the spin column was then centrifuged at 13,000 rpm for an additional 1 minute to remove trace amounts of residual ethanol. The spin column was then placed on a clean 1.5 ml MCT. 100 µl of Buffer AE (Elution Buffer) was then added to the spin column and was centrifuged at 8,000 rpm for 1 minute to elute the bound DNA. The quality and quantity of the eluted DNA was checked using Nanodrop 2000 (Thermo Scientific). Bisulfite conversion of DNA samples (from ~500ng) was done using EZ DNA methylation Gold Kit (Zymo Research, Orange, CA). The conversion reagent was added to the DNA samples and incubated in a thermal cycler for 16 cycles at 95°C for 30 seconds followed by 1 hour at 50°C and a final hold at 4°C for at least 10 minutes for denaturation and conversion of DNA. The converted samples were then added to the spin columns followed by washing using the wash buffer. The washing was then followed by desulphonation using desulphonation buffer and incubation at room temperature for 15 minutes. The spin columns were then washed twice with wash buffer to remove traces of desulphonation buffer. Finally, the converted DNA samples were eluted in 12 µl of elution buffer. The converted DNA samples were then transferred to a 0.2 ml skirted microplate and the contents were mixed properly for the amplification of the DNA samples. The Illumina hybridization oven was preheated to 37°C and the reagents MA1, RPM and MSM were thawed at room temperature and mixed by inversion. Next in a 0.8ml microplate 20 µl of MA1 was added followed by 4 µl of the converted DNA sample. 4 µl of 0.1N NaOH was then dispensed to the wells and the plate was sealed using a 96-well cap mat. The plate was then vortexed at 1600 rpm for 1 minute followed by a pulse centrifuge at 280*g. The plate was incubated at room temperature for 10 minutes. Next the cap mat was removed and 68 µl of RPM was added followed by the addition of 75 µl of MSM and the plate was again sealed. The plate was again vortexed at 1600 rpm for 1 minute followed by a pulse centrifuge at 280*g. The plate was then incubated at 37°C in the Illumina hybridization oven for 20-24 hours. This amplification step was followed by fragmentation of amplified DNA. The heat block along with midi plate insert was then preheated to 37°C and FMS reagent was thawed at room temperature and mixed by inversion. Microplate was pulse centrifuged at 280*g and the cap mat was removed. 50 µl of FMS was added to each of the well containing samples and the plate was sealed. The plate was vortexed at 1600 rpm for 1 minute followed by a pulse centrifuge at 280*g. The plate was sealed and placed on the 37°C heat block for 1 hour. Next to this plate 100 µl of PM1 was added and vortexed at 1600 rpm

for 1 minute. The plate was then incubated at 37°C for 5 minutes on a heat block following which pulse centrifugation was done. 300 µl of 100% isopropanol was then added and the plate was sealed with a new, clean and dry cap mat and mixed by inversion. The plate was incubated at 4°C for 30 minutes followed by centrifugation at 4°C at 3,000*g for 20 minutes. The cap mat was removed and the plate was quickly and immediately inverted onto absorbent pad to drain the supernatant. The inverted plate was kept undisturbed at room temperature for 1 hour to air dry the pellet. Pellets appear as blue at the bottom of the wells. Next the precipitated DNA samples were resuspended. The hybridization oven was preheated to 48°C and the heat sealer was turned on to preheat for around 20 minutes. The RA1 reagent, which was shipped frozen, was gradually warmed to room temperature. 46 µl RA1 was then added to each well containing DNA pellet and the plate was heat sealed and placed in the hybridization oven maintained at 48°C for 1 hour. The plate was then vortexed at 1800 rpm for 1 minute followed by pulse centrifugation at $280 \times g$. The fragmented and resuspended DNA samples were then dispensed onto the BeadChips and for this the plate was placed for 20 minutes on a heat block maintained at 95°C to denature the samples. Simultaneously, the BeadChips were removed from storage temperature and the Hybridization Chambers were prepared. The BeadChip Hybridization Chamber gaskets were placed into the BeadChip Hybridization Chambers and 400 µl PB2 buffer was dispensed into the humidifying buffer reservoirs in the Hybridization Chambers and the lid was placed. After the 20-minute incubation, the plate was removed from the heat block and was kept at room temperature for 30 minutes to cool. After that the plate was pulse centrifuged at $280 \times g$ and the foil seal was removed. Before loading DNA samples, plastic bags from the BeadChips were removed and were placed in a Hybridization Chamber insert. Using a multichannel precision pipette, 15 µl of each DNA sample was dispensed onto the appropriate BeadChip section and the DNA samples were allowed to disperse over the entire surface of the BeadChip. Simultaneously, on the lab tracking form the BeadChip barcode were recorded for each group of samples and in the sample sheet Satrix ID column, the BeadChip sample ID and position were also recorded. Care was also taken to load excess sample volume in the BeadChip loading port as that helped prevent low intensity areas resulting from evaporation and the remaining sample in the plate were heat-sealed with foil and stored at -80°C. The Hybridization Chamber inserts containing BeadChips were then loaded into the Illumina Hybridization Chamber and the lids were placed on the Hybridization Chamber and the chamber was sealed. The Hybridization Chamber was then placed on the Hybridization Oven maintained at 48°C for overnight. Post incubation the Hybridization Chamber was removed from the oven and was allowed to cool for 30 minutes. Next the wire handles were attached to the rack and the wash rack was submerged in the wash dish that contained 200 ml

PB1. The Hybridization Chamber inserts were removed from the Hybridization Chambers and then the BeadChip was removed from the Hybridization Chamber insert and the cover seal of the bead chip was removed. The bead chips were then slid into the wash rack carefully such that the BeadChip was submerged in the PB1 and the wash rack was moved up and down for 1 minute with gentle and slow agitation for washing the bead chips. The wash rack was then moved to another wash dish containing PB1 and the washing procedure was repeated. Next the flow chambers were assembled and the bead chip alignment fixture was filled with 150 ml of PB1 and then a black frame was placed into it. The bead chips were then placed on the black frame and a clear spacer was placed on the bead chip. The alignment bar was then placed onto the alignment fixture and a clean glass back plate was placed on top of the clear spacer that covered each bead chip. The metal clamps were then attached to the flow chambers. Trimming of the ends of the clear plastic spacers were done with clear scissors. The hybridization reservoir was washed with distilled water and a small cleaning brush to remove traces of PB2. Water was filled in the water circulator reservoir and the temperature of the water circulator was maintained at 44°C. The bubbles in the chamber rack were removed and the temperature was carefully checked to be 44°C $\pm$ 0.5°C using the Illumina Temperature Probe. Following this the single base extension was carried out by adding and incubating the reagents: 150 μ l of RA1 for 30 seconds and then repeating the step for 30 seconds, 450 μ l of XC1 for 10 minutes, 450 μ l of XC2 for 10 minutes, 200 μ l of TEM for 15 minutes, 450 μ l of 95% Formamide/ 1mM EDTA for 1 minute and the step was repeated once. Following addition of all the substances it was incubated for 5 minutes and then ramping of the chamber rack temperature was started to the temperature indicated on the STM tube. 450 μ l of XC3 was then added and incubated for a minute and the step was repeated. Into the reservoir of the flow-through chamber the following reagents were dispensed and incubated: 250 μ l STM for 10 minutes, 450 μ l XC3 for 1 minute and the step was repeated, and incubated for 5 minutes. Then again 250 μ l ATM was added and incubated for 10 minutes. Furthermore, again the following reagents were added and incubated as following: 450 μ l XC3 for 1 minute and the step was repeated once and again incubated for 5 minutes, 250 μ l STM for 10 minutes, 450 μ l XC3 for 1 minute and once repeated and incubated for 5 minutes, 250 μ l ATM for 10 minutes, 450 μ l XC3 for 1 minute and repeated once and waited for 5 minutes, 250 μ l STM for 10 minutes, 450 μ l XC3 for 1 minute and repeated once and waited for 5 minutes. The flow-through chambers was then immediately removed and placed horizontally on the lab bench at room temperature. Further the beadchips were washed and dried. The bead chips were then scanned using the Illumina iScan system. The concentrations and the purity of the DNA samples were estimated using NanoDrop 2000 (Thermo Fisher Scientific™, USA).

4.6: DNA Methylation Data Generation by 450K Bead Chip Illumina Array and Analysis

Whole genome DNA methylation profiling was done from the DNA isolated from the tumor and adjacent normal tissues of validation cohort 1 (n=7) using Infinium Human Methylation 450 K Bead Chip (Illumina). The bead chip included 485,577 CpG sites and targets 96% of CpG islands in human genome (hg19), using 4 µl of bisulfite-treated DNA, according to the manufacturer's protocol. The raw signal intensities on bead chips were read and converted to raw data (IDAT files) using the iScan platform (Illumina). The raw data including the IDAT files is available in the GEO database with the entry "GSE181740". In this current study, we have followed the analysis done by Aryee *et al.* for processing of DNA methylome data (Aryee *et al.* 2014). For each CpG site in the genome, methylation level was assessed using the established β value. The β value was calculated as the ratio of fluorescent signal intensity of the methylated (M) and total of signal intensities from the methylated (M) and unmethylated (U) alleles:

$$\beta = \frac{M}{(M + U + 100)}$$

We have used the minfi R package R version 4.0.4 to convert raw array data to β value and for data normalisation. To reduce false positive inferences from our data, we have used data quality control as described by Aryee *et al.*, Probes were excluded from further analysis if: (1) their detection p-value > 0.05, (2) they contain SNPs in their sequences, (3) they are positioned on X and Y chromosomes. With the filtered probe set, we performed both intra-array (Infinium I and Infinium II) normalisation and inter-array normalisation using quantile normalisation approach (using β values). Prior to differential methylation calculation, purity of tumor samples was estimated using "Infinium Purify R" package using R version 4.0.4. Differentially methylated positions (DMPs) between the tumor and the normal tissue samples were identified by comparing the average beta values between the study groups. The probes with average $|\Delta\beta|$ (difference between the average β -value among the tumor samples and the average β -value among the normal samples) ≥ 0.2 and Benjamini–Hochberg corrected $p < 0.05$ were considered as significantly differentially methylated positions (DMPs). hypermethylated positions (CpG sites) were DMPs with $\Delta\beta \geq 0.2$ and hypomethylated sites were DMPs with $\Delta\beta \leq -0.2$. The DMPs were annotated to genes, referred to as "gDMs".

4.7: RNA Isolation from Patient Tissue Samples

All equipment used for RNA isolation and cDNA synthesis including tips, PCR tubes, 1.5 ml centrifuge tube, and mortar-pestles, were treated with a 0.1% DEPC solution (0.1% DEPC in dH₂O) overnight to eliminate RNase contamination. On the following day, the solution was drained, and all materials were autoclaved at 15 psi for 15 minutes to remove DEPC residue. Prior to experimentation, all laboratory apparatus and equipment were rinsed with RNaseZap® Solution to remove RNase contamination. Total RNA was extracted from the psoriatic skin and adjacent normal tissue samples using RNeasy Kit (Qiagen, USA) following manufacturer's protocol. This kit employs the basic principles of organic extraction and solid-phase extraction, retaining the advantages of both while avoiding their respective shortcomings. Tissue biopsy samples collected in RNA Later (Invitrogen, Carlsbad, CA, USA) were snap-frozen in liquid nitrogen and ground to powder using mortar and pestle. In the first step of RNA isolation, powdered tissues were scraped in 500µl of a denaturation lysis buffer and mixed thoroughly till a homogeneous lysate was obtained (*Chomczynski et al. 1987*). A volume of 50µl of miRNA Homogenate Additive was added to the lysate and mixed thoroughly by vortexing or inverting the tube several times. The mixture was incubated on ice for 10 minutes. Thereafter an equal volume, i.e. 500µl of acid phenol-chloroform was added and thoroughly mixed by vortexing for 30-60 sec (*Chomczynski et al. 1987*). This was followed by centrifugation at 12000 RPM for 15 minutes to separate the aqueous and organic phases. The aqueous phase was carefully collected in a fresh tube while ensuring that the lower phase was not disturbed. 100% ethanol amounting to 1.5 volumes of the aqueous phase was added and the lysate/ethanol mixture was passed through a filter cartridge (silica column). It was subjected to centrifugation for 15 sec at 12000 RPM, following which the flow-through was discarded and process repeated till the entire lysate had passed through the filter. The column was washed with miRNA Wash solution 1 and miRNA Wash solution 2/3. Finally, RNA was eluted in 50µl nuclease-free water. RNA concentration and purity were checked in a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific). The column filter was subsequently washed with miRNA Wash Solution 1 followed by wash solution 2/3. Finally, RNA elution was carried out in 50µl nuclease-free water.

4.8: cDNA Synthesis

The cDNA synthesis of the isolated RNA was performed using PrimeScript 1st strand cDNA Synthesis Kit (Takara Bio) following manufacturer's protocol. Approximately, 1µg of total RNA was used for cDNA synthesis. Thereafter, 1µl (500ng/µl) of oligo dT primers, 4µl of 5x

buffer, 2µl (5mM) of dNTP mix, 1µl of enhancer and 1µl of reverse transcriptase enzyme mix. The remaining volume was adjusted with nuclease free H₂O to obtain a final volume of 20µl. The cDNA synthesis was carried out according to the conditions set in the thermal cycler. This comprised incubation at 42°C for 60 minutes followed by inactivation of the enzymes at 95°C for 2 minutes and then maintained indefinitely at 4°C according to manufacturer's protocol.

4.9: Gene Expression Study by Quantitative Real-Time PCR Using Sybr Green Chemistry

The synthesized cDNA product was subjected to 5 times dilution with double distilled H₂O, and approximately 10ng was used for each real time PCR reaction. For gene expression quantification, we have used the CFX Opus 96 Real-Time PCR system (Bio-Rad) using iTaq Universal SYBR Green Supermix (Bio-Rad) in a 10µl reaction with 500nM of forward and reverse exon- spanning primers. Standard PCR conditions were employed i.e. 95°C for 2 minutes, followed by 40 cycles of 95°C for 15 sec and 60°C for 30 sec. Thereafter, a dissociation curve analysis was performed at 95°C for 15 sec, 60°C for 15 sec and 95°C for 15 sec. For each primer pair, double distilled H₂O was run alongside which served as negative control to rule out primer-dimer formation or other contamination. The housekeeping gene *ACTB* and *GAPDH* was used to normalize the gene expressions to determine the fold change deregulation of the target genes in pancreatic cancer cells compared to adjacent normal cells. . RT-PCR was done using iTaq Universal SYBR Green Supermix fluorescent dye (Bio Rad Life Sciences Research). The reaction cycle included an initial denaturation step of 95°C for 10 min, 20 sec (*ACTB* and *GAPDH*) followed by 38 cycles of 95°C for 20 sec, and 60°C for 30 sec. A melting curve analysis step was carried out at the end of amplification step, consisting of denaturation at 95°C for 1 min, re-annealing at 55°C for 30 sec and a further denaturation at 95°C for 30 sec. The expression of *FAIM2* and *FOXE1* were analyzed with annealing temperature of 57°C and 57.5°C for 30 sec respectively. The expression of *NPY*, *SLITRK3*, *KCNA6* and *RASSF1A*, were analyzed with annealing temperature of 58°C for 30 sec. The expression of *SIGIRR* and *KCNJ5*, was analyzed with annealing temperature of 59°C for 30 sec. The expression of *IRF4*, *MX2* and *GALR1* were analyzed with annealing temperature of 58°C, 59°C and 57°C for 45 sec respectively. The other remaining steps for all ten genes were same as above. A melting curve analysis step was carried out at the end of amplification step, consisting of denaturation at 95°C for 1 min, re-annealing at 55°C for 30 sec and a further denaturation at 95°C for 30 sec. Target and reference gene Ct values were derived from the mean of duplicate. Delta Ct (ΔCt) values were estimated for each reaction in

comparison with the reference and were compared across groups using Wilcoxon Signed Rank test. Relative expression of targeted genes determined as $2^{-\Delta\Delta ct}$ was calculated for each of the paired sample to identify fold change. Corresponding gene Sybr green gene expression RT PCR primers are described in table 4

4.10: Bisulfite Conversion as Preprocessing Step for Methylation Specific PCR

DNA samples isolated from PDAC tissues along with their corresponding adjacent normal were used for Bisulfite Conversion. The bisulfite conversion protects the methylated cytosine residues in DNA whereas following bisulfite conversion the unmethylated cytosine residues are converted to uracil and the following are replaced by thymine in corresponding PCR reactions (Figure 18).

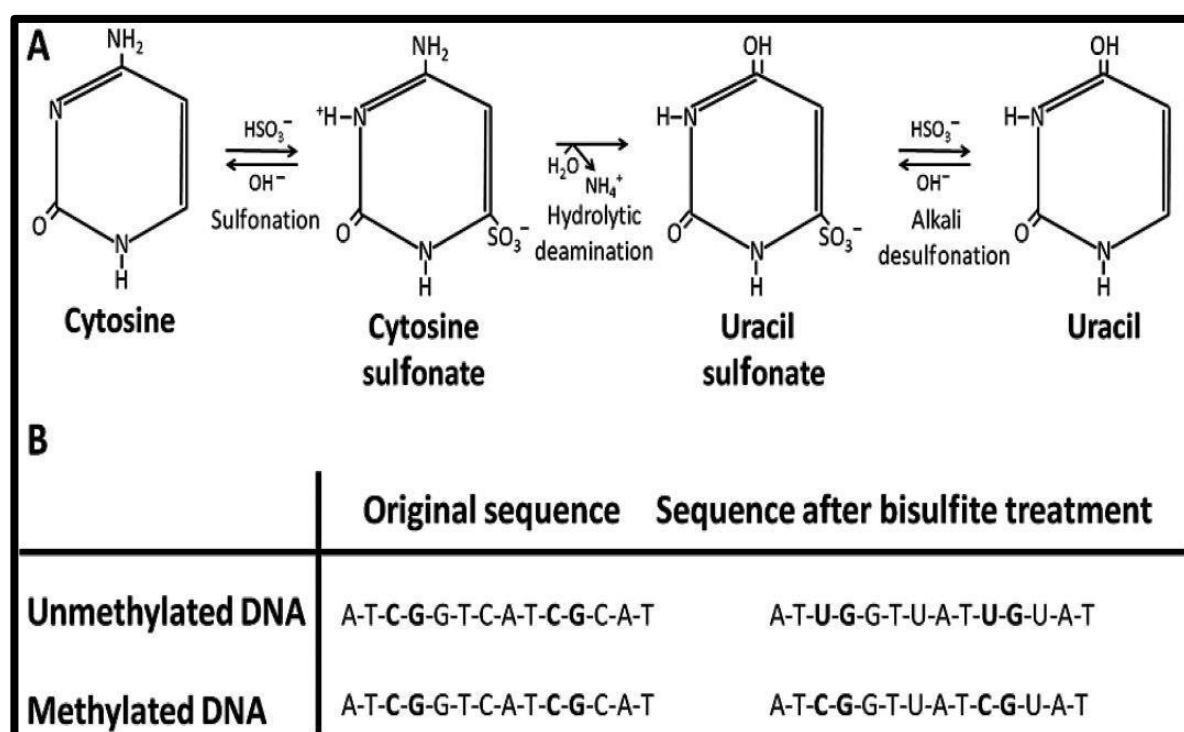


Figure 18: Methodology of the bisulfite conversion (*Kristensen and Hansen et. al., 2009*)

About 750 ng of each DNA sample was bisulfite modified. The Bisulfite conversion was done using EZ DNA methylation Gold Kit (Zymo Research) following manufacturer's protocol. Briefly, the protocol is as follows: Initially the CT conversion reagent was prepared by adding 900 μ l distilled water, 300 μ l M-Dilution Buffer and 50 μ l M-Dissolving buffer to the powdered CT Conversion Reagent and was mixed well. Approximately 750 ng of DNA was taken for the conversion and the volume of the solution was adjusted to 20 μ l with sterile distilled water. To the DNA solution, 130 μ l of the CT conversion reagent was added and was mixed well and

placed on a thermal cycler for the bisulfite conversion reaction. In the thermal cycler the reaction mixture was first incubated at 98°C for 10 minutes, 64°C for 2.5 hours and a final hold at 4°C for up to 20 hours. After incubation in the thermal cycler, the reaction mixture was added to the Zymo Spin Column. To the spin column, 600 µl M-Binding buffer was added initially. The spin columns were then centrifuged at 13,000 rpm for 1 minute. The flow through was discarded. To the spin columns 100 µl M-Wash Buffer was added and the spin columns were centrifuged at 13,000 rpm for 1 minute. The flow through was discarded and 200 µl M-Desulphonation Buffer was added very carefully and the spin columns were then incubated at room temperature for 30 minutes. The spin column was then centrifuged at 13,000 rpm for 1 minute. The flow through was discarded followed by addition of 200 µl M-wash buffer and centrifugation at 13,000 rpm for 1min. The flow through was discarded and the process was repeated thrice. The spin columns were then centrifuged at 13,000 rpm for an additional 1 minute. The spin columns were then transferred to clean 1.5 ml MCTs and 30 µl M-Elution Buffer was added at the center of the membrane of the spin column and incubated at room temperature for 5 minutes. The spin column was then centrifuged at 13,000 rpm for 1 minute. The elution process was repeated again to elute most of the bound DNA from the column. This eluted bisulfite converted DNA was then used for the BSP or q-MSP.

4.11: Validation of Differential Methylation Using Methylation Specific PCR

Association of differential methylation at key genes with pancreatic cancer, was assessed using MSP in validation cohort 2 (n=22). The input DNA concentration for MSP was 10ng/µl. Primer sequences of the six specific genes (*NPY*, *FOXE1*, *FAIM2*, *SIGIRR*, *RASSF1*, and *KCNA6*) for the methylated and unmethylated templates were designed using MethPrimer (<https://www.urogene.org/methprimer/>) (LC and R version 4.0.4 package, 2002). The MSP primer sequences are listed in table 4 . The MSP was done at 95°C for 4 min, followed by 35 cycles of 94°C for 45 s, annealing temperature (55°C to 63°C, using touchdown PCR) for 45s, and 72°C for 1 min, followed by a final extension at 72°C for 10 min in a 10-µL reaction volume containing 2 µL bisulfite-treated genomic DNA. The final PCR products were observed using electrophoresis with 1.5% agarose gel electrophoresis (1.5% concentration), with a reference of 100bp ladder. For a particular sample and a particular target region, methylation and unmethylation were detected with both bands from M and U lanes and through comparison of the band intensities. Corresponding gene primers were described in table 4.

4.12: Study of Relationship Between Differential Methylation and Gene Expression

Differential methylation at the gene promoters is often inversely associated with its expression. Gene expression data (FPKM) was downloaded for the same set of The Cancer Genome Atlas (TCGA) samples from GDC data portal (n=179). Correlations between the expressions of a gene with differential methylation at CpG site (using β values) annotated to that gene, was estimated using Pearson's correlation. To compare the expressions of selected genes between PanCa and normal pancreas tissues, we additionally investigated the gene expression data (v8) on normal pancreas tissue from GTEx database (<https://gtexportal.org/home/datasets>) (GTEx Consortium 2013). Comparison of expression of a gene between PanCa samples (from TCGA cohort) (n=179) and normal pancreatic tissues (from GTEx database V8) (n=171) was performed using Gepia V2 (Tang *et al.* 2019). Expression of selected gDMs (with promoter methylation) was studied in the tumor and normal tissue samples collected from the validation cohort 3, using RT-PCR (n=26)

4.13: Validation of Differential Methylation Using Immunohistochemical (IHC) Analysis of Pancreatic Ductal Adenocarcinoma Tissue Samples

For Immunohistochemistry staining, 5 μ m tissue sections were prepared in slides from formalin fixed, paraffin embedded PDAC tissue blocks. Routine hematoxylin and Eosin stain were carried out for all sections to ascertain histological features. Further, in separate experiments, all sections were taken for IHC following antigen retrieval in citrate buffer (pH 6.0) at 90°C for 30 min. Primary monoclonal antibody anti-NPY (Cell signalling Technology Inc., USA, 11976T, dilution 1:50), anti-FAIM2 (Novus Biologicals, LLC, USA, NBP2-24713, dilution 1:50) and anti-rabbit HRP conjugated secondary antibody (Santa Cruz Biotechnology, INC. USA, sc-2357, dilution 1:500) were used for immunohistochemical localization of proteins followed by DAB 1X (Sigma Aldrich, USA, 281751) and nuclear counterstaining was performed by routine Hematoxylin Mayer's (HIMEDIA, India, S058). Mounting of stained slides were done by DPX mount media (Sisco Research Laboratories Pvt. Ltd. India, 88147) and observed under Bright field microscope (Leica Microsystems DM750). More than 3 fields per section were chosen to derive the results. The observations were done in both 20X and 40X scopes for better in-depth observation.

4.14: Statistical Analysis

Differentially methylated positions (DMPs) were identified using F-test to compare the beta values between cancer and normal samples. Linear regression was done to identify the

association of methylation values with PanCa, after adjusting for the effects of age, gender, smoking status and tumor purity. Hierarchical clustering was done on Euclidean distance matrix to identify the subgroups of samples differing in the levels of methylation at the DMPs. The R functions “hclust” and “heatmap3” were used for heatmap preparation. Differences in expression of a gene between subgroups were evaluated using Wilcoxon Signed Rank test using GEPIA. Patients in the TCGA cohort were subdivided into high (>90th percentile) and low (<10th percentile) subgroups based on the expression levels of target genes. Overall survival (OS) of the subgroups of patients was compared using Kaplan–Meier method of survival analysis. The correlation between differential methylation and gene expression was estimated using Pearson’s correlation coefficient. The ΔCt values of gene expression data were analyzed by using Wilcoxon Rank sum test (Mann Whitney U test) by “nortest” and “ggpubr” (ggplot2) libraries in R. Overall survival analysis namely Kaplan–Meier survival analysis was done using IBM®SPSS software (IBM SPSS Statistics 27, Armonk, NY: IBM Corp) and R studio considering our Indian patient pool data. Overall and disease-free survival using TCGA data was also done using GEPIA 2.0. Correlations between gene expressions and clinico-pathological features were observed using two different methods— “Kendall tau” correlation and Pearson’s correlation test using SPSS. These survival analysis steps were done for *CD68*, *CD36*, *MIA*, *MOGAT2*, *LRRC31*, *UGT1A1*, *TFF1*, *S100P*, *MUC13*, *CALHM3*, *KCNJ5*, *NPY*, *FAIM2*, *BHLHE23*, *RASSF1A*, *SIGIRR*, *SLITRK3*, *IRF4*, *KCNA6*, *MX2*, *GALR1*, *FOXE1* and *ANKRD44*.

4.15: Validation of Differential Methylation Using Functional Enrichment

The DMPs were annotated to their nearest genes with reference genome hg19. The list of mapped genes was designated as genes with differential methylation (gDM). Enrichment of Gene Ontology (GO) terms on biological processes among the gDMs was done using Metascape (Zhou *et al.* 2019). Metascape was used to study the enrichment of pathways in the KEGG database. Functional enrichment of the DM genes. We first identified statistically enriched terms (including GO/KEGG terms, canonical pathways, hall mark gene sets, etc.,) among the gDMs through hypergeometric test. Accumulative hypergeometric *p-values* and enrichment factors were used for filtering. Remaining significant terms were then hierarchically clustered into a tree based on Kappa-statistical similarities among their gene memberships. Then 0.3 kappa score was applied as the threshold to cast the tree into term clusters. The terms within each cluster are exported in the Excel spreadsheet named “Enrichment Analysis”. We then selected a subset of representative terms from this cluster and

convert them into a network layout. More specifically, each term is represented by a circle node, where its size is proportional to the number of input genes fall into that term, and its color represent its cluster identity (i.e., nodes of the same color belong to the same cluster). Terms with a similarity score > 0.3 are linked by an edge (the thickness of the edge represents the similarity score). The network is visualized with Metascape with “force-directed” layout and with edge bundled for clarity. One term from each cluster is selected to have its term description shown as label. MCODE algorithm was then applied to this network to identify neighborhoods where proteins are densely connected. The GO enrichment analysis was applied to each MCODE network to assign “meanings” to the network component.

4.16: Validation of the Target Genes Using Publicly Available Data Source

To compare altered methylation at the gDMs (*NPY* & *FAIM2*) in PanCa samples, we further analyzed PanCa methylation database (PCMDB, <http://crdd.osdd.net/raghava/pcmdb/>), which is a comprehensive collection of differentially methylated genes (DMG) ($n=4342$) in both PanCa cell lines ($n=88$) and tissue samples ($n=3078$), previously reported to have altered methylation in PanCa.

4.17: Multi-Omics Validation Using Differentially Methylated and Methylation Altered Differentially Expressed Genes

The DMGs methylation pattern across the various CpG sites on their respective chromosomes were investigated on 35 PDAC cell lines using Cancer Cell Line Encyclopedia (CCLE) (<https://depmap.org/portal/>). DeMap portal was used to generate map representing the degree of methylation across the cell lines at various CpG locations. Correlations between gene expression and methylation data from TCGA-PAAD dataset of DMGs were observed using SMART APP (<http://www.bioinfo-zs.com/smartapp>). Protein levels of DMGs in normal tissue versus tumor tissue were analyzed using UALCAN (<http://ualcan.path.uab.edu/>) using Clinical Proteomic tumor Analysis Consortium (CPTAC) Confirmatory/Discovery dataset. Finally, Pan-cancer expression analysis of DMGs were observed using UALCAN. Mutational landscape profiling was done using Cbioportal for damaging mutations and copy number absolute changes. Multi stage analysis was also done targeting gene expression specific to expression sub types by using GEPIA 2.0. These above steps were done for DMGs and MEDEGs namely *NPY*, *FAIM2*, *CD68*, *CD36*, *MIA*, *MOGAT2*, *LRRC31*, *UGT1A1*, *TFF1*, *S100P*, *MUC13*, *CALHM3*, *KCNJ5* and *ANKRD44*.

4.18: Identification of Crucial Methylation Altered Ion Transporter Signatures

Our previous work as reported in *Chatterjee et al. 2022* showed 91 DMGs to be significant in Indian patient cohort. 91 genes underwent metascape and showed “ion transporter” to be most enriched term. We utilized the same 91 differentially methylated probes (DMPs) and did Ingenuity pathway analysis (IPA) which showcased as ion transporter to be also one of the enriched terms. When we compared our IPA observation with Metascape we obtained a list of seven differentially methylated genes found to play most significant role in ion transporter in Indian PanCa cohort.

4.19: Identification of Methylation-Regulated Differentially Expressed Genes

The differentially expressed genes (DEGs) and the DMGs were compared first to identify genes with presence in both datasets. Multiple test corrections were performed using Benjamini & Hochberg’s method and the cut-off values were set at the $FDR < 0.01$. The $|\log_2FC| > 2$ were also observed as secondary screening process as significant methylation alteration doesn’t corresponding imply a $|\log_2FC| > 2$. Then, the CpG sites corresponding to this methylation regulated differentially expressed genes (MEDEGs) were selected. Pearson correlation analysis was performed to calculate the correlation between the methylation level of DMP and the expression level of corresponding DEG. Such DMPs have a significant negative correlation and adj. $p < 0.01$ as well as adj. $p < 0.05$, which were thought to deeply influence the expression of corresponding DEGs, were selected for subsequent analysis and were referred to as MEDEG.

4.20: Whole Genome Bisulfite Sequencing Based Validation of Targeted Methylation-regulated Differentially Expressed Genes

For validation analysis of the five CpGs (*MIA*, *CD68*, and *MOGAT2*) using WGBS, we have utilized previously published data (GSE63123, URL: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63123>). We also did the above experiment using $FDR < 0.05$. We utilized WGBS data Of PanCa progression in patient samples to assess the methylation pattern of CpGs associated with *CD68* (cg05209483) and *MIA* (cg25152942) genes and the three CpG sites from *MOGAT2* (cg12531542, cg20938170, cg25255988). The WGBS data consisted of multiple tumors from same patients (4 patients were analyzed), which included normal pancreas, primary tumor, local peritoneal metastasis, and distant metastasis samples (liver and lung metastasis, data source: GSE63123).

Algorithm 1 Study Cohort

-
- 1: Assemble multiple compendium from TCGA.PAAD, ICGC-PaCa and GEO (GSE74071 and GSE 49149)
 - 2: Acquire RNA data and Methylation Array data from the Pancreatic Dataset Of (TCGA-PAAD).
 - 3: Acquire corresponding clinical information from TCGA,
 - 4: Acquire methylation data of PC patients from the ICGC dataset.
-

Algorithm 2 Identification of Methylation-regulated Differentially Expressed Genes

-
- 0: **procedure** IDENTIFICATION(*RNAseq, Methylation, clinical*)
- 1: *DEGs* $\leftarrow$ Differentially expressed genes from RNAseq data
 - 2: *DMPs* $\leftarrow$ Differentially methylated genes from Methylation data
 - 3: *overlap* $\leftarrow$ Genes with presence in both datasets
 - 4: *pvalues* $\leftarrow$ Multiple test corrections using Benjamini and Hochberg's method
 - 5: *DEGs* $\leftarrow$ DEGs with $FDR < 0.01$ and $|\log_2 FC| > 2$
 - 6: *MEDEGs* $\leftarrow$ Genes with significant negative correlation between methylation and expression
 - 7: **return** *MEDEGs*
 - 7: **end procedure**=0
-

4.21: Dimensional Reduction and Unsupervised Clustering

Principal Component Analysis, a type of unsupervised learning was implemented for dimensionality reduction by assigning correlation between multidimensional information sets. PCA was also used to make the data set easier to interpret by excluding parameter limitations. PCA followed by K-means clustering for unsupervised classification of MEDEG across TCGA- PAAD samples was done to identify the clustering pattern of the 27 MEDEG's spread across 69 Differentially methylated sites (DMS) (47 CpG's). K-means are library tidyverse (data manipulation), library cluster (clustering algorithms), library factoextra (clustering algorithms & visualization), and library ggplot2.

Algorithm 3 Dimensional Reduction and Unsupervised Clustering

-
- 1: Principal Component Analysis (PCA) was implemented for dimensionality reduction by assigning correlation between multidimensional information sets.
 - 2: PCA was also used to make the dataset easier to interpret by excluding parameter limitations.
 - 3: PCA followed by K-means clustering was done to identify the clustering pattern of the 27 MEDEGs spread across 69 DMS (47 CpGs).
 - 4: The following packages were used for K-means clustering: **tidyverse** (data manipulation), **cluster** (clustering algorithms), **factoextra** (clustering algorithms and visualization), and **ggplot2**. =0
-

4.22: Supervised Algorithm Based Machine Learning Models

Multiple predictive models such as the kNN classifier is a non-parametric supervised machine learning algorithm which is a distance-based classifier¹⁰. It is classically suitable for smaller datasets like our 47 CpG probe dataset. For handling bigger data sets and also in terms of

enhanced non-parametric approach, a random forest-based (RF) classifier was built for prediction model development. RF works on the Breiman and Cutler algorithm. Standard python packages were used which includes SciKit Learn, Pandas, Numpy, matplotlib and Seaborn for visualization.

Algorithm 3 Supervised Algorithm-based Machine Learning Models

- 1: Employ multiple predictive models including the K-Nearest Neighbors (kNN) classifier, a non-parametric supervised learning algorithm based on distance metrics, suitable for smaller datasets such as our 47 CpG probe dataset.
 - 2: For larger datasets and enhanced non-parametric modeling, implement a Random Forest (RF) classifier, an ensemble learning algorithm that constructs multiple decision trees during training and performs prediction by aggregating the outputs.
 - 3: The RF algorithm is built upon the Breiman and Cutler methodology, which utilizes bootstrap aggregating (bagging) and feature randomness to construct diverse and robust decision tree ensembles.
 - 4: Leverage the following standard Python libraries for RF classification: scikit-learn for machine learning algorithms, pandas for data manipulation, numpy for numerical operations, matplotlib and seaborn for data visualization. =0
-

4.23: Prediction Model Constructed Using Adaptive LASSO-Cox Regression Model

The construction of MEDEGs -based nomogram consists of multiple steps. The process was initiated with univariate Cox regression analysis that identified an association between the methylation level of each signature and the patient's OS in our scaled cohort ($p < 0.05$). Those sites with P-values less than 0.05 were identified as prognosis-related DMSs. Then, the Adaptive LASSO Regression method was implemented to identify the prognosis associated with signatures, and concurrently, a prime model was obtained. Calibration plots were plotted to compare the observed and predicted probabilities for the nomogram. The signature with coefficient C-index $\neq 0$ were labelled as significant variables, which led to the establishment of a risk-scoring model using the combination of weighted methylation values. The risk scores were calculated as shown in the mathematical model attached. For the purpose of classifying patients into low-risk (risk score below the mean value) or high-risk (risk score above the mean value) groups, the mean risk score value was utilized as the cut-off score. Next, using the risk score and pertinent patient characteristics assessed by multivariate Cox proportional-hazards regression, we created a nomogram based on risk scores. The signature's independent predictive value was assessed after controlling for age, sex, and stage. Additionally, it was utilized to forecast the OS of the TCGA-PAAD cohort for the first and second years.

Algorithm 4 Clinical Relevance of Methylation-regulated Differentially Expressed Genes

- 1: Construct MEDEGs-convoluted prognostic signature
- 2: Perform univariate Cox regression analysis to identify associations between methylation level of each MEDEG and patient's overall survival (OS) in scaled cohort
- 3: Identify prognosis-related DMSs with P-values less than 0.05
- 4: Implement LASSO and Adaptive LASSO Regression method to identify prognosis associated MEDEGs and obtain a prime model
- 5: Label MEDEGs with coefficient, C-index $\neq 0$ as significant variables
- 6: Establish risk scoring model using combination of weighted methylation values
- 7: Calculate risk scores using the following equation:

$$Risk_Score = \sum_{i=1}^n \beta_i \times X_i$$

- 8: Use mean risk score value as cutoff score to split patients into low-risk (risk score below mean value) or high-risk (risk score above mean value) group
- 9: Construct clinical nomogram including risk score and clinicopathological parameters of TCGA-PAAD patients evaluated by multivariate Cox proportional-hazards regression
- 10: Use nomogram to evaluate independent prognostic value of signature after adjusting for age, sex, and stage alongside predicting overall survival (1-year and 2-years) (OS) in TCGA-PAAD cohort
- 11: Evaluate discriminatory ability of nomogram by calculating concordance index (C-index)
- 12: Plot calibration plots to compare observed and predicted probabilities for nomogram =0

4.24: Tissue of Origin of Methylation-regulated Differentially Expressed Genes

Tissue specificity of selected MEDEG's were done by comparing with normal pancreas tissue and other tissue types. The data were taken from GTEx V8 dataset. The Pan-Cancer profile was developed to highlight the role of selected MEDEGs in pancreas tissue by use of TIMER 2.0. Finally, the profiles of selected CpG were compared with profiles obtained from other PanCa databases taken from the GEO database for global profiling purposes, irrespective of demography and ethnicity. These above steps were also done for MEDEGs namely *CD68*, *CD36*, *MIA*, *MOGAT2*, *LRRC31*, *UGT1A1*, *TFF1*, *S100P*, *MUC13*, *CALHM3*, *KCNJ5* and *ANKRD44*

4.25: Functional Relevance of Methylation-regulated Differentially Expressed Genes

Functional characterization of MEDEG's were done to obtain enriched gene ontology (GO) and its associated pathways Under the category, biological processes (BP) ontology that showcased the pathway enrichment which is a crucial resource for an intimate understanding of biological process from large-scale molecular cohorts produced by high throughput experimental technology. The criterion for significant enrichment was based on the Benjamini-

Hochberg adjustment ($p < 0.05$). Functional characterization leading to enriched GO were also obtained using DEGs in PanCa based on BP Ontology. The enrichment network was also obtained by MetaScape, which has its nodes colored based on the p -value. ENRICH R was used to obtain a functional characterization of the MEDEG's in the context of their role in other associated molecular processes and cellular functions. Over-representation analysis (ORA) was also used for network analysis to cross-visualize whether any reasonable output is obtained or not.

Algorithm 6 Analysis of Methylation-regulated Differentially Expressed Genes

- 1: **Tissue of Origin**
 - 2: Compare selected MEDEGs with normal pancreas tissue and other tissue types using GTEx V8 dataset
 - 3: Develop PanCancer profile to highlight role of selected MEDEGs in pancreas tissue using GEPIA
 - 4: Obtain protein level data from CPTAC database to highlight downstream protein level signature role in pancreas tissue using UALCAN
 - 5: Compare profile of selected CpGs with profile obtained from two other PanCa databases taken from GEO database for global profiling purpose irrespective of demography and ethnicity
 - 6: **Functional Relevance**
 - 7: Obtain enriched gene ontology (GO) and associated pathways under biological processes (BP) ontology using DEGs in PanCa based on BP Ontology
 - 8: Use MetaScape to obtain enrichment network with nodes colored based on p -value
 - 9: Use ENRICH to obtain functional characterization of MEDEGs in context of their role in other associated molecular processes and cellular functions ($DEG_{DMG_{Nomo}}$)
 - 10: **Immune and Cell Ecosystem Profile**
 - 11: Obtain immune and stromal cell infiltration data with selected MEDEGs using ESTIMATE and identify significance using Kruskal-Wallis
 - 12: Obtain immune profile and associated ecosystem data using deconvolution-based immune classifiers such as CIBERSORT and EcoTyper
 - 13: Use gene marker-associated immune classifier via xCell to infer 64 immune and stromal cell types
 - 14: **Pharmacogenomic Screening**
 - 15: Select top three genes for molecular dynamics lead simulation-based visualization of interaction among molecules
 - 16: Evaluate interaction between genes and drug molecules specific for pancreatic cancer tissue for suitable drug target identification of genes of interaction
 - 17: Consider drugs undergoing both clinical trial and in vitro development of small molecules =0
-

4.26: Pathway Analysis of Methylation-Driven Differentially Expressed Genes

Functional characterization of top signatures was done to obtain enriched GO and its associated pathways concerning protein-protein interactions. The criterion for significant enrichment was based on the Benjamini-Hochberg adjustment ($p < 0.05$). SHINY GO was used to obtain a functional characterization in the context of their role in other associated molecular processes and cellular functions along with biological processes (Arowolo *et al.* 2024). LinkedOmics is a highly effective publicly accessible portal that contains multi-omics data from all 32 TCGA cancer types. LinkedOmics submodules such as LinkFinder (for gene association) and LinkInterpreter (for enrichment analysis) were used to

investigate the genes that showed disparities in association with in PanCa. RNA-seq datasets with 178 sample sizes were selected from the UNC Institute's module and analyzed using Pearson correlation coefficients before being characterized in volcano plots. Genes with positive or negative correlations were screened using the p-value ($p < 0.01$). The co-expressed genes were used for gene enrichment analysis (GSEA) with the GSEA database and the KEGG Pathways plugin. The parameters for enrichment analysis were the lowest number of IDs in the 25 categories, the lowest and highest number of IDs in the categories were 3 and 2000, respectively, the top 25 significance levels, and 500 permutations.

4.27: Immune and Cell Ecosystem Profile of Methylation-regulated Differentially Expressed Genes

Immune and stromal cell infiltration data concurrent with selected MEDEG's were obtained via ESTIMATE and significance was identified using Kruskal Wallis. Immune profile and its associated ecosystem associated data were obtained using deconvolution-based immune classifiers such as CIBERSORT and ECOTYPER. Gene marker-associated immune classifier was used alongside xCell. xCell, a novel gene signature-based method, was used to infer 64 immune and stromal cell types. EcoTyper, an upgrade of CIBERSORT X (machine learning framework) was used for large-scale identification and validation of cell states and multicellular communities from bulk gene expression data. Applied to 12 major cell lineages across 16 types of human carcinoma.

4.28: Immune Function Correlation Analysis

Immune function correlation analysis was done using CIBERSORT X and was visualized using GEPIA gene specific correlation platform. Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) enables deconvolution of gene expression data for immune cell profiling study.

4.29: Identification of Candidate Small Molecules Drug Candidates Prediction

A tool called Connectivity Map (CMap, <https://clue.io/>) is designed to forecast possible medications that could cause biological states represented by particular gene expression profiles (Zhou *et al.* 2023). To investigate potential small molecule treatments, we separated the differential genes that were ultimately screened into groups that were up-regulated and down-regulated. These groups were then imported into the CMap database. A mean score that is negative suggests that the medicine has the potential to be therapeutic and that it reverses the desired biological feature ($p < 0.05$ for statistical significance). The candidate small-molecule

compounds were determined with an |enrichment score| >90. The genes that were then uploaded to CMap database were *CD36*, *TFF1*, *S100P*, *MUC13*, *UGT1A1* and *KCNJ5*.

Algorithm 5 Analysis of Methylation-regulated Differentially Expressed Genes

- 1: Tissue of Origin
- 2: Compare selected MEDEGs with normal pancreas tissue and other tissue types using GTEx V8 dataset
- 3: Develop PanCancer profile to highlight role of selected MEDEGs in pancreas tissue using GEPIA
- 4: Obtain protein level data from CPTAC database to highlight downstream protein level signature role in pancreas tissue using UALCAN
- 5: Compare profile of selected CpGs with profile obtained from other PanCa databases taken from GEO database for global profiling purpose irrespective of demography and ethnicity
- 6: Functional Relevance
- 7: Obtain enriched gene ontology (GO) and associated pathways under biological processes (BP) ontology using DEGs in PanCa based on BP Ontology
- 8: Use MetaScape to obtain enrichment network with nodes colored based on p-value
- 9: Use ENRICHR to obtain functional characterization of MEDEGs in context of their role in other associated molecular processes and cellular functions (DEGsIMGsemo)
- 10: Immune and Cell Ecosystem Profile
- 11: Obtain immune and stromal cell infiltration data with selected MEDEGs using ESTIMATE and identify significance using Kruskal-Wallis
- 12: Obtain immune profile and associated ecosystem data using deconvolution-based immune classifiers such as CIBERSORT and EcoTyper
- 13: Use gene marker-associated immune classifier via xCell to infer 64 immune and stromal cell types

4.30: Prediction of Drugs Against Kcnj5

Potential drugs against Kcnj5 were predicted using GDSC1, GDSC2, CTRP, and PRISM databases followed by the method described by *Mao et al. 2024*.

4.31: Identification of Tfs that Regulate Differentially Methylated Genes

The Transcription Factors (TFs) for *KCNJ5* were identified by combining the TF binding data of ChEA and RegNetwork . Both these databases contain only ChIP-Seq data. The common TFs in both the databases for *KCNJ5* were considered in our analysis. The genes that are then checked were *TFF1*, *S100P*, *MUC13*, and *KCNJ5*

4.32: Analysing RNA-Seq Datasets to Identify Ion Channel Genes in Pancreatic Ductal Adenocarcinoma

Initially, we combined results from two RNA-seq data sets and TCGA dataset to generate a list of DEGs for PDAC. The DEG list was prepared by considering a cut off of adjusted $p <$

0.05 and $|\log_2FC| > 0$. The two RNA-seq datasets were GSE211398 and GSE93326. We were able to detect genes by comparing 66 tumor epithelial samples with 123 stroma samples in GSE93326. In a similar manner, we compared the transcriptome data of 12 non-tumor samples with 15 PDAC samples to find the genes for GSE211398. To find DEGs in these two datasets, we utilized R 4.4.1 and DEseq2. For DEGs that were present in more than one dataset, we considered their average expression across the datasets. Next, we downloaded a list of 330 ion channel genes from HUGO Gene Nomenclature Committee (HGNC) and then analyzed their gene expression in the screened DEG list to get differentially expressed ion channel genes (DEICGs). Functional enrichment and cluster analysis of the DEICGs was done in Metascape (Zhou *et al.* 2019).

4.33: Relevant Dataset's Based Validation of *KCNJ5*'s Expression in Pancreatic Cancer.

The role of *KCNJ5* was evaluated in few other GEO based PanCa datasets, comprising of, 450K methylation data: GSE74071 (21 malignant and 7 adjacent normal) and GSE49149 (167 tumor and 27 adjacent normal) and RNA-Seq data: GSE211398 (28 tumor tissues) and GSE16515 (36 Tumor and 16 normal).

4.34: Validation Of *KCNJ5* Gene Expression in HPA Database

The Human Protein Atlas (HPA) database was used to compare the transcriptome and protein levels of the *Kcnj5* gene's expression (<https://www.proteinatlas.org/>). The gene expression of *Kcnj5* was analysed for both normal pancreatic tissue and 46 cancer cell lines. Also, the protein expression of *Kcnj5* in normal pancreas was assessed using immunohistochemistry results in this dataset.

4.35: Study Cohort for Validation of Identified Genes

The genes identified using this cohorts are *KCNJ5*, *CD36*, *TFF1*, *S100P*, *MUC13*, *ANKRD44*, *CALHM3*, *MOGAT2*, *MIA*, *CD68*, *LRRC31*, and *C2ORF54*. Initial screening and identification of signatures was done using TCGA pancreatic cancer database (RNA-Seq expression and Meth Array 450K Data). The validation cohorts used for downstream evaluation consisted of ICGC-PaCa data (Meth Array Data) along with 5 other GEO datasets.

- Methylation: GSE74071 (21 malignant and 7 adjacent normal) and GSE 49149 (167 tumor and 27 adjacent normal) Total= 188 tumor + 34 adjacent normal
- GEO datasets containing microarray expression data (GSE183795, GSE28735, and GSE15471) Total= 139 tumor with 102 adjacent normal/ 45 pairs of tumor and adjacent normal/ 36 tumors and matching normal samples = 220 tumor + 183 adjacent normal

- RNA-Seq data: GSE211398 = 28 tumor tissues only, GSE16515 = 36 tumor and 16 normal tissues, GSE93326 = 15 tumor samples

4.36 Data Availability

The datasets generated and/or analyzed in the current study are available from the corresponding author upon reasonable request. The codes are available in the following GitHub link <https://github.com/Akash-WASY/pancreatic-cancer/tree/master/paper-II>



RESULTS AND ANALYSIS



SEPTEMBER 16, 2024

PHD THESIS

Department of Life Sciences and Biotechnology, Jadavpur University

5.1: Results and Analysis

5.1.1: Characteristic of Patients Recruited in the Study

Among the 185 PanCa patients 7 were African-Americans, 11 were Asian, 160 were Caucasian population origin and 7 were from unknown source. Among 9 samples in discovery cohort 6 patients were Caucasian, 1 from Asian and 1 from African-American population origin and 1 was from unknown source. Amongst 185 PanCa patients 3 samples were excluded due to unavailability of histopathological data. Among the 182 PanCa patients in the TCGA cohort, 5 patients had stage IV PanCa (2.7%), 4 patients had stage III PanCa (2.2%), and the remaining patients had stage I + II (95.1%) (Tables 5, 6). The demographic and clinical features of the patients included in the validation cohorts were summarized in Table 6. Among the 7 patients in validation cohort1, 5 were males and 2 were females. Average age of the patients was 53.14 years. Among the males, 3 had a history of tobacco intake, either through smoking or chewing habits. None of the females had any history of tobacco intake. Histopathological analysis showed different grades of tumor among the patients with increasing severity: well-differentiated adenocarcinoma (WDA: $n=2$), moderately differentiated adenocarcinoma (MDA: $n=2$) and poorly differentiated adenocarcinoma (PDA: $n=3$). The validation of methylation was done among validation cohort 2 with 22 PDAC patients. Among them, 18.2% ($n=4$) of the patients had PDA PDAC, 27.3% ($n=6$) had MDA PDAC and 54.5% ($n=12$) had WDA PDAC. The validation of gene expression was done among validation cohort 3 with 26 PDAC patients. Among them, 30.8% ($n=8$) of the patients had PDA PDAC, 23.1% ($n=6$) had MDA PDAC and 46.2% ($n=12$) had WDA PDAC. The validation of target genes associated protein level expression by IHC experiment was done among validation cohort 4 with 20 PDAC patients. Among them, 25% ($n=5$) of the patients had PDA PDAC, 55% ($n=11$) had MDA PDAC and 20% ($n=4$) had WDA PDAC (Table7).

5.1.2: Differential Methylation in Pancreatic Cancer

The TCGA-PAAD cohort dataset included DNA methylome data on tumor and normal tissue samples from 9 patients. Global differential methylation marks were identified by comparing the 9 PanCa samples with their solid tissue normal (Table 8). A total of 7832 DMPs ($|\Delta\beta| > 0.2$, $p < 0.05$) were identified between these tumor and normal samples in the TCGA cohort (Table 8, Figure 19 A). Among the DMPs, 54% were hypermethylated and 46% were hypomethylated (Figure 19 B). The 7832 DMPs could correctly classify cancer and normal samples in separate clusters (Figure 19 A). To compare the differential methylation in the

TCGA PanCa samples, we selected the study by *Mishra et al. 2017*, where the authors submitted a list of DMPs in all PanCa samples in the TCGA cohort. All the 7832 DMPs were also significantly differentially methylated in PDAC in their study. The delta beta values (i.e. differences in the beta (β) values between cancer and normal samples), estimated in our study were highly correlated with their study results (Figure 19 C).

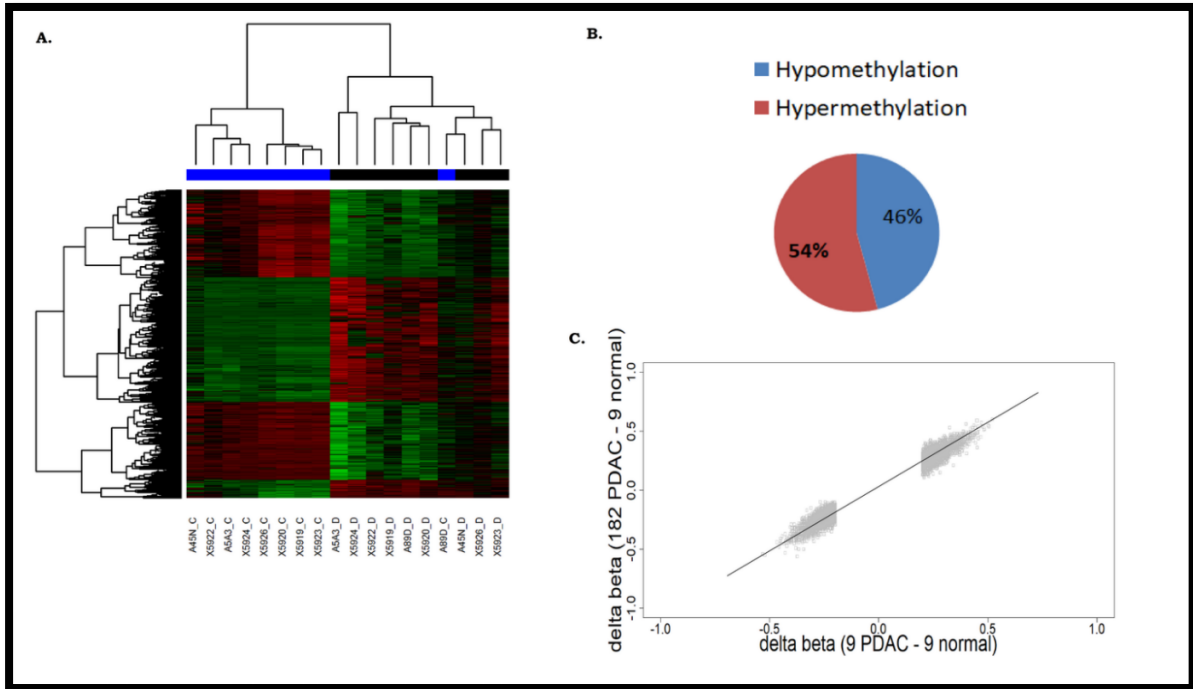


Figure 19: Differentially methylated positions identified in the pancreatic cancer samples of the TCGA cohort. DNA methylation data from cancer and solid tissue normal samples was downloaded for 9 PAAD patients in the TCGA cohort.

(A) 7832 DMPs were identified: Hierarchical clustering analysis showed separate classification of cancer and solid tissue normal samples, based on 7832 DMPs. (B) 54% were hypermethylated and 46% were hypomethylated. (C) Delta beta correlation plot correlating 9 cancer and solid tissue normal against 182 cancer and 9 solid tissue normal.

We have performed the differential methylation analysis in the validation cohort 1 ($n=7$), with two aims: (1) the DMPs in PanCa identified from the TCGA cohort and the Indian cohort should be concordant and (2) novel differential methylation marks could be identified specific to the Indian patient cohort, due to ethnicity and cultural distinctions. We identified a total of 156 DMPs ($|\Delta\beta| > 0.2$), mapped to 91 genes in PDAC tumor compared to adjacent normal counter parts from 7 PDAC patients (Table 9) after Benjamini Hochberg multiple testing correction ($q < 0.05$). Hierarchical clustering of the samples with these 156 DMPs, showed separate clustering of cancer and the normal samples (Figure 20A). Among the 156 DMPs, 37.2% ($n=58$) were hypermethylated and 62.8% ($n=98$) were hypomethylated

among cancer samples (Figure 20B). Among the previously identified 7832 DMPs in the TCGA cohort, 68 DMPs overlap with DMPs, identified from our data (Figure 20C, Table 6). The differences in beta values between the paired tumor and normal samples at these 68 DMPs in the TCGA data (for 9 PAAD patients) showed a positive correlation with our study (correlation coefficient = 0.98) (Figure 20D, Table10).

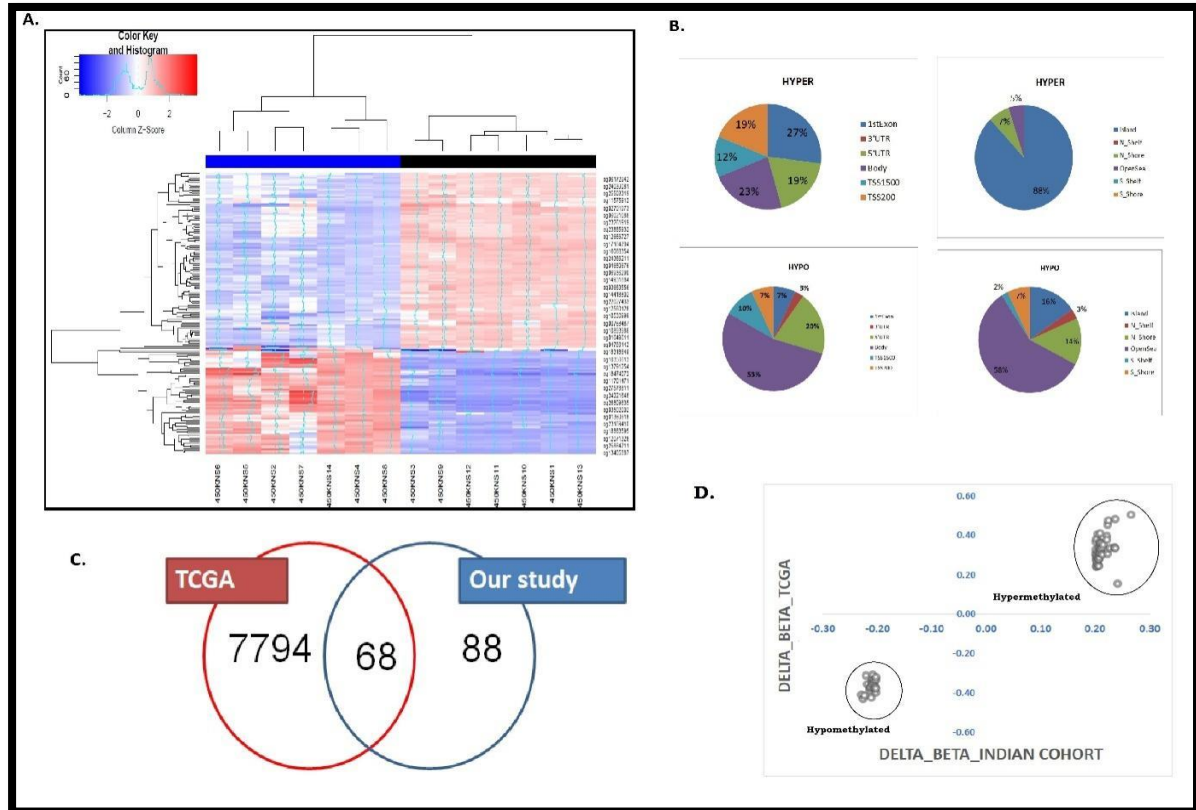


Figure 20: Illustration of 450K array-based data clustering

(A) Hierarchical clustering of the samples based on the top-ranked DMPs ($n = 156$): Illustrative heat map denoting the Hierarchical Clustering of the PDAC samples ($n = 7$) based on the top-ranked DMPs ($n = 156$). Blue to red denoted increase in beta value (hypermethylation). The cancer samples were marked as Blue and the normal samples were marked as Black (B) Genomic annotations of DMPs-CpG islands (Islands, Shores (± 2 KB from the boundaries of the islands), Shelves (± 2 KB from the boundaries of the shores) and Open Sea) or the transcription start site (5' UTR, Exon 1, Promoter, Body, Non-genic). Distribution of the hypomethylated and hypermethylated DMPs across different segments of the genome. (C) Venn diagram showing common and uncommon DMPs in the TCGA cohort and Indian cohort. Among the 7832 DMPs, 68 DMPs were common to our study finding. (D) Correlation plot between the delta beta values at these 68 DMPs in TCGA data (for 9 PAAD patients) and in Indian cohort ($n = 7$)

Excluding the 68 DMPs, methylation at the remaining 88 DMPs (out of 156 DMPs) was significantly associated with PDAC in validation cohort 1 (Table11). We have looked into the hemi-methylation at the 156 DMPs in the cancer samples of our cohort ($n = 7$) and also in the TCGA cohort. Our results showed that the average beta values at the DMPs between 0.2 and

0.8 were 100% in our cohort and 91.6% in the TCGA cohort (Figure 21).

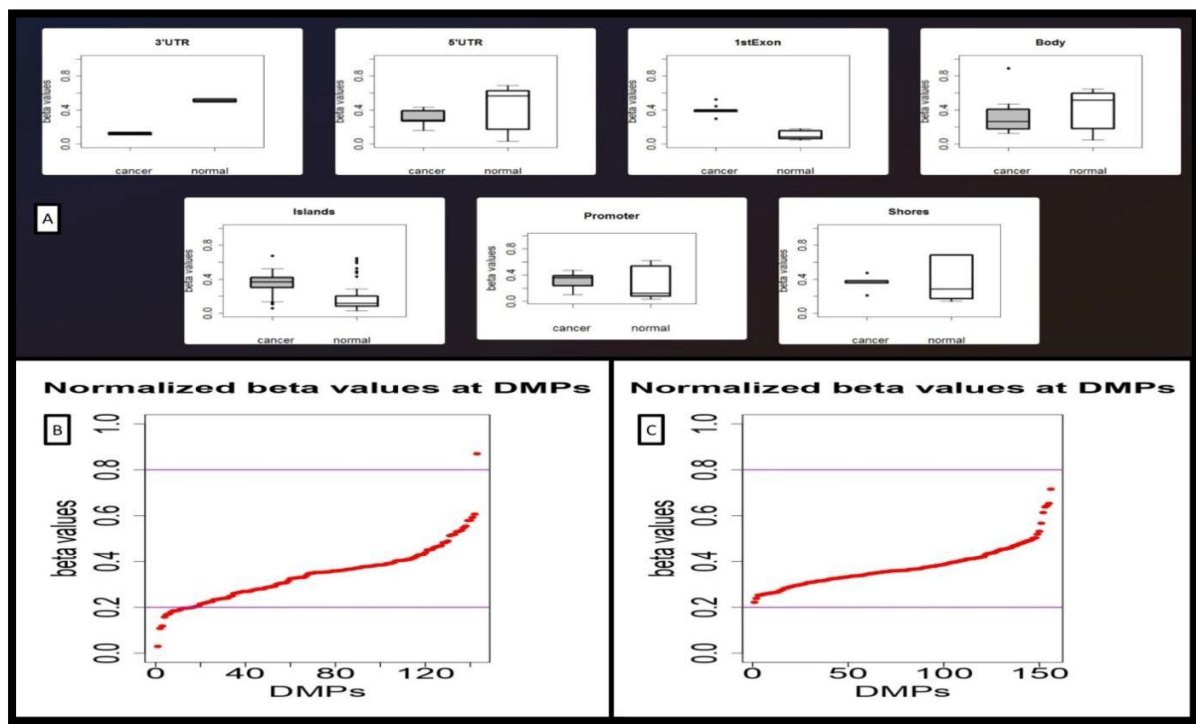


Figure 21: Representative images of 156 differentially methylated positions

plotted based on: (A) DMP's have been plotted based on the distribution of beta values at the 156DMPs across different annotated regions of the genome. (B & C) The average beta values lie within the range of 0.2-0.8.

5.1.3: Change of Methylation Status at the DMPs with PDAC Cancer Prognosis

Methylation marks in the DNA can be associated with the severity of cancer stages. We hypothesized that differential methylation signatures at the DMPs were related to poor prognosis in PDAC. We have estimated the average beta value for each cancer stage in the TCGA cohort (stages IV, III, II and I) (Figure 22). For the hypomethylated DMPs, the cancer samples with stage I showed the highest average beta values and stage IV samples showed the lowest average beta values (Figure 22A). Reversely, for the hypermethylated DMPs, the highest levels of methylation were observed in cancer samples of stages IV and III compared to stage I and stage II (Figure 22B).

Gene methylation was often related inversely to gene expression. We further compared the expressions of the gDMs between the TCGA cohort of PanCa samples and the GTEx cohort of normal pancreas tissues. Out of 91 gDMs, 41 gDMs showed significant differences in expression in the TCGA PanCa samples compared to the pure healthy pancreas tissues in the GTEx cohort (Table12). For the hypomethylated gDMs, higher expressions were observed in the TCGA cancer samples compared to the GTEx cohort samples (Table12). The reverse

scenario was observed for hypermethylated gDMs, where PanCa samples in TCGA had lower expression than the normal pancreatic tissues in the GTEx cohort (Table12).

We have used GEPIA2 to identify the significant changes in expression of the 41 gDMs in the TCGA-PDAC cohort. A total 17 genes among the 41 gDMs showed significant differences in expression across cancer stages in the TCGA cohort (Table9). hypomethylated gDMs like *MX2*, *OAS2*, *SPRED2*, *ADAP1*, and *SLC17A9* genes showed significant increased expression in cancer stage IV compared to stage I (Figure 22C). Although expressions of 9 hypermethylated genes were significantly different across cancer stages in the TCGA cohort, reduced expressions with increased cancer severity were observed for 3 hypermethylated genes- *SLITRK3*, *FOXE1*, and *TWIST1* (Figure 22D).

We have further investigated the survival of the patients with different levels of expressions of gDMs in the TCGA cohort, using GEPIA2 (Tang et al. 2019). The survival of patients with higher expression of hypomethylated gDMs was worse than those with low expression (*APOL3*, *RASSF1*, *OAS2*, and *MX2*). A reverse scenario was observed for hypermethylated gDMs (*CPEB1*, *KCNA3*, *FAIM2*, and *CACNB2*) (Figure 22E).

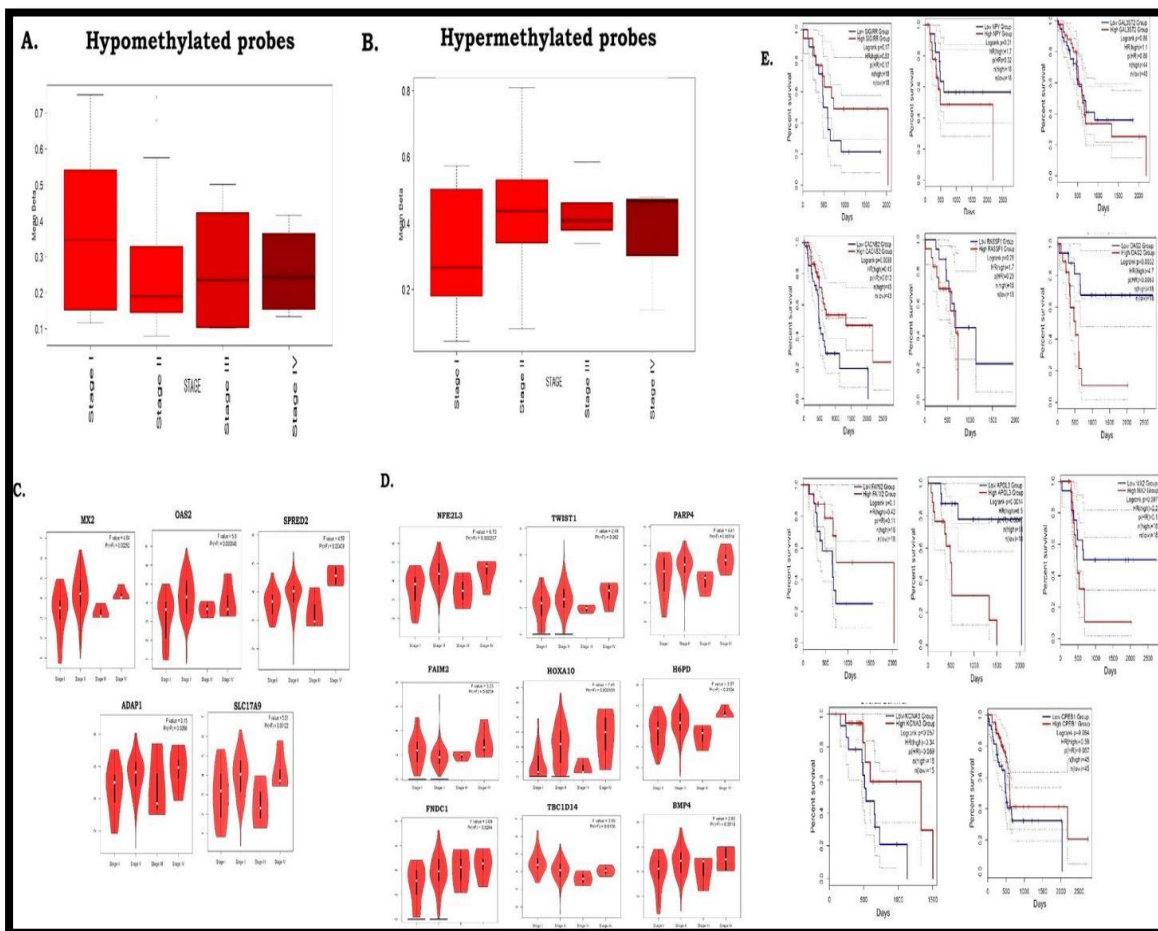


Figure 22: Changes of methylation with increasing severity of cancer phenotypes in the TCGA cohort.

(A) Changes in methylation at the hypomethylated DMPs from Stage I to Stage IV of cancer in the TCGA cohort. (B) Changes in methylation at the hypermethylated DMPs from mild to severe stages of cancer in the TCGA cohort. (C) Higher expression of hypomethylated gDMs observed from Stage I to Stage IV cancer samples. (D) Changes in expression of hypermethylated gDMs across stages of TCGA cohort. (E) Survival plots showing differences in disease prognosis among the TCGA cohort patients with high and low expressions of the gDMs.

To further validate our findings of the association of 91 gDMs with PDAC, we searched PCMDB database. Among the 91 gDMs, methylation data of 59 genes were retrieved from the database. For most of the hypermethylated gDMs, we obtained concordant results from both the tissue samples and cell lines (Tables 13).

5.1.4: Relation Between Promoter Methylation at DMPs with Gene Expression

Hypermethylation at the promoter can cause DNA compaction, giving limited access to the binding sites of the transcription factors and vice versa. Consequently, promoter hypermethylation can cause reduced gene expression and vice versa. In our study results, among the hypomethylated DMPs and the hypermethylated DMPs, 37% and 50% reside on the gene promoters (1500 bp upstream of transcription start site [TSS]), 5'UTR, respectively (Data previously shown). Aberrant methylation at the CpG islands and surrounding shores (± 2 kb from CpG islands) was found to be associated with tissue-specific and cancer-specific methylation signatures. Among the hypermethylated DMPs, 88% resided on the annotated CpG islands and 7% on the CpG shores. Comparatively, hypomethylated DMPs were less common in the CpG islands (16%) and shores (21%).

When compared with the normal pancreas tissue ($n = 171$) from GTEx database we observed genes with promoter methylation to be differentially expressed in the PanCa samples of the TCGA cohort (Table 12). Up-regulated expressions of the hypomethylated gDMs in cancer were observed for *OAS2*, *MX2*, *APOL3*, *ABHD8*, *RASSF1*, *SIGIRR*, and *SPRED2* (Figure 23). For the hypermethylated gDMs, down-regulated expression in cancer samples was observed for *CACNB2* and *ZIC1* (Figure 23).

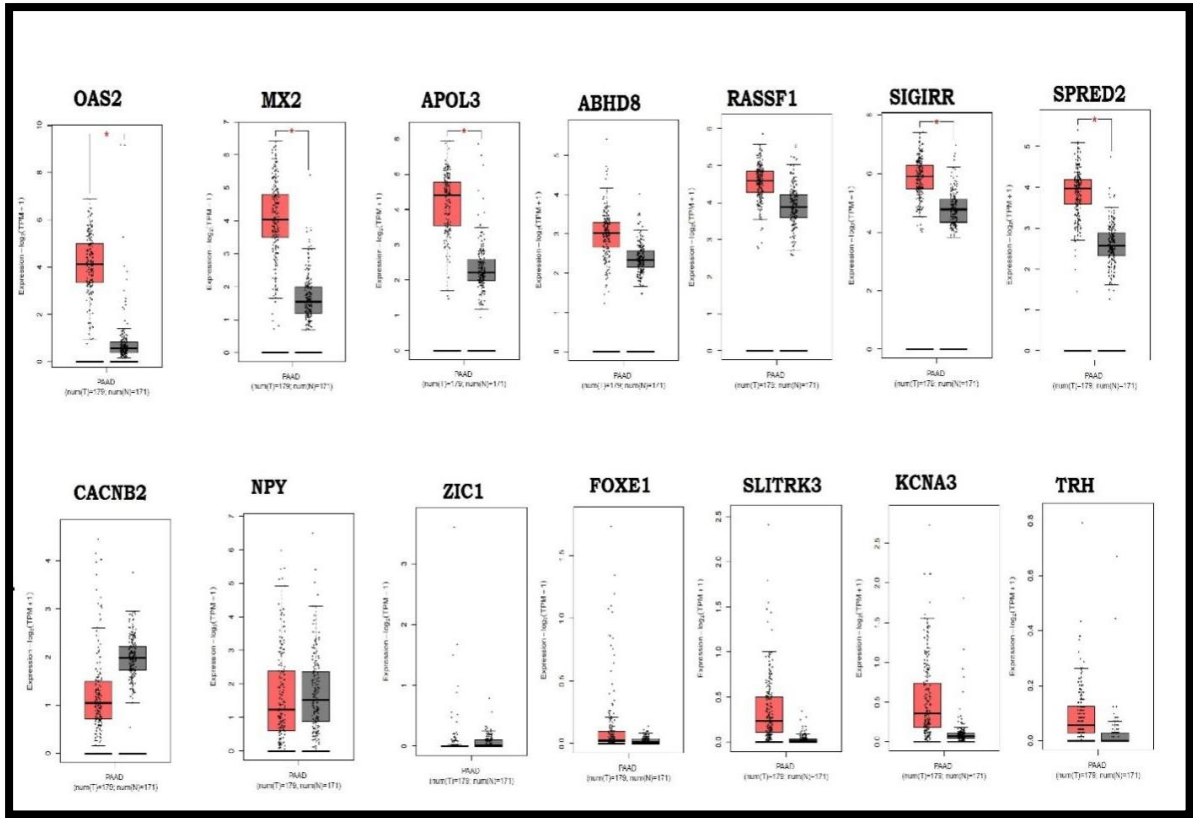


Figure 23: Differences in expressions of genes with promoter hypomethylation and hypermethylation.

Boxplots showing expressions of gDMs among the pancreatic cancer samples in TCGA cohort (FPKM) and normal pancreas tissue samples from GTEx database V8. This data represents the gene expression data developed from the RNA Seq databases from both the respective datahubs. 178 TCGA tumor samples were compared against 171 GTEx normal pancreas tissue samples. “*” indicated significant differences.

To validate promoter associated methylation, MSP was done for 3 promoter CpG sites at 3 gDMs- *RASSF1*, *SIGIRR* and *KCNA6* in validation cohort 2 (n=22). Our results showed, hypermethylation at *KCNA6* promoter and hypomethylation at *RASSF1* and *SIGIRR* promoter in PDAC samples compared to the normal samples, in the validation cohort 2 (Figure 24A). We have further validated differential promoter methylation at *FAIM2*, *NPY*, and *FOXE1* in validation cohort 2 using MSP. Interestingly, data on MSP from the validation cohort showed an increased percentage of promoter methylation among the cancer samples compared to the normal samples (Figure 24A, B). Furthermore, we consider the *RASSF1*, we observed a higher percentage of unmethylation than methylation signals in cancer. For *SIGIRR*, we observed, a reduced amount of methylation in cancer compared to the controls. Both observations suggested hypomethylation of *RASSF1* and *SIGIRR* in PDAC (Figure 24C). For the remaining gDMs (*NPY*, *FOXE1*, *FAIM2*, and *KCNA6*), we found hypermethylation in

PDAC, as confirmed by the higher extent of methylation in cancer samples (Figure 24C).

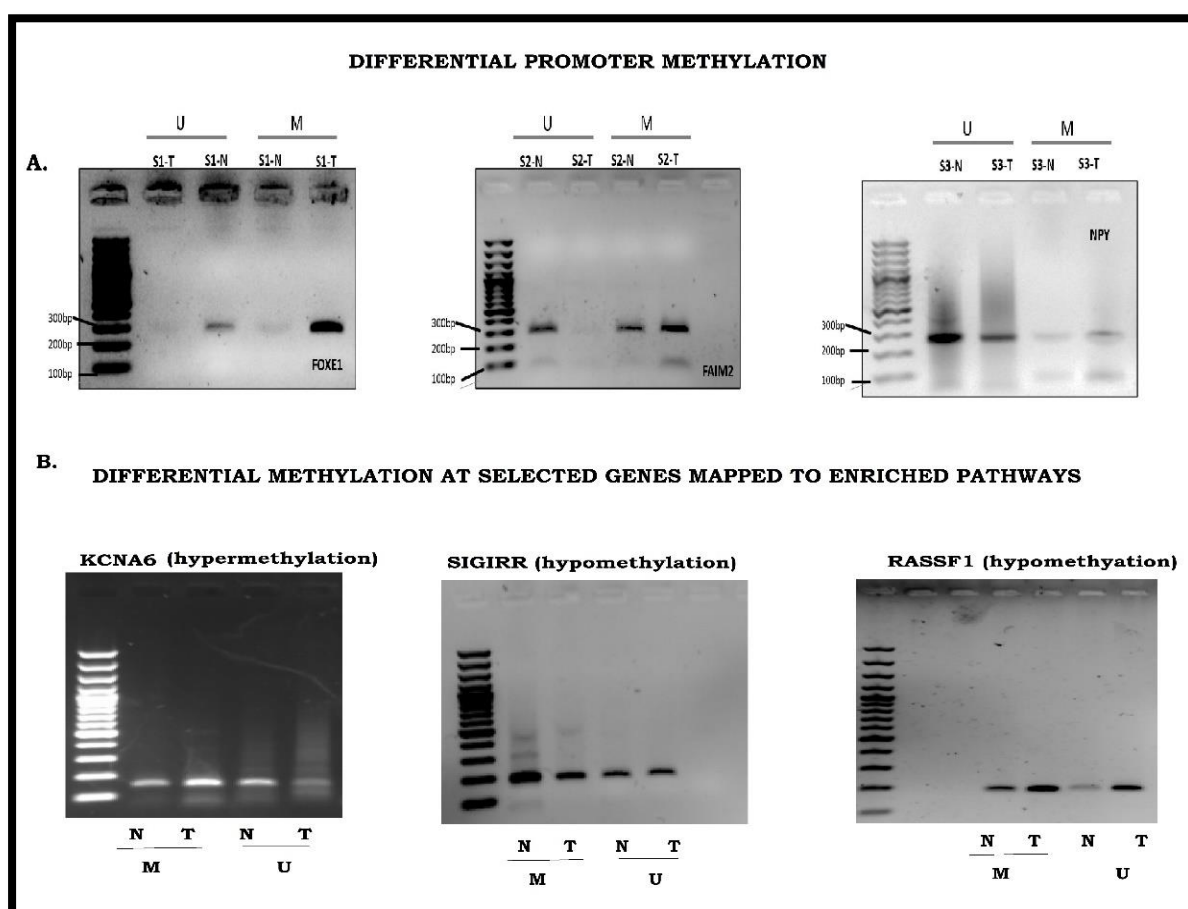


Figure 24: Agarose gel electrophoresis of methylation specific PCR products:

In all sub figure first lane represents 100 bp ladder, next two-lane methylated tumor and normal PCR products and proceeding two lanes are unmethylated tumor and normal PCR products. A Representative agarose gel images for the methylation status of pathway enriched loci namely KCNA6 (M:145 bp, U:147 bp), RASSF1 (M:193 bp, U:197 bp) and SIGIRR (M:192 bp, U:192 bp). B Representative agarose gel images for the differential promoter methylation status of FAIM2(M:278 bp, U:286 bp), NPY (M:264 bp, U:266 bp) and FOXE1(M:275 bp, U:277 bp). For each subset the same patient's tumor and normal has been used. C Percentage of methylation in cancer and normal samples was shown in bar graph. Triple independent validation of the respective methylation status of above genes across all 22 (Adjacent tumor and normal) paired samples were done by Agarose gel electrophoresis of methylation specific PCR products. "M" represents methylated PCR product; "U" represents, unmethylated PCR products; "N" represents Adjacent Normal Sample; "T" represents Tumor sample

We validated differential expressions of 3 hypomethylated gDMs (RASSF1, SIGIRR, and MX2) (Figure 25A) and 3 hypermethylated gDMs (KCNA6, GALR1, and IRF4) (Figure 25B) in validation cohort 3 (n=26). The extended promoter methylation in cancer samples of validation cohort 1 was observed. Using the distribution of beta values at all the CpG sites annotated to the gene promoters containing DMPs, we generated density plots that

showed differences in average beta values in cancer compared to controls (difference in median $|\text{beta value}| \geq 0.1$) at each CpG site in the promoters of six genes (*FAIM2*, *NPY*, *GSC2*, *BHLHE23*, *SLITRK3* and *FOXE1*) (Figure 25C, 26). The expression levels of *FAIM2*, *NPY*, and *FOXE1* were also compared in the cancer samples with normal samples in validation cohort 3 and a negative relationship between methylation and gene expression was observed (Figure 25C, Tables 14, 15).

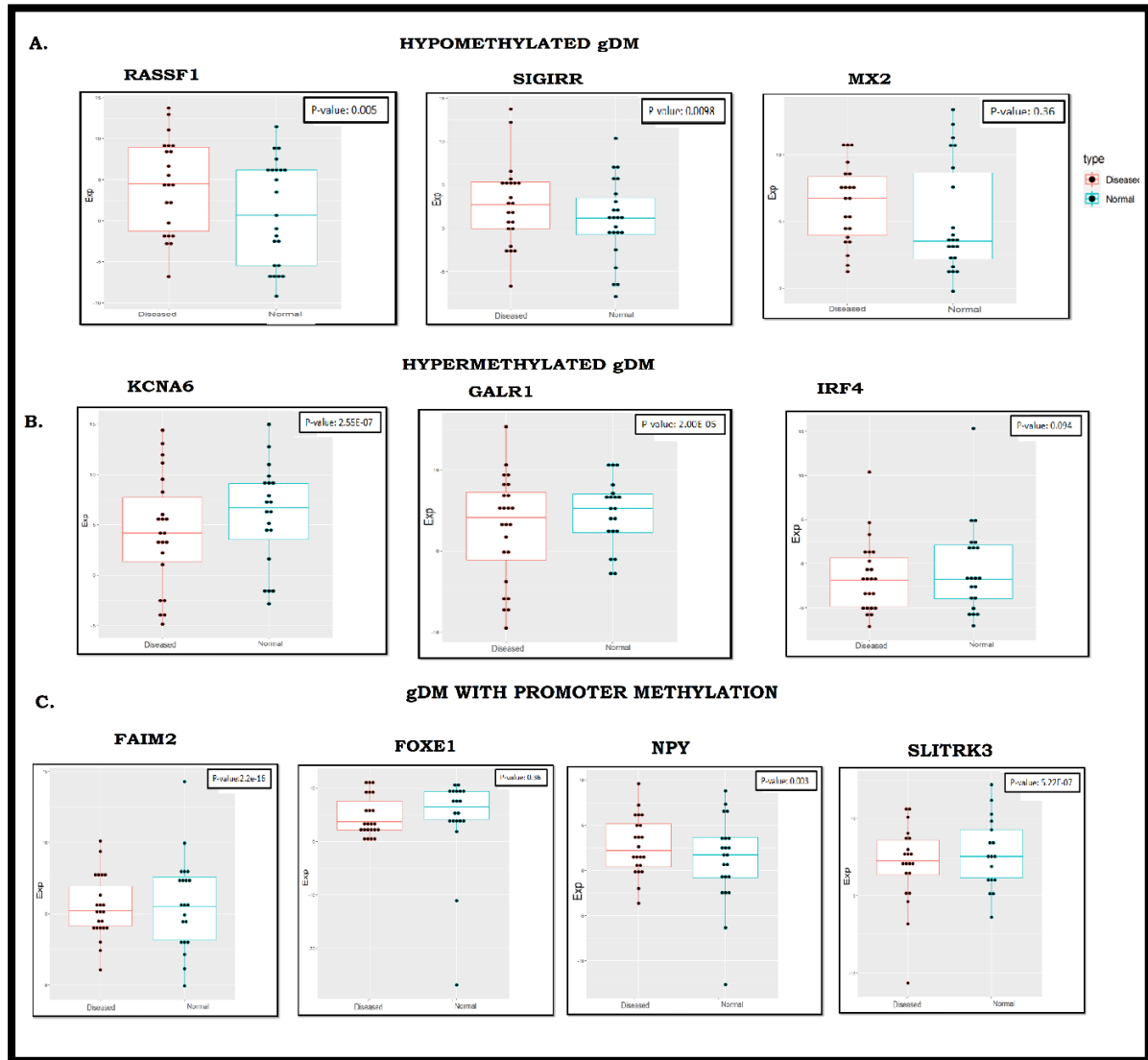


Figure 25: Validation of differential expressions of gDMs in validation cohort 3 (n = 26).

(A) hypomethylated gDMs. (B) hypermethylated gDMs. (C) gDMs with broad promoter methylation

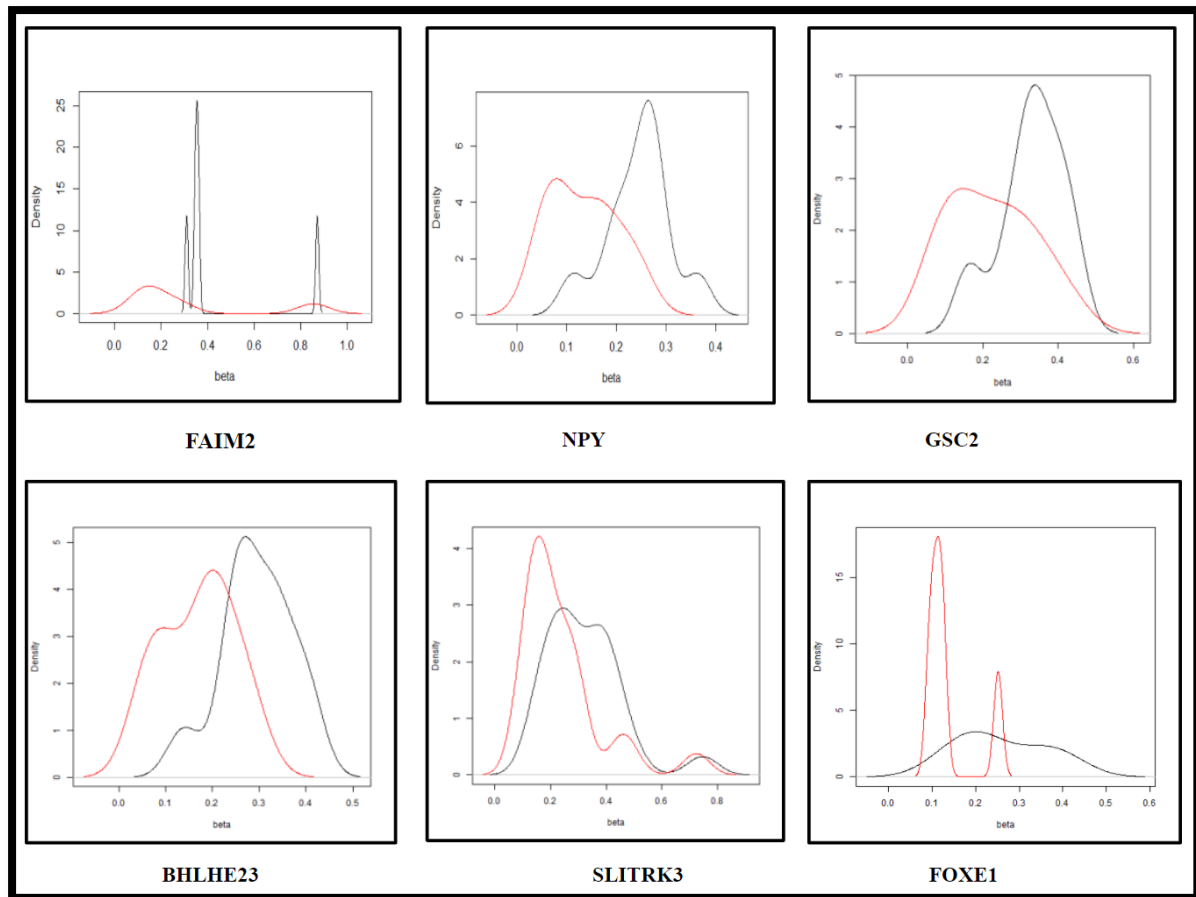


Figure 26. Distribution of average beta values at the promoters of 6 target genes between tumor and adjacent normal.

Tumor is represented as the red curve against the black curve of adjacent normal.

5.1.5: Enriched Pathway Analysis Using of Differentially Methylation Genes in PDAC

The 156 DMPs identified in our study were annotated to 91 genes with differential methylation (gDMs) (Table 9). Enrichment analysis with GO/KEGG terms, canonical pathways, hall mark gene sets, among the 91 genes identified “regulation of ion transport”, “interferon alfa/beta signalling”, “morphogenesis and development” and “transcriptional misregulation” as the 4 most statistically significant enriched terms (Figure 27A). The genes included in each enriched term were listed in Table16. Voltage-gated ion channels are integral membrane proteins, which selectively transport ions and are activated upon a change of membrane potential. Channel activation enables transportation of potassium ions down their electrochemical gradient. The low expression of Kv1.3 in PDAC can be explained by a hypermethylation of the *KCNA3* gene promoter. The ion channel regulation was seen to have a significant role in regulating cell apoptosis, evasion and survival along with invasion and progression in PDAC (Tables 16, 17).

Network analysis with selected enriched terms with a similarity score > 0.3 were linked with

an edge and showed several interactions (Figure 27B, C). Enriched terms with higher similarity, thicker were the defined edges (Figure 27B). Stronger interactions were observed between— (1) “regulation of ion transport”, (2) “Interferon alfa/beta signalling”, (3) morphogenesis and development and (4) transcriptional misregulation in cancer. We next ran the MCODE algorithm in Metascape on this network to identify neighborhoods, where proteins were found to be densely connected. MCODE algorithm identifies two major clusters (Table 18)—MCODE1 (including *HLA-A*, *MX2*, *IRF4*, *PML* and *OAS2*) (Figure 27D), which were mostly hypomethylated in PanCa and MCODE2 (including *ADCY3*, *NPY*, *GRM7*, *GALR1* and *OPRK1*) which were mostly hypermethylated in PanCa (Figure 27D). MCODE1 cluster proteins function in Interferon signalling whereas the MCODE2 cluster proteins function in cell-signalling (Table 18). Differential survival of genes annotated in MCODE1 (*MX2*, *OAS2*, *IRF4*, *PML*, and *HLA-A*) and MCODE2 (*ADCY3*, *GALR1*, *NPY*, and *OPRK1*), in the TCGA PanCa cohort further showed that patients with lower expressions of genes annotated to MCODE1 and higher expressions of genes annotated to MCODE2 showed better disease prognosis (Figure 28).

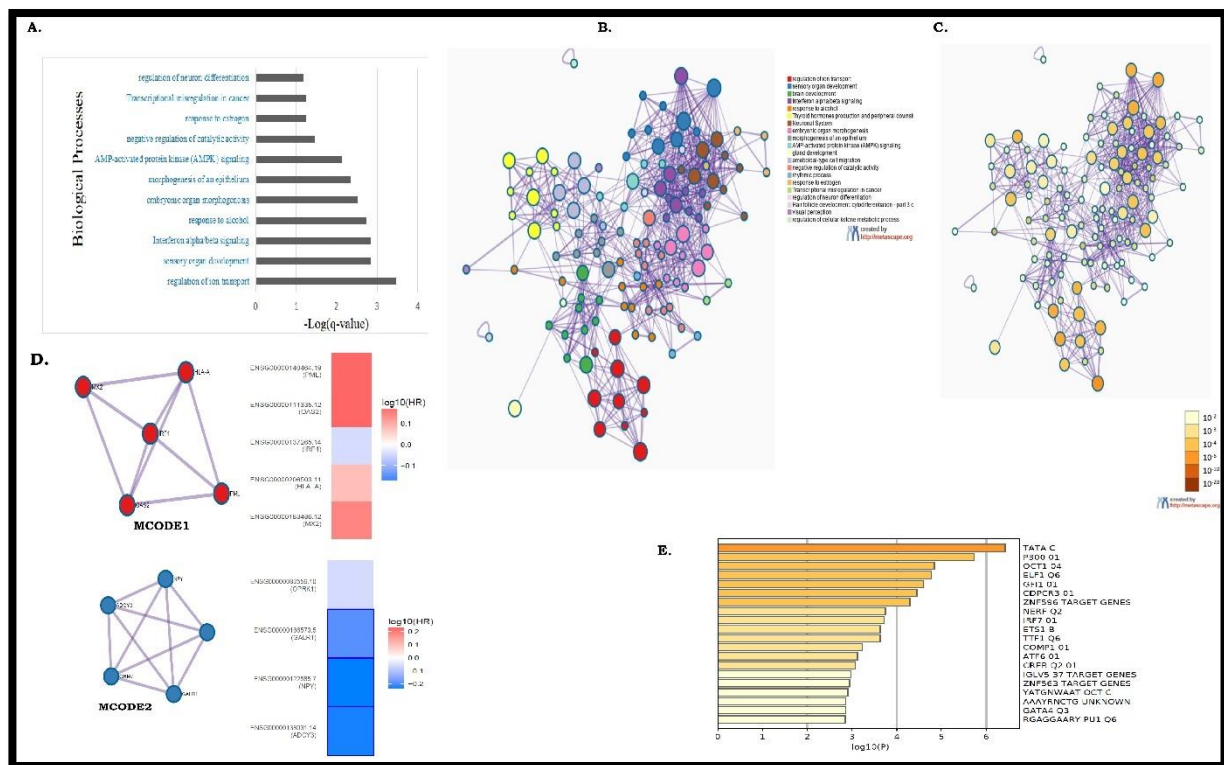


Figure 27: Functional annotations of gDMs and identified enriched pathways using Metascape.

(A) Enrichment of biological processes among the 91 gDMs. (B) and (C) Network analysis with Metascape, where each enriched biological process was coloured distinctly. Each term in the network is represented by a circle node, where its size is proportional to the number of input genes fall into that term, and its colour

represent its cluster identity (i.e., nodes of the same color belong to the same cluster). Terms with a similarity score > 0.3 are linked by an edge (the thickness of the edge represents the similarity score). (D) MCODE networks showing interconnection between proteins. (E) Enrichment of transcription factors, known to regulate subsets of gDMs

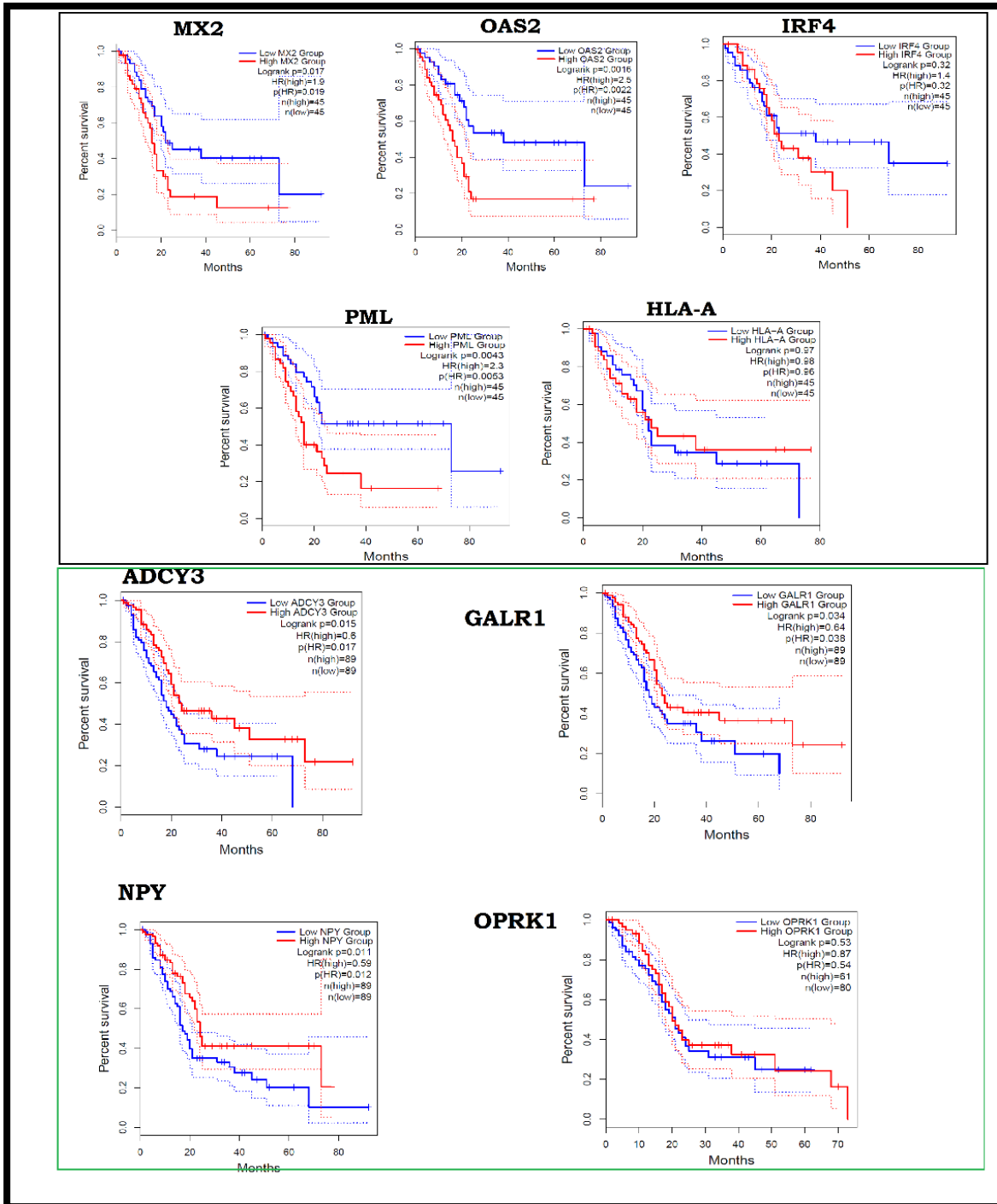


Figure 28. Differential survival of genes annotated in Metascape.

MCODE1(MX2, OAS2, IRF4, PML, HLA-A) and MCODE2 (ADCY3, GALR1, NPY, OPRK1), in the TCGA pancreatic cancer cohort. Patients with lower expressions of genes annotated to MCODE1 and higher expressions of genes annotated to MCODE2 showed better disease prognosis.

Altered DNA methylation can affect the binding of the transcription factors (TFs), thus altering gene expressions. We searched for common TFs that could regulate the 91 gDMs. Enrichment analysis showed enrichment of several TFs including *IRF7*, *OCT1*, *CREBP*, *ETS1*, *ELF1* and *NERF* ($-\log P < -3$) (Figure 27).

5.1.6: Assessment of Methylation Along with Gene and Protein Expression of Npy and Faim2 Genes in PDAC

In regards to our all-previous observations starting from cohort 1 till cohort 3 we speculate and might interpreted that the selection of two novel and significantly hypermethylated genes in our Indian patient population for downstream further validation and reclaimed our prior observation using cohort 4, our Indian patient survival dataset and other global data bases-based approaches. In cohort 1 that underwent 450 K methylation 6 genes were chosen out of which some genes underwent methylation and gene expression validation based on their survival information. These genes are not reported previously in TCGA which mostly contains Western world data and are reportedly novel also in terms of Asian (preferably Indian patient population) scenarios. Amongst these genes Neuropeptide Y (*NPY*) and Fas Apoptotic Inhibitory Molecule 2 (*FAIM2*) showed the highest alteration in differential gene expression. In terms of their methylation status by using MSP the gel images showed significant observation in case of both *NPY* and *FAIM2*. In addition to that, the changes in methylation were notably higher for these two genes in normal adjacent tissues with respect to their tumor counterparts. Furthermore, we validated our observations for both *NPY* and *FAIM2* genes by using global databases based *in silico* approaches and its outcome was torchbearer of our assessment post which we further evaluated our observations in cohort 4.

The *NPY* is involved in cell motion, invasion and cell proliferation and associated with multiple carcinomas. hypermethylation of *NPY* is also observed in certain carcinomas. Methylated *NPY* in circulating tumor DNA is currently a major focus for cancer biomarker detection in colorectal cancers. *FAIM2* also plays a significant role in apoptosis inhibition by inhibiting Fas/CD95-mediated apoptosis.

We observed significant hypermethylation of *NPY* and *FAIM2* both in the TCGA cohort (*NPY*: $\Delta\beta/p$: 0.28/1.54E⁻⁰⁵, *FAIM2*: $\Delta\beta/p$: 0.33/8.23E⁻⁰⁷) and validation cohorts (*FAIM2*: p :- 2.46E⁻⁰⁶/ $\Delta\beta$ - 0.209; *NPY*: p : -4.07E⁻⁰⁷/ $\Delta\beta$ - 0.205) (Tables 13, 19), and data previously shown). We also observed that in validation cohort 3, down-regulated expressions of *NPY* gene in 86.9% patients whereas, *FAIM2* gene in 82.6% of PDAC patient. Further

assessment were also re-investigated, patients showing lower expression had poor disease prognosis (log rank $p=0.01$) than patients with higher expressions of *NPY* and *FAIM2* in their tumors, which could be indirectly controlled by differential methylation (Figure 29A). Negative correlation between the methylation and gene expressions of *NPY* ($p: 0.00093$, correlation coefficient $= -0.24$) but not for *FAIM2* ($p: 0.82$, $R: -0.017$), in TCGA cohort was observed (Figure 29B). The normal tissue samples in GTEx cohort had higher average expressions of both *NPY* and *FAIM2* than in comparison with cancer samples in the TCGA cohort (Figure 29C). Significant correlation was observed between *NPY* and *FAIM2* expressions ($p<0.01$, $R=0.14$) (Figure 29D). The expression of the two above genes also showed association with progressive stages of PDAC (Figure 29E).

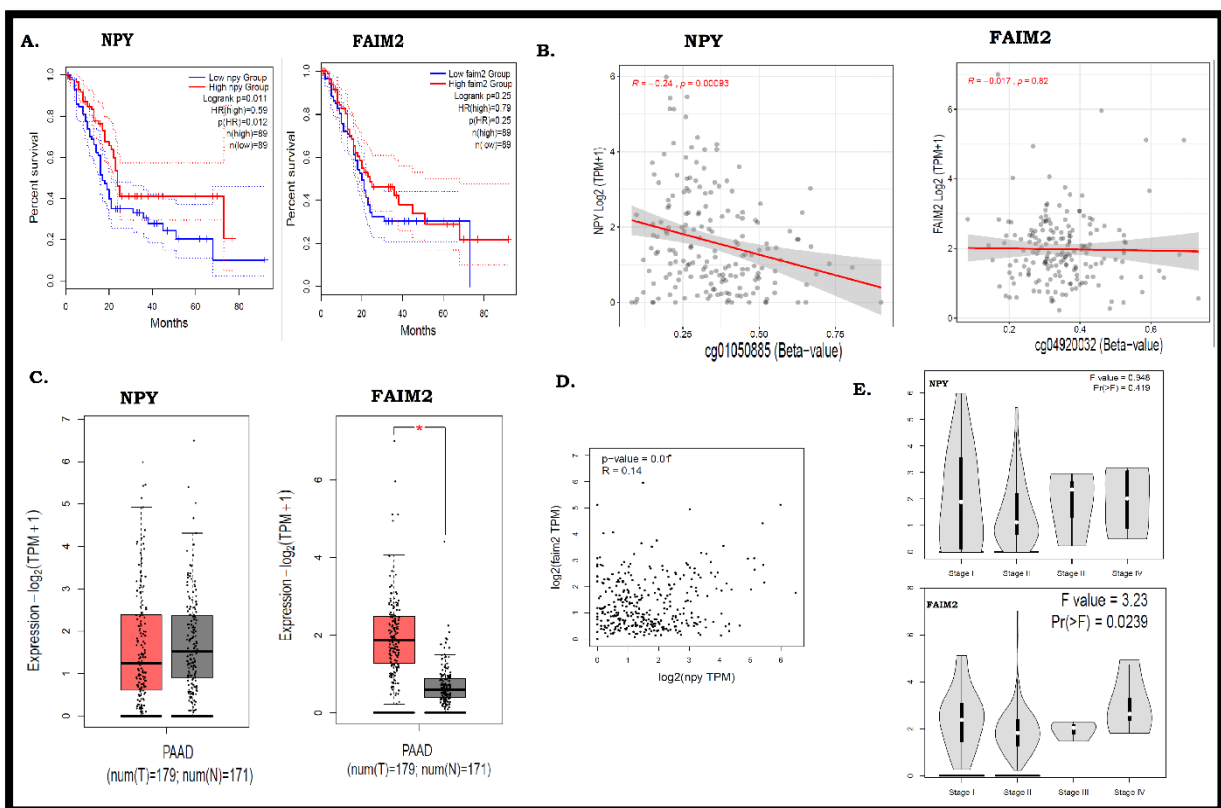


Figure 29: Differential methylation and expressions of *NPY* and *FAIM2* in the TCGA cohort.

(A) Differential survival of patients in the TCGA cohort with “high” and “low” expressions of *NPY* and *FAIM2*. (B) Correlations between beta value and gene expression of *NPY* and *FAIM2*, among the cancer samples of the TCGA cohort. (C) Differences in expressions of *NPY* and *FAIM2* genes between cancer and normal samples of the TCGA cohort. (D) Correlation between the expressions of *FAIM2* and *NPY* genes in the TCGA cohort cancer samples. (E) Expression levels of *FAIM2* and *NPY* genes across stages of cancer.

Finally, we have further data mined for methylation in *NPY* gene in the CCLE database, which included methylation data on 35 PanCa cell lines. The results showed significant

hypermethylation in *NPY* gene in 35 PanCa cell lines (Figure 30A). Pan-cancer analysis also showed down-regulation of the two genes in cancer. The expression level was also estimated by using the TCGA RNA seq PAAD dataset by using UALCAN (Figure 30B).

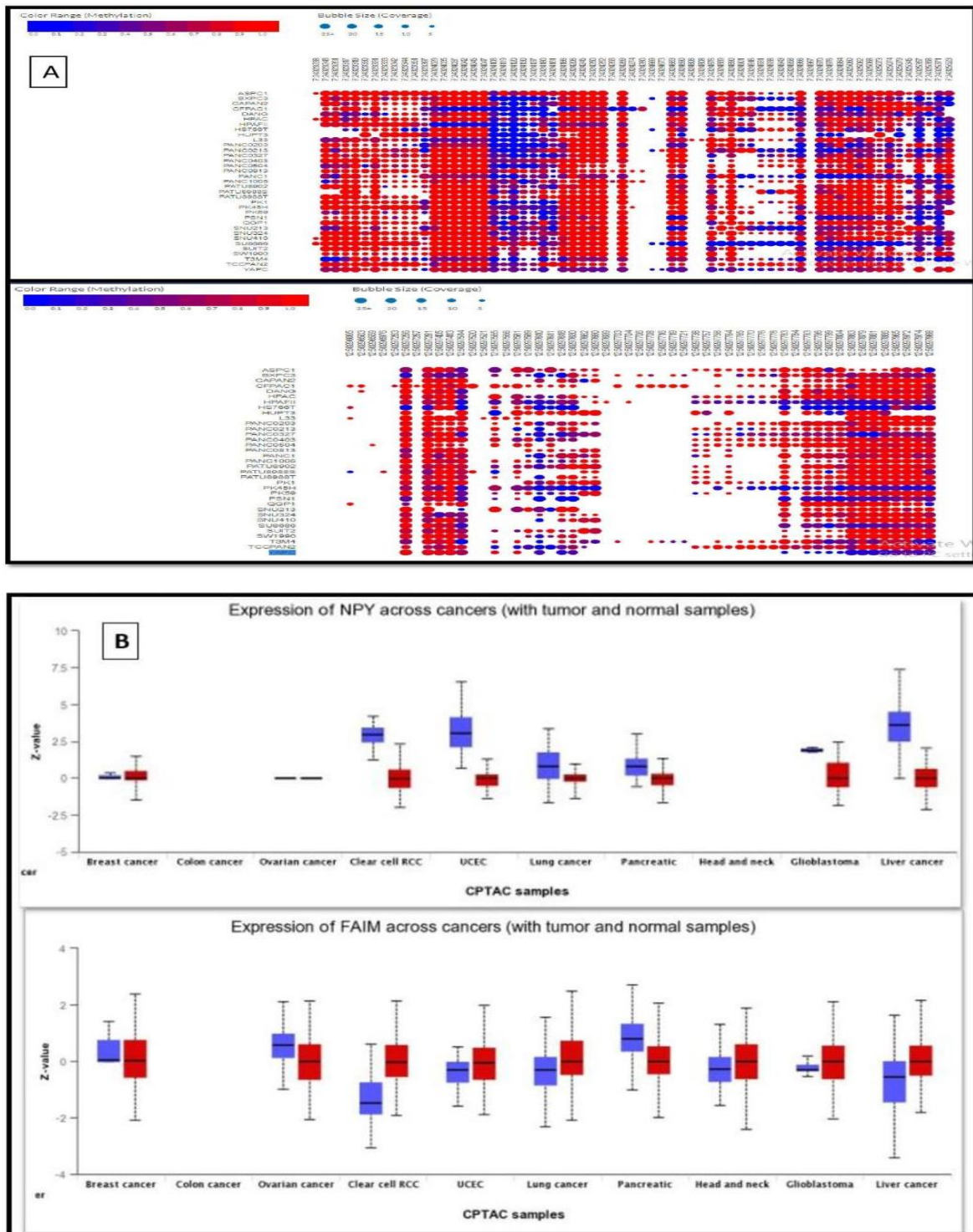


Figure 30: Differential methylation of *NPY* and *FAIM2* in CCLE cell lines along with Pan cancer expression.

(A) Methylation patterns at the respective *NPY* and *FAIM2* genes, spanning various CpG sites across 35 different pancreatic cancer cell lines. This bubble plot has been generated using CCLE database and its

associated DeMap Portal. (B) Pan-cancer analysis of NPY and FAIM2 expression level was also done using TCGA RNA seq PAAD data by using UALCAN

5.1.7: Cross-validation of Both *FAIM2* and *NPY* Hypermethylation by IHC Experiment in Another Validation Set

Using IHC method, in separate validation cohort 4, we further showed that the protein expressions of both Faim2 and Npy were lower in tumor tissues compared to adjacent normal counterpart sections. After assessments of all sections and stained with anti-Npy antibody it has been shown that 84% down-regulation in tumor tissues and 86% up-regulation in normal (Figure 31A, B). Furthermore, we have also observed that Faim2-expression was down-regulated in tumor tissues by 87% compared to 94% upregulation in adjacent normal part of normal tissues (Figure 31C, D). The results were closely monitored in both 20× and 40× magnifications. Cell morphology was correlated with pathology slides of clinical samples for both tumor and normal sections (Figure 31E, F). These results were in concordance with the data obtained from MSP and gene expression experiments previously. Identifying more specific PDAC markers that outperform the currently available markers would be very useful in establishing a diagnosis in patients with challenging histopathology. Both Faim2 and Npy protein expressions across all samples supported this motivation.

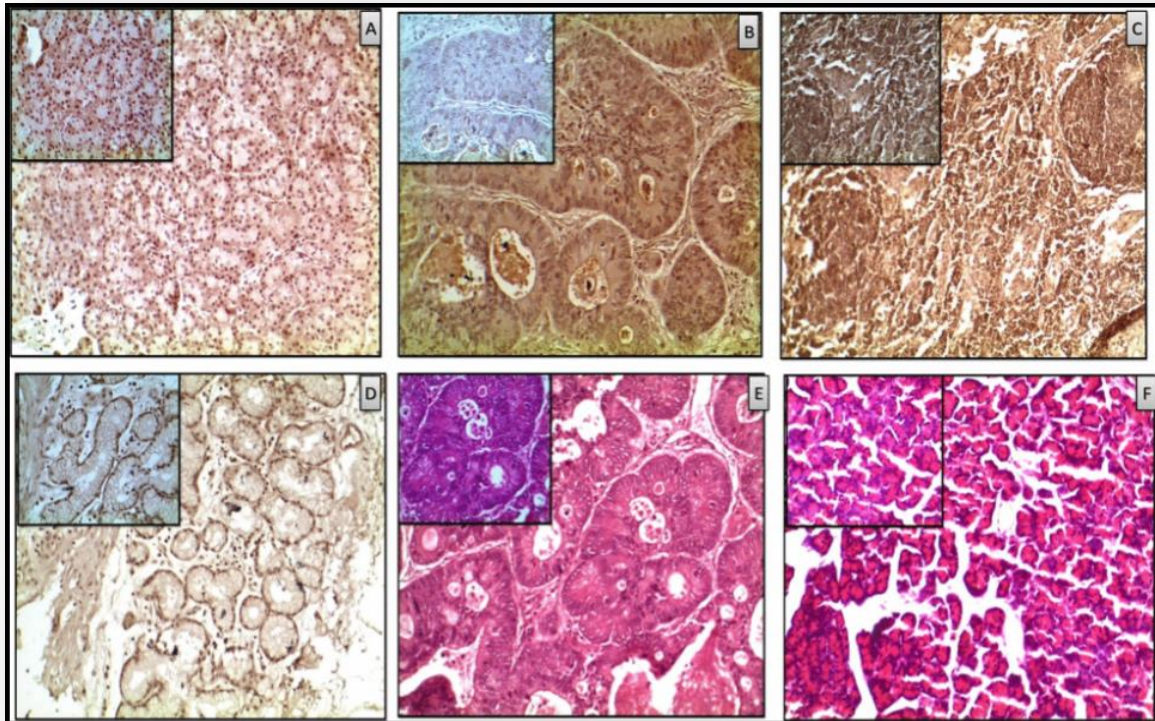


Figure 31: Representative images of immunohistochemistry (IHC) showing expressions of Faim2 and Npy proteins in PDAC and paired control tissues (n = 20):

Antibody specific to *Npy* and *Faim2*, were used for comparison and images were clicked at both 20× and 40×. Every Slide consisted of 3 independent sections for IHC. The representation is expression of *Npy* in normal tissues (A) and in cancer (B). Expression of *Faim2* in normal (C) and in cancer (D). A and C took more browner spots in normal cells than (B) and (D) of the proteins in compared to cancer tissues. (E) and (F) H&E section of both tumor and normal were taken at similar resolutions. The bigger Figure were taken in 20× and the sub-Figure in smaller boxes were taken in 40×. Every sample had 3 sections on a slide with similar results to avoid technical biasness

5.1.8: Association of *NPY* and *FAIM2* Expression with Clinicopathological Risk Factors Among PDAC Patients

Cancer progression and its associated diagnosis and treatment are based on a plethora of parameters taking into consideration clinical, subclinical, and genomic parameters. In addition to that, in correlation to mRNA expression levels, protein levels of both *Npy* and *Faim2* in normal pancreatic tissues were significantly higher than in PDAC tumor tissues, in the CPTAC Confirmatory/Discovery dataset (Figure 32A), further validating down-regulation in PDAC (Figure 30B). Overall survival analysis of validation cohort 3 showed that joint down-regulation of *NPY* and *FAIM2* is associated with poor prognosis ($p: 0.065$) (Figure 32B). Additionally, patients with *NPY* down-regulation ($p: 0.041$) and high alcohol consumption had poor OS ($p: 0.039$) (Figure 32C, D). Down-regulations of both *FAIM2* and *NPY* were been strong negatively correlated with diabetes ($R/p: -0.484/0.026$). Dual down-regulation of *FAIM2* and *NPY* has been observed to be correlated with *KCNA6* down-regulation ($R/p: 0.592/0.016$) and *MX2* up-regulation ($R/p: 0.476/0.045$). *FAIM2* regulation also correlates with *SLITRK3* ($R/p: 0.545/0.019$). *RASSF1A* has a very strong correlation with the pancreatic lesion ($R/p: 1.00/0.00$). Lastly, *GALR1* has also been documented to multiple correlations with tumor stages, lymph node status and the expression levels of *SLITRK3* and *SIGIRR* genes (Table 19, 20).

Classical Pearson's correlation has been done along with "Kendall tau" correlation after considering the non-linearity of the subclinical and clinical data of patients. Starting from an ethnic point of view sex has a significant negative correlation with smoking as the abundance of smokers is higher in the Indian males than females ($R/p: -0.419/0.037$). Negative data correlation between sex and smoking has also been reported after doing "Kendall tau" correlation ($p=0.40$). The same has been observed between alcohol and smoking ($R/p: 0.577/0.003$). Considering the clinico-pathological parameters significant correlation has been documented between tumor stages with jaundice and lymph node ($R/p: 0.559/0.038$). In

our sample pool, a significant correlation has been observed between tumor stages (TNM grading) with lymph node metastasis ($R/p : 1.0/0.00$).

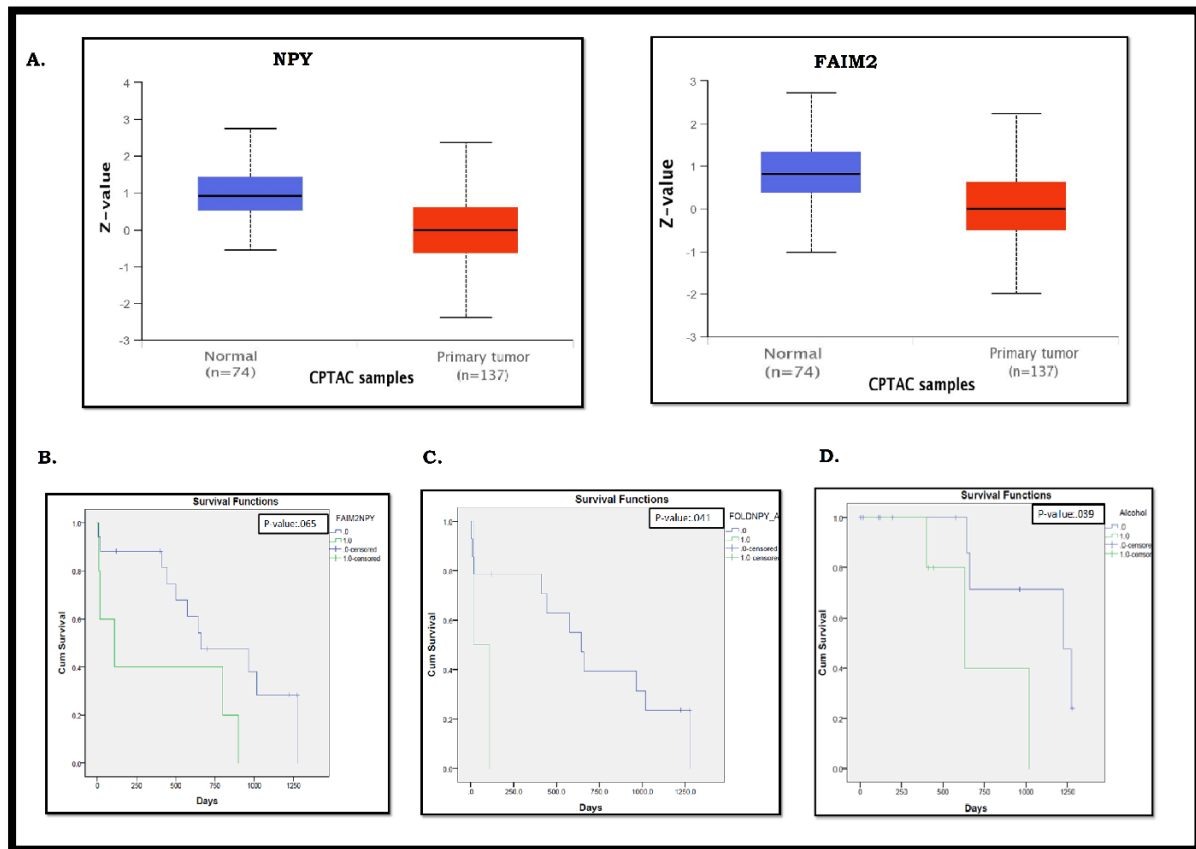


Figure 32: Protein expression of *NPY* and *FAIM2* along with joint survival patterns.

(A) Differential protein expression of *NPY* and *FAIM2* in the cancer samples compared to control samples in the TCGA cohort. (B–D) Overall survival analysis showing the difference in disease prognosis among patients in the validation cohort 3, with different levels of expressions of *NPY* and *FAIM2*

5.2 Identification of Ion Channel Transporters Using Differentially Expressed Genes (DEGs) in PDAC

Ninety-one differentially methylated genes (gDMs) of interest, identified in our previous study, were subjected to enrichment of GO terms on biological processes using Metascape. Enrichment pathways in KEGG database, showed enrichment of voltage gated ion-channel mediated pathways, particularly, hypermethylation of potassium (*KCNA3*, and *KCNA6*) and calcium (*CACNB2*) ion channels in PDAC. These genes were further assessed through (i) epigenomics roadmap project, showing most significant correlation with DNA packaging protein, histone 3 (H3K27meH9) (Figure 33A), (ii) KEGG 2021 Human showing most significant correlation with progesterone mediated oocyte maturation (Figure 33B), and (iii) Reactome 2022 showing most significant correlation with g2 phase r-hsa-68911 (Figure 33C)

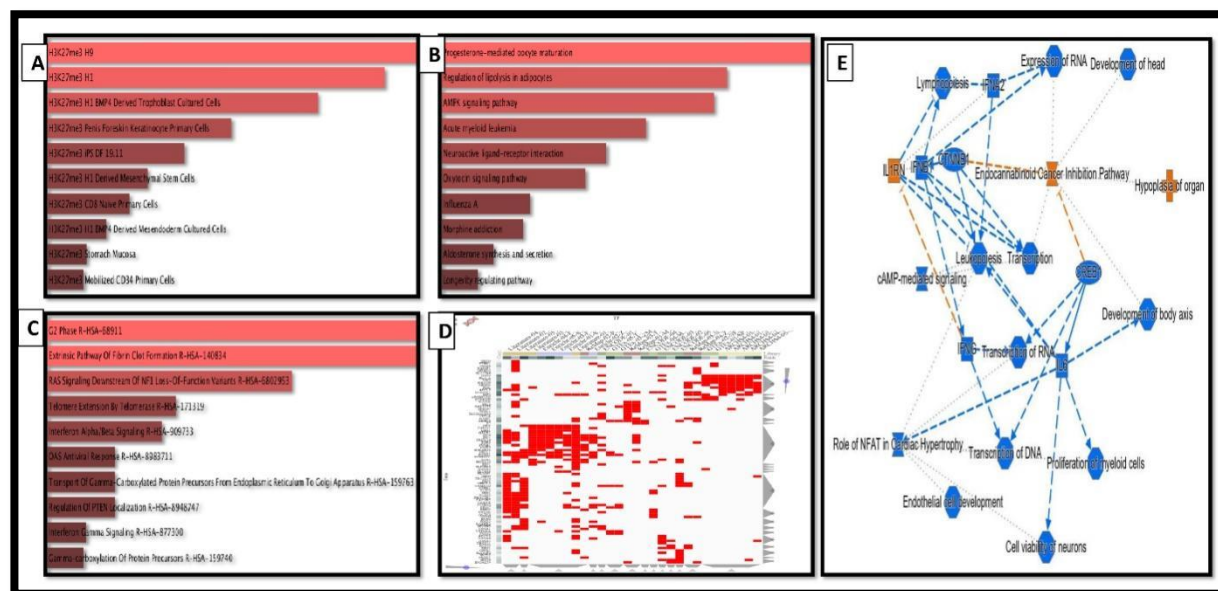


Figure 33: Pathway associated landscape evaluation: Differentially methylated 91 gene signature status evaluation through

(A) epigenomics roadmap project, (B) KEGG 2021 Human, (C) proteomic signature through Reactome 2022, (D) CHEA3 gene signature density-based heat map density mapping transcription factor, (E) IPA pathway analysis showing respective interactions. Orange annotation: significant, Blue annotation: non-significant.

Metascape data from our previous study and IPA analysis resulted in identification of 7 ‘ion transport regulation’ associated gDMs: *RASA3*, *CLIC5*, *KCNA3*, *KCNA6*, *KCNJ5*, *CACNB2* and *GABRR2*. Pan cancer profiling of these gDMs were done targeting all cancer subtypes (tumor vs normal) (Figure 34). Expression profile of *KCNJ5*, *KCNA3*, *CLIC5*, *CACNB2*, and *GABRR2* genes, assessed through GTEx (normal database) targeting all 32 tissue types, demonstrated low expression of *KCNJ5*, *CACNB2* and *RASA3*, while the rest, *CLIC5*, *KCNA3*, *KCNA6* and *GABRR2* showed negligible expression in normal pancreas, (Figure 35).

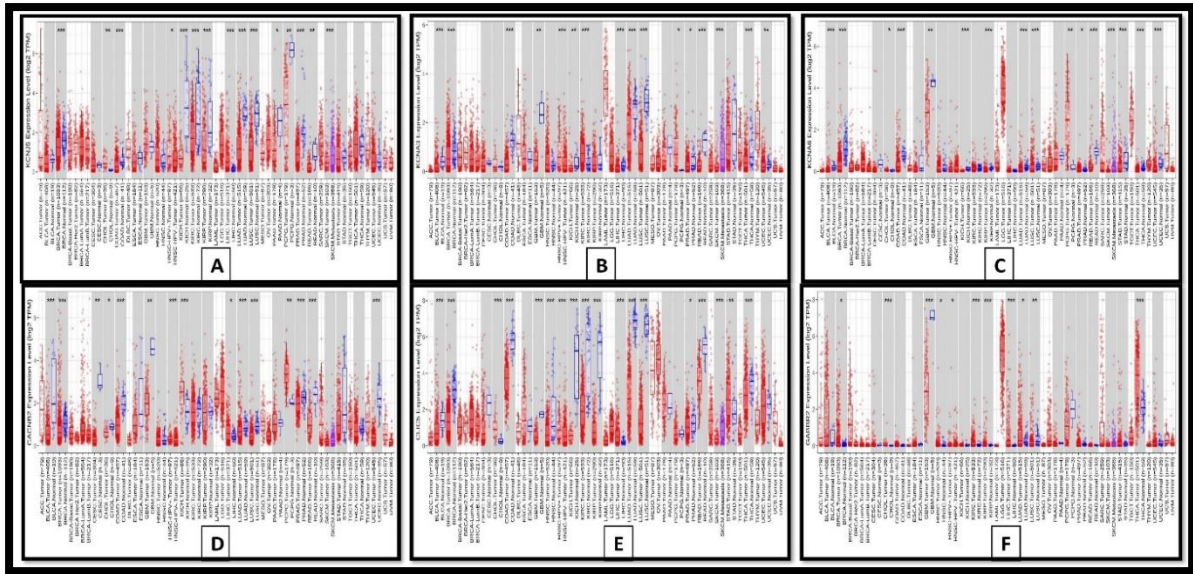


Figure 34. Pan cancer profiling of 6 different methylated genes targeting all cancer subtypes (tumor vs normal) and respective significance.

(A) *KCNJ5*, (B) *KCNA3*, (C) *KCNA6*, (D) *CACNB2*, (E) *CLIC5*, and (F) *GABBR2*.

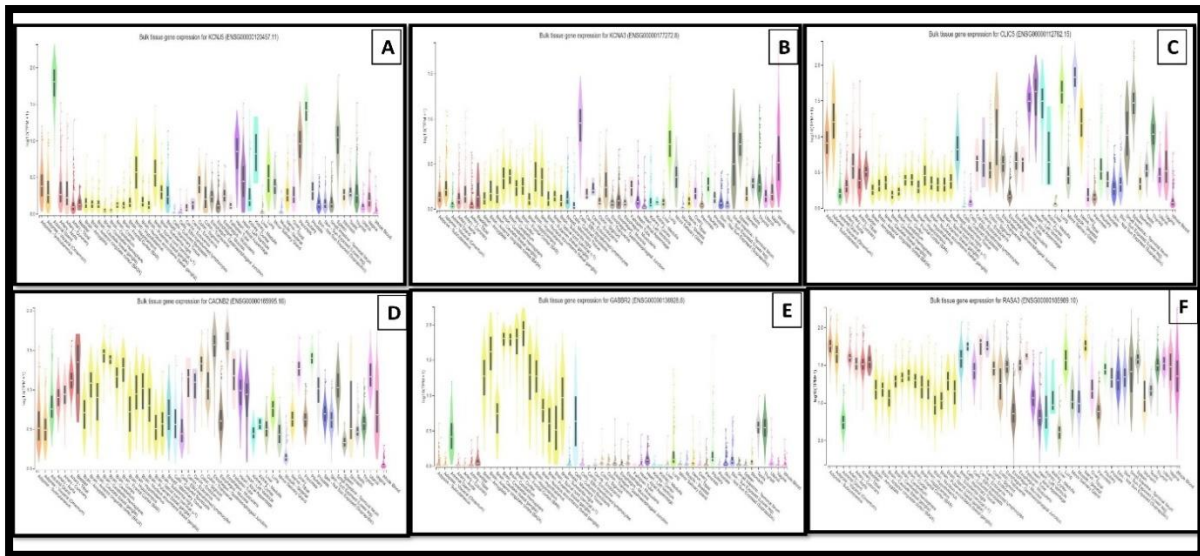


Figure 35: Normal tissue expression profile using GTEx database.

(A) *KCNJ5*, (B) *KCNA3*, (C) *CLIC5*, (D) *CACNB2*, (E) *GABBR2* and (F) *RASA3* using GTEx (normal database) targeting all 32 tissue types.

Differential mRNA expression in pancreatic tumor (TCGA dataset) and normal pancreas (GTEx dataset) using Wilcoxon signed rank test showed significant elevated expressions of *KCNJ5* (Figure 36A) and non-significant high expression of *CACNB2* (Figure 37A) in normal pancreas with respect to PanCa, whereas, *RASA3* (Figure 36B) and *CLIC5* (Figure 36C) showed significant over-expression in PanCa in comparison to that of normal pancreas. *KCNA3* was also over-expressed (not significant), in PanCa (Figure 37B). Extent of methylation of both *CACNB2* and *KCNJ5* was significantly higher in PanCa tissue as

compared to that of normal pancreas (Figure 36D, E) in TCGA cohort. Gene expression of *KCNJ5*, *KANA6* and *GABBR2* showed no significant correlation with methylation in TCGA cohort. *CACNB2*, *CLICK5* and *KCNA3* showed negative correlation between methylation and gene expressions with *R-values* of -0.43 , -0.16 , and -0.52 respectively. Only *RASA3* expression demonstrated a non-significant positive correlation with methylation (*R-value*= 0.12) (Figure 38). Stage-wise expression of *CACNB2* and *GABBR2* genes have shown significant over-expression with advancement of PanCa stages, I-IV, while, *KCNJ5* illustrated only marginal upregulation (Figure 39). Assessing *Kcnj5* protein (Kir 3.4) expression evaluated through CPTAC in patient samples (tumor vs normal) demonstrated its down-regulation in PDAC in comparison to normal pancreas (Figure 36F). Mutational status of these 7 genes, assessed through cBioPortal (TCGA, Firehose Legacy) have identified missense mutation of all the 7 genes in PDAC. *CACNB2* and *GABBR2* showed deep deletion, while, *KCNA6* demonstrated amplification in PDAC along with mis-sense mutation (Figure 40, 41). Kaplan–Meier survival curves, demonstrated that OS of low *CACNB2* group conferred poor prognosis as compared to high *CACNB2* group (Figure 36G), while OS of both high and low *KCNJ5* groups were comparable (Figure 36H).

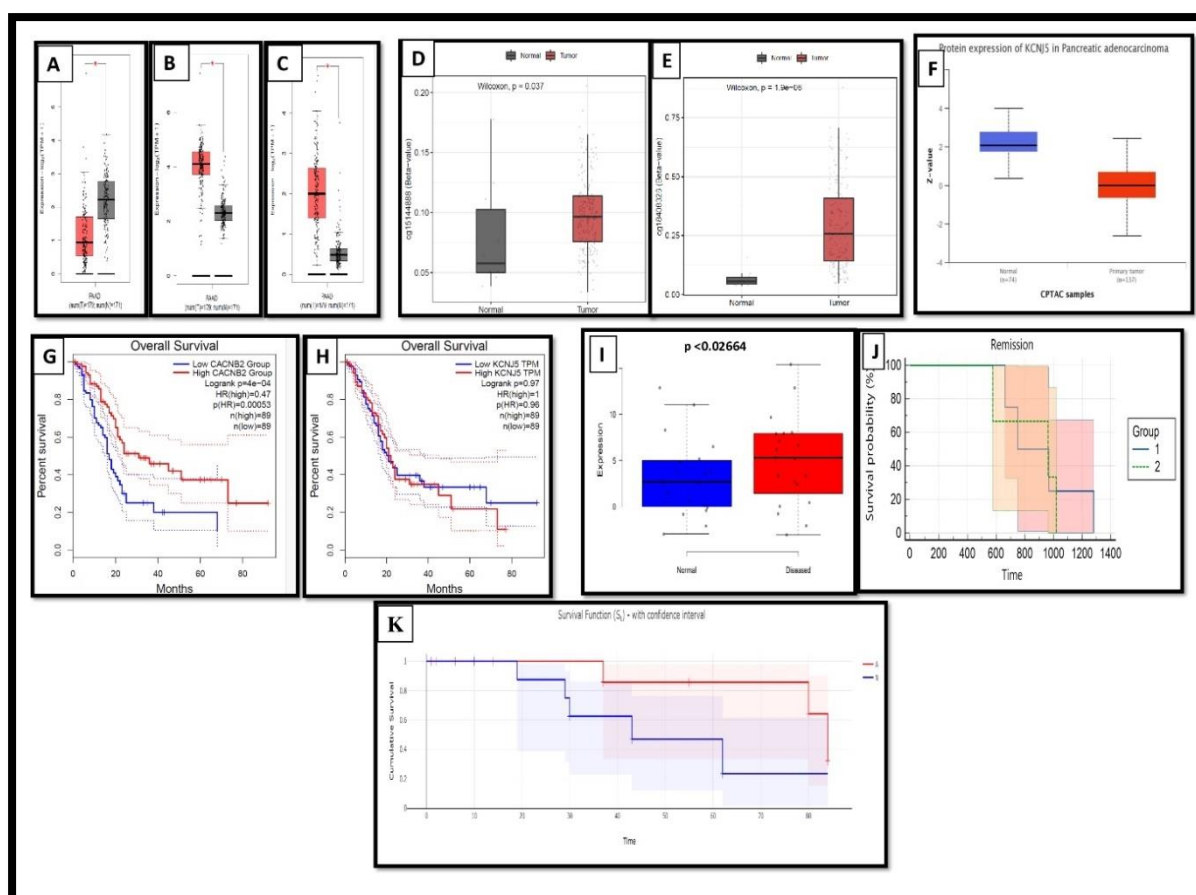


Figure 36: Clinical relevance of significant differentially methylated genes.

Differential expression (TCGA, tumor Vs GTEx, normal) of (A) *RASA3* (B) *KCNJ5* (C) *CLIC5*. Differential methylation pattern of *KCNJ5* (D), *CACNB2* (E) (TCGA, tumor Vs TCGA, normal) using Wilcoxon Rank sum test; $p=0.037$. (F) *Kcnj5* protein expression was evaluated in CPTAC patient samples (tumor vs normal). Overall survival shown through Kaplan-Meier plot across TCGA patient samples post splitting the samples in low and high-risk groups for (G) *CACNB2*, $p=4e-04$, (H) *KCNJ5*, $p=0.97$. (I) Differential expression of *KCNJ5* on PDAC paired tumor samples ($n=20$, $p < 0.026$). (J) Overall survival using Kaplan Meier survival analysis of *KCNJ5* using gene expression data done on 20 patient samples in Indian cohort (1= high *KCNJ5* group, 2 = low *KCNJ5* group). Time is measured in days.

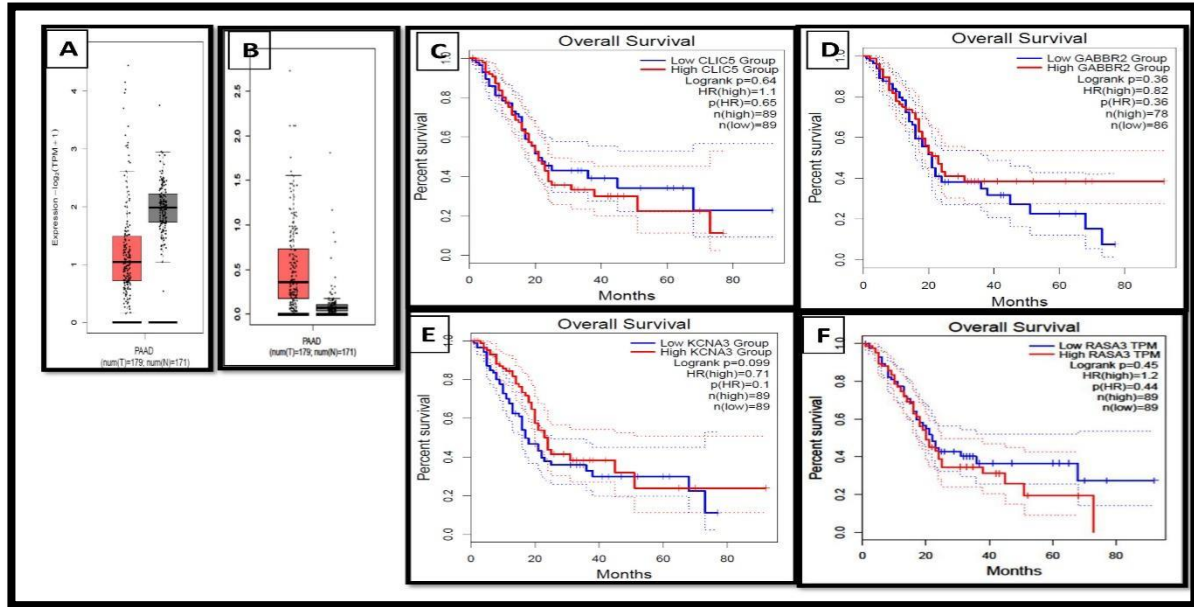


Figure 37: Clinical relevance of differentially methylated genes
Differential methylation of (A) *CACNB2* (B) *KCNJ5* in normal and cancerous pancreas. Overall survival of (C) *CLIC5* (D) *GABBR2* (E) *KCNJ5* and (F) *RASA3*.

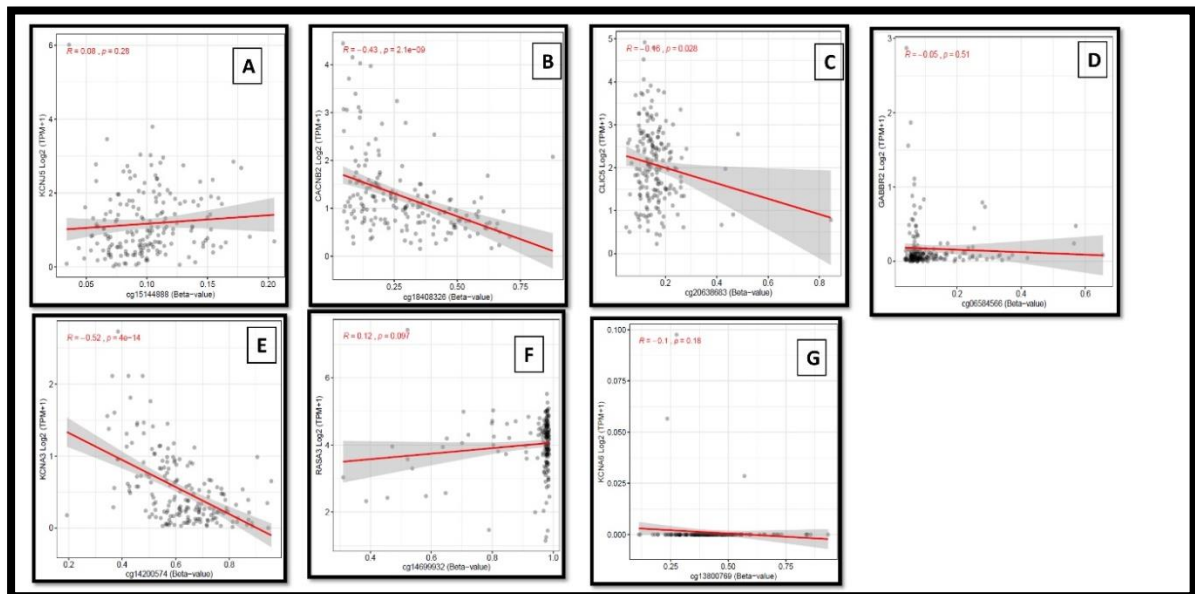


Figure 38: Correlation between differential methylation and expressions of ion transporter genes

(A) *KCNJ5*, (B) *CACNB2*, (C) *CLIC5*, (D) *GABBR2*, (E) *KCNA3*, (F) *RASA3* and (G) *KCNA6* in the TCGA cohort.

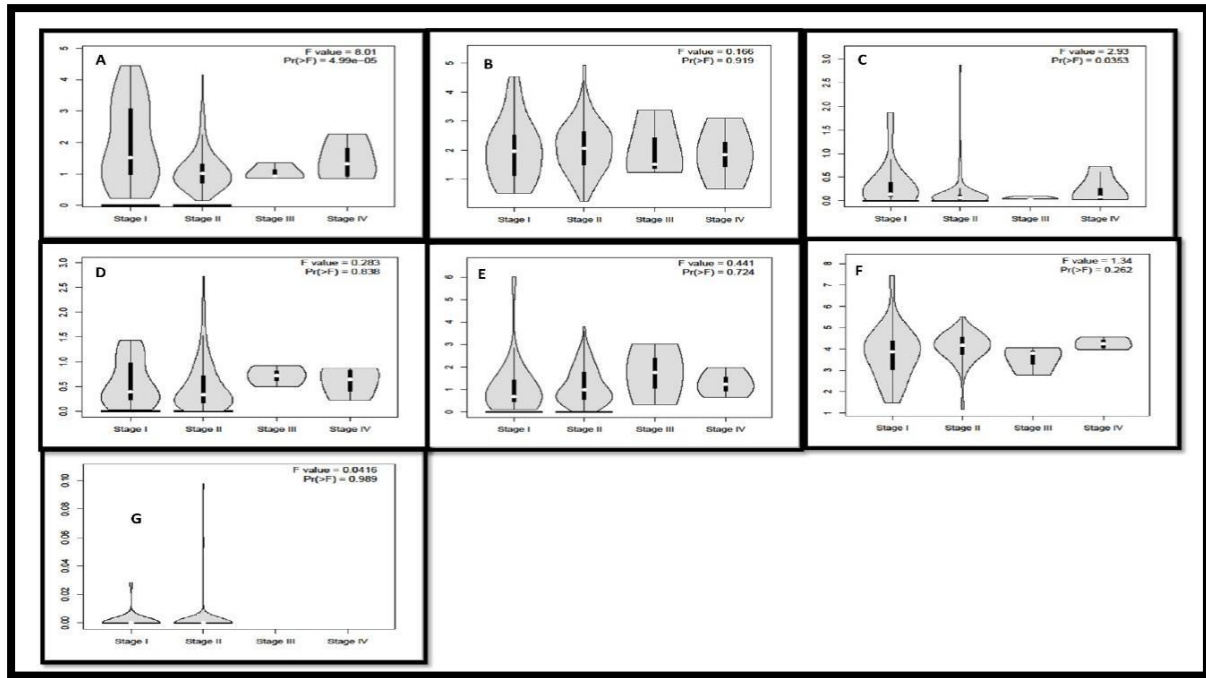


Figure 39: Stage wise expression of ion transporter genes

(A) *CACNB2*, (B) *CLIC5*, (C) *GABBR2*, (D) *KCNA3*, (E) *KCNJ5*, (F) *RASA3*, (G) *KCNA6* in Pancreatic Cancer using GEPIA in PAAD dataset.

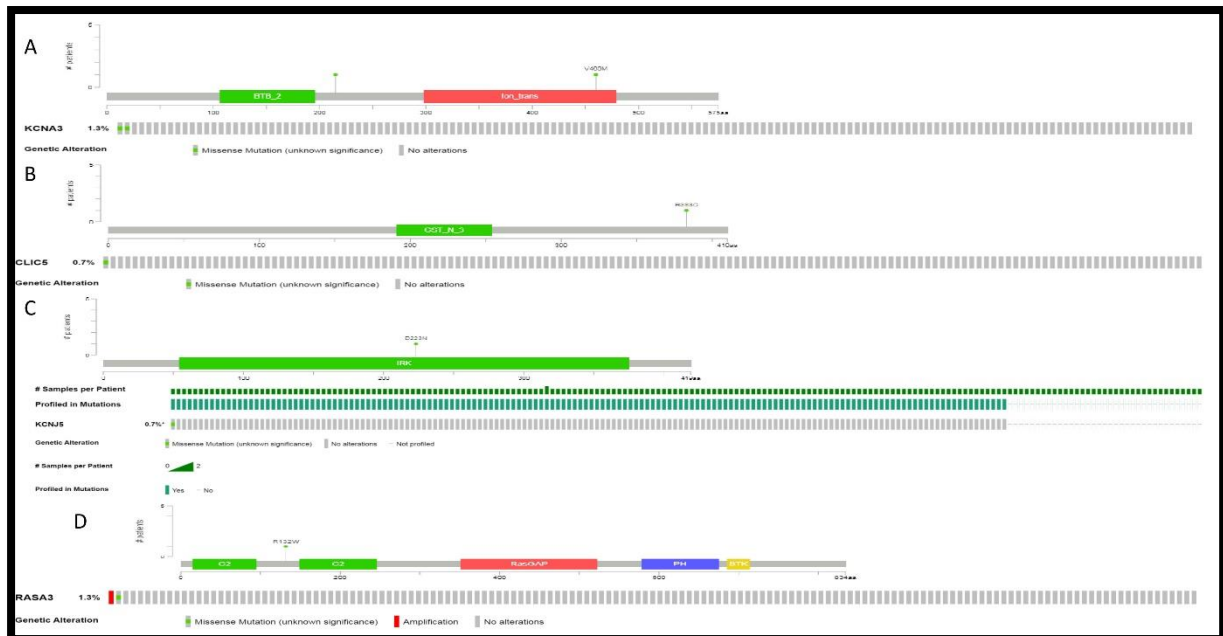


Figure 40: Mutational status of 4 ion transporter genes

(A) *KCNA6*, (B) *CLIC5*, (C) *KCNJ5* and (D) *RASA3*, in PDAC (TCGA, Firehose Legacy) using cBioPortal.

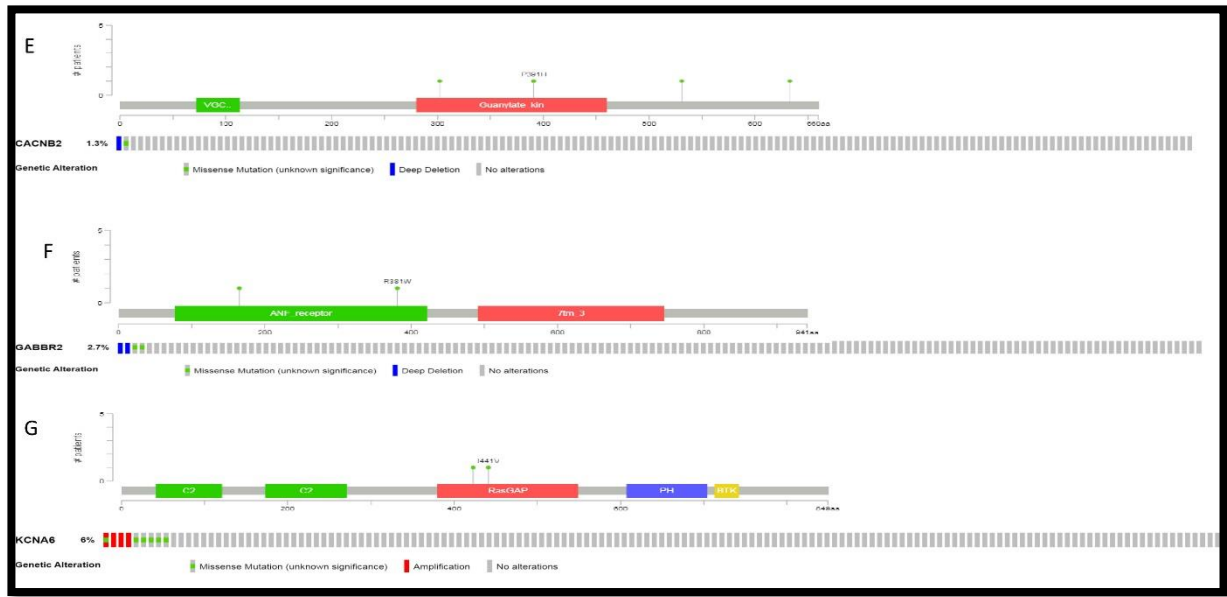


Figure 41: Mutational status of 3 ion transporter genes

(E) *CACNB2*, (F) *GABBR2* and (G) *KCNA6* in PDAC (TCGA, Firehose Legacy) using cBioPortal.

5.2.2: *KCNJ5* Expression Validation and Kaplan-Meier Overall Survival Analysis.

Post validation using gene expression levels along with methylation status check followed by their protein validation showcased *KCNJ5* as the most significant gene across all omics. OS of both high and low *CLIC5*, *RASA3* and *KCNA3* group were comparable, while low *GABBR2* group have shown adverse prognosis as compared to high *GABBR2* group (Figure 37C-F). Literature based primers of *KCNJ5* were used to screen 20 PDAC samples. Post normalization, *GAPDH* was used as a housekeeping gene to compare tumor and normal pancreas data. Overexpression of *KCNJ5* was shown in tumor as compared to normal pancreatic tissue in our Indian PanCa cohort (Figure 36I). OS within high and low *KCNJ5* group, in Indian PanCa cohort showed significant difference ($p=0.0019$), with high *KCNJ5* group showing poorer prognosis as compared to low *KCNJ5* group (Figure 36J). Kaplan-Meier was done also post comparison between the alcohol and non-alcoholic groups and was found to be border line significant ($p<0.07$) (Figure 36K). Further validation of *KCNJ5* expression using 4 publicly available GEO datasets, observations demonstrated association of hypermethylation with down-regulation (fold change -1.9) in concordance with findings of TCGA dataset and we found it to be significantly downregulated which is in concordance with our TCGA and other GEO datasets observations (Table 21).

5.2.3: Correlation of Immune Cell With Differentially Methylated Genes

Immune profiling using LM22 cell types of *KCNJ5*, *CACNB2* revealed a positive correlation between both *KCNJ5* and *CACNB2* with resting mast cells, while showing negative

correlation with undifferentiated macrophage (M0) and memory B cells. *KCNJ5* demonstrated a positive correlation with CD8+ T cells, CD4+ memory T cells, pro-inflammatory macrophages (M1), while *CACNB2* was positively correlated with activated NK cells, anti-inflammatory macrophages (M2) and activated mast cells. A negative correlation was observed in case of *KCNJ5* and dendritic and activated mast cells. Also, *CACNB2* showed negative correlation with regulatory T cells (Figure 42A and B).

5.2.4: Cross -Validation of *KCNJ5* Expression in HPA Database

On comparing the gene expression of *KCNJ5* in normal and PanCa cell lines, it was found that the mean nTPM value is 14.5 and 0.6 for normal pancreatic tissue and PanCa respectively (Figure 42C). Immunohistochemistry results reported that *KCNJ5* is highly and moderately expressed at protein level in normal exocrine and endocrine cells of pancreas respectively (Figure 42D). Hence, *KCNJ5* showed low expression in PanCa compared to normal pancreatic tissue.

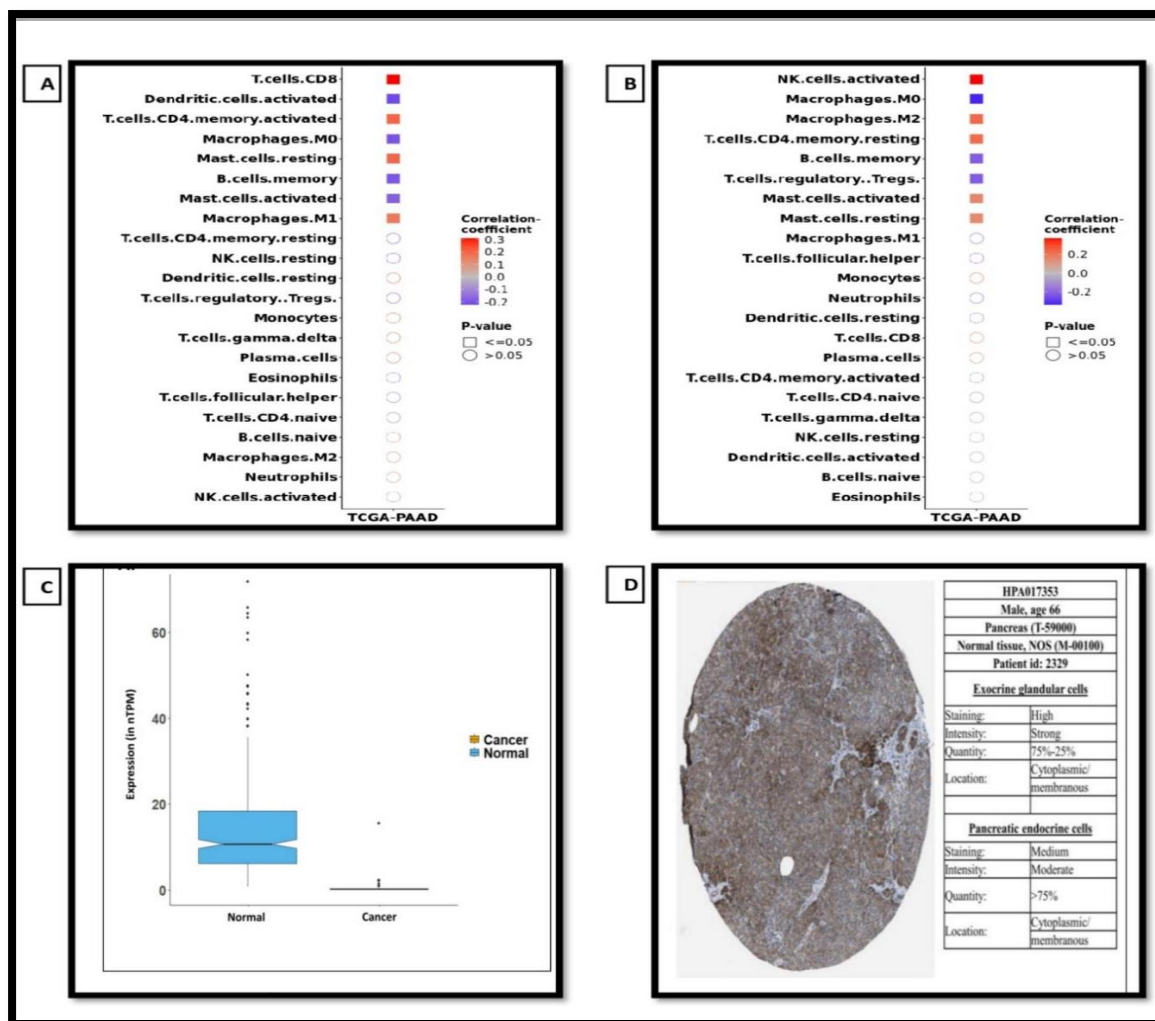


Figure 42: Deconvolution immune profiling using LM22 cell types along with expression status profile of cell lines present in HPA

(A) *KCNJ5*, (B) *CACNB2* expression correlation in different immune cell types (C) HPA obtained cell line expression differences (D) HPA obtained *KCNJ5* associated normal pancreas histopathology status. Significant to insignificant correlation coefficient shown by blue to red. Expression of *KCNJ5* in Human Protein Atlas A. Whisker plot comparing the RNA expression (in nTPM) of *KCNJ5* in 328 GTEx normal pancreatic tissue samples and 46 PanCa cell lines B. Immunohistochemistry of normal pancreatic tissue depicting high expression of *KCNJ5*.

5.2.5: Functional Network Analysis of *KCNJ5*

Using an interaction score threshold of 0.7 (high confidence), the STRING PPI analysis (Figure 43) resulted a highly clustered network, indicating significantly more interaction than expected for a random set of similar size drawn from the genome (enrichment $p < 0.001$). Enrichment analysis of *KCNJ5* based on PPI using GO, represented significant involvement of inwardly rectifying potassium channel subfamily J: *KCNJ3*, *KCNJ5*, *KCNJ6*, *KCNJ9* and *KCNJ15*, G protein subunit γ and γ aminobutyric acid type B receptor subunit 2 components in GO biological pathways (Figure 43A), molecular functions (Figure 43B) and cellular components (Figure 43C) terms in the network. Similar enrichment analysis of *KCNJ5* in our Indian cohort using GO, depicted concordant observations in GO biological pathways (Figure 43D).

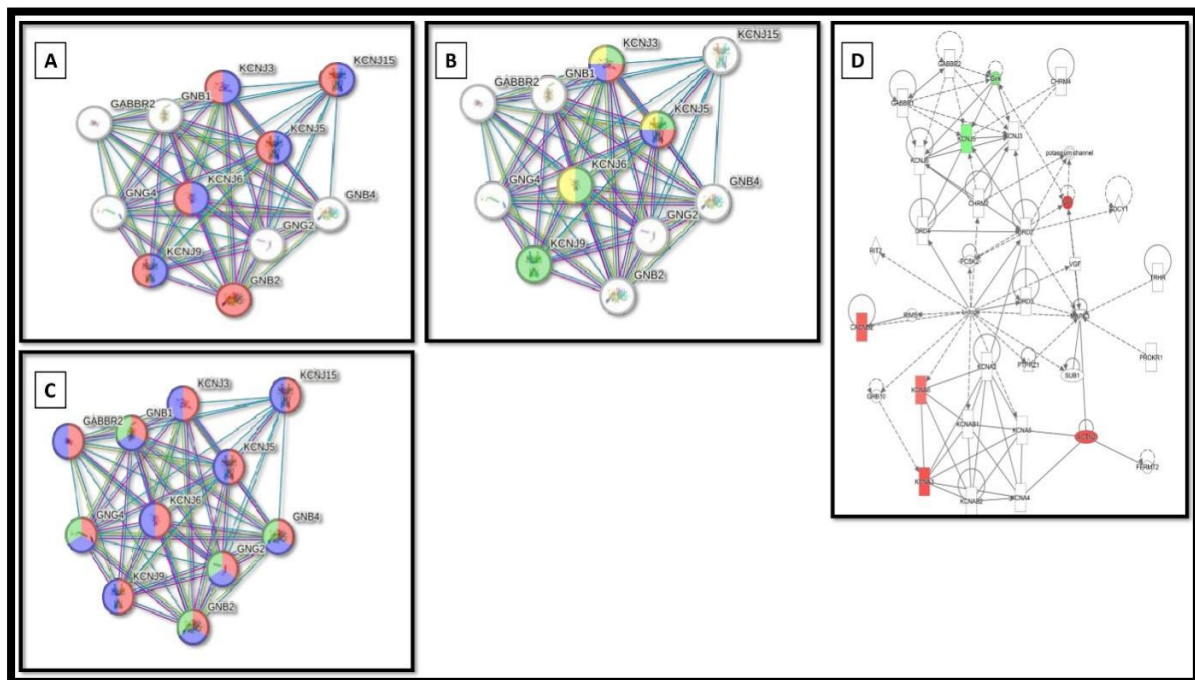


Figure 43: Functional validation of *KCNJ5* gene.

Gene Ontology based on (A) biological processes using STRING (B) molecular functions (C) cellular components. (D) Interaction between all ion channel genes, including *KCNJ5*, *CACNB2*, *KCNA3*, *KCNA6*, and *GABBR2*.

5.2.6: Functional Enrichment of *KCNJ5* and Associated Coexpressed Genes

Linked-Omic, based functional and pathways enrichment analysis of *KCNJ5* and its associated genes in a cohort of 178 patients, having total 19,774 genes were enriched in this study, of them, 11,057 genes reported negative correlations, while the remaining 8717 genes showed significant positive correlation with *KCNJ5* (Figure 44A). The most positive and negative correlated genes associated with *KCNJ5* are represented in the heatmap (Figure 44B, C). For functional and pathway enrichment of the correlated genes with *KCNJ5*, enrichment studies were executed using GSEA method and KEGG pathways category in the Link Interpreter submodule of Linked Omics. As exhibited in Figure 44D, the *KCNJ5* correlated genes show a positive association with oxidative phosphorylation, DNA repair, proteasome, spliceosome related pathways, etc. with ribosome linked pathways, being the most prominent. While, IgA production Th 1,2,17 differentiations, chemokine signalling etc. Documented negative association of *KCNJ5*, with hematopoietic cell lineage pathways showing most negative association. The enrichment plot of ribosome-associated (Figure 44E) and hematopoietic cell lineage (Figure 44F) pathways were demonstrated.

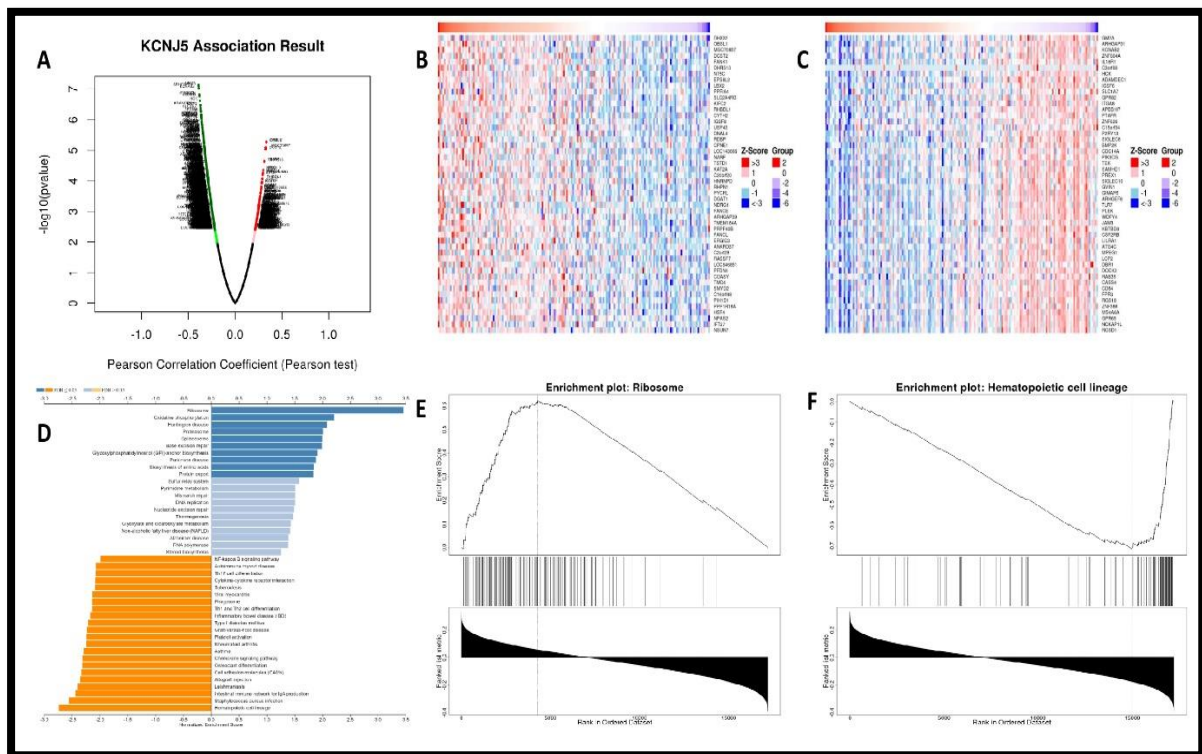


Figure 44: Functional enrichment of *KCNJ5*

(A) Volcanic graph of *KCNJ5* representing positive (red) and negative (green) correlated genes. (B) Heatmap of *KCNJ5* associated positive correlated genes. (C) Heatmap of *KCNJ5* associated negative correlated genes. (D) Functional and pathways enrichment analysis of *KCNJ5* associated positive (blue) and negative (yellow) correlated genes. (E) Enrichment plot for ribosome associated pathways. (F) Enrichment plot for hematopoietic cell lineage pathways.

5.2.7: Identification of Differentially Expressed Ion Channel Gene for PDAC and Protein-Protein Interaction Enrichment Analysis

A list of 8208 DEGs obtained by combining the results from GSE211398 and GSE93326 and TCGA, was further analyzed to acquire 129 DEICGs. The expression ($\log_2FC$) of 41 up-regulated and 88 down-regulated DEICGs are shown in Figure 45A. Next, we generated a PPIN of DEICGs in Metascape. Subsequently, the functional enrichment and MCODE cluster analysis of this PPIN (Figure 45B, C, D) in Metascape reported that GO:0098660 (inorganic ion transmembrane transport) is the most enriched GO Biological Process and there are 8 significant clusters in this PPIN (Figure 45D).

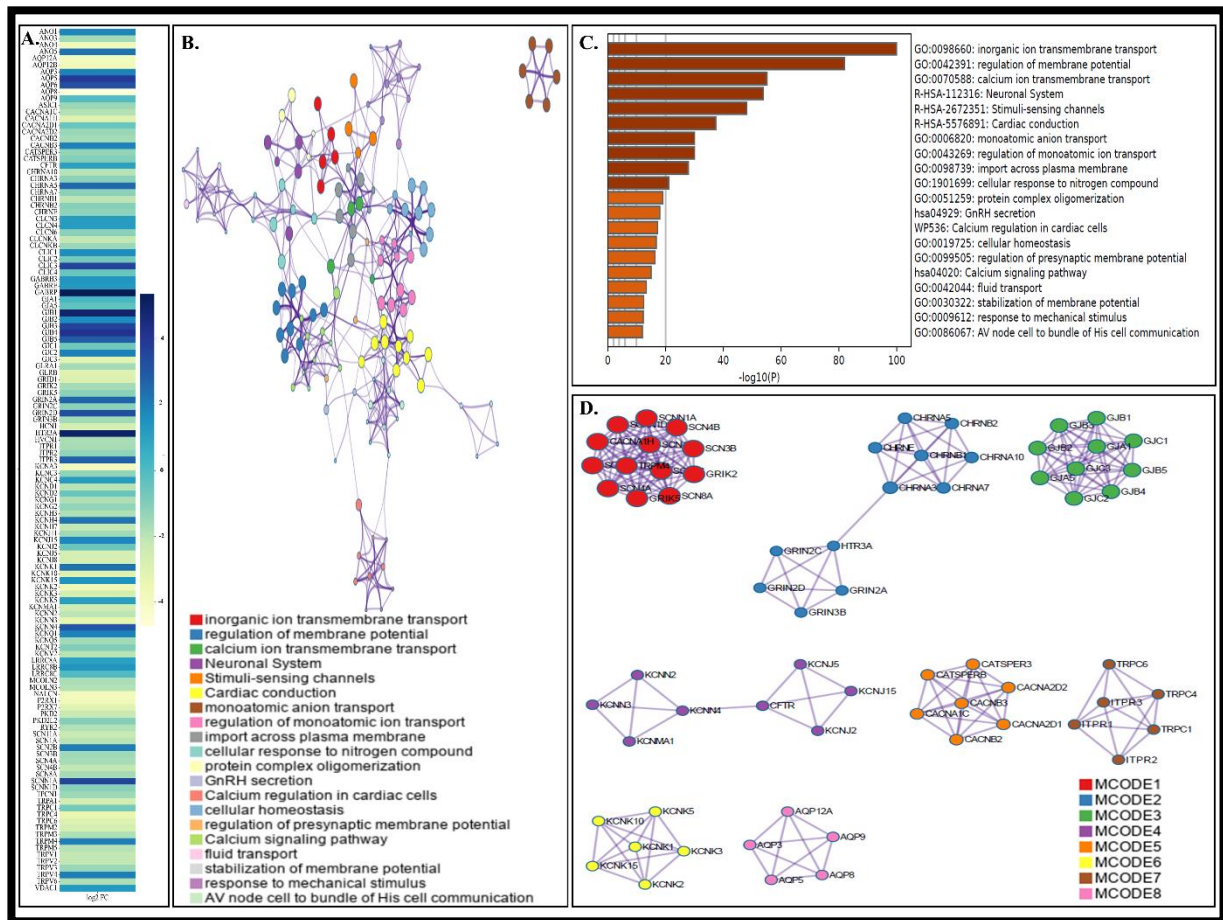


Figure 45: Identification, functional enrichment and cluster analysis of protein-protein interaction network of DEICGs in PDAC.

(A) Heat map depicting the average gene expression ($\log_2FC$) of 129 DEICGs in three datasets -GSE211398, GSE93326 and TCGA (B) Functional enrichment network of DEICGs that is cluster ID-coloured, with nodes with the same cluster ID being closer to one another (C) Bar plot of Top 20 significant GO Biological Processes and KEGG pathways of DEICGs (D) MCODE clusters identified in the network of protein-protein interactions of DEICGs.

5.2.8: Revalidation of Methylation Status in 450K Dataset from Indian Cohort

To further validate our observations concerning methylation status of seven genes, we utilized our own dataset GSE181740 with PanCa patient population of Indian origin. *KCNJ5*, *CLIC5*, *GABRR2*, and *RASA3* were found to be hypomethylated and *CACNB2*, *KCNA3*, while *KCNA6* were hypermethylated based on the delta beta value, obtained post normalization using 450K methylome array (Figure 46). However, all together in Indian cohort of PanCa, *CACNB2*, *KCNA3* and *KCNA6* were hypermethylated, while *KCNJ5*, *CLIC5*, *GABRR2* and *RASA3* were hypomethylated (Figure 46).

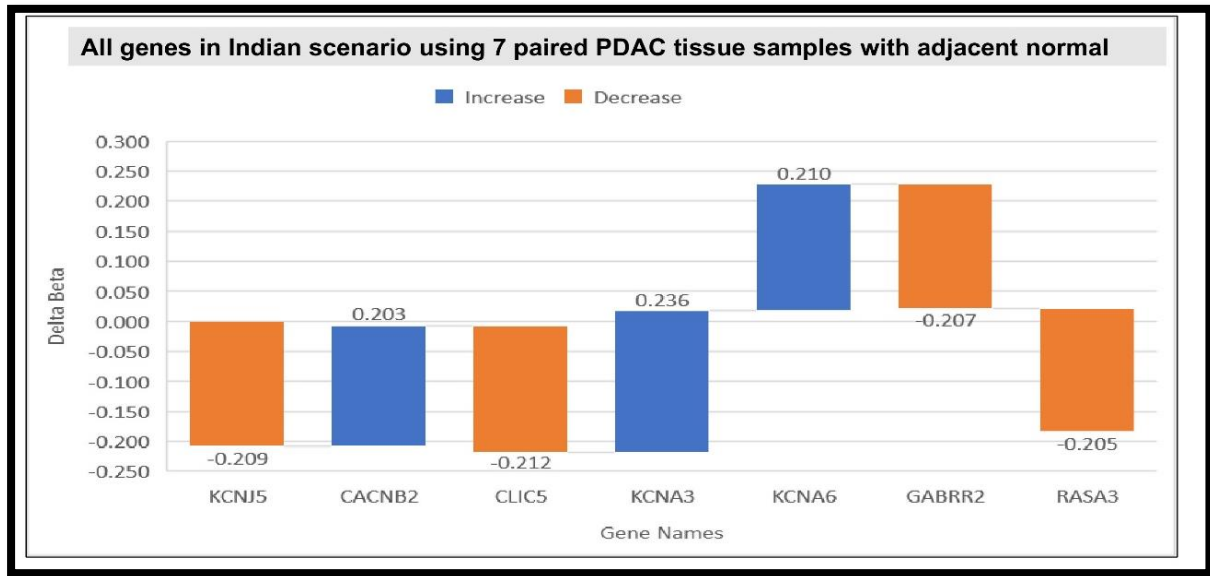


Figure 46: Methylation status in Indian pancreatic cancer dataset

Methylation status of *KCNJ5*, *CACNB2*, *CLIC5*, *GABRR2*, *KCNA3*, *RASA3* and *KCNA6* in Indian cohort of PanCa.

5.2.9: Transcription Factors Regulate the Expression of *KCNJ5*

By combining the TF-binding data of ChEA and RegNetwork, we identified two transcription factors, *POU2F1* and *POU3F1*, both of which might regulate the overexpression of *KCNJ5* in PanCa.

5.2.10: *KCNJ5* Expression in Various PanCa Cell Lines

The CCLE DepMap portal have shown that median expression of *KCNJ5* is low across 50 PanCa cell lines (Figure 47A). But, on examining its expression in individual PanCa cell line in CCLE DepMap portal it was found that, *KCNJ5* is significantly over-expressed in SU86.86, YAPC and HUPT4 PanCa cell line (Figure 47B.), while no cell line documented its protein level expression so far. Similarly, Human Protein Atlas also showed that *KCNJ5* is overexpressed in SU86.86, YAPC and HUPT4 (Figure 47C). It is of interest to note that the

origin of SU86.86 is from Caucasian population while that of YAPC and HUPT4 are from Asian population.

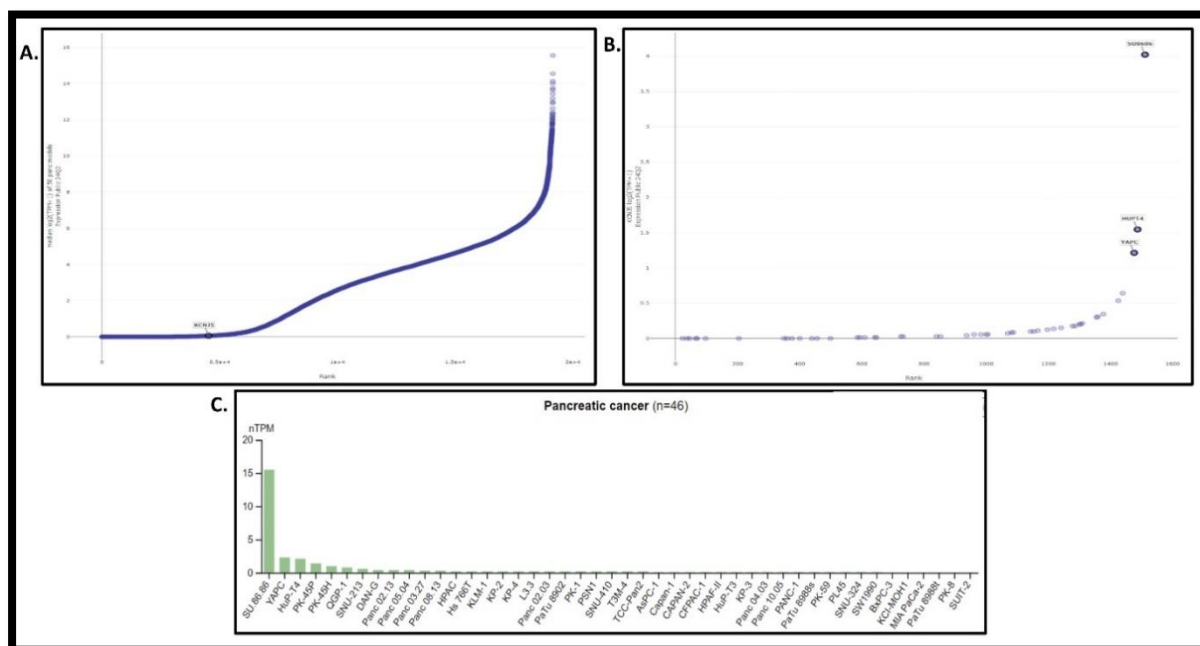


Figure 47: Gene expression of *KCNJ5* in pancreatic cancer cell lines

(A) Median gene expression of *KCNJ5* across 50 PanCa cell lines in Dep Map portal (B) Gene expression of *KCNJ5* in 50 PanCa cell lines in Dep Map portal. (C) Gene expression of *KCNJ5* in 46 PanCa cell lines in Human Protein Atlas.

5.2.11: Identification of Small Molecules and Drug Targeting Kcnj5 Protein

Several small molecule/ drugs targeting Kcnj5 identified through CMap database based on their high enrichment score are Amiloride (enrichment score 1.31) and Vernakalant hydrochloride (enrichment score 1.97) (Table 22). FDA approved, Amiloride is a sodium ion channel blocker, while, sodium and potassium ion channel blocker, Vernakalant hydrochloride, is not approved by FDA. However, Vernakalant hydrochloride (enrichment score 1.97) is approved by European Medicines Agency and in regular clinical practice in European Union and the United Kingdom since 2010. Few other small molecules/ drugs targeting Kcnj5, identified through depmap portal are Dalfampridine, Glyburide and Levcromakalim. Potassium channel blocker, Dalfampridine, and second-generation sulfonylurea and ATP channel blocker, Glyburide are approved by FDA. However, potassium channel activator, Levcromakalim is under clinical assessment through trial studies. Targets were also checked using GDSC1/2 and PRISM but no significant observations were noticed.

5.3: Methylation-regulated Differentially Expressed Genes Identification

5.3.1: Identification of Methylation-regulated Differentially Expressed Genes in PAAD

The workflow of the entire study for the identification of MEDEGs is presented in Figure 48A. We initially performed differential expression analysis and differential methylation analysis to select DEGs and DMGs between PAAD and adjacent normal pancreatic tissue in TCGA-PAAD dataset. With cut-off criteria of $FDR \leq 0.05$ and $|\log_2FC| > 2.0$, a total of 120 DEGs and 69 DMGs were identified. The DEGs and DMGs gene lists obtained were compared using $FDR \leq 0.01$ to generate 27 gene lists, which were found to be common in both lists, therefore showcasing the genes as methylation-regulated differentially expressed genes (Figure 48B). The correlation between 27 DMGs and 27 DEGs with significant negative correlation was visualized (Figure 48C) and also crosschecked once the 27 genes were truly categorized under $Adj. p \leq 0.01$ (Figure 48D) (Table23).

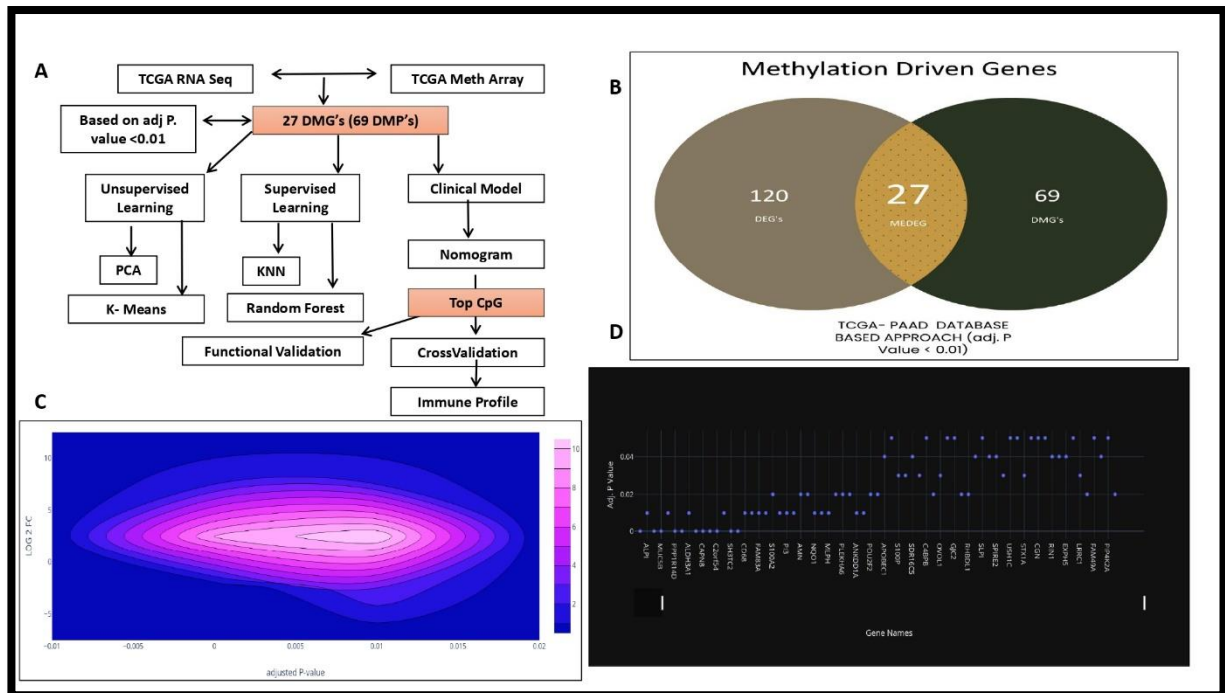


Figure 48: Identification of the methylation driven differentially expressed genes:

(A) The systemic work-flow of our study. (B) The distinguishing Venn diagram of the MEDEG's. TCGA-PAAD dataset contained 120 DEG's ($p < 0.05$) and 69 DMG's ($p < 0.05$). 27 MEDEG's were identified based on $FDR < 0.01$. (C) Correlation between DMGs and DEG's with significant negative correlation and $adj. P < 0.01$ leading to 27 MEDEG's identification. (D) Identification of 27 DEG's out of 69 based on $adj. P < 0.01$.

5.3.2: Clustering of Methylation-regulated Differentially Expressed Genes Across Global Pancreatic Cancer Databases

Prior to clustering and dimensional reduction, the 27 genes were compared across tumor and normal using TCGA-PAAD database to generate the pattern of expression alteration based on the fold change (Figure 49A). Principal Component Analysis (PCA) was done for dimensional reduction using TCGA-PAAD database and GTEx Pancreas normal data. Significant clustering was observed as normal pancreas data from GTEx clustered individually when visualized against TCGA-PAAD tumor tissue data showing significance of methylation data to generate cluster identities (Figure 49B). Prior to PCA clustering, the optimum variance was also checked. Spearman's correlation was also done across 27 genes to obtain correlation patterns across genes (Figure 49C). The methylation values of the corresponding CpGs sites were successful in clustering TCGA- tumor sample from TCGA adjacent normal sample and from GTEx pancreas normal sample into 3 identifiable regions. Clustering was also done using an unsupervised learning algorithm to obtain clusters in an unlabelled process corresponding to 27 MEDEGs (47 CpGs). Prior to K-means clustering, the optimum number of clusters was obtained using Euclidean distance measurement and silhouette calculations (Figure 49D, E). Four clusters (K2-K5) were checked before concluding the optimum cluster size. 2 clusters (K-2) were found to be optimum (Figure 49F, G). This demonstrated success of methylation values in generating clusters based on the β values.

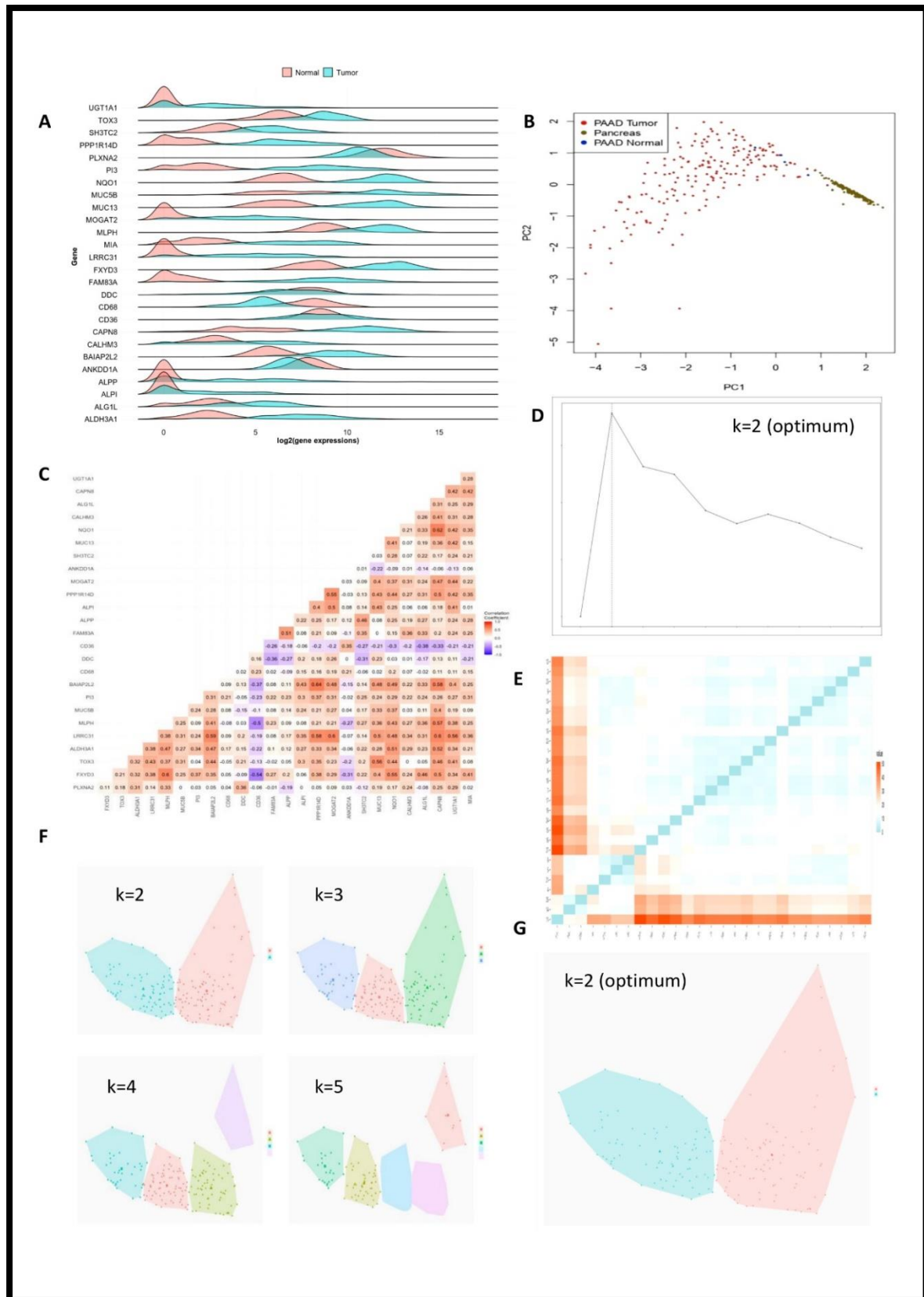


Figure 49: Pattern understand and clustering of DNA methylation driven genes across global PanCa databases: (A) Differential expression pattern of 27 genes across Tumor and Normal Samples. (B) Principal Component Analysis undertaking TCGA-PAAD and GTEx Pancreatic Cancer Database. (C)

Spearman Correlation to identify correlation patterns across 27 genes. (D) Silhouette methodology to identify optimum clusters. (E) Euclidean distance calculation prior to K-means algorithm-based calculations. (F) K-means clustering considering four varied cluster sizes (K2-K5). (G) Optimum cluster identification (K-2) using 27 genes mapped to 47 CpG's.

5.3.3: Implementation of Based Random Forest Classification Using Methylation-regulated Differentially Expressed Genes

A Random Forest algorithm was implemented to forecast the number of days until patients pass away by utilizing an extensive collection of characteristics, which encompassed MEDEGs. The RF modelling process was executed in the subsequent manner:

The algorithm can be mathematically represented as follows:

Given a dataset $(\mathbf{x}_1, \mathbf{y}_1, \mathbf{d}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n, \mathbf{d}_n)$, where $\mathbf{x}_i$ is a vector of features, $\mathbf{y}_i$ is the time of event or censoring and $\mathbf{d}_i$ is the event indicator (1 if event, 0 if censoring).

Let $\mathbf{h}_i(\mathbf{x})$ be a survival function estimated by the i^{th} tree, the random survival forest can be represented as:

$$\hat{\mathbf{S}}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \mathbf{h}_i(\mathbf{x})$$

Where n is the number of trees in the forest, and $\mathbf{h}_t(\mathbf{x})$ is the survival function predicted by the t -th tree for observation $\mathbf{x}$.

In addition, RSF also provides an estimate of the standard error of the predictions, which can be represented as:

$$\hat{\mathbf{se}}(\mathbf{x}) = \sqrt{\frac{1}{n-1} \sum_{t=1}^n \left(\mathbf{h}_t(\mathbf{x}) - \hat{\mathbf{S}}(\mathbf{x}) \right)^2}$$

Where n is the number of trees in the forest, and $\mathbf{h}_t(\mathbf{x})$ is the survival function predicted by the t -th tree for observation $\mathbf{x}$, and $\hat{\mathbf{S}}(\mathbf{x})$ is the overall survival function predicted by the forest for observation $\mathbf{x}$.

5.3.3.1: Data Splitting

In order to evaluate the model's responses to novel data, the dataset was partitioned into training and testing sets.

5.3.3.2: Feature Selection

The selection of MEDEGs and other pertinent features was predicated on their predictive value for survival outcomes. This was accomplished by employing techniques such as recursive feature elimination and the RF algorithm's feature importance scores.

5.3.3.3: Model Training

The training set was utilized to train RF models, with a predetermined quantity of trees (n). In order to cultivate each tree, the data were divided recursively in accordance with the optimal split point, which was identified using a dividing criterion like the log-rank test.

5.3.3.4: Cross-Validation

During the training phase, k-fold cross-validation (5-fold or 10-fold) was implemented to guarantee robustness and prevent *overfitting*. The process comprised dividing the data into k subsets and validating the model on the remaining fold after training it on k-1 folds. This procedure was repeated in order to derive average performance metrics.

5.3.3.5: Performance Metrics Indicators

The area under the receiver operating characteristic curve (AUC-ROC) served as the principal metric for assessing the RF model's discriminatory capability. The AUC-ROC metric offers an exhaustive evaluation of the model's capability to differentiate among various survival outcomes.

5.3.3.6: The Determination of the Cut-off

In order to ascertain the most effective threshold for categorizing patients as high-risk or low-risk, sensitivity, specificity, or other pertinent criteria were considered. The selection of this cut-off was intended to optimize the ratio of true positives to false positives.

5.3.3.7: Concluding Model Assessment

The efficacy of the ultimate RF model was evaluated on an independent testing set in order to assess its ability to generalize. In addition to AUC-ROC, various metrics, including accuracy, precision, recall, and F1-score, were taken into account. By performing the RF analysis 1523 times in total, various subsets of the dataset were incorporated to ensure the results' stability and dependability. Quantifying the predictive performance of the Random Forest model in predicting days to death for patients diagnosed with PanCa, the final AUC-ROC, precision-recall curves, and other pertinent metrics were disclosed. We have estimated the classifier score of 0.924, which suggests that our Random Forest classifier is performing very well on the given dataset, making accurate predictions for the majority of the instances. Keep in mind that accuracy alone might not be sufficient to evaluate the performance of our model, especially in the presence of imbalanced classes or other specific considerations. It's often a good idea to complement accuracy with other evaluation metrics such as precision, recall, F1-score, or confusion matrix, depending on the nature of your classification problem. So, we

went for the other matrices. The precision, or positive predictive value, of 0.769 indicates that 76.9% of the instances indicated as having PanCa were correctly identified as positive cases. Practically speaking, this implies that when the model makes a prediction of a positive result, such the detection of PanCa, it is accurate around 76.9% of the time. The recall, also known as sensitivity or true positive rate, is 0.748. This means that our model correctly detected and collected 74.85% of all actual positive cases of PanCa. This statistic highlights the classifier's capacity to accurately identify instances of PanCa, suggesting its sensitivity to the existence of the disease.

The F1-score, which is the harmonic mean of precision and recall, is 0.9546. This indicates a well-balanced performance between precision and recall. This statistic encompasses both false positives and false negatives, yielding a comprehensive assessment of the model's capacity to precisely forecast PanCa. The high F1-score indicates a strong equilibrium between accurately detecting positive situations and avoiding both false positives and false negatives. To summarize of this part, our Random Forest classifier shows good performance in predicting PanCa, with high precision, recall, and a well-balanced F1-score. The results demonstrate the model's efficacy in reliably detecting positive cases and reducing the likelihood of incorrect predictions, which is vital in the context of diagnosing and predicting the outcome of PanCa (Figure 50).The variance in the test data was seen to be 0.1091 which ensures that our outcome doesn't get biased due to test data's variance.

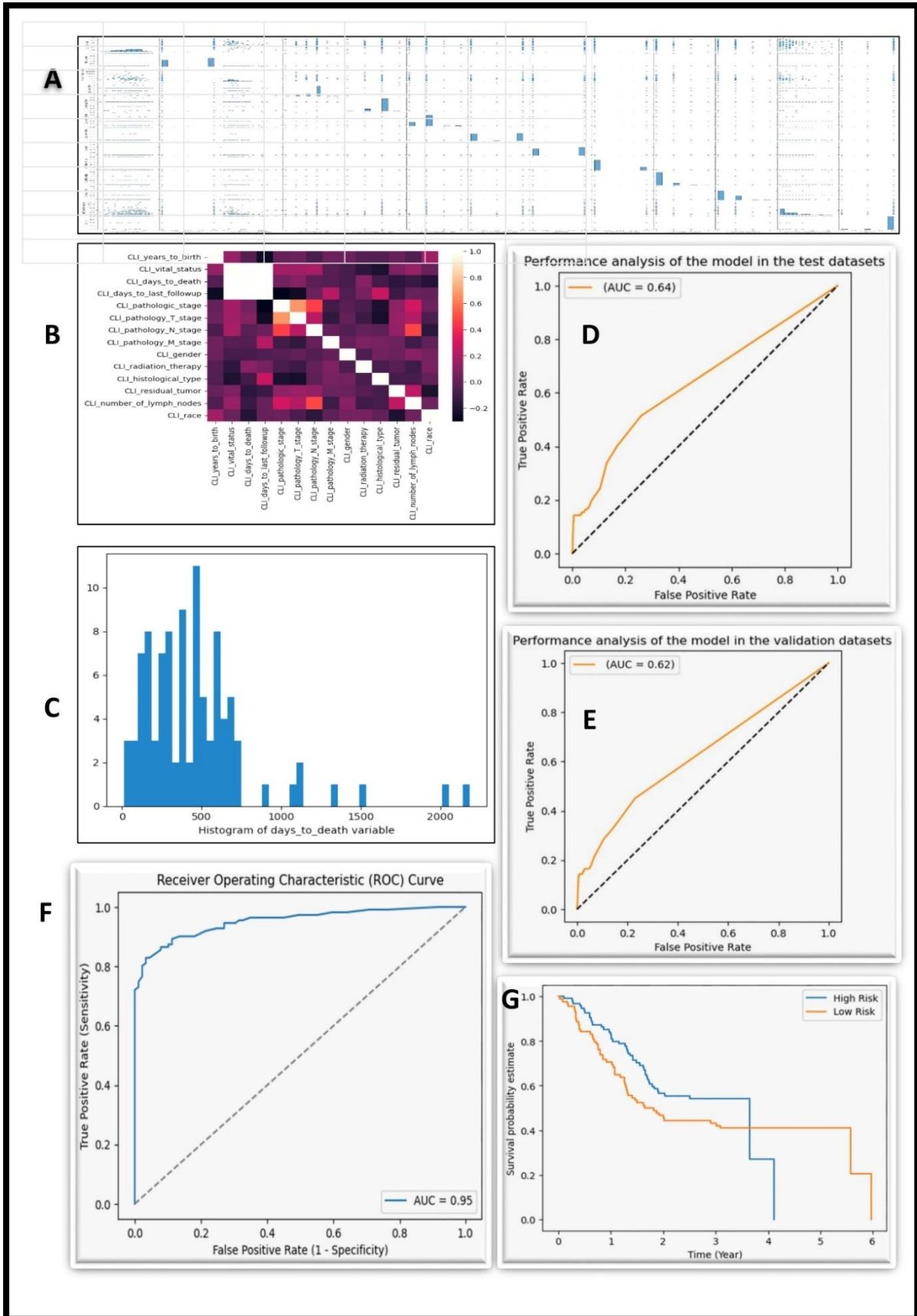


Figure 50: Supervised clustering to develop methylation driven predictive models: (A) Pairwise-pattern plot across all significant CpG sites using Random Forest Classifier. (B) Multivariate

Correlation across Clinicopathological Parameters using Random Forest Classifier. (C) Pattern identification in context of Random Forest Classifier event, Days to Death.

5.3.4: Construction of Methylation-regulated Differentially Expressed Genes Based Nomogram

The univariate Cox regression analysis was first performed to calculate the association between the methylation level of each DMS and the patient's OS in the training cohort. Those sites with *p-values* less than 0.05 were identified as prognosis-related MEDEGs. Here $\log(\lambda)$ value is 1.145. This suggests that the hazard ratio associated with the covariates in your model is approximately $e^{1.145}$ i.e 3.144. As the hazard ratio is greater than 1, it indicates an increased risk of the event (CLI Death). Now we explore the variable selection issue of the Cox proportional hazards model and suggest a unified model selection and estimate method with the desired theoretical features and ease of computation. The Adaptive LASSO (Least Absolute Shrinkage and Selection Operator) estimator is produced by the novel approach, which is based on a penalized log partial likelihood with the adaptively weighted L_1 penalty on regression coefficients. The approach uses varying penalties for various coefficients. Insignificant variables are subject to greater penalties than important ones, which results in a tendency for important variables to be kept in the selection process while unimportant variables are more likely to be eliminated. The test is done on the estimator's theoretical characteristics, including consistency and rate of convergence. Then, using Adaptive LASSO method to screen the prognosis-related MEDEGs and obtain an optimal model subsequently, the prognosis-related MEDEGs with a coefficient not equal to 0 were retained as significant variables, and a risk scoring model was established using the combination of weighted methylation values.

Using the median risk score (-4.307) as the cut-off point, patients were divided into low-risk (risk score below the median value) or high-risk (risk score above the median value) group correspondingly. We obtained the optimal value of the parameter λ , which controlled the degree of Adaptive LASSO regression complexity, and selected the significant variables through multiple cross-validation. We found that the parameter λ reached the optimal value, when the number of variables was five. Therefore, combining the regression coefficients of three DMSs under the optimal λ value, we constructed a 5 MEDEGs (*MOGAT2*, *MIA*, *CD68*, *LRRC31*, and *C2ROF54*) risk score model to guide the prognosis of PAAD patients. The optimal lambda value of 0.7588, in the context of Adaptive LASSO for Cox regression, signifies the strength of the regularization applied to the coefficients in your model. In

LASSO regularization, the lambda parameter controls the degree of shrinkage applied to the estimated coefficients. This lambda value implies:

5.3.4.1: Regularization Strength:

A higher lambda value corresponds to stronger regularization. In the context of LASSO, higher lambda values lead to more coefficients being shrunk to exactly zero, effectively performing variable selection by excluding less relevant features from the model.

5.3.4.2: Sparsity in Coefficients:

The optimal lambda value of 0.7588 suggests that the model has found a balance between fitting the data well and keeping the model simple. Some coefficients in your model may have been shrunk to zero, indicating that certain variables are not contributing significantly to the prediction of survival outcomes.

5.3.4.3: Model Interpretability:

Adaptive LASSO, in particular, adapts the penalty weights for each variable based on their estimated importance. The adaptive nature helps in variable selection while still allowing important variables to have non-zero coefficients. AUC of 0.66 suggests that the model has the ability to discriminate between positive and negative instances, but the discrimination may not be very strong.

Our optimal lambda value of 0.758 reflects the degree of regularization applied to your Cox regression model. It indicates a good balance between model complexity and goodness of fit, with certain coefficients potentially being shrunk to zero for variable selection. We have developed a nomogram that combines the risk score and clinical risk factors of PAAD patients assessed by multivariate Cox proportional-hazards regression to predict the 3- and 5-year overall survival in the TCGA-PAAD cohort, in order to apply the prognostic value of the MEDEGs-based signature in clinical settings. The discriminative capacity of the nomogram was assessed by computing the concordance index (C-index), a metric for discriminatory power. In order to compare the observed and projected probabilities for the nomogram, calibration plots were generated. (Figure 51).

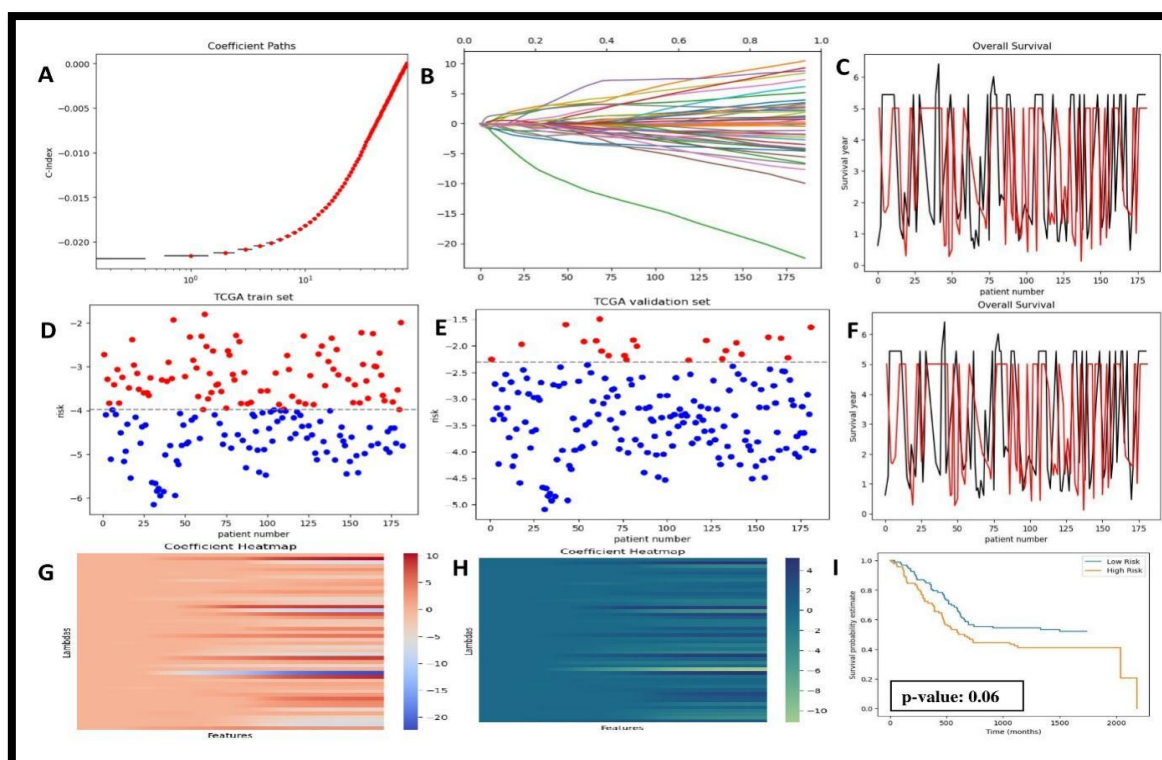


Figure 51: Development of a seven-CpG-sites-based prognostic prediction model of pancreatic cancer using Adaptive Lasso regression.

(A) The λ value of the lasso model was decided by cross-validation with minimal misclassification using C -index. (B) Five CpG sites and their corresponding coefficients were utilised to construct the prognostic risk model (D & E) Scatter plot of the distribution pattern of risk score between high- and low-risk groups in the training set and validation set. (C & F) Overall survival analysis across the patient sets between the high- and low-risk groups in the validation set. (G & H) Coefficient derived Heat map of the seven CpG sites across the training set and Validation set.

5.3.5: Identification of Biological Process Involving the Methylation-regulated Differentially Expressed Genes

To further investigate the biological processes in which the MEDEGs might be involved in, we performed GO annotation BP Ontology considering Benjamini-Hochberg adjustment in overall biological processes and we also used MEDEG in PanCa based on BP Ontology considering Benjamini-Hochberg adjustment. The top significant terms emerging from the gene ontology enrichment analysis were shown in Figure 52A, B. The results indicated that these MEDEGs played key roles in multiple cancer-related pathways and further indicated that the MEDEGs might be involved in PAAD progression by regulating the corresponding gene expression. Functional enrichment was also done using Metascape though no significant interactions were observed (Figure 52 C). Enrichment was also done using Enrich R targeting ChEA data 2022 along with ENCODE_Histone_Modifications_2015, ENCODE_TF_ChIP-seq_2015, Epigenomics_Roadmap_HM_ChIP-seq, FANTOM6_lncRNA_KD_DEGs,

TF_Perturbations_Followed_by_Expression, TF-LOF_Expression_from_GEO. Few significant observations were documented in all the above databases (Figure 52 D - J).

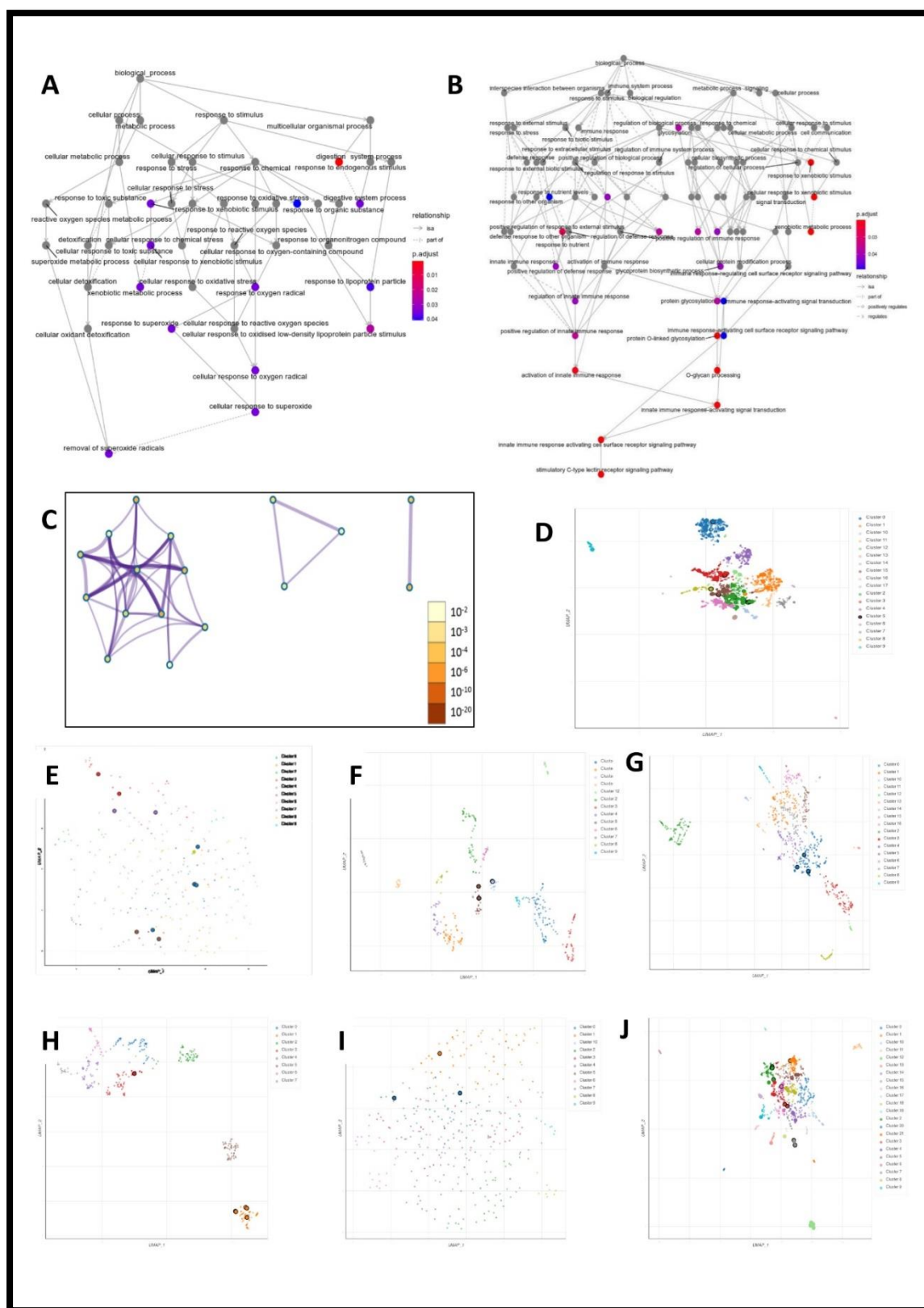


Figure 52: Functional characterization of 27 methylation driven differentially expressed Genes;

(A)GO Plot based on BP Ontology considering Benjamini-Hochberg adjustment. (B)GO plot obtain using DEG in PanCa based on BP Ontology considering Benjamini-Hochberg adjustment. (C)Enrichment network obtained by MetaScape which has its nodes colored by p-value. The darker the color, the more statistically significant the node is. (D-J) Enrich R based targeting ChEA data 2022 along with ENCODE_Histone_Modifications_2015, ENCODE_TF_ChIP-seq_2015, Epigenomics_Roadmap_HM_ChIP-seq, FANTOM6_lncRNA_KD_DEGs, TF_Perturbations_Followed_by_Expression, TF-LOF_Expression_from_GEO.

5.3.6: Targeted Functional Enrichment of *MOGAT2*, *MIA*, and *CD68*

Post both supervised and unsupervised algorithm based evaluation followed by nomogram and expression status validation across multiple platforms and dataset we observed that out of 5 only 3 MEDEGs namely *CD68* (cg05209483) and *MIA* (cg25152942) genes and the three CpG sites from *MOGAT2* (cg12531542, cg20938170, cg25255988) were the most significant MEDEGs. With a sample size of 178 patients, Linked Omic was utilized to look into the pathways and functional enrichment analysis of *MOGAT2*, *MIA*, and *CD68* as well as the genes that are linked to them. Red dots indicated a negative correlation with the corresponding gene and green dots a positive correlation as was done in case of *MOGAT2*, *MIA* and *CD68* (Figure 53 A-C). A heatmap shows the most positively and negatively correlated genes with *MOGAT2*, *MIA*, and *CD68* (Figure 53 D-I). Using the GSEA method and the KEGG pathways category in the Link Interpreter submodule of Linked Omics, enrichment studies were carried out for the functional and pathway enrichment of the correlated genes with *MOGAT2*, *MIA*, and *CD68* (Figure 53 J-L). Most enriched pathway namely *Staphylococcus aureus* infection and morphine addiction were showcased in Figure 53 M, N.

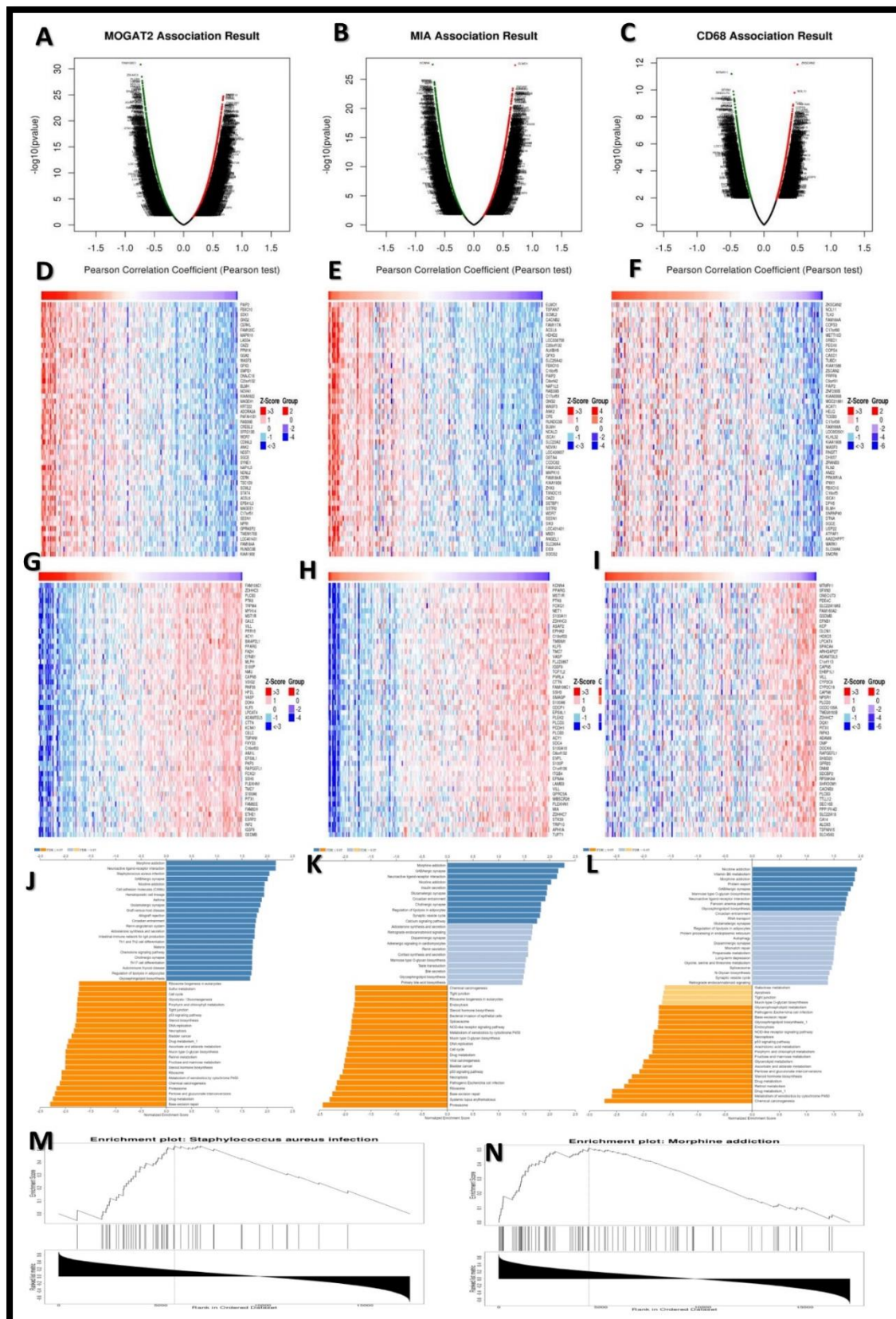


Figure 53: Functional enrichment of MOGAT2, MIA and CD68

(A-C) Volcanic graph of MOGAT2, MIA and CD68 representing positive (red) and negative (green) correlated genes. (D-F) Heatmap of MOGAT2, MIA and CD68 associated positive correlated genes. (G-I) Heatmap of

MOGAT2, *MIA* and *CD68* associated negative correlated genes. (J-L) Functional and pathways enrichment analysis of *MOGAT2*, *MIA* and *CD68* associated positive (blue) and negative (yellow) correlated genes. (M-N) Enrichment plot of important enriched functional pathways.

5.3.7: Cross Validation of Methylation -regulated Differentially Expressed for 5 Genes

For further validation of methylation markers, we also utilized WGBS data of PanCa progression in patient samples to assess the methylation pattern of CpGs associated with *CD68* (cg05209483) and *MIA* (cg25152942) genes and the three CpG sites from *MOGAT2* (cg12531542, cg20938170, cg25255988). The WGBS data consisted of multiple tumors from same patients (four patients were analyzed), which included normal pancreas, primary tumor, local peritoneal metastasis, and distant metastasis samples (liver and lung metastasis, data source: GSE63123). For *CD68*, we found the cg05209483 is hypomethylated in primary tumor and also in liver, lung metastatic samples ((median methylation =0) and peritoneum tumor (median methylation =0.142, i.e., 14.2%) compared to the normal pancreas (median methylation =0.33, Figure 54). For *MIA* (cg25152942), we found it to be hypomethylated in primary tumor (median methylation =0.428) compared to normal pancreas (median methylation =0.66). However, lung, liver and peritoneum tumors showed varying degrees of methylation (0.52 to 0.94). For *MOGAT2* CpG sites, we found these are highly methylated in normal pancreas (0.66 to 0.75) and for all CpG sites of *MOGAT2*, we found they are completely hypomethylated in the liver metastasis tumors. Interestingly, for the lung metastasis, peritoneum and primary tumors, we found consistent hypermethylation (Figure 54).

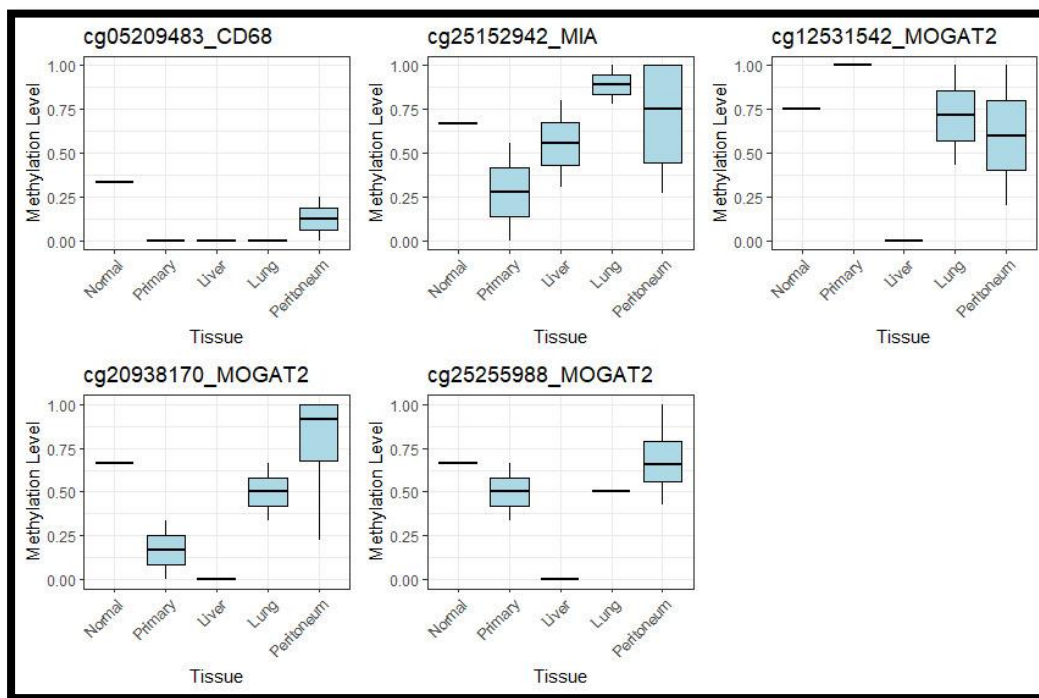


Figure 54: Methylation markers (*MIA*, *MOGAT2*, and *CD68*) evaluated using WGBS data.

This consisted of multiple tumors from same patients, which included normal pancreas, primary tumor, local peritoneal metastasis and distant metastasis samples.

This suggests that the identified CpGs of *MOGAT2* could be a specific marker of PanCa to liver metastasis. In pan tissue expression analysis, we also found that *MOGAT2* has the highest expression in the liver (median transcript per million= 57.92, RNA-Seq data, sample size= 226, data source: GTEx, Figure 55 A).

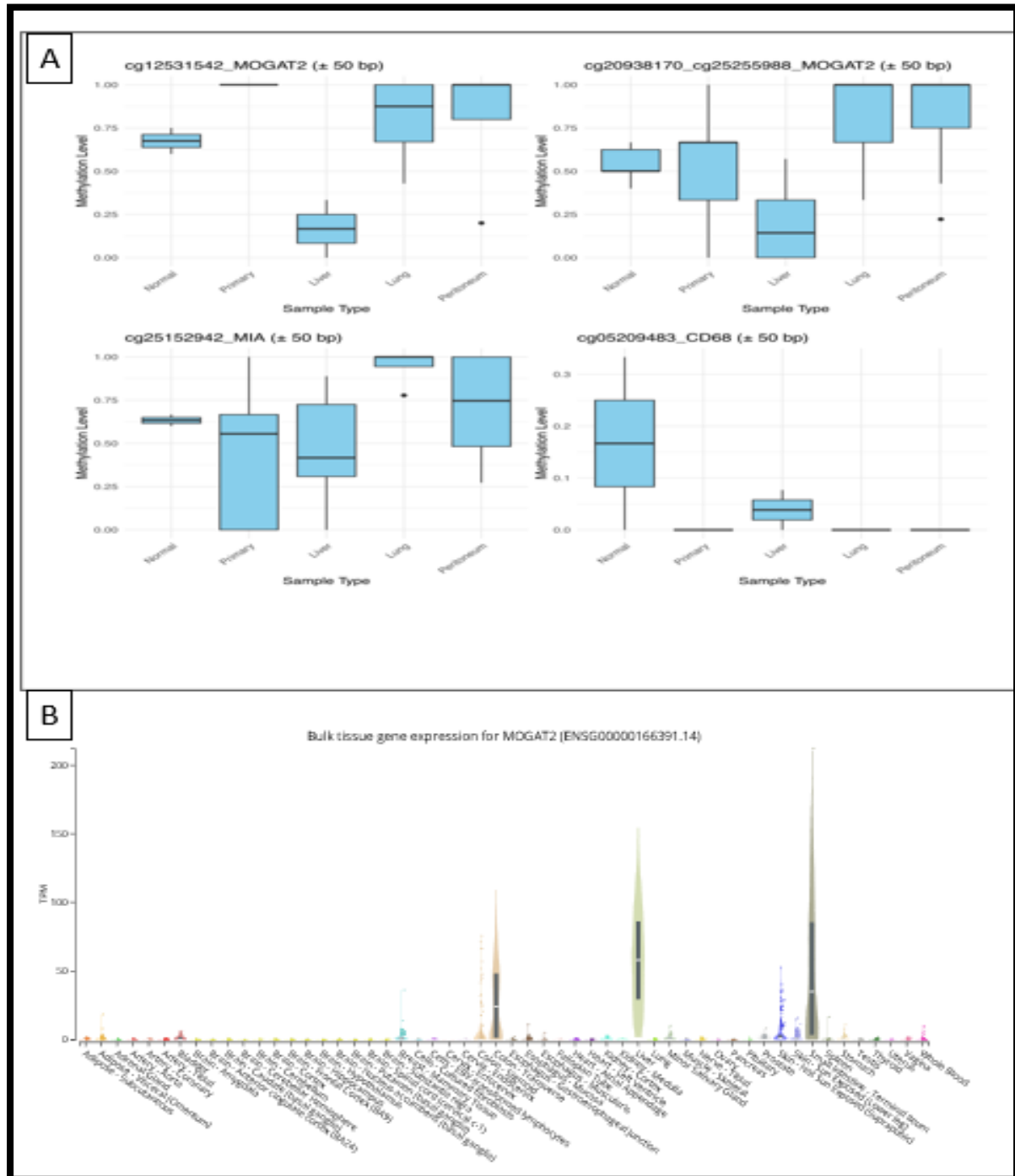


Figure 55: Tissue of origin targeting for respective methylation driven differentially expressed genes

(A) WGBS validation of CpG sites across liver to corresponding metastasis. (B) Bulk Tissue expression of *MOGAT2*.

However, future methylation studies with large sample numbers containing multiple metastatic sites will be needed to confirm this observation. As we have obtained WGBS data, we aimed to assess the adjacent regions of the CpGs (by expanding the analysis to 50 bp on each side, i.e., of 100 bp window, Figure 55 B). However, there are no adjacent CpGs for *MIA* and *CD68*, and for *MOGAT2*, cg20938170 and cg25255988 were located next to each other (19 bp apart). *MIA*, *MOGAT2* and *CD68* also have shown presence in CCLE datasets containing 35 PanCa cell lines. *MOGAT2* also showed its significance when seen in GO, BP and also in ORA as shown in the networks obtained. (Figure 56).

5.3.8: Validation of CpGs (MEDEGs) of Interest in Independent GEO Datasets Containing Methylation and Expression Data of Pancreatic Cancer.

We then compared our data to 3 GEO methylation datasets (450K) and 3 other expression datasets (RNA seq and micro-array). Two methylation datasets were from western culture and 1 GEO methylation dataset (GSE181740) is our own dataset obtained from Indian patient population. 2 western world methylation datasets showed our target genes of interest to be significant (*MOGAT2*: $p - 0.002$ / *CD68*: $p - 0.001$) but in Indian patient population it was not found which can be due to epigenetic tissue specific expression which can be diverse due to change in demographic, ethnic, and lifestyle differences. The total 7 CpGs were found to be significant in the western GEO methylation datasets and also was found to be significant when observed in both RNA seq and micro-array expression datasets. (Figure 56)

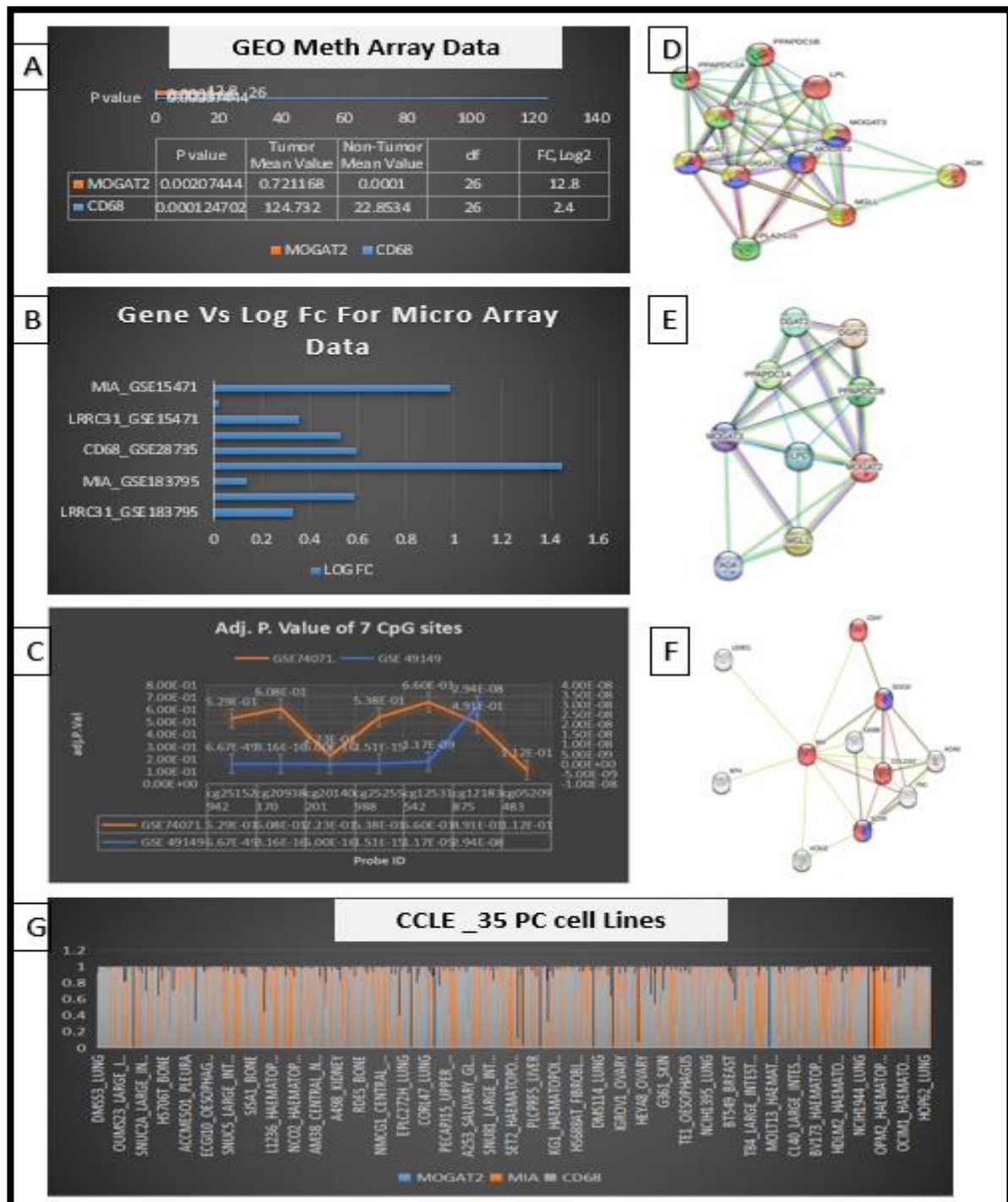


Figure 56: Independent database associated validation of methylation driven differentially expressed genes

(A-C) Profile validation of 7 CpG sites/ 5Genes across GEO databases containing Methylation array and RNA seq expression data of Pancreatic Cancer. (D) MOGAT2 gene ontology considering biology process found to be enriched via STRING. (E) MOGAT2 gene ontology considering molecular process found to be enriched via STRING. (F) ORA was used to showcase MIA network.

5.3.9: Parallel Screening of Methylation-regulated Differentially Expressed of 5 Genes Using Functionally Derived Databases

All five genes were checked in all above cases and non-significant observations were depicted in Figure 57 and Figure 58. The OS was also evaluated using Kaplan-Meier and by splitting the TCGA sample set into high risk and low risk group. ANOVA was used to evaluate individual probes representing the genes across 4 cancer stages to evaluate the significance of our desired genes. Apart from this work we tried to evaluate the joint survival of 3 genes which individually were found to be significant. The test group met each of the subsequent requirements: $MIA > 0.50$, $CD68 > 0.50$, and $MOGAT2 > 0.50$. The control group met each and every one of the subsequent requirements: $CD68 \leq 0.50$, $MIA \leq 0.50$, and $MOGAT2 \leq 0.50$. The percentiles (0.00 to 1.00) for each of the aforementioned thresholds relate to other patients in this study for whom data are available. Cases with insufficient data or those that did not meet the aforementioned criteria were eliminated. Of the 185 cases for which clinical data were available, 26 met the test criteria and 23 met the control criteria. A total of 136 cases were disqualified. With a corresponding $p < 0.6519$, the test group's hazard ratio compared to the control group was 1.215.

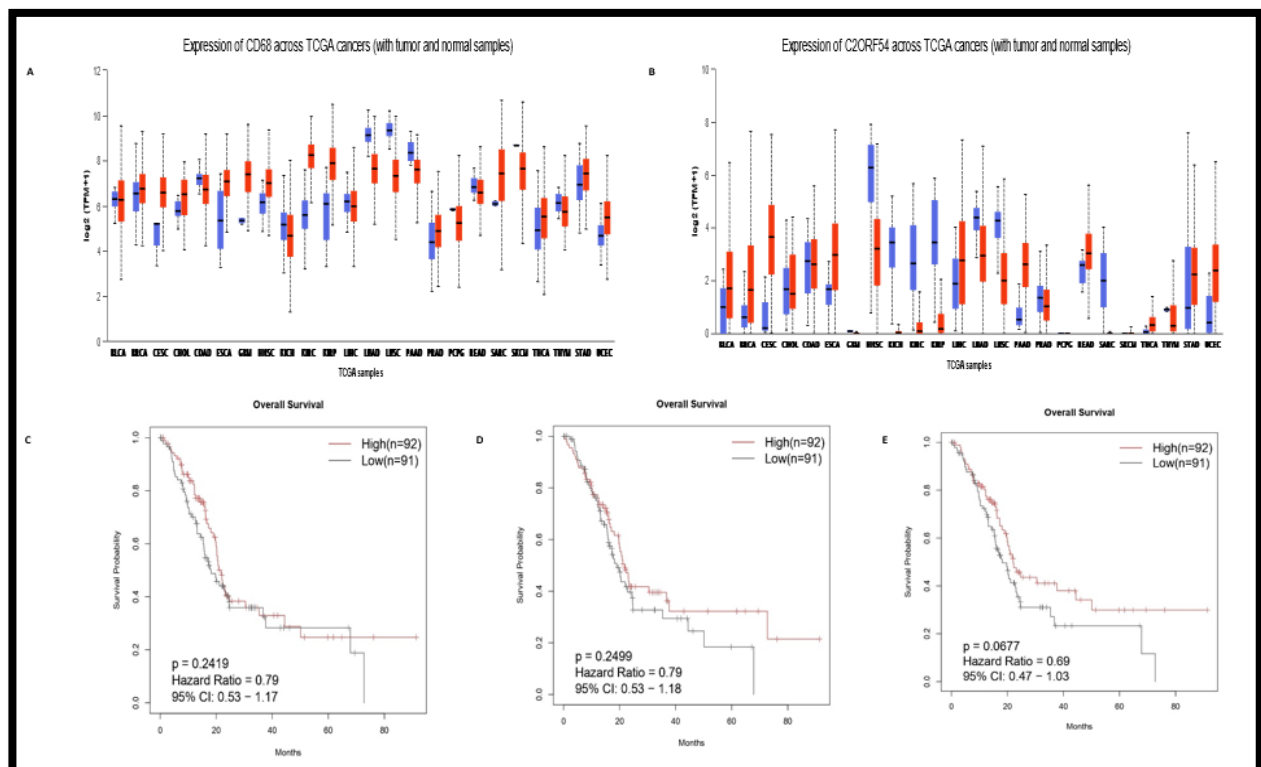


Figure 57: Survival associated screening of methylation driven differentially expressed genes

(A)Pan-Cancer validation of CD68 across various cancer types.(B)Pan-Cancer validation of C2ORF54 across

various cancer types.(C)Kaplan-Meier plot depicting Overall survival for patients based on *cg20140201* (*C2orf54*) expression. (D)Kaplan-Meier plot depicting Overall survival for patients based on *cg05209483* (*CD68*) expression. (E)Kaplan-Meier plot depicting Overall survival for patients based on *cg25152942* (*MIA*) expression.

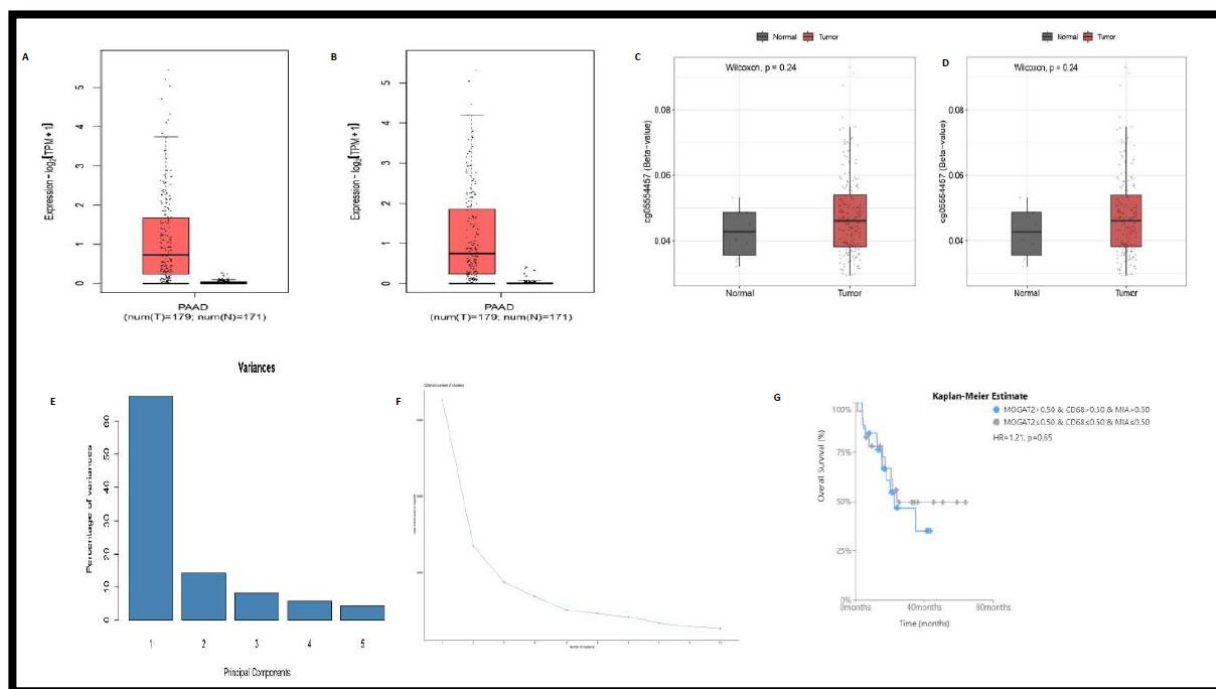


Figure 58: Parallel screening of gene sets using alternative screening processes

(A) *LRRC31* expression validation across TCGA and GTEx. (B) *MOGAT2* expression validation across TCGA and GTEx. (C-D) Cross validation of top CpG across tumor and normal using Wilcoxon Rank Sum Test.(E)Variance identification prior to PCA. (F) ELBOW method for identification of optimum cluster identification prior to K-means. (G) The mRNA expression (RNA Seq V2 RSEM) data was used to select each group by combining the percentiles of each gene's expression in the set (*MOGAT2*, *MIA*, *CD68*) within the study population.

TCGA PanCa samples were also compared with GTEx normal pancreas tissue to check change in differential expression. Wilcoxon test was used to compare change in differential methylation of the probes across tumor vs normal. *MOGAT2*, *MIA* and *CD68* reported significance/ border line significance across all the above-mentioned tests. *LRRC31* and *C2ORF54* showed no significance in the downstream evaluations mentioned in above tests. Protein level expression of *MOGAT2*, *CD68* and *MIA* were found to be significant when observed across CPTAC database. pan-cancer Validation of *MOGAT2*, *MIA*, and *LRRC31* showing highest fold change. *MIA*, *MOGAT2* and *CD68* expressions were observed in GTEx database to identify tissue of origin and their relevant expression in normal pancreas tissue (Figure59, 60). The mutation profile of *MIA*, *MOGAT2* and *CD68* was also checked which showed *MIA* to have highest percentage of amplification (Figure 61).

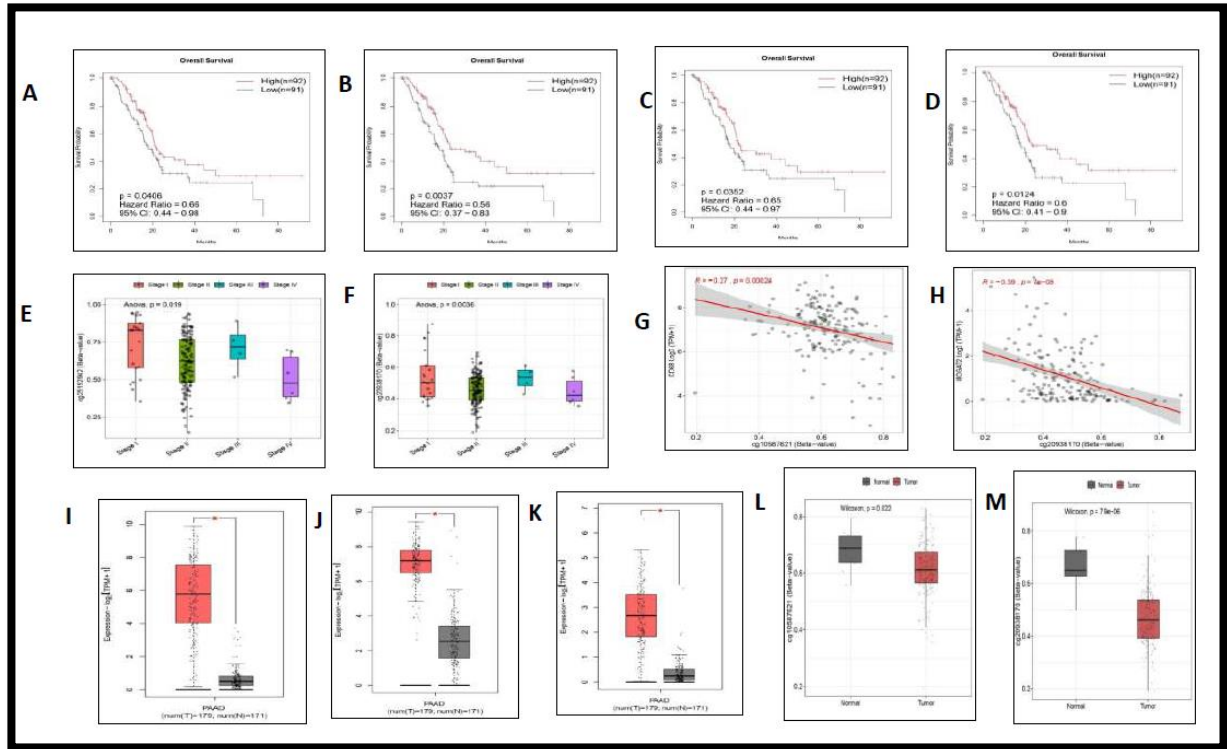


Figure 59: Cross Validation of methylation driven differentially expressed genes and their CpG's considering their clinical significance:

(A-D) Kaplan Meir survival plots of top 4 genes considering TCGA- PAAD Database. (E-F) Annova based identification of top CpG's across 4 stages of PanCa using TCGA- PAAD Database. (G-H) Correlation of top 2 genes with their corresponding CpG's. (I-K) Expression comparison of top 3 genes across TCGA-PAAD and GTEx normal Pancreatic tissue data. (L-M) Cross validation of top CpG across tumor and normal using Wilcoxon Rank Sum Test.

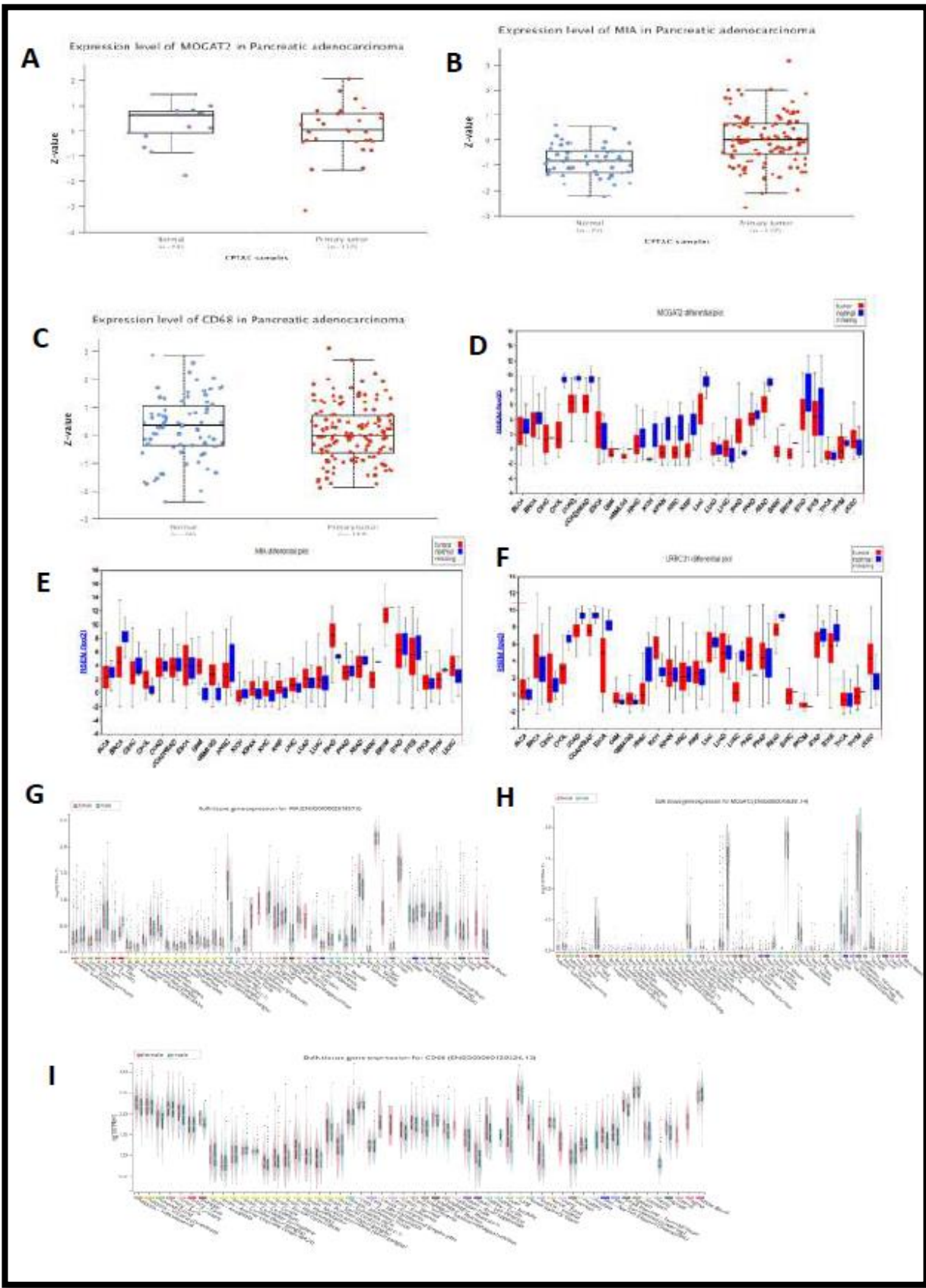


Figure 60: Validation of 5 significant genes across varied platforms:

(A) Protein level expression of MOGAT2 observed across CPTAC database. (B) Protein level expression of MIA observed across CPTAC database. (C) Expression status of CD68 in pancreas normal tissue obtained from

GTEx Database (D) Protein level expression of CD68 observed across CPTAC database. (E) pan-cancer Validation of MOGAT2 showing highest fold change. (F) Expression status of MIA in pancreas normal tissue obtained from GTEx Database. (G) pan-cancer Validation of MIA showing comparatively higher fold change. (H) Expression status of MOGAT2 in pancreas normal tissue obtained from GTEx Database. (I) pan-cancer Validation of LRRC31 showing comparatively higher fold change.

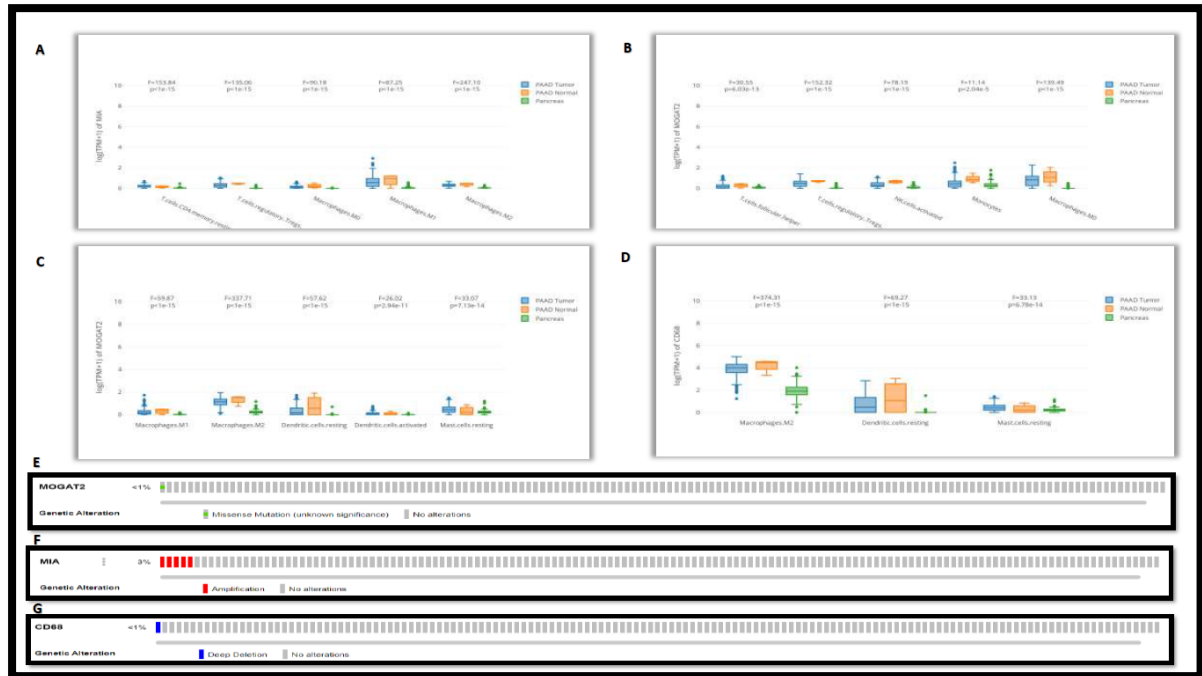


Figure 61: Immune profiling to associate immune dysregulation associated with methylation driven differentially expressed genes

(A) Immune deconvolution of MIA gene associated expression using CIBERSORT. (B) Immune deconvolution of MOGAT2 gene associated expression. (C) Immune deconvolution of MOGAT2 gene associated expression. (D) Immune deconvolution of CD68 gene associated expression. (E-G) Mutation profile of MOGAT2, MIA and CD68.

The immune profile deconvolution was also done using CIBERSORT, xCELL and ESTIMATE was also used to obtain the immune and stromal score (Figure 61). Correlation across 64 cell types were done using xCell. xCell was also used to obtain cell types association with the cell ecosystem based on the expression data, and a heat map was plotted to evaluate the correlation. ESTIMATE obtained scores were used to split the sample into HR and LR followed by plotting their respective KM. Lastly, CIBERSORT was used to identify levels of gene expression corresponding to cell types and were compared against TCGA PC cells vs TCGA Adjacent normal pancreas samples and with GTEx normal samples. MOGAT2 and CD68 have shown significance and concordance with the observation across the varied immune profile experiments done (Figure 62).

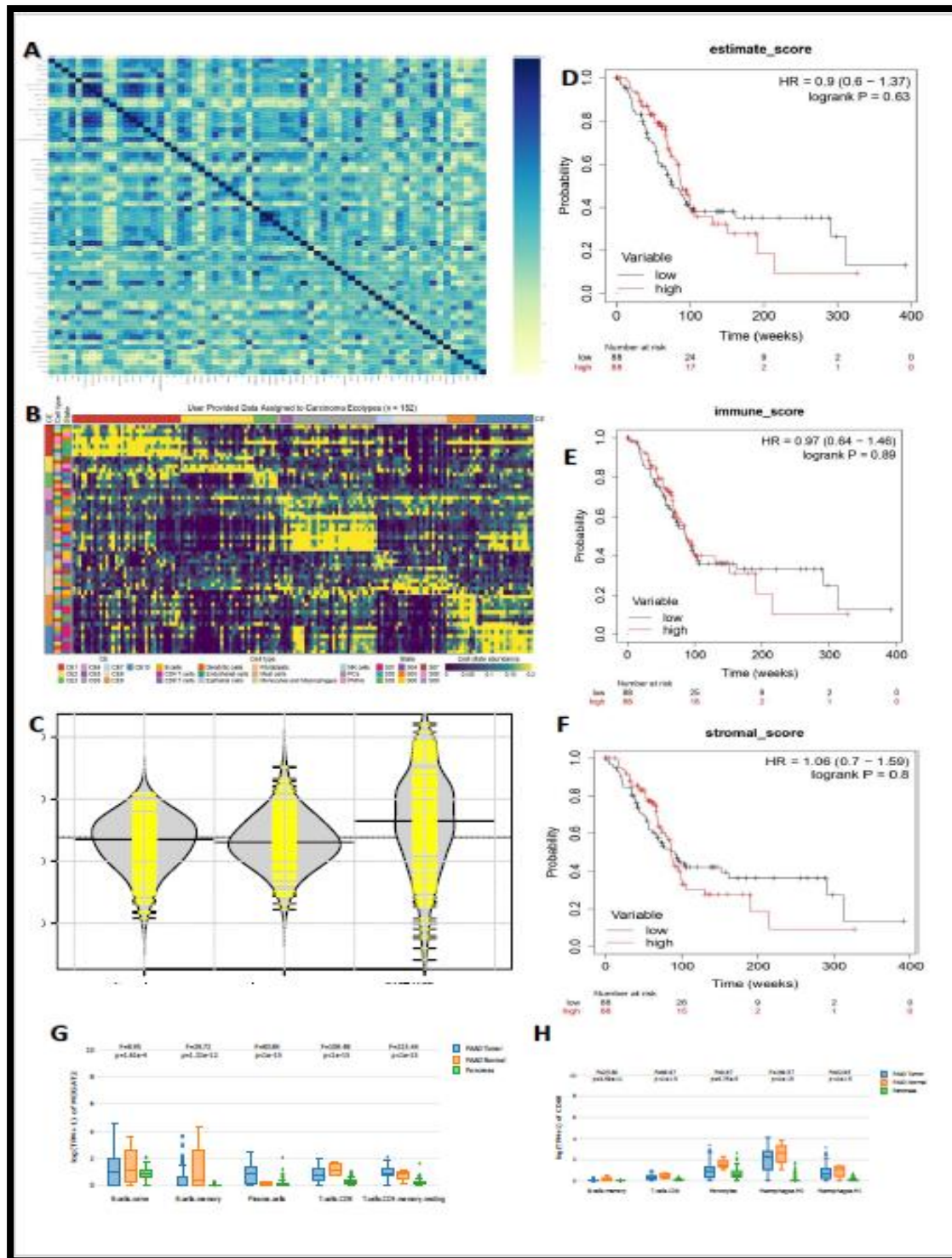


Figure 62: Immune landscape evaluation using multiple cell type on sample pool

(A) Correlation across all 64 cell types using XCell. (B) Cell Type associated ecosystem correlation based on expression data. (C) The optimal cutoff identification for Estimate, Stromal and Immune scores. (D) Kaplan-Meier plot depicting Overall survival for patients with high vs. low Estimate scores. (E) Kaplan-Meier plot depicting Overall survival for patients with high vs. low Immune scores. (F) Kaplan-Meier plot depicting Overall survival for patients with high vs. low Stromal scores. (G) Significant cell type signatures in regards to MOGAT2 gene expression alteration based on TCGA and GTEx database. (H) Significant cell type signatures in regards to CD68 gene expression alteration based on TCGA and GTEx database.

5.4 Targeting of Methylation-regulated Differentially Expressed Genes

5.4.1: Targeting Concurrently Significant Methylation-regulated Differentially Expressed Genes in PAAD

The workflow of the entire study for the identification of Methylation-regulated Differentially Expressed Genes is presented in Figure 63. We initially performed differential expression analysis and differential methylation analysis to select DEGs and DMGs between PAAD and adjacent normal pancreatic tissue in the TCGA-PAAD dataset. With cut-off criteria of $FDR \leq 0.05$ and $|\log_2FC| > 2.0$, a total of 120 DEGs and 181 DMPs were identified. The DEGs and DMGs gene lists obtained were compared using $FDR \leq 0.05$ to generate 69 gene lists, which were found to be common in both lists, therefore showcasing the genes as methylation-regulated differently expressed genes.

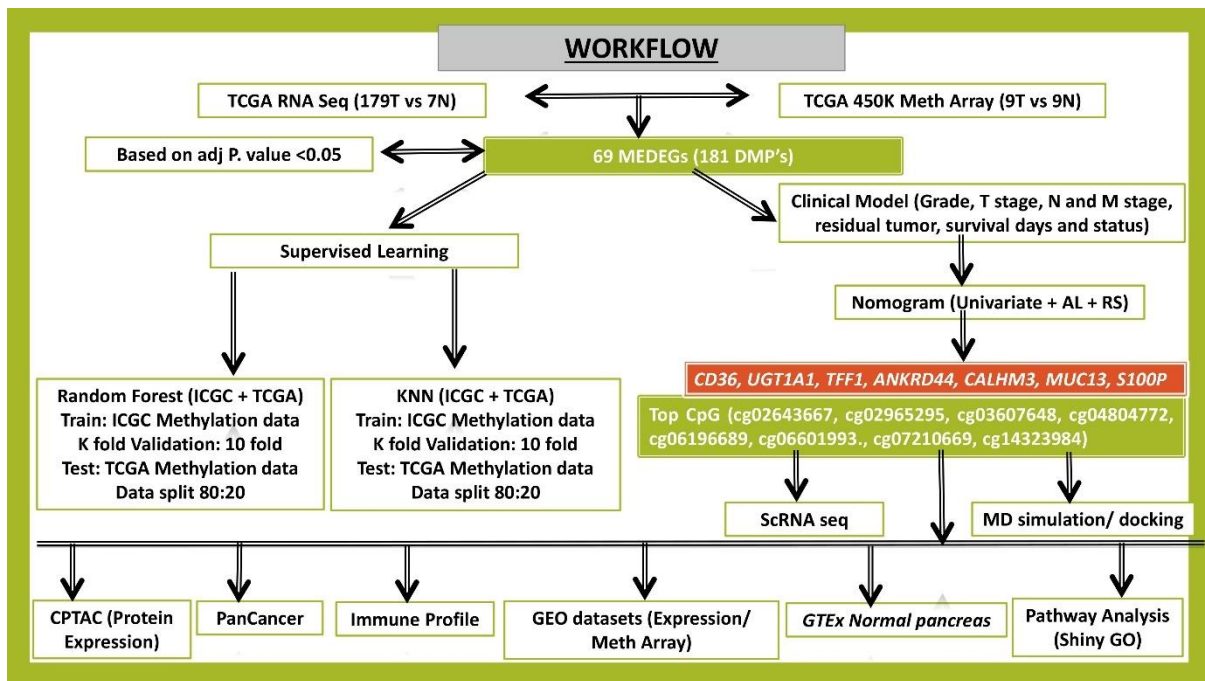


Figure 63: Detailed workflow of the study to establish methylation driven differentially expressed genes as potential pancreatic cancer diagnostic biomarkers.

5.4.2: Calibration and Clinical Application of Lasso-Cox Rregression

Methylation-regulated differently expressed genes obtained were used for in generating the nomogram model (Figure 64A). First, the correlation between each DMS's methylation level and the patient's OS in the training cohort was determined using univariate Cox regression analysis (Figure 64B). Sites associated to prognosis were identified as MEDEGs with p-values less than 0.05. In this case, $\log(\lambda)$ equals 1.275 (Figure 64C). This indicates that the estimated hazard ratio of the covariates in your model is roughly $e^{1.275}$, or 3.5787. Since the hazard ratio is higher than 1, there is a greater chance that the event (CLI Death) will occur (Figure 64D).

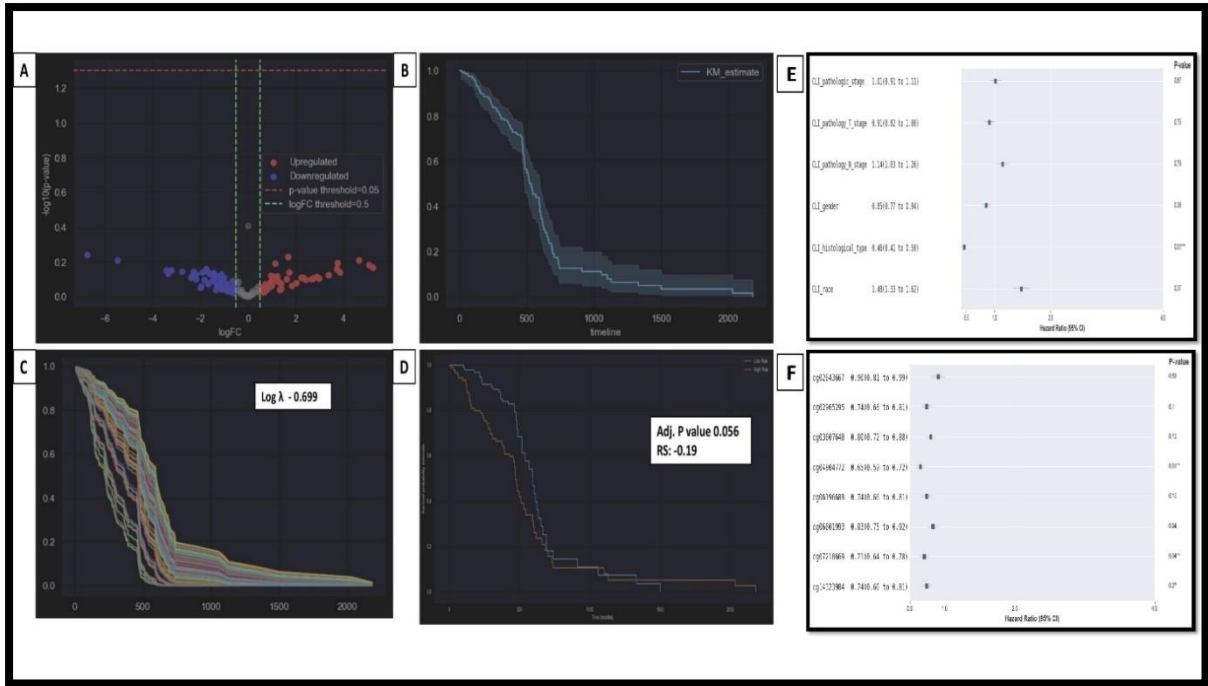


Figure 64: Clinical model development for identification of prognostic biomarkers

(A) Joint distribution of 69 genes based on their regulation pattern using log fold change as axis (B) Cumulative KM estimate using Univariate Cox regression considering TCGA PAAD patient profiles (C) Lambda parameter estimation prior to doing Adaptive LASSO (D) Kaplan-Meier survival analysis post splitting the TCGA patient pool into high and low risk using risk score generated using methylation value as mean point (E) The Figures show each variable along with its associated Hazard Ratio (HR) and p-value. CLI_pathologic_stage: Higher stages are typically associated with worse outcomes. If the HR is significantly > 1, it confirms this expectation. CLI_pathology_T_stage, CLI_pathology_N_stage: These are components of the TNM staging system. Higher T and N stages usually indicate more advanced disease and worse prognosis. CLI_gender: This shows if there's a significant difference in outcomes between genders. CLI_histological_type: Different histological types may have different prognoses. CLI_race: This indicates if there are significant differences in outcomes among different racial groups. (F) The top 8 CpGs obtained from Nomogram along with HR and P value is showcased. $P < 0.05$ is typically considered statistically significant.

To summarize, the degree of regularization performed to your Cox regression model is reflected in our optimal lambda value of 0.6999. It shows a decent trade-off between goodness of fit and model complexity, with some coefficients perhaps being reduced to zero for variable selection. Heatmap depicting the outcome is shown in Figure 65A. A nomogram was created for the TCGA-PAAD cohort to predict the 1- and 2-year OS, taking into account the clinical risk factors and risk score of PAAD patients as assessed by multivariate Cox proportional-hazards regression (Figure 64E, F and 65B). This was done in order to convert the prognostic value of the MeDEG based signature into a clinical application. The concordance index (C-index), a measure of discrimination, was computed to assess the

nomogram's discriminatory power. To compare the observed and anticipated probabilities for the nomogram, calibration plots were created.

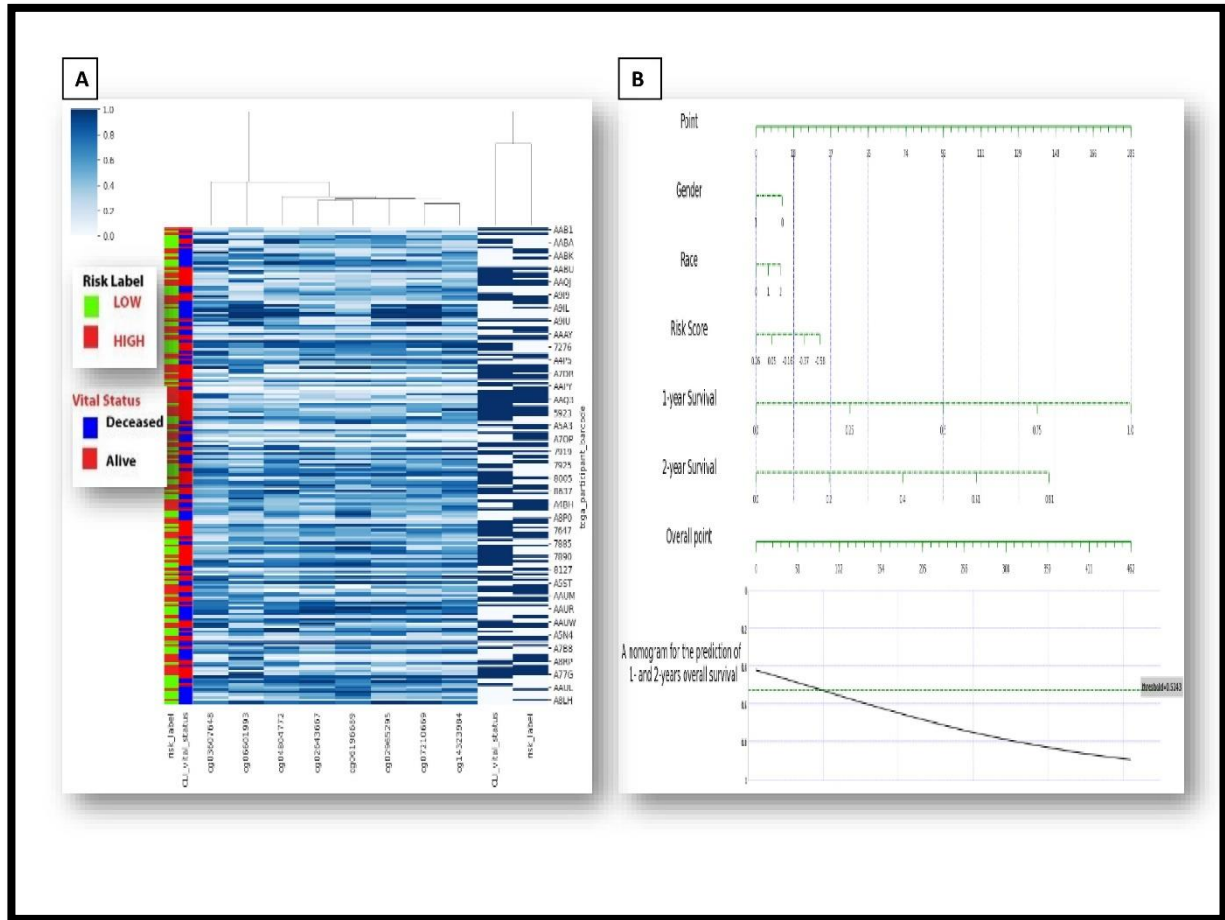


Figure 65: Clinical visualization of the 7 top gene set

(A) Methylation based heatmap of the 7 top CpG sites along with TCGA patient labels on right hand side. The risk score and the vital status of the TCGA patients are also depicted along with its correlation in the map (B) Nomogram for the prediction of 1- and 2-years overall survival in PAAD patients along with the risk score obtained.

To find the DNA methylation sites connected to CLI_vital_status, LASSO regression was used. The accuracy of the model was 0.5676. The LASSO algorithm identified the following DNA methylation sites: cg09060908, cg20340346, cg26683025, cg02965295. An Adaptive LASSO technique was used to further refine the model, selecting 8 variables and eliminating the remaining 125. cg02643667 (*TFF1*), cg02965295 (*ANKRD44*), cg03607648 (*UGT1A1*), cg04804772 (*CALHM3*), cg06196689 (*MUC13*), cg06601993 (*CD36*), cg07210669 (*S100P*), and cg14323984 (*S100P*) are the DNA methylation sites that were chosen (Table 24). At 1.0, the Adaptive LASSO model's accuracy was flawless. This suggests that the model has a strong prediction ability and offers insightful information about the connection between DNA methylation patterns and CLI_vital_status.

5.4.3: Implementation of Supervised Learning Algorithms Using Methylation-regulated Differentially Expressed Genes

With a precision of 0.815, or positive predictive value, 81.5% of the cases reported as having PanCa were accurately classified as positive cases. In practical terms, this means that the model predicts good outcomes, such the identification of PanCa, with an accuracy rate of approximately 81.5%. Recall, which is 0.788, is often referred to as sensitivity or true positive rate. This indicates that 78.8% of all real positive instances of PanCa were successfully identified and gathered by our model (Figure 66A, B). This Figure illustrates how well the classifier can detect PanCa cases, indicating that it is sensitive to the disease's presence (Figure 66C). The harmonic mean of recall and precision, called the F1-score, is 0.8012. This suggests a performance that strikes a balance between recall and precision. This metric includes erroneous positives as well as false negatives, providing a thorough evaluation of the model's ability to accurately predict PanCa. A strong balance between correctly identifying positive situations and minimizing false positives and false negatives is shown by the high F1-score. Our Random Forest classifier performs well in PanCa prediction, exhibiting excellent recall, precision, and a balanced F1-score. The outcomes show how effective the model is in accurately identifying positive cases and lowering the probability of inaccurate forecasts, both of which are critical when it comes to diagnosing PanCa and forecasting its course. The test data's variance was found to be 0.1091, ensuring that the volatility in the test data won't skew our results. Post this we used the same dataset for KNN model construction and we obtain K optimum to be 9 and accuracy to be 0.92 (Figure 66D)

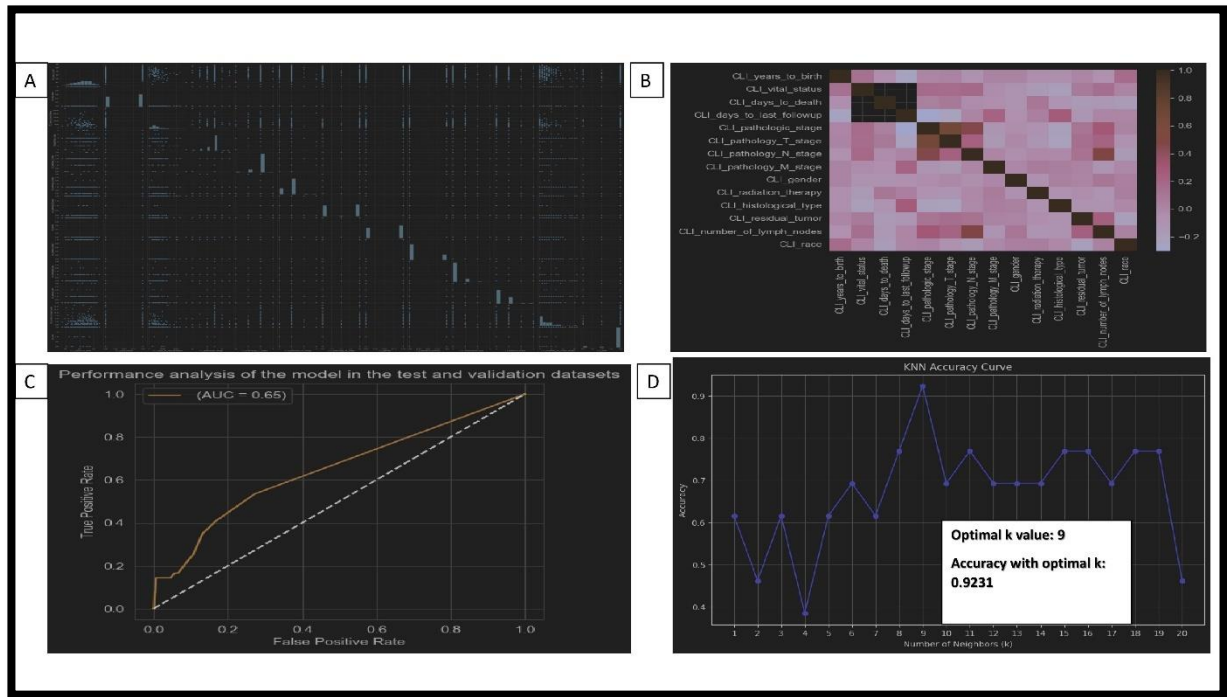


Figure 66: Prediction model development using supervised learning algorithms

(A) Pairwise correlations between the random forest model's key predictors prior to proceeding with classifier (B) Multivariate correlation amongst key variables involved in the supervised learning process and also used in nomogram model (C) Area Under Curve (AUC) of the random forest model post training and validation of the model (AUC=0.65) using ICGC dataset (D) KNN model accuracy (0.9231) curve and optimal K obtained in this process ($k=9$).

5.4.4: Validation of 7 Signatures Using Global Obtained Datasets

The top 7 genes along with their corresponding 8 CpGs were validated first using 2 450K datasets. GSE74071 reported *S100P* to be most significant whereas GSE49149 showed all 7 genes to be significant ($p < 0.05$) (Table 25). RNA seq data obtained from GSE211398 have shown *S100P* and *MUC13* to be most significant (Table 26). Following this we further evaluated our data using microarray datasets of PanCa. GSE183795 showed *S100P*, *MUC13*, *CD36*, *TFF1*, *CALHM3* to significantly dysregulated. GSE28735 validated *S100P*, *TFF1*, *CD36*, *UT1A1* to significantly dysregulated. GSE15471 finally demonstrated *UGT1A1*, *CD36*, *S100P*, *CALHM3*, *ANKRD44*, *TFF1* to significantly dysregulated. Consistently, we revalidated our results using data generated by Mayo clinic and *UGT1A1*, *CD36*, *S100P*, *MUC13*, and *TFF1* were seen to significantly dysregulated and in similar concordance in terms of gene expression (Table 27).

5.4.5: *TFF1* and *S100P* Showcasing Transition from Pre-cancerous to Metastatic Stage

Tu et al. 2021 did a network-based analysis which offers gene networks that are progressively activated in the course of CP, PanCa, and metastatic PanCa. It might aid in the logical development of anti-cancer medications. They obtained RNA seq DEG data containing CP, PanCa and metastatic PanCa samples amongst which all showed *TFF1* to be most DEG along with *S100P* (Table 26) (*Tu et al. 2021*).

5.4.6: Evaluation the Methylation-regulated Differentially Expressed Genes Role in Translational Way

We then cross evaluated the 7 genes based on survival status using expression data and *CD36* was found to be most significant ($p < 0.04$) (Figure 67A). This was followed by *UGT1A1* based on survival status (Figure 67B). The rest 6 cg IDs mapped genes were further checked for survival using TCGA dataset using expression data (Figure 68 A-E) and Pearson correlation was also done to check correlation between every gene methylation and expression values (Figure 68 F, G and 69A-E). Following this we also checked joint survival of 2 most survival based relevant genes using expression data on patients (percentile > 50%) from TCGA dataset and obtained $p = 0.94$ (Figure 68F). Based on this outcome we analyzed a pan-cancer profiling of *CD36* which showed similar observation across other cancer types (Figure 67C). We then subsequently checked the protein level status of *CD36* and *UGT1A1* based on CPTAC dataset and we found both of them have significant Z-values and the expression status is in concordance with our previous observations (Figure 67D-E). Finally, we checked *UGT1A1* pan-cancer status for further validation and got concordance with our observation (Figure 67F). Disease free survival was also visualized obtaining similar results (Figure 67G-H).

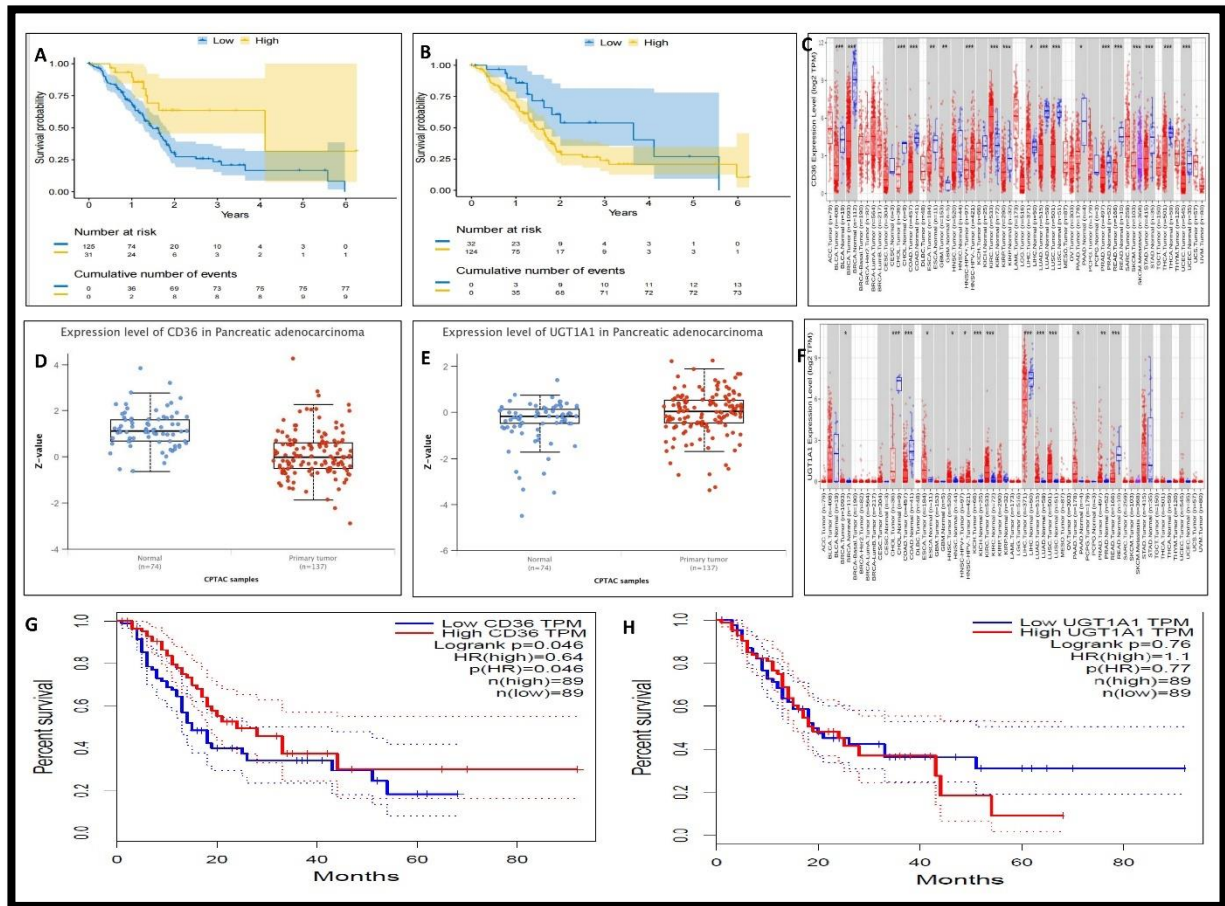


Figure 67: CD36 and UGT1A1 clinical relevance various cancer types

(A) CD36 gene-based survival analysis post splitting the data into high risk and low risk by utilising the expression data (B) UGT1A1 gene-based survival analysis post splitting the data into high risk and low risk by utilising the expression data (C) CD36 expression (log2 TPM) based pan cancer profiling (D) CD36 protein level expression across CPTAC dataset (E) UGT1A1 protein level expression across CPTAC dataset (F) UGT1A1 expression (log2 TPM) based pan cancer profiling. (G) CD36 based overall survival post splitting sample pool (TCGA PAAD samples) into HR and LR based on median expression value. (H) UGT1A1 based overall survival post splitting sample pool (TCGA PAAD samples) into HR and LR based on median expression value.

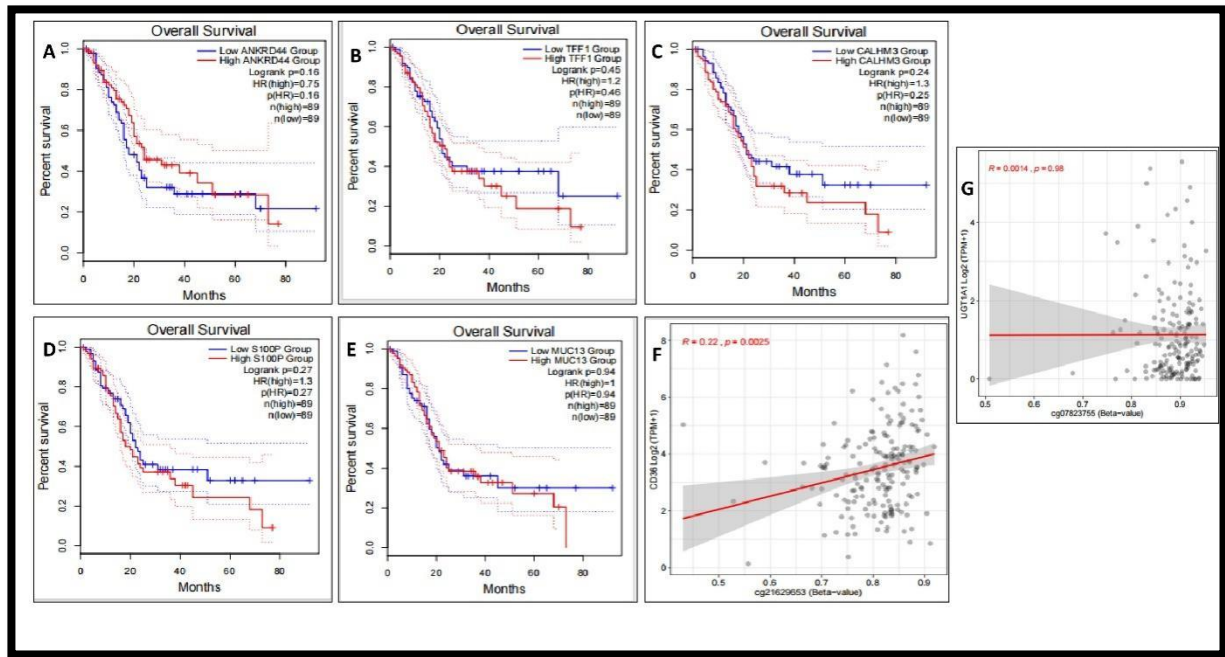


Figure 68: Overall survival and correlation analysis of nomogram identified gene sets

(A-E) Overall survival deduced based on expression derived risk groups concerning genes ANKRD44, TFF1, CALHM3, S100P and MUC13 (F-G) Correlation using genes expression value and methylation beta value of the gene specific probe for CD36 and UGT1A1.

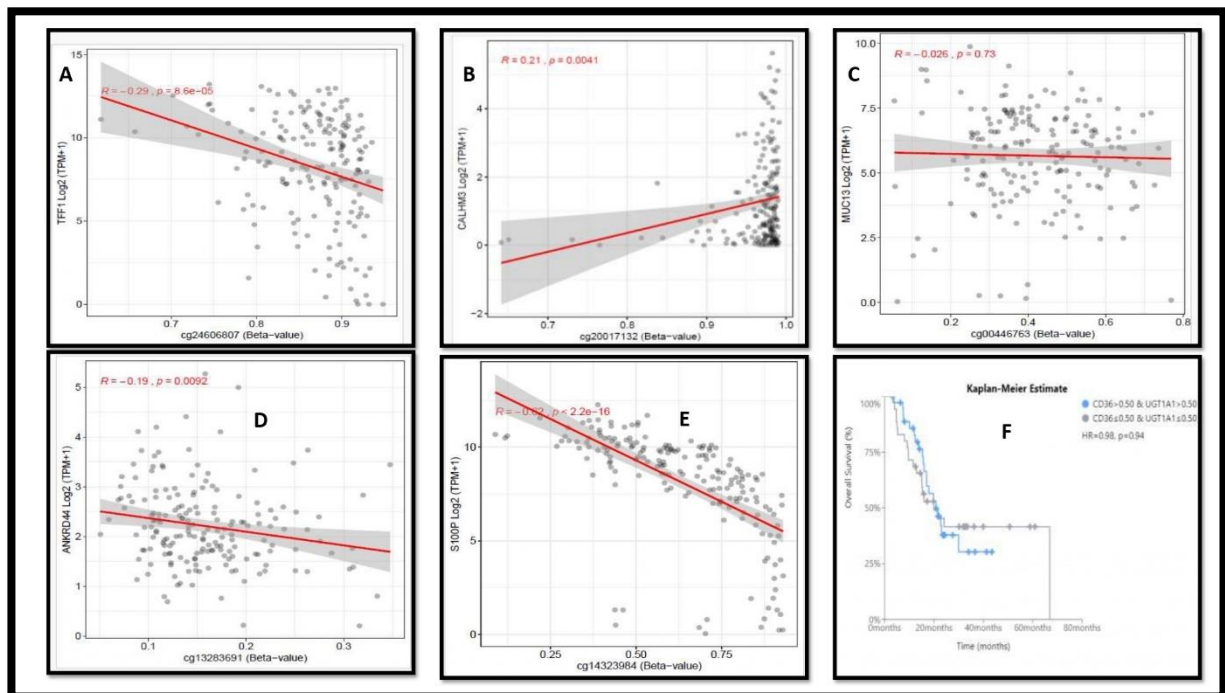


Figure 69: Correlation analysis of nomogram identified gene sets along with joint OS

(A-E) Correlation using genes expression value and methylation beta value of the gene specific probe for ANKRD44, TFF1, CALHM3, S100P and MUC13 (F) Double correlation using both CD36 and UGT1A1 ($p < 0.94$) involving patients with presence of both genes >50%.

We further checked the protein level status of 4 other genes excluding *UGT1A1* as it was not in CPTAC database (*ANKRD44*, *MUC13*, *TFF1*, and *S100P*) and were found to showcase concordant outcomes corresponding to their expression and methylation status (Figure 70A-D).

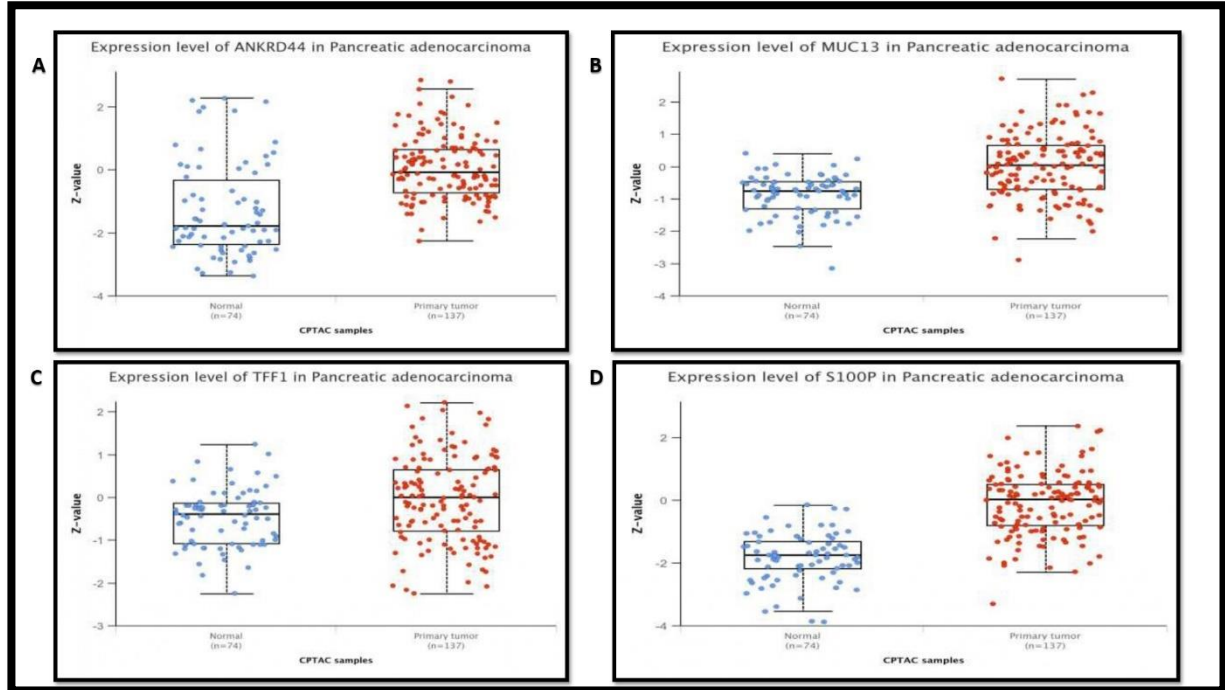


Figure 70: Protein level expression of 4 gene sets

(A-D) CPTAC based protein expression profile based on Z -scores target for genes namely *ANKRD44*, *TFF1*, *S100P* and *MUC13*.

We also generated pan-cancer profiling of remaining 5 genes (*ANKRD44*, *MUC13*, *TFF1*, *S100P* and *CALHM3*) and the signatures have shown significant outcomes when compared in a tumor vs normal status (Figure 71A-E).

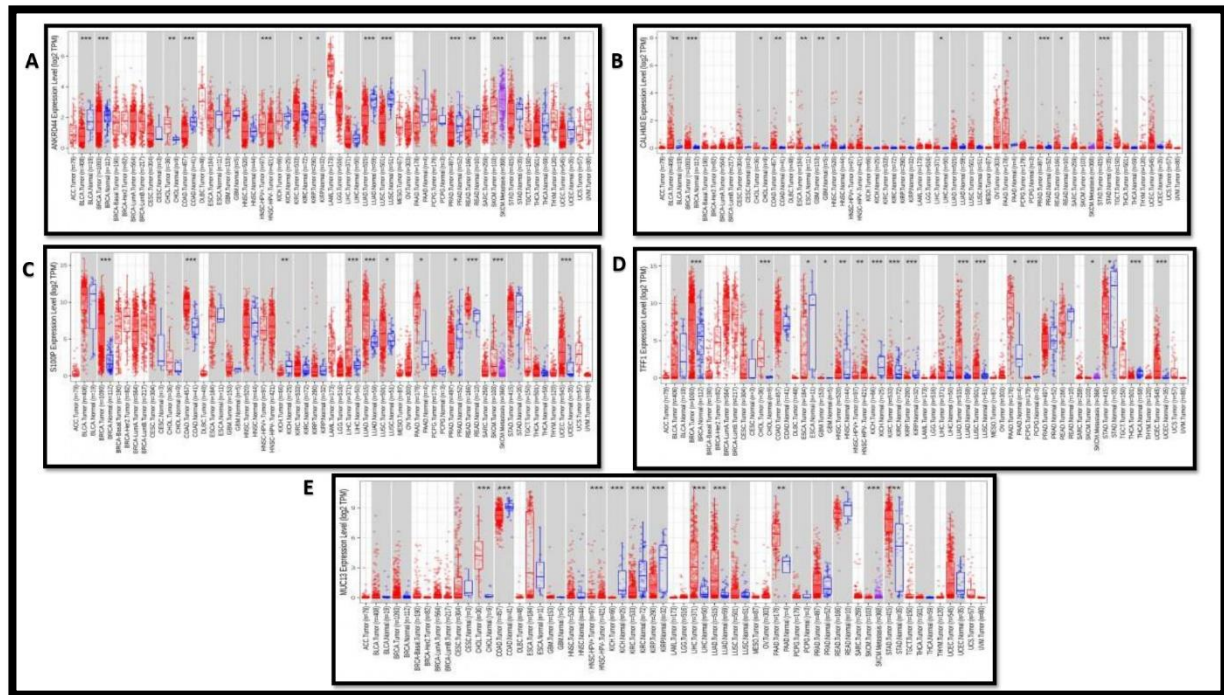


Figure 71: Pan-cancer profiling of nomogram identified gene sets

(A-E) pan-cancer analysis of *ANKRD44*, *TFF1*, *CALHM3*, *S100P* and *MUC13*.

We also checked mutational profile of 5 genes excluding *UGT1A1* as it didn't had a mutational significance as per Cbioportal concerning TCGA dataset (PAAD) and the result showed presence of mutational signatures like amplification and deep deletion (Figure 72).

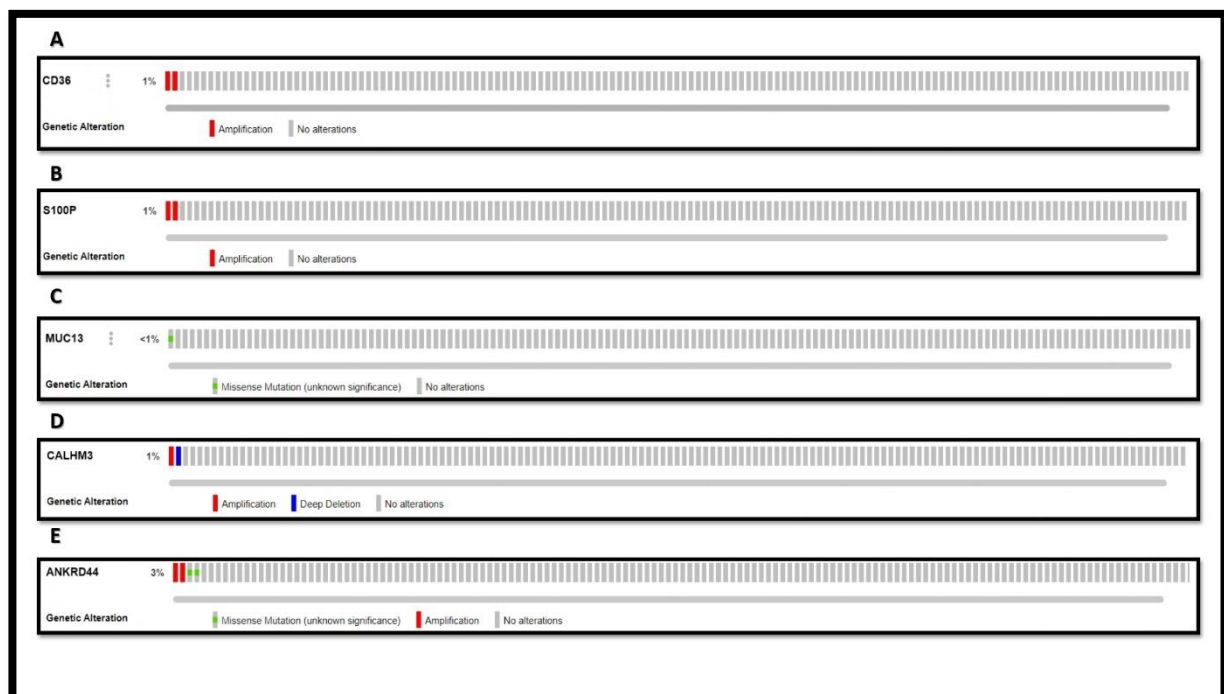


Figure 72: Mutational profiling of nomogram identified gene sets

C-bioportal based mutational profile (percentage based) of genes namely (A) *ANKRD44*, (B) *CD36*, (C) *CALHM3*, (D) *S100P* and (E) *MUC13*.

We also checked the interaction of the 7 genes with biological processes using gene ontology platforms with integration of path DB gene lists. We obtained protein-protein interaction data which highlighted mostly on pathways concerning gastrointestinal system which validates the targeted function of our markers (Figure 73A). Gene ontology biological processes highlighted the significance of terms such “Maintenance of Gastrointestinal Epithelium”, “Digestive System Homeostasis” (Figure 73B). Gene Ontology Molecular processes illustrated similar outcomes as they highlighted fatty acid and lipoprotein-based systems (Figure 73C). Finally, we evaluated our 7 genes using an enrichment chart for the cumulative outcome observation (Figure 73D).

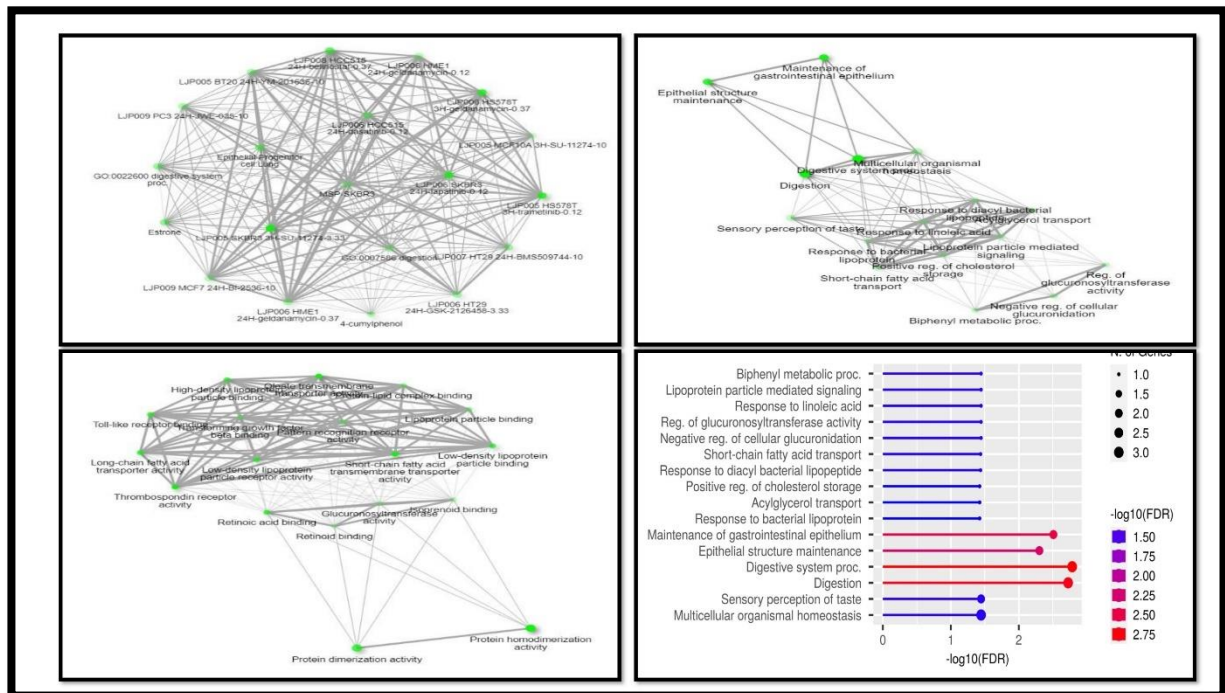


Figure 73: Pathway analysis using seven genes of interest

(A) It shows mapping of 7 genes of interest based on all gene sets available in pathway DB and GO digestion along with few other were significantly enriched. (B) GO BP based PPI were plotted which showed enrichment of “Maintenance of GI epithelium”, “Digestive system” (C) GO MF based PPI showed HDL and LDL particle binding, fatty acid transporter enrichment. (D) Enrichment chart obtained using 7 genes of interest based on -log₁₀ FDR concerning all biological processes.

5.4.7: Immune Profile Deconvolution Concerning CD36 as Top Clinically Relevant Signature.

Since CD36 was found to universally significant across all omics we checked its immune profile to correlate our observation with immune cells expression status. We generated a pan-cancer profile across this immune status deconvolution process. CD36 showed correlation with T Cells CD8+, Treggs, B cell, T Cell CD4+, Cancer associated fibroblasts and

hematopoietic stem cells. Few sub-types of T Cells and B Cells were also seen to be correlated. We undertake this process concerning immune profile multiple datasets namely EPIC, XCELL, TIMER, QUANTISEQ, TIDE, MCP COUNTER, CIBERSORT and CIBERSORT ABS (Figure 74A-D).

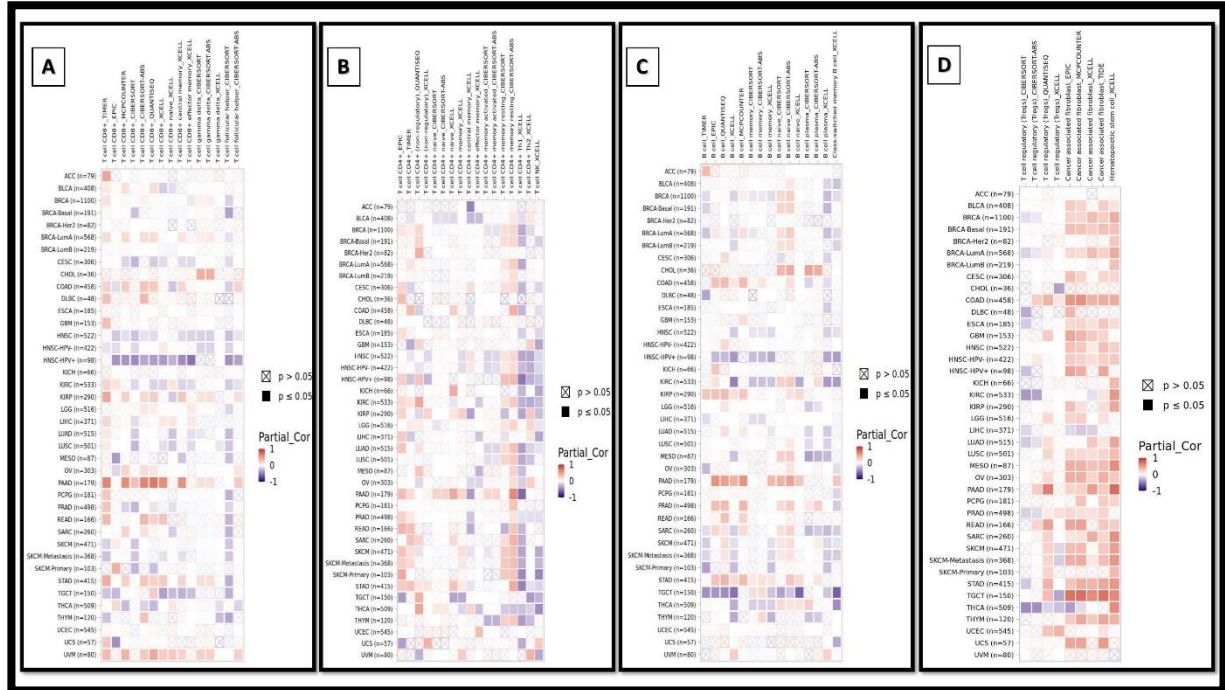


Figure 74: Immune cell types dysregulation based on *CD36* expression

(A-D) *CD36* tumor cell immune profiling considers multiple cancer types and multiple immune deconvolution datasets showcasing the correlation amongst cancer type and datasets used. Various immune cell types such as T cell and B cell subtypes are considered along with cancer associated fibroblasts and hematopoietic stem cells.

5.4.8: Multistage Evaluation of Methylation-regulated Differentially Expressed Genes

All sub-stage of TCGA PanCa patient pool were evaluated using targeted gene expression. *TFF1* showed dysregulation across all stages and substages ($p < 0.0007$). *S100P* also showed dysregulation across stages and substages ($p < 0.000$). *CD36* which is the most clinically relevant gene showed dysregulation across stages and substages ($p < 0.06$). *UGT1A1* showed dysregulation across stages and substages ($p < 0.02$). *CALHM3* and *ANKRD44* showed dysregulation across stages and substages ($p < 0.2$ and $p < 0.9$) (Figure 75 A-F).

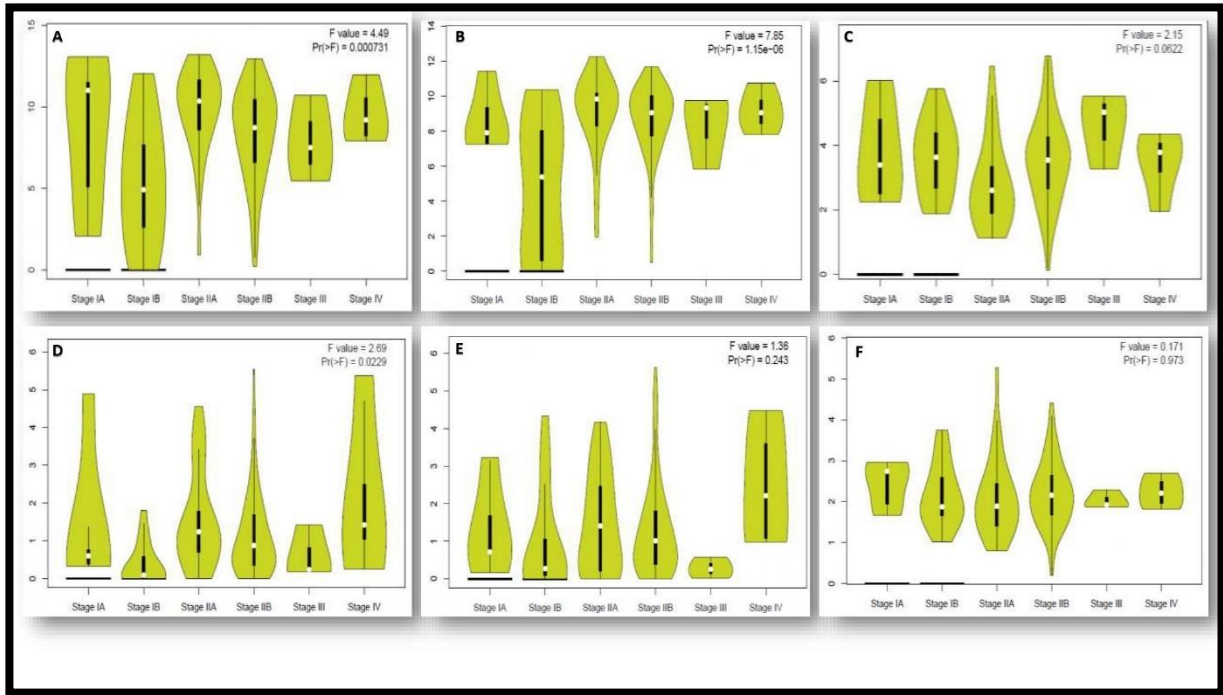


Figure 75: Stage specific targeted gene expression across TCGA pancreatic cancer patient pool

(A) *TFF1* showed dysregulation across stages and substages ($p < 0.0007$) (B) *S100P* showed dysregulation across stages and substages ($p < 0.000$) (C) *CD36* showed dysregulation across stages and substages ($p < 0.06$) (D) *UGT1A1* showed dysregulation across stages and substages ($p < 0.02$) (E) *CALHM3* showed dysregulation across stages and substages ($p < 0.2$) (F) *ANKRD44* showed dysregulation across stages and substages ($p < 0.9$)

5.4.9: Functional Enrichment of Methylation-regulated Differentially Expressed Genes

Linked Omic was utilized to examine the pathways and functional aspects of MEDEGs in a sample size of 178 patients. A heatmap shows the majority of correlated, positive, and negative genes that are linked to *CD36* (Figure 76 A). Using the KEGG pathways category and the GSEA method, enrichment studies were carried out for the functional and pathway enrichment of the correlated genes with *CD36* in the Link Interpreter submodule of Linked Omics. As exhibited in Figure 76 B, C, *CD36* have positive and negative involvement in multiple processes. The enrichment plots of some important enriched functional and pathways are demonstrated in Figure 76D, E. Similar sets of results were recorded for *TFF1* (Figure 77), *UGT1A1* (Figure 78), and *S100P* (Figure 79).

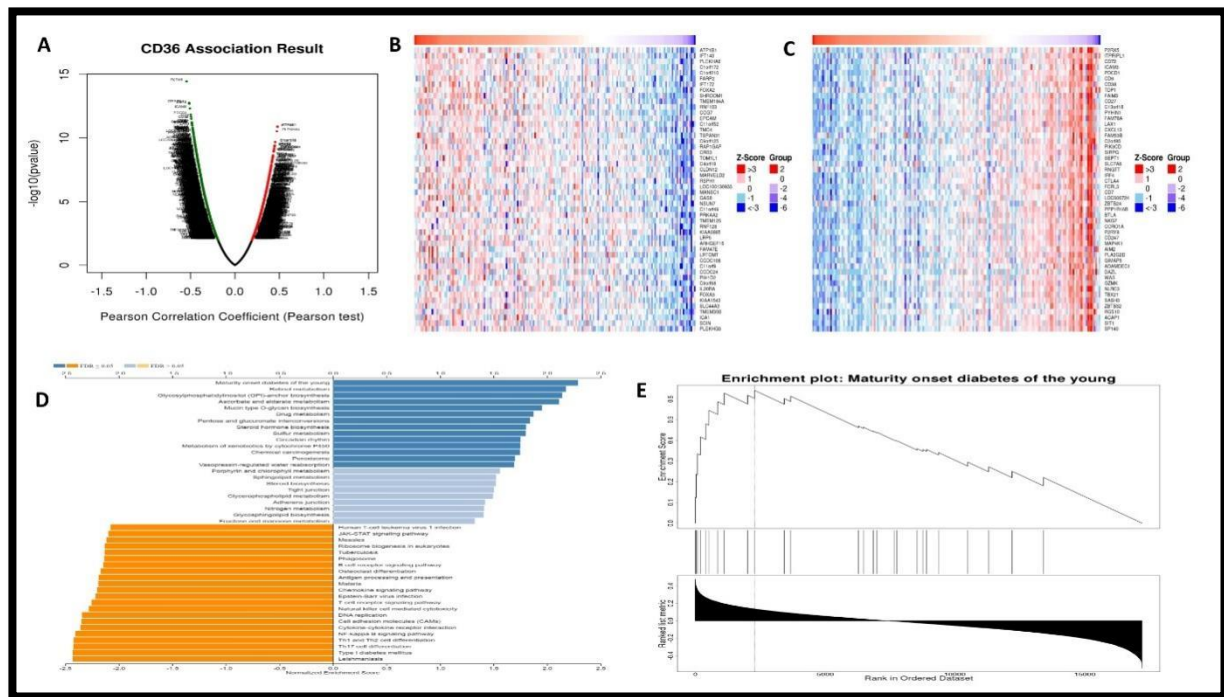


Figure 76: Functional enrichment of CD36

(A) Volcanic graph of CD36 representation positive (red) and negative (green) correlated genes. (B) Heatmap of CD36 associated positive correlated genes. (C) Heatmap of CD36 associated negative correlated genes. (D) Functional and pathways enrichment analysis of CD36 associated positive (blue) and negative (yellow) correlated genes. (E) Enrichment plot of important enriched functional pathways.

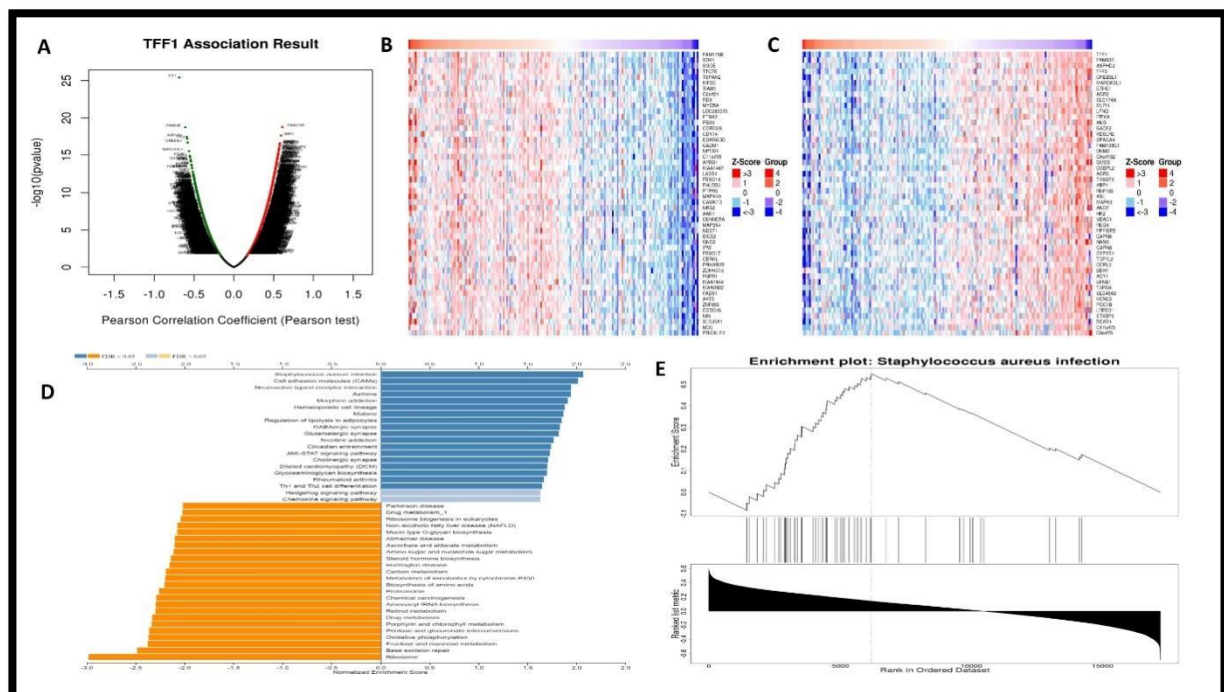


Figure 77: Functional enrichment of TFF1

A) Volcanic graph of TFF1 representing positive (red) and negative (green) correlated genes. B) Heatmap of TFF1 associated positive correlated genes. C) Heatmap of TFF1 associated negative correlated genes. D) Functional and pathways enrichment analysis of TFF1 associated positive (blue) and negative (yellow) correlated genes. E) Enrichment plot of important enriched functional pathways.

correlated genes. E) Enrichment plot of important enriched functional pathways.

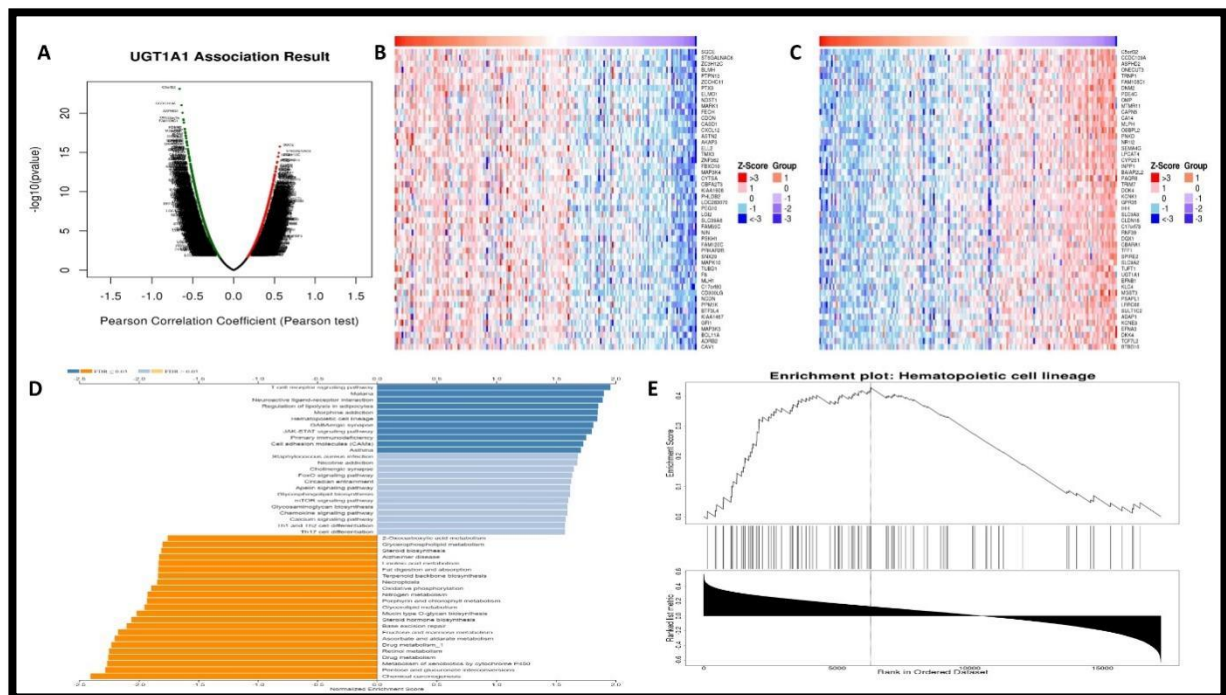


Figure 78: Functional enrichment of UGT1A1

A) Volcanic graph of UGT1A1 representing positive (red) and negative (green) correlated genes. B) Heatmap of UGT1A1 associated positive correlated genes. C) Heatmap of UGT1A1 associated negative correlated genes. D) Functional and pathways enrichment analysis of UGT1A1 associated positive (blue) and negative (yellow) correlated genes. E) Enrichment plot of important enriched functional pathways.

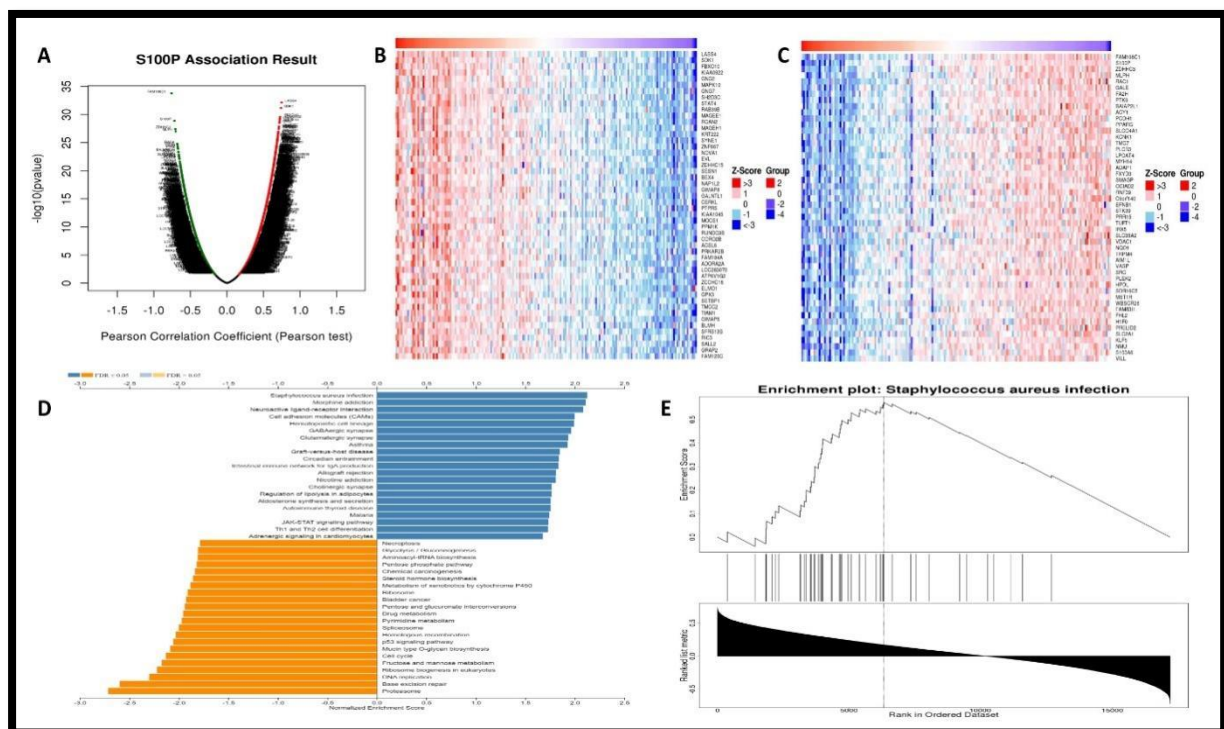


Figure 79: Functional enrichment of S100P

A) Volcanic graph of S100P representing positive (red) and negative (green) correlated genes. B) Heatmap of

S100P associated positive correlated genes. C) Heatmap of *S100P* associated negative correlated genes. D) Functional and pathways enrichment analysis of *S100P* associated positive (blue) and negative (yellow) correlated genes. E) Enrichment plot of important enriched functional pathways.

5.4.10: Evaluating Prognostic Implications of *CD36* and *UGT1A1*

We checked the role of *CD36* in human protein atlas to identify its implications as prognostic marker which was found to be significant ($p < 0.001$) (Figure 80A). Then we have checked *CD36* expression in TCGA tumor vs TCGA normal data ($p < 0.05$) and then in TCGA tumor vs GTEx normal data (Figure 80B, C). Then we redistributed our expression-based data concerning *CD36* ($p < 0.04$) and *UGT1A1* ($p < 0.11$) into male and female to observe the distribution while doing a Kaplan- Meier (Figure 80D, E). Finally, we checked the role of *CD36* and *UGT1A1* in normal tissues using bulk RNA seq based expression data across all 32 tissue types which showed results in concordance with our observation i.e. *CD36* Overexpression in normal and *UGT1A1* lower expression in normal (Figure 81A, B).

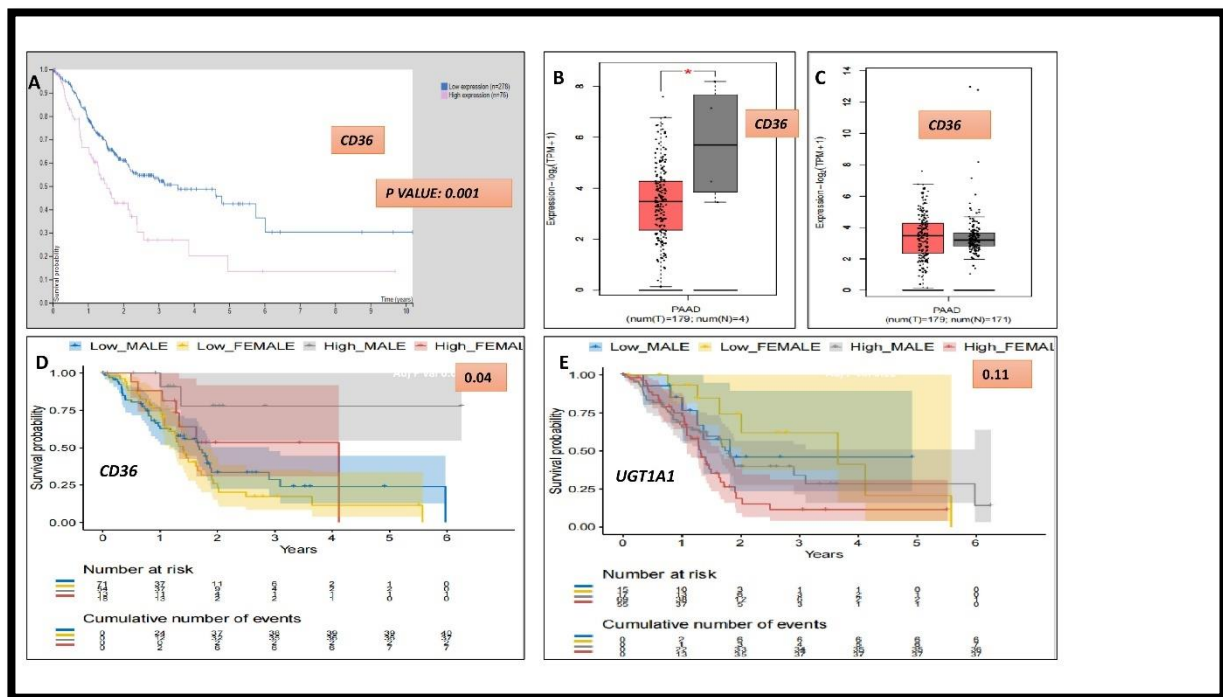


Figure 80: Survival and expression evaluation of *CD36* and *UGT1A1*

(A) Showcasing prognostic role of *CD36* in stomach cancers by using human protein atlas (B) *CD36* shown to be significantly differentially expressed when TCGA tumor samples were compared with TCGA normal pancreas tissues (C) *CD36* shown to be differentially expressed when TCGA tumor samples are compared with GTEx normal pancreas tissues (D) *CD36* expression data based survival curve post risk grouping followed by gender clustering (E) *UGT1A1* expression data based survival curve post risk grouping followed by gender clustering.

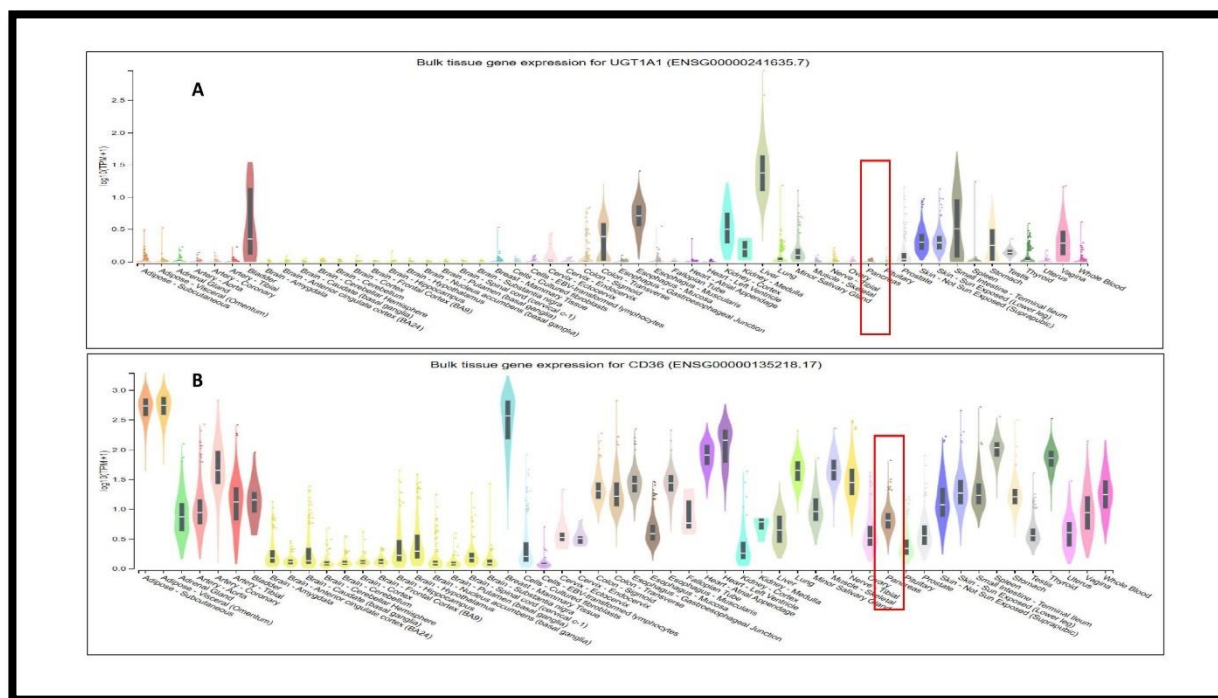


Figure 81: Targeting tissue specific expression using normal tissue datasets

(A) Bulk tissue gene expression of *UGT1A1* across all 32 tissue types and concordance result in Pancreas normal tissues (B) Bulk tissue gene expression of *CD36* across all 32 tissue types and concordance result in Pancreas normal tissues

5.4.11: *CD36*, *TFF1*, *S100P*, *MUC13* and *UGT1A1* Expression in Various PanCa Cell Lines

The CCLE DepMap portal showed that median expression of *CD36*, *TFF1*, *S100P*, *MUC13* and *UGT1A1* across 50 PanCa cell lines. But, on examining its expression in individual PanCa cell line in CCLE DepMap portal it was found that all 50 PanCa cell lines don't have expression consistent expression of these genes but the median expression and pattern of expression is in consistent without prior observations hence stating data concordance (Figure 82).

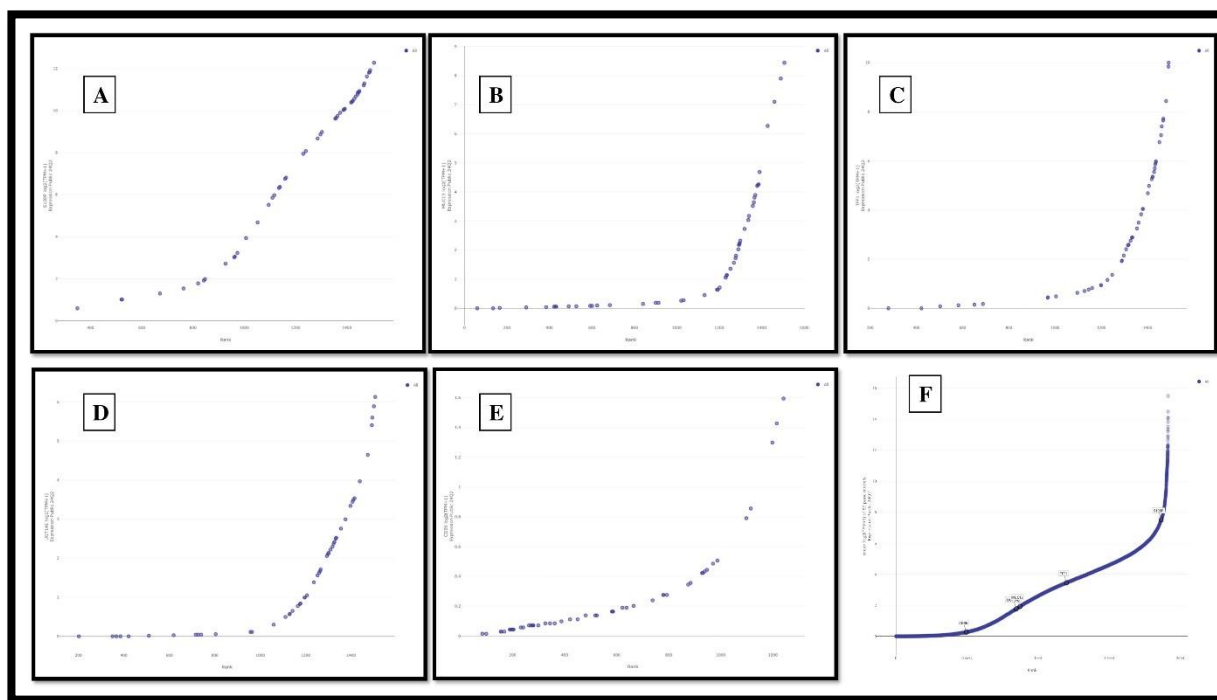


Figure 82: Gene expression of methylation driven differentially expressed genes in 50 pancreatic cancer cell lines

(A) Gene expression of *S100P* in 50 PanCa cell lines in Dep Map portal (B) Gene expression of *MUC13* in 50 PanCa cell lines in Dep Map portal (C) Gene expression of *TFF1* in 50 PanCa cell lines in Dep Map portal (D) Gene expression of *UGT1A1* in 50 PanCa cell lines in Dep Map portal (E) Gene expression of *CD36* in 50 PanCa cell lines in Dep Map portal (F) Median gene expression of *CD36*, *UGT1A1*, *TFF1*, *MUC13* and *S100P* across 50 PanCa cell lines in Dep Map portal

5.4.12: Drug Prediction on *TFF1*, *MUC13* and *S100P*

Prediction on *Tff1*, *Muc13*, and *S100p* was performed in GDSC1/2, CTRP, and PRISM databases but showed no significant results illustrating concordance with our observations. No relevant data was obtained from CMap dataset while targeting from small molecules.

5.4.13: Transcription Factor- mRNA Relationship for *TFF1*, *MUC13* and *S100P*

MUC13 was associated with TFs namely *ATF3* and *SPI1*. *SPI1* was also seen to be upstream of *MUC13* when evaluated using ChEA3. *ATF4* and *JUN* were found to be upstream of *TFF1* and *S100P* when evaluated using multiple datasets. This showcases that these 3 genes have multiple TF upstream modulating their expression. The TFs obtained upstream for *TFF1*, *MUC13*, and *S100P* according to the four respective datasets.



DISCUSSION



SEPTEMBER 16, 2024

PHD THESIS

Department of Life Sciences and Biotechnology, Jadavpur University

6.0: Discussion

Globally, PanCa, especially ductal adenocarcinoma, is one of the most prevalent cancer types. Late diagnosis, treatment failure, and loco-regional recurrences all contribute to its poor prognosis. Along with genetic alterations, studies in the past decade showed that epigenetic alterations contribute to PDAC development and pathogenesis (*Henriksen et al. 2021, Biankin et al. 2012, and Bararia et al. 2023*). Recent findings have shown abnormal DNA methylation can mark the spectrum of cancer progression, including the precancerous lesions, thus serving as biomarkers for diagnosis and prognosis (*Hansen et al. 2011, and Fujisawa et al. 2020*). Thus, a specific abnormal methylation profile can serve as a biomarker for that cancer type. Since methylation marks can be reversed to an unmethylated state, epigenetic marks can act as a lucrative therapeutic target. The effects of methylation on gene expression, can vary across different ethnic populations with varied risk factors (*Chatterjee et al. 2022*). Epigenetic and genetic signatures of PDAC partially vary across the globe based on expression frequency and alternated driver behaviors. According to our previous study, *Saha et al. 2020* explained *KRAS* hotspot mutation was observed in low frequency in the Indian PDAC population whereas other parts of the world, like USA/Canada/European countries and Australia based studies documented *KRAS* as a high-frequency gene around 90% in the same disease. This partially variable profile may be contributed by the patient pool led by combined factors such as ethnicity, environmental, lifestyle and occupational risk factors (*Saha et al. 2020*). As already reported in our previous findings in *Saha et al.* it was observed that demographic effects are in play while screening genomic landscape similar reports were seen in our methylome data. In our current study, correlation between sex, smoking and alcohol consumption has been documented which is specific to our demographic and ethnic beliefs. To gain a clear understanding of the epigenetic landscapes of PanCa, our study strongly suggests a novel architecture of epigenetic landmarks and insight into potential epigenetic outcome on PDAC in the Indian patient population. In our 450K methylome study, it has been identified that DNA methylation changes occurred in 91 gDMs associated with PDAC in the Indian population. Among them, differential methylation in 47gDMs in PDAC concordant with that in the TCGA PAAD cohort. The differential methylation marks were associated with progressive cancer stages and prognosis, both in our cohort and the TCGA cohort. Thus, our results showed both common and unique differential methylation marks in PDAC patients in the Indian cohort (*Saha et al. 2020, and Chatterjee et al. 2022*).

Among the 156 DMPs, 37.2% (n=58) were hypermethylated and 62.8% (n=98) were hypomethylated among cancer samples. Among the previously identified 7832 DMPs in the TCGA cohort, 68 DMPs overlap with DMPs, identified from our data. The differences in beta values between the paired tumor and normal samples at these 68 DMPs in the TCGA data (for 9 PAAD patients) showed a positive correlation with our study (correlation coefficient = 0.98). Excluding the 68 DMPs, methylation at the remaining 88 DMPs (out of 156 DMPs) was significantly associated with PDAC in validation cohort. We have looked into the hemimethylation at the 156 DMPs in the cancer samples of our cohort (n=7) and also in the TCGA cohort. Our results showed that the average beta values at the DMPs between 0.2 and 0.8 were 100% in our cohort and 91.6% in the TCGA cohort. The details of DMPs common and novel with TCGA cohort is shared in table. Top hypomethylated genes were *MX2*, *OAS2*, *SPRED2*, *ADAP1*, *SLC17A9* and top hypermethylated genes were *FAIM2*, *HOXA10*, *TWIST1*, *PARP4*, *H6PD*. Top 7 genes identified based on survival are *KCNA6* (-), *RASSF1*(-), *SIGIRR* (-), *FOXE1*(+), *FAIM2*(+), *NPY* (+) and *BHLHE23* (+). Promoter hypomethylation have been observed in *IRF4*, *PML*, *MX2*, *OAS2*, *HLA-A*, and *SIGIRR* which are involved in the interferon signalling pathway (MCODE1). however, the MCODE2 (*ADYC3*, *GALR1*, *NPY*, and *OPRK1*) cluster proteins function in cell-signalling. Differential survival of genes annotated in MCODE1 (*MX2*, *OAS2*, *IRF4*, *PML*, and *HLA-A*) and MCODE2 (*ADYC3*, *GALR1*, *NPY*, and *OPRK1*), in the TCGA PanCa cohort further showed that patients with lower expressions of genes annotated to MCODE1 and higher expressions of genes annotated to MCODE2 showed better disease prognosis. In our study, we have observed hypomethylation of MCODE1, suggesting activation of immune pathways in PDAC. A recent study showed that over-expression of *MX2* reduced cell proliferation, migration, and invasion via *ERK/P38/NF- κ B* signalling pathway in glioblastoma cells. Up-regulated expression of *OAS2* was previously reported in PanCa. Interferon induced cell killing is a potent anti-tumor immune response in cancer. Previously a study had shown that induction of Type I Interferon signalling made PanCa vulnerable to innate immune systems and showed better response with immune checkpoint therapy (Kim *et al.* 2018). The interferon signalling pathway inhibits cell proliferation and cell migration in PanCa. Up-regulated expression of these genes may thus signify the active anti-tumor immune response in PDAC (Fujisawa *et al.* 2020). Hypermethylation and downregulation of TSG, DNA repair genes are observed in tumors, which aid in cancer progression and metastasis (Natale *et al.* 2019). In current study, we observed hypermethylation of (1) TSGs including *LOC645323*, *FOXE*, and *TCERG1L*, (2) transcription regulator genes (*BHLHE23*, *GSC2*, *FOXE1*,

HOXA10 and *TWIST1*), (3) Ion transporters—*KCNA3*, *KCNA6*, *CACNB2* and (4) Immune regulators including *HLA-A*, *IRF4*. Previous studies have reported hypermethylation of these transcription regulators in multiple cancers including PanCa and reduced expression of these genes was associated with a worse prognosis of PAAD cancer in the TCGA cohort. *FOXE1* is one of the most frequently hypermethylated TSGs in PanCa (Kinugawa *et al.* 2017, and Natale *et al.* 2019). hypermethylation of another previously reported TSG *TCERG1L* was observed in PDAC compared to normal tissues (Vincent *et al.* 2011). *FOXE1*, *TWIST1* and other hypermethylated genes involved in cell cycle regulation were observed in PDAC cancers, thus suggesting that their down-regulation may encourage uncontrolled cell proliferation in PDAC. Deregulations of ion transporters (Calcium, Potassium and Sodium) were observed in many cancers including PanCa's and often represented as biomarkers (Anderson *et al.* 2019). Ion transporters can help in angiogenesis and cancer metastasis. We observed hypermethylation of potassium (*KCNA3*, and *KCNA6*) and calcium (*CACNB2*) ion channels in PDAC. Studies on animal models showed that epigenetic modification of *KCNA3* gene limits T-cell activation (Patel *et al.* 2019). The Ca^{2+} and K^{+} ion channels thus can inhibit T-cell activation and proliferation upon antigen recognition by epigenetic modulation, creating an immunosuppressive tumor microenvironment in PDAC, which is again associated with poor prognosis and recurrence. hypermethylation of potassium channels may thus alter the immune context in PDAC and limit anti-tumor immunity (Alshalalfa *et al.* 2019). One of the recent findings documented epigenetic dysregulation (hypermethylation) of Ca^{2+} ion transporters in PDAC . Additionally, hypermethylation of *HLA-A* also suggested suppression of antigen presentation and antitumor activity (Gregório *et al.* 2020). In this current study, we observed hypermethylation of two genes—*NPY* and *FAIM2* in Indian PanCa samples, which were also observed in the TCGA cohort and in the PanCa cell lines in the CCLE database. hypermethylation of both *NPY* and *FAIM2* was correlated with reduced expression and was associated with poor survival, both in the TCGA and the Indian patient cohorts. The significant poorer survival also observed when *FAIM2* and *NPY* act jointly. Our data strongly represents that the *NPY* and *FAIM2* cumulatively contributed to significant poorer survival in the Indian Cohort. In IHC study, validation cohort 4 we also reclaimed down-regulation of protein expression for both genes in PanCa samples with respect to normal counterparts. The result is consistent with the gene expression and promoter methylation. All our patient cohorts including discovery and validation cohorts have been revealed that hypermethylation and down-regulation properties in mRNA and finally in protein levels of *NPY* and *FAIM2*. Considering CPTAC database it has been further proved that both Npy and Faim2 protein

levels were also shown down-regulated in PanCa. All our data and from publicly available data sets strongly suggests and support that *NPY* and *FAIM2* are the novel potential frequently hypermethylated genes and trigger the PanCa progression and development in our patient cohort (Indian PanCa patient population). To our knowledge, hypermethylation of *FAIM2* was previously observed in ductal carcinoma of breast cancer, where the authors have shown an association of DNA methylation with progressive stages (Johnson *et al.* 2015). *FAIM2* is a member of the transmembrane BAX inhibitor motif-containing (TMBIM) family which comprised of 6 anti-apoptotic proteins. *FAIM2* can suppress cell death by regulating Calcium ion homeostasis in the endoplasmic reticulum (Hong *et al.* 2019). Thus, hypermethylation and downregulation of *FAIM2* can help the tumor to evade apoptosis. However, in current scenario, *FAIM2* role in PanCa lacks research and further study is needed but in other cancer types *FAIM2* has shown to be significant. The role of *FAIM2* in obesity has already been documented which is a prime risk factor in PDAC development. It has also been reported by Kang *et al.* 2016, that *FAIM2* acts as an novel biomarker in small cell lung carcinoma (SCLC) therefore emphasizing *FAIM2*'s role as cancer biomarkers (Kang *et al.* 2016). Another hypermethylated gene in PDAC was a *NPY*, which was included in MCODE2 along with *ADCY3*, *GALR1*, *OPRK1*. *NPY* stimulates cell proliferation and has been implicated as growth-promoting factor in various malignancies, but little is known about the effects of *NPY* on PanCa. *NPY* promoter have reported frequent hypermethylated in colorectal, and rectal cancers, surprisingly, in prostate and cholangiocarcinoma, *NPY* overexpression was observed but no correlation with hypomethylation has been found (Alshalalfa *et al.* 2019). A physiological function of *NPY* is to regulate food intake and increase fat storage which is a risk factor for PanCa. In addition, our study have shown that *NPY* downregulation ($p=0.041$) and high alcohol consumption were associated with poor OS. All previous findings, including our data strongly support that in PDAC, hypermethylation of *NPY* promoter might be correlated with inactivation of gene expression and might promote carcinogenesis (Zhang *et al.* 2011). Considering PDAC aggressive nature and its associated high mortality rate, the urgency of novel therapeutic strategies to combat is of priority. Advances in the methylation landscape are crucial to understand the development of epigenetically targeted biomarker (Zhang *et al.* 2011). Our novel findings in this ongoing study can state that a combination of epigenetic drugs along with existing targeted therapies will be a determinant targeted therapeutic approach that emphasizes the importance of our 450 K methylome study in PDAC patients across our nation (India) (Sikdar *et al.* 2018). The reversible nature of methylation signatures is a key to the development of markers targeting

disease diagnosis and prognosis followed by therapeutic aids. Decoding the methylation landscape using 450 K methylome and parallel comparison with global datasets from TCGA makes our understanding even magnified and clinically more suitable for biomarker development (*Alshalalfa et al. 2019*). Two novel genes namely *NPY* and *FAIM2* have been observed from our work to be frequently hypermethylated significantly in all Indian cohorts and the result are in concordance with mRNA and protein expression, and protein databases making them translational importance. A statistically significant correlation with clinical and subclinical parameters also concludes to the fact that demographic and ethnic reasons do play a role in the epigenetic marker development. The methylation status of both *NPY* and *FAIM2* and correlation with the TCGA validates our data against Western world data but the variance in the level of expression and methylation also concludes the demographic and ethnic inputs specific to our patient pool making it a significant study (*Chatterjee et al. 2022*). This helps in gaining a clear understanding of the epigenetic landscapes of PanCa, as our study strongly suggests and support novel architecture of epigenetic landmarks including therapeutic strategies, and insight into potential epigenetic outcome on PDAC in the Indian patient population. There is a biological theory that long-term exposure to heavy metals in the environment causes epigenetic alterations, with associated health risks. However, researchers are still trying to determine whether the level or type of heavy metal ion exposure is linked to epigenetic changes among subjects who are affected by serious environmental contaminants (*Antwi et al. 2023*). Further, Ni, Cr, As and Cd were ranked above the upper limit suggested by the United States Environmental Protection Agency for carcinogenic risk rating (*Aendo et al. 2022*). Some exploration on this mining region peoples blood levels for essential and nonessential elements is yet to be done along with potential relationship between prolonged metal exposures and epigenetic dysregulation. Understanding why these heavy metals have elevated levels in pancreatic juice and what this means for pancreatic cells would go a long way in diagnosing, treating or even preventing PanCa. Thus, abnormal changes to the epigenome play a major role in the angiogenesis and progression of metal-induced cancer. It is noteworthy that much research is still needed to fully understand the function that epigenetic pathways play in the carcinogenesis and angiogenesis caused by heavy metals (*Hong et al. 2020*). Specifically, a variety of heavy metals can change the methylation of DNA, resulting in either hyper- or hypomethylation. A study by *Zhang et al.* revealed that Cd led to promoter hypermethylation of *DAPK1* and *RASSF1A*, reducing their expression and increasing cell proliferation. Moreover, low quantities of Cd may affect DNA methylation, which could be the method by

which Cd proliferates (Zhang *et al.* 2022). The findings of a clinic-based case-control study by Antwi *et al.* revealed that there may be a connection between PanCa risk and routine exposure to pesticides, asbestos, benzene, and chlorinated hydrocarbons along with nickel and chromium (Antwi *et al.* 2023). This showcases the fact that there might be a correlation amongst our observed signatures with heavy metal interactions.

Based on 450 K methylome data, a previous study from our group showed that epigenetic and genetic signatures of Indian cohort of PDAC patients diverge partially based on expression frequency and alternated driver behaviors, across the globe. This wide variation in methylation landscape between TCGA, and our cohort, included 156 differentially methylated positions annotated to 91 genes having differential methylation (gDMs). Similarly, our previous data also showed variance in various parameters like, *KRAS* mutation frequency, *NPY* and *FAIM2* differential methylation pattern in Indian PDAC samples and Indian data were different with respect to western patient cohort (Anderson *et al.* 2019). The possible reason of this variability were ethnicity, environment, lifestyle and occupational hazards. Enrichment analysis using GO/KEGG terms, canonical pathways and hall mark gene sets and network analysis of these 91 gDMs, pointed “regulation of ion transport”, as the most statistically significant enriched term (Zhou *et al.* 2019). Metascape data from our previous study and IPA data from this study resulted in the identification of 7 ‘ion transport regulation’ associated gDMs: *RASA3*, *CLIC5*, *KCNA3*, *KCNA6*, *KCNJ5*, *CACNB2* and *GABRR2*. Significant differential expression in normal tissue samples in GTEx cohort versus cancer samples in the TCGA cohort was shown in case of *RASA3*, *KCNJ5* and *CLIC5*, while that of *CACNB2* was not significant. *KCNJ5*, was hypomethylated and overexpressed in our Indian cohort while hypermethylated and underexpressed in the TCGA cohort. Kaplan-Meier survival data of Indian cohort also showed significant poor survival in high *KCNJ5* group as compared to low *KCNJ5* group. This finding is further strengthened by its over-expression significantly in PanCa cell lines, SU86.86, YAPC and HuP T4, while its protein, Kir3.4 expression is not demonstrated in any cell lines so far, showing concordance with TCGA based *KCNJ5* expression patterns (Zhou *et al.* 2019). *KCNJ5*, coding for inward rectifying potassium channel protein, Kir3.4 is regulated by G proteins, usually detected at low levels in normal pancreas. Kir channels are associated with heart rate, hormone secretion, renal functions, atrial fibrillation, hypertension, and type II diabetes. Similarly, Ca^{2+} Voltage-Gated Channel Auxiliary Subunit Beta 2 (*CACNB2*) is mutated in lung cancer and Hodgkin lymphoma and showed chemoresistance in esophageal squamous cell carcinoma, while suppressed head and neck squamous cell carcinoma (Hibino *et al.* 2010).

Various kinds of Kir channels are associated with all the hallmarks of cancer in vast array of malignancies. A group-member, Kir3.4 (*KCNJ5*) promoted cell proliferation in aldosterone-producing adenoma. It showed association with restructuring of immune microenvironment in colon cancer, facilitating G-protein/ phospholipase C/ DAG mediated M2 macrophage infiltration and recruitment both *in vitro* and *in vivo* (xenograft animal models) (Wang *et al.* 2020). However, our LM22 based immune profiling data demonstrated a positive correlation with M1 TAMs, acting as a tumor suppressor, mediated by the release of pro-inflammatory cytokines and tumor necrosis factors. *KCNJ5* boosted the aggressiveness of triple-negative breast cancer and was elevated in MGC-803, gastric cancer cells (Li *et al.* 2018). Similarly, we documented the upregulation of *KCNJ5* in our Indian PanCa cohort, showing its credibility as a drug target in these devastating diseases. TFs of *KCNJ5*, identified in this study, *POU2F1* and *POU3F1* possibly regulating *KCNJ5* expression, are previously reported to be associated with malignancy of head and neck, lung and thyroid. Thus, the regulation of *KCNJ5* through modification of *POU2F1* and *POU3F1*, could help in devising advanced therapeutic strategies. Our study predicted that onco-channel *KCNJ5* can be targeted by several other ion channel blockers, Amiloride, Vernakalant hydrochloride, Dalfampridine, Glyburide and Levocromakalim. Among these, antikaliuretic-diuretic, Amiloride, synergized the chemosensitivity of (i) Erlotinib in Bxpc-3 and PANC-1 PanCa cells and (ii) Sunitinib in renal cancer Caki-1 cells (Zheng *et al.* 2015). It also suppressed growth and proliferation of H6 hepatoma cells, and induced apoptosis in cell line and animal models of multiple myeloma. Anti-diabetic drug, Glibenclimide also showed anti-neoplastic effects. Thus, these predicted synthetic anti-Kcnj5 compounds demand further assessment for their therapeutic potential in Indian patients' suffering from PanCa. (Al-Karagholi *et al.* 2019). PanCa, a highly aggressive disease, progresses rapidly and resist available treatments, leading to devastating outcome. Though numerous studies are highlighting the intricate mechanisms of disease progression, translation of this knowledge into PanCa therapy is lacking. Thus, novel concepts in this field are the only way out. Embedded on cell membrane, ion-channels are able to alter tumor and its microenvironment through manipulating various signalling cascades. Externally located ion channels make them an attractive and ideal drug target. Thus, our data demonstrating over-expression of Kcnj5 can be targeted in disease management of Indian PanCa patients. The small sample size (n=20) of our Indian cohort, indicating the association of enhanced *KCNJ5* with PanCa, however needs to be confirmed in a larger cohort couple with immunohistochemistry and further functional studies in PanCa cell lines. Owing to the critical and complex role of potassium ion channels in cancer signalling

pathways, *KCNJ5*, identified as a novel oncochannel overexpressed and associated with poor patient survival in Indian cohort of PanCa, indicates its essential role is disease physiology. Several lines of therapeutic strategies specific for Indian patient population that can be developed based on our findings are (i) targeting two TFs *POU2F1* and *POU3F1* controlling the expression of *KCNJ5*, (ii) prediction of Amiloride, Vernakalant hydrochloride, Dalfampridine, Glyburide and Levromakalim as small molecules acting as anti-Kcnj5 and (iii) targeting Kcnj5 with natural product protodioscin. The role of *KCNJ5* is also demographic/ethnicity specific which is validated across several observations and this uplifts the role of biomarker development specific to our population. The data would help to develop remedial measures in facilitating PanCa management in Indian patients

The mortality-to-incidence ratio has not significantly decreased despite significant advancements in PanCa research. In addition to tumor heterogeneity, adaptability, and aggressive metastasis, which manifest in over 50% of identified patients, this is primarily caused by a lack of early diagnostic distinctive symptoms and trustworthy biomarkers.

PanCa systems biology research depends on integrating omics data from several biological levels. A comprehensive understanding of the disease's mechanisms and increased understanding of potential diagnostic or prognostic biomarkers and drug targets were made possible by the systems biology approach, which addressed the often-arising challenges in cancer diagnosis and treatment due to the disease's complex pathogenic landscape and cellular heterogeneity. There are encouraging and ongoing initiatives to integrate multi-omic data in the direction of personalized treatment. The initial step in precision medicine is to stratify patients based on their clinical characteristics, biomarkers, and disease subtypes. Or statistics. Later, other characteristics were added to the stratification process, including the use of the surroundings, past medical history, actions, and routines results in smaller collectives. Theoretically, this categorization method should prevent clinical trial failures because small patient populations receive tailored therapies and the appropriate diagnosis, or straight to people. Rather than using "one-size-fits-all" methods of treatment, the most effective therapy options or drugs for every person or a small group can be obtained by illness classification, followed by personalization through the use of multi-omics networks (Marques *et al.* 2024). In general, machine learning methods, such as supervised learning algorithms like random forests and support vector machines, as well as unsupervised learning methods like dimensionality reduction and clustering, are highly beneficial for PanCa research. They make it possible for researchers to anticipate patient outcomes, enhance diagnostic accuracy, draw meaningful conclusions from complicated datasets, and support

individualized treatment plans (Turanli et al. 2021). Machine learning can be used to predict PanCa patients' survival rates with a high degree of prediction performance. Enhancing survival prediction can aid in improved therapy selection as well as the planning of social and care requirements (Tripathi et al. 2024). Differential expressed genes were compared with differential methylated genes obtained from TCGA PAAD tumor and adjacent normal pancreatic tissue with cut-off criteria of $FDR \leq 0.01$ leading to identification of top MEDEGs namely Monoacylglycerol O-acyltransferase 2 (*MOGAT2*), *MIA*, *CD68*, *LRRC31*, and *C2ORF54*. Our research in PanCa has unearthed significant insights into the roles of various genes, particularly (*MOGAT2*, and its potential linkage to PanCa. *MOGAT2*, a central component in the monoacylglycerol pathway for triacylglycerol synthesis, is known for its high expression in the human liver and small intestine, playing a vital role in lipid metabolism. Intriguingly, recent studies have positioned *MOGAT2* as an attractive therapeutic target for obesity, type 2 diabetes mellitus, and nonalcoholic steatohepatitis.. This is further compounded by findings from Lian et al. who suggest *MOGAT2*'s potential as a biomarker in colorectal cancer metastasis (Lian et al. 2020). Furthermore, a differentially methylated gene in inflammatory breast cancer was found to be *MOGAT2*. Heterogeneous sub clonal populations arise during the evolutionary course of PDAC and are responsible for the growth of the main tumor, its regional dissemination, its distant metastasis, and patient death (Wang et al. 2021¹). Nonetheless, in patients who are not receiving treatment, the genetics of metastases closely resembles that of the original tumor, and all subclones have PDAC driver mutations. This suggests that an epigenetic mechanism could be at work during metastasis. *MOGAT2* was also found to have a crucial role in PanCa metastasis to nearby organs such as the liver as documented in WGBS data (Lian et al. 2020). Given these multifaceted roles, our study extends the conversation, suggesting a hitherto underexplored yet possibly significant link between *MOGAT2* and PDAC, thereby opening new avenues for research and therapeutic intervention. The role of *CD68* in PDAC also garners significant attention in our study. As a protein integral to various physiological and pathological processes, including atherosclerosis, inflammation, autoimmunity, bone resorption, and tumor progression, *CD68*'s prominence in the tumor microenvironment of PDAC is notable (Auwera et al 2010). It is particularly overexpressed in TAMs and tumor cells. These findings align with the broader oncology research, suggesting that TAMs play a pivotal role in tumor progression and could be key players in the immune response within the PDAC tumor microenvironment. Yuan et al. showed that patients with breast cancer, showcase combined high expression of *CD47* and *CD68* which is a unique prognostic factor (Yuan et al. 2019). According to this

study, there was a considerable increase in *CD68*-labeled TAMs in breast cancer, and advanced TNM stage, histological grade, Lymph Node Metastasis (LNM), Eestrogen Receptor (ER) status, Progesterone Receptor (PR) status, and recurrence were all linked to high macrophage infiltration *Zhang et al.* showed the involvement of *CD68* in tumor immunity across all types of cancer too (*Zhang et al. 2022*). Different tumor forms have different clinical prognoses and immunological responses to high expression levels of *CD68*. For many different types of tumors, inhibiting *CD68*-dependent signalling may be a useful therapeutic approach for immunotherapy. This established role of *CD68* henceforth showcases the probability of *CD68* having crucial role in PanCa as a biomarker (*Zhang et al. 2022*). The investigation into *MIA* offers another intriguing aspect of our study. Recognized as a protein associated with melanoma metastasis, Intriguingly, alterations in the expression of *MIA* in lung cancer cells lead to significant changes in apoptotic pathways, notably affecting the levels of pro-apoptotic marker *Bax* and anti-apoptotic marker *Bcl-2*, as well as influencing the PI3K/mTOR signalling pathway (*Sasahira et al. 2016*). This highlights the potential of *MIA* as not only a marker for cancer progression but also as a player in the oncogenic process, raising the question of its role in PDAC (*Gu et al. 2020*). This dual role of *LRRC31*, impacting both DNA repair mechanisms and tumor progression, presents it as a novel and promising gene target in the context of PDAC treatment (*Chen et al. 2020*). *MIA* ($p < 0.001$) in breast cancer and *CD68* ($p < 0.001$) in renal cancer were found significant as prognostic markers in human protein atlas. We also cross checked the functional enrichment of *MOGAT2*, *CD68* and *MIA* which showed its significant association with multiple important pathways signifying its role. While our study has revealed several novel gene (*MOGAT2*, *CD68* and *MIA*) associations in PDAC, it is imperative to emphasize the need for extensive validation before these findings can be translated into clinical practice. The integration of ML methods in our research has undoubtedly deepened our understanding of PDAC, yet the translation of these insights into practical, clinical applications must be approached with rigorous validation processes to ensure their effectiveness and reliability in improving patient outcomes in PDAC (*Yokoyama et al. 2020*).

Differential expression analysis and differential methylation analysis combination was used to select DEGs and DMGs between PAAD and adjacent normal pancreatic tissue in the TCGA-PAAD dataset with cut-off criteria of $FDR \leq 0.05$ and $|\log_2FC| > 2.0$ leading to identification of 7 significant MEDEGs namely *CD36*, *UGT1A1*, *TFF1*, *S100P*, *MUC13*, *CALHM3* and *ANKRD44*. *CD36* was discovered intracellularly, in the plasma membrane, and in conjunction with insulin granules in the pancreas. Reduced *CD36* expression in PDAC is

linked to larger tumor sizes and lower survival rates, but it is not as strongly linked to TNM staging (Tanase et al. 2020). According to a study by Jia et al. PanCa patients' levels of *CD36* gene expression are low. Reduced TNM staging and CA19-9 levels are predicted by reduced *CD36* expression, but bigger tumor size and a worse prognosis for survival are. It has the potential to predict prognosis and clinicopathological traits, but more research is need to fully understand the discrepancy (Jia et al. 2018). This was supported by our observation of *CD36* down regulation across all datasets evaluated concerning multiple omics spaces. According to the findings of Feng et al. *UGT1A1* may be a viable target and predictive biomarker for PanCa patients. Future research is necessary to corroborate our results and advance the *UGT1A1*'s clinical relevance in PanCa (Feng et al. 2021). Umemoto et al. suggest that in advanced PanCa patients of Japanese origin (n=31) with *UGT1A1*-DV treated with FOLFIRINOX, irinotecan should be provided at a starting dose of ≤ 120 mg/m² (Umemotoe et al. 2021) . This dyregulation of *UGT1A1* was in concordance with our study and also suggested us to do drug targeting analysis using molecular dynamics which lead us to finding of apigenin-7-O-glucuronide which is more suitable drug target for *UGT1A1* and has shown higher efficiency in silico than the above-mentioned drugs (Umemotoe et al. 2021). *S100P* expression in pancreatic, cervical, and prostate cancer is regulated by epigenetic modifications. The anti-histaminic medication cromolyn inhibits the *S100P*–RAGE interaction, which decreases the growth of pancreatic tumors and boosts the efficacy of the chemotherapy medication gemcitabine (Bresnick et al. 2015). *S100P* is a Ca²⁺ binding protein that has been identified as a putative tumor biomarker due to its overexpression in a number of malignancies (Jiang et al. 2012). *S100P* was shown to be broadly expressed in both normal and malignant tissues in a study by Parkkila et al. (Parkkila et al. 2008). Given that it is highly expressed in certain malignancies, this molecule could be a viable target for upcoming medical diagnostic and therapeutic uses. Excellent stability is a must for a successful biomarker. In this regard, it has been demonstrated that the expression of the *S100P* gene or protein is correlated with patient survival in cases of lung and breast cancer, and it has been suggested that this gene or protein is an early embryonic marker of pancreatic carcinogenesis (Jiang et al. 2012). Using gene specific priming and quantitative real-time RT-PCR, *S100P* expression in cancer has been studied. *S100P* expression in PanCa and IPMN is notably higher than in non-neoplastic pancreatic tissue. Pancreatic juice from IPMN and cancer tissue has a higher concentration of *S100P* than juice from pancreatitis. Utilizing receiver operating characteristic curve analysis, the measurement of *S100P* in pancreatic juice can effectively differentiate between neoplastic disease and chronic pancreatitis (area under

the curve = 0.837; 95% CI: 0.749– 0.903). The expression of *S100P* was detected in 32 cases of PDAC and cases with dubious diagnoses (n = 6). However, only one out of 14 cases of benign or reactive ductal epithelium showed positive *S100P* expression, and no *S100P* expression was detected in the endocrine tumor. These findings are from a different survey on the screening of PDAC using fine-needle aspiration biopsy (FNAB) (Ohuchida *et al.* 2006). Our results based on *S100P* upregulation followed by single cell observation reporting signature presence in ductal cell type highlights the above-mentioned observations. Debernardi *et al.* conducted a study wherein they examined the efficacy of an enhanced urine biomarker panel consisting of *LYVE1*, *REG1B*, and *TFF1* in a broader set of specimens (Debernardi *et al.* 2020). The results showed that the panel could reliably distinguish patients with early stages of PDAC from both control subjects and patients having benign hepatobiliary diseases. PDAC precursor lesions, also known as PanINs, have been found to have increased levels of this protein, which is linked to cancer growth and metastasis. Shen *et al.* used a transgenic model of PanINs to show that early lesions and lymphatic vessels are linked to lymph angiogenesis and endothelial remodeling (Shen *et al.* 2019). Previously, we found that *TFF1* overexpression was present in early PanINs in FPC before cancer development. According to a study by Jahan *et al.* Tissue and serum tests confirmed a considerable rise in *TFF* levels in precursor lesions and early stages of PanCa (Jahan *et al.* 2019). Previous research has revealed *TFF1* to be a functional biomarker in a number of different tumor types, including gastric, oesophageal squamous cell, and breast cancer. *TFF1* was found to show similar results in our multi-omics-based observations too demonstrating validation of outcome using our models and probable utility as biomarker. Mucin 13 (*MUC13*) is a glycoprotein that is surface-expressed and plays a role in the carcinogenesis of several cancers, such as colorectal, renal, and PanCas. According to research by Pang *et al.*, *MUC13* may function as an oncogenic glycoprotein to speed up the development of lung cancer by abnormally activating the *ERK/JNK/p38* signalling pathway (Pang *et al.* 2021). It may also be a useful target for lung cancer therapy. Sheng *et al.*'s demonstration that *MUC13* increases *NF-κB* signalling in response to TNF and chemicals that damage DNA offers a novel molecular target for the targeted suppression of *NF-κB* activation. These findings suggest that treating metastatic malignancies with a combination of chemotherapy and *MUC13* antagonist may be more effective (Sheng *et al.* 2017). The ductal cells had the highest expression, followed by the exocrine cells and the mixed cell types. All things considered, our result supports the significance of *tMUC13* in pancreatic pathobiology as well as its potential as a biomarker and therapeutic target of PanCa (Dhasmana *et al.* 2023).

Followed by this we obtained a very prominent data concerning again *TFF1* and *S100P* as we found them to be dysregulated across precancerous lesions to PanCa to metastasis. These genes are likely to be grouped together in the same module if they exhibit a progressive upregulation in pancreatitis, metastatic PanCa, and normal pancreas. Hence this shows that our data obtained have significant potential to be used as diagnostic cell type specific biomarkers. When we identified biomarker with cut off $p < 0.01$, we compared our data to three other GEO methylation datasets (450K) and three other expression datasets (RNA seq and micro-array). 2 methylation datasets were from western culture and 1 GEO methylation dataset (GSE181740) is our own dataset obtained from Indian patient population. 2 western world methylation datasets showed *MOGAT2* to be significant but in Indian patient population it was not found which can be due to epigenetic tissue specific expression which can be diverse due to change in demographic, ethnic, and lifestyle differences. The total 7 CpGs were found to be significant in the western GEO methylation datasets and also was found to be significant when observed in both RNA seq and micro-array expression datasets. When we identified biomarker with cut off $p < 0.05$, then the top 7 genes along with their corresponding 8 CpGs were validated first using two 450K datasets. GSE74071 showed *S100P* to be most significant whereas GSE 49149 showed all 7 genes to be significant. RNA seq data obtained from GSE211398 showed *S100P* and *MUC13* to be most significant. Following this we further evaluated our data using microarray datasets of PanCa. GSE183795 showed *S100P*, *MUC13*, *CD36*, *TFF1*, *CALHM3* to significantly dysregulated. GSE28735 showed *S100P*, *TFF1*, *CD36*, *UT1A1* to significantly dysregulated. GSE15471 finally demonstrated *UGT1A1*, *CD36*, *S100P*, *CALHM3*, *ANKRD44*, *TFF1* to significantly dysregulated. In case of *KCNJ5* GEO datasets and the observations showed hypermethylation along with down regulation which is in concordance with the observation seen in TCGA dataset. With a low 5-year overall survival rate because to early lymph node metastases and difficulties with early diagnosis, PanCa presents a substantial challenge. All aspects of PanCa patient care can benefit from the combination of DL and ML applications, including early diagnosis, enhanced biomarker accuracy, therapy response monitoring, prognostication of perioperative and survival outcomes, and resource optimization (Patel et al. 2024). Difference in signatures expression and methylation status has been observed across demographic regions, ethnic variations also including occupational and lifestyles pattern differences. Previously our publications showed difference in *KRAS* mutation profile when we compared our Indian patient population data with western world data (Saha et al. 2020). Similar sort of observation was observed in our previous finding where *NPY* and *FAIM2*

methylation status showed difference in our Indian patient population data when we compared with TCGA datasets. This hence discloses the fact that prognostic markers can sometimes vary based on demographic and ethnic backgrounds of the population under study.



LIMITATIONS



SEPTEMBER 16, 2024

PHD THESIS

Department of Life Sciences and Biotechnology, Jadavpur University

7.0: Limitations of The Study

The sample size of our Indian cohort used for 450K analysis was only 7 paired PDAC samples. The tumor purity was also a limitation as primary tissue obtained had normal tissue contamination as the tissues were not frozen dissected tissues. Association of *KCNJ5* with poor prognosis is with Indian ethnicity, environment, and lifestyle, enables identification of novel targets and disease management of PanCa patients of India, with the need of a larger cohort-based study. All Multi-omics derived biomarkers need to validate in Indian patients' samples (gene level expression and immunohistochemistry methods)



BIBLIOGRAPHY



SEPTEMBER 16, 2024

PHD THESIS

Department of Life Sciences and Biotechnology, Jadavpur University

8.0 Bibliography

1. Abulwerdi F, Liao C, Liu M, Azmi AS, Aboukameel A, Mady AS, Gulappa T, Cierpicki T, Owens S, Zhang T, Sun D, Stuckey JA, Mohammad RM, Nikolovska-Coleska Z. A novel small-molecule inhibitor of mcl-1 blocks pancreatic cancer growth in vitro and in vivo. *Mol Cancer Ther.* 2014 Mar;13(3):565-75. doi: 10.1158/1535-7163.MCT-12-0767. Epub 2013 Sep 9. PMID: 24019208; PMCID: PMC4174574.
2. Aendo P, Netvichian R, Thiendedsakul P, Khaodhiar S, Tulayakul P. Carcinogenic Risk of Pb, Cd, Ni, and Cr and Critical Ecological Risk of Cd and Cu in Soil and Groundwater around the Municipal Solid Waste Open Dump in Central Thailand. *J Environ Public Health.* 2022 Feb 28;2022:3062215. doi: 10.1155/2022/3062215. PMID: 35265140; PMCID: PMC8901317.
3. Albaradei S, Thafar M, Alsaedi A, Van Neste C, Gojobori T, Essack M, Gao X. Machine learning and deep learning methods that use omics data for metastasis prediction. *Comput Struct Biotechnol J.* 2021 Sep 4;19:5008-5018. doi: 10.1016/j.csbj.2021.09.001. PMID: 34589181; PMCID: PMC8450182.
4. Al-Karagholi MA, Ghanizada H, Hansen JM, Skovgaard LT, Olesen J, Larsson HBW, Amin FM, Ashina M. Levromakalim, an Adenosine Triphosphate-Sensitive Potassium Channel Opener, Dilates Extracerebral but not Cerebral Arteries. *Headache.* 2019 Oct;59(9):1468-1480. doi: 10.1111/head.13634. Epub 2019 Sep 18. PMID: 31535367.
5. Alshalalfa M, Nguyen PL, Beltran H, Chen WS, Davicioni E, Zhao SG, Rebbeck TR, Schaeffer EM, Lotan TL, Feng FY, Mahal BA. Transcriptomic and Clinical Characterization of Neuropeptide Y Expression in Localized and Metastatic Prostate Cancer: Identification of Novel Prostate Cancer Subtype with Clinical Implications. *Eur Urol Oncol.* 2019 Jul;2(4):405-412. doi: 10.1016/j.euo.2019.05.001. Epub 2019 Jun 2. PMID: 31164324; PMCID: PMC7597937.
6. Anderson EM, Thomassian S, Gong J, Hendifar A, Osipov A. Advances in Pancreatic Ductal Adenocarcinoma Treatment. *Cancers (Basel).* 2021 Nov 3;13(21):5510. doi: 10.3390/cancers13215510. PMID: 34771675; PMCID: PMC8583016.
7. Anderson KJ, Cormier RT, Scott PM. Role of ion channels in gastrointestinal cancer. *World J Gastroenterol.* 2019 Oct 14;25(38):5732-5772. doi: 10.3748/wjg.v25.i38.5732. PMID: 31636470; PMCID: PMC6801186.
8. Anderson KJ, Cormier RT, Scott PM. Role of ion channels in gastrointestinal cancer. *World J Gastroenterol.* 2019 Oct 14;25(38):5732-5772. doi: 10.3748/wjg.v25.i38.5732. PMID: 31636470; PMCID: PMC6801186.
9. Anfossi S, Giordano A, Gao H, Cohen EN, Tin S, Wu Q, Garza RJ, Debeb BG, Alvarez RH, Valero V, Hortobagyi GN, Calin GA, Ueno NT, Woodward WA, Reuben JM. High serum miR-19a levels are associated with inflammatory breast cancer and are predictive of favorable clinical outcome in patients with metastatic HER2+ inflammatory breast cancer. *PLoS One.*

- 2014 Jan 8;9(1):e83113. doi: 10.1371/journal.pone.0083113. PMID: 24416156; PMCID: PMC3885405.
10. Antequera F, Bird A. Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A*. 1993 Dec 15;90(24):11995-9. doi: 10.1073/pnas.90.24.11995. PMID: 7505451; PMCID: PMC48112.
 11. Antwi MB, Dumitriu G, Simón-Santamaria J, Romano JS, Li R, Smedsrød B, Vik A, Eskild W, Sørensen KK. Liver sinusoidal endothelial cells show reduced scavenger function and downregulation of Fc gamma receptor IIb, yet maintain a preserved fenestration in the Glmpgt/gt mouse model of slowly progressing liver fibrosis. *PLoS One*. 2023 Nov 1;18(11):e0293526. doi: 10.1371/journal.pone.0293526. PMID: 37910485; PMCID: PMC10619817.
 12. Appleby PN, Crowe FL, Bradbury KE, Travis RC, Key TJ. Mortality in vegetarians and comparable nonvegetarians in the United Kingdom. *Am J Clin Nutr*. 2016 Jan;103(1):218-30. doi: 10.3945/ajcn.115.119461. Epub 2015 Dec 9. PMID: 26657045; PMCID: PMC4691673.
 13. Araki H, Tazawa H, Kanaya N, Kajiwara Y, Yamada M, Hashimoto M, Kikuchi S, Kuroda S, Yoshida R, Umeda Y, Urata Y, Kagawa S, Fujiwara T. Oncolytic virus-mediated p53 overexpression promotes immunogenic cell death and efficacy of PD-1 blockade in pancreatic cancer. *Mol Ther Oncolytics*. 2022 Sep 13;27:3-13. doi: 10.1016/j.omto.2022.09.003. PMID: 36212775; PMCID: PMC9513735.
 14. Arowolo O, Suvorov A. Underexplored Molecular Mechanisms of Toxicity. *J Xenobiot*. 2024 Jul 18;14(3):939-949. doi: 10.3390/jox14030052. PMID: 39051348; PMCID: PMC11270369.
 15. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014 May 15;30(10):1363-9. doi: 10.1093/bioinformatics/btu049. Epub 2014 Jan 28. PMID: 24478339; PMCID: PMC4016708.
 16. Bararia A, Chakraborty P, Roy P, Chattopadhyay BK, Das A, Chatterjee A, Sikdar N. Emerging role of non-invasive and liquid biopsy biomarkers in pancreatic cancer. *World J Gastroenterol*. 2023 Apr 21;29(15):2241-2260. doi: 10.3748/wjg.v29.i15.2241. PMID: 37124888; PMCID: PMC10134423.
 17. Bararia A, Das A, Mitra S, Banerjee S, Chatterjee A, Sikdar N. Deoxyribonucleic acid methylation driven aberrations in pancreatic cancer-related pathways. *World J Gastrointest Oncol*. 2023 Sep 15;15(9):1505-1519. doi: 10.4251/wjgo.v15.i9.1505. PMID: 37746645; PMCID: PMC10514732.
 18. Bararia A, Dey S, Gulati S, Ghatak S, Ghosh S, Banerjee S, Sikdar N. Differential methylation landscape of pancreatic ductal adenocarcinoma and its precancerous lesions. *Hepatobiliary Pancreat Dis Int*. 2020 Jun;19(3):205-217. doi: 10.1016/j.hbpd.2020.03.010. Epub 2020 Apr 8. PMID: 32312637.

19. Bararia A, Dey S, Gulati S, Ghatak S, Ghosh S, Banerjee S, Sikdar N. Differential methylation landscape of pancreatic ductal adenocarcinoma and its precancerous lesions. *Hepatobiliary Pancreat Dis Int.* 2020 Jun;19(3):205-217. doi: 10.1016/j.hbpd.2020.03.010. Epub 2020 Apr 8. PMID: 32312637.
20. Bartsch DK, Gress TM, Langer P. Familial pancreatic cancer--current knowledge. *Nat Rev Gastroenterol Hepatol.* 2012 Aug;9(8):445-53. doi: 10.1038/nrgastro.2012.111. Epub 2012 Jun 5. PMID: 22664588.
21. Ben Q, Wang K, Yuan Y, Li Z. Pancreatic cancer incidence and outcome in relation to ABO blood groups among Han Chinese patients: a case-control study. *Int J Cancer.* 2011 Mar 1;128(5):1179-86. doi: 10.1002/ijc.25426. PMID: 20473916.
22. Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol.* 1988 Oct 20;203(4):971-83. doi: 10.1016/0022-2836(88)90122-2. PMID: 3210246.
23. Bian B, Bigonnet M, Gayet O, Loncle C, Maignan A, Gilabert M, Moutardier V, Garcia S, Turrini O, Delpero JR, Giovannini M, Grandval P, Gasmi M, Ouaisi M, Secq V, Poizat F, Nicolle R, Blum Y, Marisa L, Rubis M, Raoul JL, Bradner JE, Qi J, Lomberg G, Urrutia R, Saul A, Dusetti N, Iovanna J. Gene expression profiling of patient-derived pancreatic cancer xenografts predicts sensitivity to the BET bromodomain inhibitor JQ1: implications for individualized medicine efforts. *EMBO Mol Med.* 2017 Apr;9(4):482-497. doi: 10.15252/emmm.201606975. PMID: 28275007; PMCID: PMC5376755.
24. Bian J, Almhanna K. Pancreatic cancer and immune checkpoint inhibitors--still a long way to go. *Transl Gastroenterol Hepatol.* 2021 Jan 5;6:6. doi: 10.21037/tgh.2020.04.03. PMID: 33409400; PMCID: PMC7724172.
25. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, Miller DK, Wilson PJ, Patch AM, Wu J, Chang DK, Cowley MJ, Gardiner BB, Song S, Harliwong I, Idrisoglu S, Nourse C, Nourbakhsh E, Manning S, Wani S, Gongora M, Pajic M, Scarlett CJ, Gill AJ, Pinho AV, Rooman I, Anderson M, Holmes O, Leonard C, Taylor D, Wood S, Xu Q, Nones K, Fink JL, Christ A, Bruxner T, Cloonan N, Kolle G, Newell F, Pinese M, Mead RS, Humphris JL, Kaplan W, Jones MD, Colvin EK, Nagrial AM, Humphrey ES, Chou A, Chin VT, Chantrell LA, Mawson A, Samra JS, Kench JG, Lovell JA, Daly RJ, Merrett ND, Toon C, Epari K, Nguyen NQ, Barbour A, Zeps N; Australian Pancreatic Cancer Genome Initiative; Kakkar N, Zhao F, Wu YQ, Wang M, Muzny DM, Fisher WE, Brunicardi FC, Hodges SE, Reid JG, Drummond J, Chang K, Han Y, Lewis LR, Dinh H, Buhay CJ, Beck T, Timms L, Sam M, Begley K, Brown A, Pai D, Panchal A, Buchner N, De Borja R, Denroche RE, Yung CK, Serra S, Onetto N, Mukhopadhyay D, Tsao MS, Shaw PA, Petersen GM, Gallinger S, Hruban RH, Maitra A, Iacobuzio-Donahue CA, Schlick RD, Wolfgang CL, Morgan RA, Lawlor RT,

- Capelli P, Corbo V, Scardoni M, Tortora G, Tempero MA, Mann KM, Jenkins NA, Perez-Mancera PA, Adams DJ, Largaespada DA, Wessels LF, Rust AG, Stein LD, Tuveson DA, Copeland NG, Musgrove EA, Scarpa A, Eshleman JR, Hudson TJ, Sutherland RL, Wheeler DA, Pearson JV, McPherson JD, Gibbs RA, Grimmond SM. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature*. 2012 Nov 15;491(7424):399-405. doi: 10.1038/nature11547. Epub 2012 Oct 24. PMID: 23103869; PMCID: PMC3530898.
26. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan JB, Shen R. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011 Oct;98(4):288-95. doi: 10.1016/j.ygeno.2011.07.007. Epub 2011 Aug 2. PMID: 21839163.
27. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, Gunderson KL. Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics*. 2009 Oct;1(1):177-200. doi: 10.2217/epi.09.14. PMID: 22122642.
28. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Vollmer E, Goldmann T, Seifart C, Jiang W, Barker DL, Chee MS, Floros J, Fan JB. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res*. 2006 Mar;16(3):383-93. doi: 10.1101/gr.4410706. Epub 2006 Jan 31. PMID: 16449502; PMCID: PMC1415217.
29. Blockus H, Chédotal A. Slit-Robo signalling. *Development*. 2016 Sep 1;143(17):3037-44. doi: 10.1242/dev.132829. PMID: 27578174.
30. Bock C. Analysing and interpreting DNA methylation data. *Nat Rev Genet*. 2012 Oct;13(10):705-19. doi: 10.1038/nrg3273. PMID: 22986265.
31. Bosetti C, Rosato V, Li D, Silverman D, Petersen GM, Bracci PM, Neale RE, Muscat J, Anderson K, Gallinger S, Olson SH, Miller AB, Bas Bueno-de-Mesquita H, Scelo G, Janout V, Holcatova I, Lagiou P, Serraino D, Lucenteforte E, Fabianova E, Baghurst PA, Zatonski W, Foretova L, Fontham E, Bamlet WR, Holly EA, Negri E, Hassan M, Prizment A, Cotterchio M, Cleary S, Kurtz RC, Maisonneuve P, Trichopoulos D, Polesel J, Duell EJ, Boffetta P, La Vecchia C, Ghadirian P. Diabetes, antidiabetic medications, and pancreatic cancer risk: an analysis from the International Pancreatic Cancer Case-Control Consortium. *Ann Oncol*. 2014 Oct;25(10):2065-2072. doi: 10.1093/annonc/mdu276. Epub 2014 Jul 23. PMID: 25057164; PMCID: PMC4176453.
32. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024 May-Jun;74(3):229-263. doi: 10.3322/caac.21834. Epub 2024 Apr 4. PMID: 38572751.
33. Bresnick AR, Weber DJ, Zimmer DB. S100 proteins in cancer. *Nat Rev Cancer*. 2015 Feb;15(2):96-109. doi: 10.1038/nrc3893. PMID: 25614008; PMCID: PMC4369764.

-
34. Brozos-Vázquez E, Toledano-Fonseca M, Costa-Fraga N, García-Ortiz MV, Díaz-Lagares Á, Rodríguez-Ariza A, Aranda E, López-López R. Pancreatic cancer biomarkers: A pathway to advance in personalized treatment selection. *Cancer Treat Rev.* 2024 Apr;125:102719. doi: 10.1016/j.ctrv.2024.102719. Epub 2024 Mar 12. PMID: 38490088.
35. Butcher LM, Beck S. Nano-MeDIP-seq Methylome Analysis Using Low DNA Concentrations. *Methods Mol Biol.* 2017;1589:115-138. doi: 10.1007/7651_2015_259. PMID: 26138987.
36. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med.* 2003 Apr 24;348(17):1625-38. doi: 10.1056/NEJMoa021423. PMID: 12711737.
37. Cao L, Huang C, Cui Zhou D, Hu Y, Lih TM, Savage SR, Krug K, Clark DJ, Schnaubelt M, Chen L, da Veiga Leprevost F, Eguez RV, Yang W, Pan J, Wen B, Dou Y, Jiang W, Liao Y, Shi Z, Terekhanova NV, Cao S, Lu RJ, Li Y, Liu R, Zhu H, Ronning P, Wu Y, Wyczalkowski MA, Easwaran H, Danilova L, Mer AS, Yoo S, Wang JM, Liu W, Haibe-Kains B, Thiagarajan M, Jewell SD, Hostetter G, Newton CJ, Li QK, Roehrl MH, Fenyö D, Wang P, Nesvizhskii AI, Mani DR, Omenn GS, Boja ES, Mesri M, Robles AI, Rodriguez H, Bathe OF, Chan DW, Hruban RH, Ding L, Zhang B, Zhang H; Clinical Proteomic Tumor Analysis Consortium. Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell.* 2021 Sep 16;184(19):5031-5052.e26. doi: 10.1016/j.cell.2021.08.023. PMID: 34534465; PMCID: PMC8654574.
38. Chatterjee A, Bararia A, Ganguly D, Mondal PK, Roy P, Banerjee S, Ghosh S, Gulati S, Ghatak S, Chattopadhyay BK, Basu P, Chatterjee A, Sikdar N. DNA methylome in pancreatic cancer identified novel promoter hypermethylation in NPY and FAIM2 genes associated with poor prognosis in Indian patient cohort. *Cancer Cell Int.* 2022 Nov 3;22(1):334. doi: 10.1186/s12935-022-02737-1. PMID: 36329447; PMCID: PMC9635159.
39. Chatterjee A, Rodger EJ, Eccles MR. Epigenetic drivers of tumorigenesis and cancer metastasis. *Semin Cancer Biol.* 2018 Aug;51:149-159. doi: 10.1016/j.semcancer.2017.08.004. Epub 2017 Aug 12. PMID: 28807546.
40. Chatterjee S, Sanjeev BS. Over-representation analysis of angiogenic factors in immunosuppressive mechanisms in neoplasms and neurological conditions during COVID-19. *Microb Pathog.* 2023 Dec;185:106386. doi: 10.1016/j.micpath.2023.106386. Epub 2023 Oct 19. PMID: 37865274.
41. Chauhan N, Dhasmana A, Jaggi M, Chauhan SC, Yallapu MM. miR-205: A Potential Biomedicine for Cancer Therapy. *Cells.* 2020 Aug 25;9(9):1957. doi: 10.3390/cells9091957. PMID: 32854238; PMCID: PMC7564275.
42. Chen Y, Jiang T, Zhang H, Gou X, Han C, Wang J, Chen AT, Ma J, Liu J, Chen Z, Jing X, Lei H, Wang Z, Bao Y, Baqri M, Zhu Y, Bindra RS, Hansen JE, Dou J, Huang C, Zhou J. LRR31 inhibits DNA repair and sensitizes breast cancer brain metastasis to radiation therapy. *Nat*

- Cell Biol. 2020 Oct;22(10):1276-1285. doi: 10.1038/s41556-020-00586-6. Epub 2020 Oct 1. PMID: 33005030; PMCID: PMC7962994.
43. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem.* 1987 Apr;162(1):156-9. doi: 10.1006/abio.1987.9999. PMID: 2440339.
44. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 1994 Aug 11;22(15):2990-7. doi: 10.1093/nar/22.15.2990. PMID: 8065911; PMCID: PMC310266.
45. Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M. DNA methylation: bisulphite modification and analysis. *Nat Protoc.* 2006;1(5):2353-64. doi: 10.1038/nprot.2006.324. PMID: 17406479.
46. Colella S, Shen L, Baggerly KA, Issa JP, Krahe R. Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *Biotechniques.* 2003 Jul;35(1):146-50. doi: 10.2144/03351md01. PMID: 12866414.
47. Debernardi S, O'Brien H, Algahmdi AS, Malats N, Stewart GD, Plješa-Ercegovac M, Costello E, Greenhalf W, Saad A, Roberts R, Ney A, Pereira SP, Kocher HM, Duffy S, Blyuss O, Crnogorac-Jurcevic T. A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer: A case-control study. *PLoS Med.* 2020 Dec 10;17(12):e1003489. doi: 10.1371/journal.pmed.1003489. PMID: 33301466; PMCID: PMC7758047.
48. Del Poggetto E, Ho IL, Balestrieri C, Yen EY, Zhang S, Citron F, Shah R, Corti D, Diaferia GR, Li CY, Loponte S, Carbone F, Hayakawa Y, Valenti G, Jiang S, Sapio L, Jiang H, Dey P, Gao S, Deem AK, Rose-John S, Yao W, Ying H, Rhim AD, Genovese G, Heffernan TP, Maitra A, Wang TC, Wang L, Draetta GF, Carugo A, Natoli G, Viale A. Epithelial memory of inflammation limits tissue damage while promoting pancreatic tumorigenesis. *Science.* 2021 Sep 17;373(6561):eabj0486. doi: 10.1126/science.abj0486. Epub 2021 Sep 17. PMID: 34529467; PMCID: PMC9733946.
49. Delpu Y, Hanoun N, Lulka H, Sicard F, Selves J, Buscail L, Torrisani J, Cordelier P. Genetic and epigenetic alterations in pancreatic carcinogenesis. *Curr Genomics.* 2011 Mar;12(1):15-24. doi: 10.2174/138920211794520132. PMID: 21886451; PMCID: PMC3129039.
50. Demir Karaman E, Işık Z. Multi-Omics Data Analysis Identifies Prognostic Biomarkers across Cancers. *Med Sci (Basel).* 2023 Jun 27;11(3):44. doi: 10.3390/medsci11030044. PMID: 37489460; PMCID: PMC10366886.
51. Deng Z, Fan T, Xiao C, Tian H, Zheng Y, Li C, He J. TGF- β signalling in health, disease, and therapeutics. *Signal Transduct Target Ther.* 2024 Mar 22;9(1):61. doi: 10.1038/s41392-024-01764-w. PMID: 38514615; PMCID: PMC10958066.

52. Dhasmana A, Dhasmana S, Agarwal S, Khan S, Haque S, Jaggi M, Yallapu MM, Chauhan SC. Integrative big transcriptomics data analysis implicates crucial role of MUC13 in pancreatic cancer. *Comput Struct Biotechnol J*. 2023 May 1;21:2845-2857. doi: 10.1016/j.csbj.2023.04.029. PMID: 37216018; PMCID: PMC10192752.
53. Dimitrakopoulos C, Hindupur SK, Häfliger L, Behr J, Montazeri H, Hall MN, Beerenwinkel N. Network-based integration of multi-omics data for prioritizing cancer genes. *Bioinformatics*. 2018 Jul 15;34(14):2441-2448. doi: 10.1093/bioinformatics/bty148. PMID: 29547932; PMCID: PMC6041755.
54. Ding G, Xu X, Li D, Chen Y, Wang W, Ping D, Jia S, Cao L. Fisetin inhibits proliferation of pancreatic adenocarcinoma by inducing DNA damage via RFXAP/KDM4A-dependent histone H3K36 demethylation. *Cell Death Dis*. 2020 Oct 22;11(10):893. doi: 10.1038/s41419-020-03019-2. PMID: 33093461; PMCID: PMC7582166.
55. Distler M, Aust D, Weitz J, Pilarsky C, Grützmann R. Precursor lesions for sporadic pancreatic cancer: PanIN, IPMN, and MCN. *Biomed Res Int*. 2014;2014:474905. doi: 10.1155/2014/474905. Epub 2014 Mar 24. PMID: 24783207; PMCID: PMC3982269.
56. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Gräf S, Johnson N, Herrero J, Tomazou EM, Thorne NP, Bäckdahl L, Herberth M, Howe KL, Jackson DK, Miretti MM, Marioni JC, Birney E, Hubbard TJ, Durbin R, Tavaré S, Beck S. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotechnol*. 2008 Jul;26(7):779-85. doi: 10.1038/nbt1414. PMID: 18612301; PMCID: PMC2644410.
57. Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D, Danenberg PV, Laird PW. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res*. 2000 Apr 15;28(8):E32. doi: 10.1093/nar/28.8.e32. PMID: 10734209; PMCID: PMC102836.
58. Eberle MA, Pfützer R, Pogue-Geile KL, Bronner MP, Crispin D, Kimmey MB, Duerr RH, Kruglyak L, Whitcomb DC, Brentnall TA. A new susceptibility locus for autosomal dominant pancreatic cancer maps to chromosome 4q32-34. *Am J Hum Genet*. 2002 Apr;70(4):1044-8. doi: 10.1086/339692. Epub 2002 Feb 27. PMID: 11870593; PMCID: PMC379100.
59. Ehrlich M, Lacey M. DNA hypomethylation and hemimethylation in cancer. *Adv Exp Med Biol*. 2013;754:31-56. doi: 10.1007/978-1-4419-9967-2_2. PMID: 22956495.
60. Elrakaybi A, Ruess DA, Lübbert M, Quante M, Becker H. Epigenetics in Pancreatic Ductal Adenocarcinoma: Impact on Biology and Utilization in Diagnostics and Treatment. *Cancers (Basel)*. 2022 Nov 30;14(23):5926. doi: 10.3390/cancers14235926. PMID: 36497404; PMCID: PMC9738647.
61. Esteller M. Epigenetic changes in cancer. *F1000 Biol Rep*. 2011;3:9. doi: 10.3410/B3-9. Epub 2011 May 3. PMID: 21655338; PMCID: PMC3100810.

-
62. Fan JB, Gunderson KL, Bibikova M, Yeakley JM, Chen J, Wickham Garcia E, Lebruska LL, Laurent M, Shen R, Barker D. Illumina universal bead arrays. *Methods Enzymol.* 2006;410:57-73. doi: 10.1016/S0076-6879(06)10003-8. PMID: 16938546.
63. Fávoro W.J., dos Santos M.M., Pereira M.M., Garcia P.V., Durán N. Effects of P-MAPA immunotherapy associated with gemcitabine on chemically-induced pancreatic cancer in animal model: New therapeutic perspectives. *Biointerface Res. Appl. Chem.* 2022;12:7540–7555
64. Feng HY, Chen YC. Role of bile acids in carcinogenesis of pancreatic cancer: An old topic with new perspective. *World J Gastroenterol.* 2016 Sep 7;22(33):7463-77. doi: 10.3748/wjg.v22.i33.7463. PMID: 27672269; PMCID: PMC5011662.
65. Feng L, Wang Y, Qin J, Fu Y, Guo Z, Zhang J, He G, Jiang Z, Xu X, Zhou C, Gao Y. UGT1A Gene Family Members Serve as Potential Targets and Prognostic Biomarkers for Pancreatic Cancer. *Biomed Res Int.* 2021 Sep 20;2021:6673125. doi: 10.1155/2021/6673125. PMID: 34595239; PMCID: PMC8478536.
66. Fiorentino G, Visintainer R, Domenici E, Lauria M, Marchetti L. MOUSSE: Multi-Omics Using Subject-Specific SignaturEs. *Cancers (Basel).* 2021 Jul 8;13(14):3423. doi: 10.3390/cancers13143423. PMID: 34298641; PMCID: PMC8304726.
67. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A.* 1992 Mar 1;89(5):1827-31. doi: 10.1073/pnas.89.5.1827. PMID: 1542678; PMCID: PMC48546.
68. Fujikura K, Alruwaili ZI, Haffner MC, Trujillo MA, Roberts NJ, Hong SM, Macgregor-Das A, Goggins MG, Roy S, Meeker AK, Ding D, Wright M, He J, Hruban RH, Wood LD. Downregulation of 5-hydroxymethylcytosine is an early event in pancreatic tumorigenesis. *J Pathol.* 2021 Jul;254(3):279-288. doi: 10.1002/path.5682. Epub 2021 May 21. PMID: 33870509; PMCID: PMC9516434.
69. Fujisawa M, Kanda T, Shibata T, Sasaki R, Masuzaki R, Matsumoto N, Nirei K, Imazu H, Kuroda K, Sugitani M, Takayama T, Moriyama M. Involvement of the Interferon signalling Pathways in Pancreatic Cancer Cells. *Anticancer Res.* 2020 Aug;40(8):4445-4455. doi: 10.21873/anticancer.14449. PMID: 32727774.
70. Fukushima N, Sato N, Ueki T, Rosty C, Walter KM, Wilentz RE, Yeo CJ, Hruban RH, Goggins M. Aberrant methylation of preproenkephalin and p16 genes in pancreatic intraepithelial neoplasia and pancreatic ductal adenocarcinoma. *Am J Pathol.* 2002 May;160(5):1573-81. doi: 10.1016/S0002-9440(10)61104-2. PMID: 12000709; PMCID: PMC1850889.
71. Fukutake N, Ueno M, Hiraoka N, Shimada K, Shiraishi K, Saruki N, Ito T, Yamakado M, Ono N, Imaizumi A, Kikuchi S, Yamamoto H, Katayama K. A Novel Multivariate Index for Pancreatic Cancer Detection Based On the Plasma Free Amino Acid Profile. *PLoS One.* 2015
-

- Jul 2;10(7):e0132223. doi: 10.1371/journal.pone.0132223. PMID: 26133769; PMCID: PMC4489861.
72. Funaki NO, Tanaka J, Kasamatsu T, Ohshio G, Hosotani R, Okino T, Imamura M. Identification of carcinoembryonic antigen mRNA in circulating peripheral blood of pancreatic carcinoma and gastric carcinoma patients. *Life Sci.* 1996;59(25-26):2187-99. doi: 10.1016/s0024-3205(96)00576-0. PMID: 8950323.
73. Gad MM, Găman MA, Saad AM, Al-Husseini MJ, Shehata OA, Saleh MA, Nelson AD, Simons-Linares CR. Temporal trends of incidence and mortality in Asian-Americans with pancreatic adenocarcinoma: an epidemiological study. *Ann Gastroenterol.* 2020 Mar-Apr;33(2):210-218. doi: 10.20524/aog.2020.0450. Epub 2020 Feb 12. PMID: 32127743; PMCID: PMC7049244.
74. Gaidhani RH, Balasubramaniam G. *An epidemiological review of pancreatic cancer with special reference to India.* *Indian J Med Sci* 2021;73(1):99-109.
75. Gandhi S, de la Fuente J, Murad MH, Majumder S. Chronic Pancreatitis Is a Risk Factor for Pancreatic Cancer, and Incidence Increases With Duration of Disease: A Systematic Review and Meta-analysis. *Clin Transl Gastroenterol.* 2022 Mar 28;13(3):e00463. doi: 10.14309/ctg.0000000000000463. PMID: 35142721; PMCID: PMC8963838.
76. Gansauge S, Gansauge F, Ramadani M, Stobbe H, Rau B, Harada N, Beger HG. Overexpression of cyclin D1 in human pancreatic carcinoma is associated with poor prognosis. *Cancer Res.* 1997 May 1;57(9):1634-7. PMID: 9134998.
77. Garcia-Gomez A, Rodríguez-Ubreva J, Ballestar E. Epigenetic interplay between immune, stromal and cancer cells in the tumor microenvironment. *Clin Immunol.* 2018 Nov;196:64-71. doi: 10.1016/j.clim.2018.02.013. Epub 2018 Mar 6. PMID: 29501540.
78. Genkinger JM, Spiegelman D, Anderson KE, Bergkvist L, Bernstein L, van den Brandt PA, English DR, Freudenheim JL, Fuchs CS, Giles GG, Giovannucci E, Hankinson SE, Horn-Ross PL, Leitzmann M, Männistö S, Marshall JR, McCullough ML, Miller AB, Reding DJ, Robien K, Rohan TE, Schatzkin A, Stevens VL, Stolzenberg-Solomon RZ, Verhage BA, Wolk A, Ziegler RG, Smith-Warner SA. Alcohol intake and pancreatic cancer risk: a pooled analysis of fourteen cohort studies. *Cancer Epidemiol Biomarkers Prev.* 2009 Mar;18(3):765-76. doi: 10.1158/1055-9965.EPI-08-0880. Epub 2009 Mar 3. PMID: 19258474; PMCID: PMC2715951.
79. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004;5(10):R80. doi: 10.1186/gb-2004-5-10-r80. Epub 2004 Sep 15. PMID: 15461798; PMCID: PMC545600.

-
80. Gregório C, Soares-Lima SC, Alemar B, Recamonde-Mendoza M, Camuzi D, de Souza-Santos PT, Rivero R, Machado S, Osvaldt A, Ashton-Prolla P, Pinto LFR. Calcium signalling Alterations Caused by Epigenetic Mechanisms in Pancreatic Cancer: From Early Markers to Prognostic Impact. *Cancers (Basel)*. 2020 Jun 30;12(7):1735. doi: 10.3390/cancers12071735. PMID: 32629766; PMCID: PMC7407273.
81. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013 Jun;45(6):580-5. doi: 10.1038/ng.2653. PMID: 23715323; PMCID: PMC4010069.
82. Gu QH, Li D, Xie ZH, Shen QB. The clinical significance of MIA gene in tumorigenesis of lung cancer. *Neoplasma*. 2020 May;67(3):660-667. doi: 10.4149/neo_2020_190511N422. Epub 2020 Jan 27. PMID: 31986890.
83. Guler GD, Ning Y, Ku CJ, Phillips T, McCarthy E, Ellison CK, Bergamaschi A, Collin F, Lloyd P, Scott A, Antoine M, Wang W, Chau K, Ashworth A, Quake SR, Levy S. Detection of early stage pancreatic cancer using 5-hydroxymethylcytosine signatures in circulating cell free DNA. *Nat Commun*. 2020 Oct 19;11(1):5270. doi: 10.1038/s41467-020-18965-w. PMID: 33077732; PMCID: PMC7572413.
84. Gupta S, Wang F, Holly EA, Bracci PM. Risk of pancreatic cancer by alcohol dose, duration, and pattern of consumption, including binge drinking: a population-based study. *Cancer Causes Control*. 2010 Jul;21(7):1047-59. doi: 10.1007/s10552-010-9533-6. Epub 2010 Mar 27. PMID: 20349126; PMCID: PMC2883092.
85. Hagihara A, Miyamoto K, Furuta J, Hiraoka N, Wakazono K, Seki S, Fukushima S, Tsao MS, Sugimura T, Ushijima T. Identification of 27 5' CpG islands aberrantly methylated and 13 genes silenced in human pancreatic cancers. *Oncogene*. 2004 Nov 11;23(53):8705-10. doi: 10.1038/sj.onc.1207783. PMID: 15467763.
86. Hansen KD, Timp W, Bravo HC, Sabuncian S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, Briem E, Zhang K, Irizarry RA, Feinberg AP. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*. 2011 Jun 26;43(8):768-75. doi: 10.1038/ng.865. PMID: 21706001; PMCID: PMC3145050.
87. Hassan MM, Bondy ML, Wolff RA, Abbruzzese JL, Vauthey JN, Pisters PW, Evans DB, Khan R, Chou TH, Lenzi R, Jiao L, Li D. Risk factors for pancreatic cancer: case-control study. *Am J Gastroenterol*. 2007 Dec;102(12):2696-707. doi: 10.1111/j.1572-0241.2007.01510.x. Epub 2007 Aug 31. PMID: 17764494; PMCID: PMC2423805.
88. Henriksen SD, Thorlacius-Ussing O. Cell-Free DNA Methylation as Blood-Based Biomarkers for Pancreatic Adenocarcinoma-A Literature Update. *Epigenomes*. 2021 Apr 9;5(2):8. doi: 10.3390/epigenomes5020008. PMID: 34968295; PMCID: PMC8594668.
89. Heo YJ, Hwa C, Lee GH, Park JM, An JY. Integrative Multi-Omics Approaches in Cancer Research: From Biological Networks to Clinical Subtypes. *Mol Cells*. 2021 Jul 31;44(7):433-443. doi: 10.14348/molcells.2021.0042. PMID: 34238766; PMCID: PMC8334347.
-

90. Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A*. 1996 Sep 3;93(18):9821-6. doi: 10.1073/pnas.93.18.9821. PMID: 8790415; PMCID: PMC38513.
91. Herremans KM, Riner AN, Winn RA, Trevino JG. Diversity and Inclusion in Pancreatic Cancer Clinical Trials. *Gastroenterology*. 2021 Dec;161(6):1741-1746.e3. doi: 10.1053/j.gastro.2021.06.079. Epub 2021 Aug 17. PMID: 34416276; PMCID: PMC8840813.
92. Hezel AF, Kimmelman AC, Stanger BZ, Bardeesy N, Depinho RA. Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev*. 2006 May 15;20(10):1218-49. doi: 10.1101/gad.1415606. PMID: 16702400.
93. Hibino H, Inanobe A, Furutani K, Murakami S, Findlay I, Kurachi Y. Inwardly rectifying potassium channels: their structure, function, and physiological roles. *Physiol Rev*. 2010 Jan;90(1):291-366. doi: 10.1152/physrev.00021.2009. PMID: 20086079.
94. Hong CJ, Yeon J, Yeo BK, Woo H, An HK, Heo W, Kim K, Yu SW. Fas-apoptotic inhibitory molecule 2 localizes to the lysosome and facilitates autophagosome-lysosome fusion through the LC3 interaction region motif-dependent interaction with LC3. *FASEB J*. 2020 Jan;34(1):161-179. doi: 10.1096/fj.201901626R. Epub 2019 Nov 20. PMID: 31914609.
95. Hruban RH, Fukushima N. Pancreatic adenocarcinoma: update on the surgical pathology of carcinomas of ductal origin and PanINs. *Mod Pathol*. 2007 Feb;20 Suppl 1:S61-70. doi: 10.1038/modpathol.3800685. PMID: 17486053.
96. Hruban RH, Klimstra DS. Adenocarcinoma of the pancreas. *Semin Diagn Pathol*. 2014 Nov;31(6):443-451. doi: 10.1053/j.semdp.2014.08.004. Epub 2014 Sep 2. PMID: 25441308; PMCID: PMC4313868.
97. Hruban RH, Petersen GM, Goggins M, Tersmette AC, Offerhaus GJ, Falatko F, Yeo CJ, Kern SE. Familial pancreatic cancer. *Ann Oncol*. 1999;10 Suppl 4:69-73. PMID: 10436789.
98. Hu JX, Zhao CF, Chen WB, Liu QC, Li QW, Lin YY, Gao F. Pancreatic cancer: A review of epidemiology, trend, and risk factors. *World J Gastroenterol*. 2021 Jul 21;27(27):4298-4321. doi: 10.3748/wjg.v27.i27.4298. PMID: 34366606; PMCID: PMC8316912.
99. Huang J, Lucero-Prisno DE 3rd, Zhang L, Xu W, Wong SH, Ng SC, Wong MCS. Updated epidemiology of gastrointestinal cancers in East Asia. *Nat Rev Gastroenterol Hepatol*. 2023 May;20(5):271-287. doi: 10.1038/s41575-022-00726-3. Epub 2023 Jan 11. PMID: 36631716.
100. Iguchi E, Safgren SL, Marks DL, Olson RL, Fernandez-Zapico ME. Pancreatic Cancer, A Mis-interpretor of the Epigenetic Language. *Yale J Biol Med*. 2016 Dec 23;89(4):575-590. PMID: 28018146; PMCID: PMC5168833.
101. Illingworth RS, Bird AP. CpG islands--'a rough guide'. *FEBS Lett*. 2009 Jun 5;583(11):1713-20. doi: 10.1016/j.febslet.2009.04.012. Epub 2009 Apr 18. PMID: 19376112.

-
102. Iodice S, Gandini S, Maisonneuve P, Lowenfels AB. Tobacco and the risk of pancreatic cancer: a review and meta-analysis. *Langenbecks Arch Surg.* 2008 Jul;393(4):535-45. doi: 10.1007/s00423-007-0266-2. Epub 2008 Jan 12. PMID: 18193270.
 103. Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet.* 1994 Aug;7(4):536-40. doi: 10.1038/ng0894-536. PMID: 7951326.
 104. Issa JP. CpG island methylator phenotype in cancer. *Nat Rev Cancer.* 2004 Dec;4(12):988-93. doi: 10.1038/nrc1507. PMID: 15573120.
 105. Iwasaki E, Fukuhara S, Horibe M, Kawasaki S, Seino T, Takimoto Y, Tamagawa H, Machida Y, Kayashima A, Noda M, Hayashi H, Kanai T. Endoscopic Ultrasound-Guided Sampling for Personalized Pancreatic Cancer Treatment. *Diagnostics (Basel).* 2021 Mar 8;11(3):469. doi: 10.3390/diagnostics11030469. PMID: 33800161; PMCID: PMC8000964.
 106. Jahan R, Ganguly K, Smith LM, Atri P, Carmicheal J, Sheinin Y, Rachagani S, Natarajan G, Brand RE, Macha MA, Grandgenett PM, Kaur S, Batra SK. Trefoil factor(s) and CA19.9: A promising panel for early detection of pancreatic cancer. *EBioMedicine.* 2019 Apr;42:375-385. doi: 10.1016/j.ebiom.2019.03.056. Epub 2019 Apr 5. PMID: 30956167; PMCID: PMC6491718.
 107. Jia S, Zhou L, Shen T, Zhou S, Ding G, Cao L. Down-expression of CD36 in pancreatic adenocarcinoma and its correlation with clinicopathological features and prognosis. *J Cancer.* 2018 Jan 1;9(3):578-583. doi: 10.7150/jca.21046. PMID: 29483963; PMCID: PMC5820925.
 108. Jiang H, Hu H, Tong X, Jiang Q, Zhu H, Zhang S. Calcium-binding protein S100P and cancer: mechanisms and clinical relevance. *J Cancer Res Clin Oncol.* 2012 Jan;138(1):1-9. doi: 10.1007/s00432-011-1062-5. Epub 2011 Sep 24. PMID: 21947242.
 109. Johansen D, Stocks T, Jonsson H, Lindkvist B, Björge T, Concin H, Almquist M, Häggström C, Engeland A, Ulmer H, Hallmans G, Selmer R, Nagel G, Tretli S, Stattin P, Manjer J. Metabolic factors and the risk of pancreatic cancer: a prospective analysis of almost 580,000 men and women in the Metabolic Syndrome and Cancer Project. *Cancer Epidemiol Biomarkers Prev.* 2010 Sep;19(9):2307-17. doi: 10.1158/1055-9965.EPI-10-0234. PMID: 20826833.
 110. Johnson KC, Koestler DC, Fleischer T, Chen P, Jenson EG, Marotti JD, Onega T, Kristensen VN, Christensen BC. DNA methylation in ductal carcinoma in situ related with future development of invasive breast cancer. *Clin Epigenetics.* 2015 Jul 25;7(1):75. doi: 10.1186/s13148-015-0094-0. PMID: 26213588; PMCID: PMC4514996.
 111. Jonckheere N, Skrypek N, Van Seuningem I. Mucins and pancreatic cancer. *Cancers (Basel).* 2010 Oct 25;2(4):1794-812. doi: 10.3390/cancers2041794. PMID: 24281201; PMCID: PMC3840449.
 112. Jones BA, Varambally S, Arend RC. Histone Methyltransferase EZH2: A Therapeutic Target for Ovarian Cancer. *Mol Cancer Ther.* 2018 Mar;17(3):591-602. doi: 10.1158/1535-7163.MCT-17-0437. PMID: 29726819; PMCID: PMC5939583.

-
113. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet.* 2002 Jun;3(6):415-28. doi: 10.1038/nrg816. PMID: 12042769.
114. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet.* 2012 May 29;13(7):484-92. doi: 10.1038/nrg3230. PMID: 22641018.
115. Kagan AB, Garrison DA, Anders NM, Webster JA, Baker SD, Yegnasubramanian S, Rudek MA. DNA methyltransferase inhibitor exposure-response: Challenges and opportunities. *Clin Transl Sci.* 2023 Aug;16(8):1309-1322. doi: 10.1111/cts.13548. Epub 2023 Jun 21. PMID: 37345219; PMCID: PMC10432879.
116. Kang HC, Kim JI, Chang HK, Woodard G, Choi YS, Ku JL, Jablons DM, Kim IJ. FAIM2, as a novel diagnostic maker and a potential therapeutic target for small-cell lung cancer and atypical carcinoid. *Sci Rep.* 2016 Sep 28;6:34022. doi: 10.1038/srep34022. PMID: 27677402; PMCID: PMC5039724.
117. Kanwal R, Gupta S. Epigenetic modifications in cancer. *Clin Genet.* 2012 Apr;81(4):303-11. doi: 10.1111/j.1399-0004.2011.01809.x. Epub 2011 Dec 8. PMID: 22082348; PMCID: PMC3590802.
118. Karandish F, Mallik S. Biomarkers and Targeted Therapy in Pancreatic Cancer. *Biomark Cancer.* 2016 Apr 26;8(Suppl 1):27-35. doi: 10.4137/BiC.s34414. PMID: 27147897; PMCID: PMC4847554.
119. Karczewski KJ, Snyder MP. Integrative omics for health and disease. *Nat Rev Genet.* 2018 May;19(5):299-310. doi: 10.1038/nrg.2018.4. Epub 2018 Feb 26. PMID: 29479082; PMCID: PMC5990367.
120. Karpińska M, Czauderna M. Pancreas-Its Functions, Disorders, and Physiological Impact on the Mammals' Organism. *Front Physiol.* 2022 Mar 30;13:807632. doi: 10.3389/fphys.2022.807632. PMID: 35431983; PMCID: PMC9005876.
121. Kato S, Honda K. Use of Biomarkers and Imaging for Early Detection of Pancreatic Cancer. *Cancers (Basel).* 2020 Jul 19;12(7):1965. doi: 10.3390/cancers12071965. PMID: 32707720; PMCID: PMC7409286.
122. Kereszturi E, Szmola R, Kukor Z, Simon P, Weiss FU, Lerch MM, Sahin-Tóth M. Hereditary pancreatitis caused by mutation-induced misfolding of human cationic trypsinogen: a novel disease mechanism. *Hum Mutat.* 2009 Apr;30(4):575-82. doi: 10.1002/humu.20853. PMID: 19191323; PMCID: PMC2663013.
123. Kim GE, Bae HI, Park HU, Kuan SF, Crawley SC, Ho JJ, Kim YS. Aberrant expression of MUC5AC and MUC6 gastric mucins and sialyl Tn antigen in intraepithelial neoplasms of the pancreas. *Gastroenterology.* 2002 Oct;123(4):1052-60. doi: 10.1053/gast.2002.36018. PMID: 12360467.
124. Kim JC, Ha YJ, Tak KH, Roh SA, Kwon YH, Kim CW, Yoon YS, Lee JL, Park Y, Kim SK, Kim SY, Cho DH, Kim YS. Opposite functions of GSN and OAS2 on colorectal cancer metastasis,

- mediating perineural and lymphovascular invasion, respectively. *PLoS One*. 2018 Aug 27;13(8):e0202856. doi: 10.1371/journal.pone.0202856. PMID: 30148861; PMCID: PMC6110496.
125. Kim S, Park BK, Seo JH, Choi J, Choi JW, Lee CK, Chung JB, Park Y, Kim DW. Carbohydrate antigen 19-9 elevation without evidence of malignant or pancreatobiliary diseases. *Sci Rep*. 2020 Jun 1;10(1):8820. doi: 10.1038/s41598-020-65720-8. PMID: 32483216; PMCID: PMC7264353.
 126. Kinugawa Y, Uehara T, Sano K, Matsuda K, Maruyama Y, Kobayashi Y, Nakajima T, Hamano H, Kawa S, Higuchi K, Hosaka N, Shiozawa S, Ishigame H, Ota H. Methylation of Tumor Suppressor Genes in Autoimmune Pancreatitis. *Pancreas*. 2017 May/Jun;46(5):614-618. doi: 10.1097/MPA.0000000000000804. PMID: 28196014.
 127. Kloesch B, Ionasz V, Paliwal S, Hruschka N, Martinez de Villarreal J, Öllinger R, Mueller S, Dienes HP, Schindl M, Gruber ES, Stift J, Herndler-Brandstetter D, Lomberg GA, Seidler B, Saur D, Rad R, Urrutia RA, Real FX, Martinelli P. A GATA6-centred gene regulatory network involving HNFs and Δ Np63 controls plasticity and immune escape in pancreatic cancer. *Gut*. 2022 Apr;71(4):766-777. doi: 10.1136/gutjnl-2020-321397. Epub 2021 Apr 12. PMID: 33846140; PMCID: PMC9733634.
 128. Kobayashi K, Einama T, Takihata Y, Yonamine N, Fujinuma I, Tsunenari T, Kouzu K, Nakazawa A, Iwasaki T, Ueno H, Kishi Y. Therapeutic efficacy of dose-reduced adjuvant chemotherapy with S-1 in patients with pancreatic cancer: a retrospective study. *BMC Cancer*. 2022 Sep 30;22(1):1028. doi: 10.1186/s12885-022-10116-2. PMID: 36180830; PMCID: PMC9524130.
 129. Kohi S, Sato N, Cheng XB, Koga A, Higure A, Hirata K. A novel epigenetic mechanism regulating hyaluronan production in pancreatic cancer cells. *Clin Exp Metastasis*. 2016 Mar;33(3):225-30. doi: 10.1007/s10585-015-9771-9. Epub 2015 Nov 20. PMID: 26589701.
 130. Koni M, Pinnarò V, Brizzi MF. The Wnt signalling Pathway: A Tailored Target in Cancer. *Int J Mol Sci*. 2020 Oct 18;21(20):7697. doi: 10.3390/ijms21207697. PMID: 33080952; PMCID: PMC7589708.
 131. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2014 Nov 15;13:8-17. doi: 10.1016/j.csbj.2014.11.005. PMID: 25750696; PMCID: PMC4348437.
 132. Kumar L, Kumar S, Sandeep K, Patel SKS. Therapeutic Approaches in Pancreatic Cancer: Recent Updates. *Biomedicines*. 2023 Jun 1;11(6):1611. doi: 10.3390/biomedicines11061611. PMID: 37371705; PMCID: PMC10296116.
 133. Kuroki T, Tajima Y, Kanematsu T. Role of hypermethylation on carcinogenesis in the pancreas. *Surg Today*. 2004;34(12):981-6. doi: 10.1007/s00595-004-2858-6. PMID: 15580378.
 134. Laird PW. The power and the promise of DNA methylation markers. *Nat Rev Cancer*. 2003 Apr;3(4):253-66. doi: 10.1038/nrc1045. PMID: 12671664.

135. Larsson SC, Wolk A. Red and processed meat consumption and risk of pancreatic cancer: meta-analysis of prospective studies. *Br J Cancer*. 2012 Jan 31;106(3):603-7. doi: 10.1038/bjc.2011.585. Epub 2012 Jan 12. PMID: 22240790; PMCID: PMC3273353.
136. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, Wei CL. Dynamic changes in the human methylome during differentiation. *Genome Res*. 2010 Mar;20(3):320-31. doi: 10.1101/gr.101907.109. Epub 2010 Feb 4. PMID: 20133333; PMCID: PMC2840979.
137. Lee ES, Lee JM. Imaging diagnosis of pancreatic cancer: a state-of-the-art review. *World J Gastroenterol*. 2014 Jun 28;20(24):7864-77. doi: 10.3748/wjg.v20.i24.7864. PMID: 24976723; PMCID: PMC4069314.
138. Lee T, Teng TZJ, Shelat VG. Carbohydrate antigen 19-9 - tumor marker: Past, present, and future. *World J Gastrointest Surg*. 2020 Dec 27;12(12):468-490. doi: 10.4240/wjgs.v12.i12.468. PMID: 33437400; PMCID: PMC7769746.
139. Leonhardt CS, Kinney-Köster B, Hank T, Habib JR, Shoucair S, Klaiber U, Cameron JL, Hackert T, Wolfgang CL, Büchler MW, He J, Strobel O. Resected Early-Onset Pancreatic Cancer: Practices and Outcomes in an International Dual-Center Study. *Ann Surg Oncol*. 2023 Apr;30(4):2433-2443. doi: 10.1245/s10434-022-12901-6. Epub 2022 Dec 7. PMID: 36479659; PMCID: PMC10027827.
140. Leontiou CA, Hadjidaniel MD, Mina P, Antoniou P, Ioannides M, Patsalis PC. Bisulfite Conversion of DNA: Performance Comparison of Different Kits and Methylation Quantitation of Epigenetic Biomarkers that Have the Potential to Be Used in Non-Invasive Prenatal Testing. *PLoS One*. 2015 Aug 6;10(8):e0135058. doi: 10.1371/journal.pone.0135058. PMID: 26247357; PMCID: PMC4527772.
141. Li D, Morris JS, Liu J, Hassan MM, Day RS, Bondy ML, Abbruzzese JL. Body mass index and risk, age of onset, and survival in patients with pancreatic cancer. *JAMA*. 2009 Jun 24;301(24):2553-62. doi: 10.1001/jama.2009.886. PMID: 19549972; PMCID: PMC2760963.
142. Li D, Tang H, Hassan MM, Holly EA, Bracci PM, Silverman DT. Diabetes and risk of pancreatic cancer: a pooled analysis of three large case-control studies. *Cancer Causes Control*. 2011 Feb;22(2):189-97. doi: 10.1007/s10552-010-9686-3. Epub 2010 Nov 21. PMID: 21104117; PMCID: PMC5312666.
143. Li J, Yuan S, Norgard RJ, Yan F, Sun YH, Kim IK, Merrell AJ, Sela Y, Jiang Y, Bhanu NV, Garcia BA, Vonderheide RH, Blanco A, Stanger BZ. Epigenetic and Transcriptional Control of the Epidermal Growth Factor Receptor Regulates the Tumor Immune Microenvironment in Pancreatic Cancer. *Cancer Discov*. 2021 Mar;11(3):736-753. doi: 10.1158/2159-8290.CD-20-0519. Epub 2020 Nov 6. PMID: 33158848; PMCID: PMC7933070.
144. Li X, Wang B, Tang L, Zhang Y, Chen L, Gu L, Zhang F, Ouyang J, Zhang X. GSTA1 Expression Is Correlated With Aldosterone Level in KCNJ5-Mutated Adrenal Aldosterone-

- Producing Adenoma. *J Clin Endocrinol Metab.* 2018 Mar 1;103(3):813-823. doi: 10.1210/jc.2017-00950. PMID: 29165597.
145. Lian W, Jin H, Cao J, Zhang X, Zhu T, Zhao S, Wu S, Zou K, Zhang X, Zhang M, Zheng X, Peng M. Identification of novel biomarkers affecting the metastasis of colorectal cancer through bioinformatics analysis and validation through qRT-PCR. *Cancer Cell Int.* 2020 Mar 30;20:105. doi: 10.1186/s12935-020-01180-4. PMID: 32256214; PMCID: PMC7106634.
146. Liggett T, Melnikov A, Yi QL, Replogle C, Brand R, Kaul K, Talamonti M, Abrams RA, Levenson V. Differential methylation of cell-free circulating DNA among patients with pancreatic cancer versus chronic pancreatitis. *Cancer.* 2010 Apr 1;116(7):1674-80. doi: 10.1002/cncr.24893. PMID: 20143430.
147. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 2009 Nov 19;462(7271):315-22. doi: 10.1038/nature08514. Epub 2009 Oct 14. PMID: 19829295; PMCID: PMC2857523.
148. Lomberg G, Dusetti N, Iovanna J, Urrutia R. Emerging epigenomic landscapes of pancreatic cancer in the era of precision medicine. *Nat Commun.* 2019 Aug 28;10(1):3875. doi: 10.1038/s41467-019-11812-7. PMID: 31462645; PMCID: PMC6713756
149. Longnecker, Daniel S. 2021. Anatomy and Histology of the Pancreas. *Pancreapedia: Exocrine Pancreas Knowledge Base*, DOI: 10.3998/panc.2021.01
150. Lowenfels AB, Maisonneuve P, DiMaggio EP, Elitsur Y, Gates LK Jr, Perrault J, Whitcomb DC. Hereditary pancreatitis and the risk of pancreatic cancer. International Hereditary Pancreatitis Study Group. *J Natl Cancer Inst.* 1997 Mar 19;89(6):442-6. doi: 10.1093/jnci/89.6.442. PMID: 9091646.
151. Luo W, Wang J, Chen H, Ye L, Qiu J, Liu Y, Wang R, Weng G, Liu T, Su D, Tao J, Ding C, You L, Zhang T. Epidemiology of pancreatic cancer: New version, new vision. *Chin J Cancer Res.* 2023 Oct 30;35(5):438-450. doi: 10.21147/j.issn.1000-9604.2023.05.03. PMID: 37969957; PMCID: PMC10643340.
152. Lüttges J, Klöppel G. Precancerous conditions of pancreatic carcinoma. *J Hepatobiliary Pancreat Surg.* 2000;7(6):568-74. doi: 10.1007/s005340070005. PMID: 11180888.
153. Maitra A, Hruban RH. Pancreatic cancer. *Annu Rev Pathol.* 2008;3:157-88. doi: 10.1146/annurev.pathmechdis.3.121806.154305. PMID: 18039136; PMCID: PMC2666336.
154. Mao Y, Wang W, Yang J, Zhou X, Lu Y, Gao J, Wang X, Wen L, Fu W, Tang F. Drug repurposing screening and mechanism analysis based on human colorectal cancer organoids. *Protein Cell.* 2024 Apr 1;15(4):285-304. doi: 10.1093/procel/pwad038. PMID: 37345888; PMCID: PMC10984622.

-
155. Marques L, Costa B, Pereira M, Silva A, Santos J, Saldanha L, Silva I, Magalhães P, Schmidt S, Vale N. Advancing Precision Medicine: A Review of Innovative In Silico Approaches for Drug Development, Clinical Pharmacology and Personalized Healthcare. *Pharmaceutics*. 2024 Feb 27;16(3):332. doi: 10.3390/pharmaceutics16030332. PMID: 38543226; PMCID: PMC10975777.
156. Matsubayashi H, Canto M, Sato N, Klein A, Abe T, Yamashita K, Yeo CJ, Kalloo A, Hruban R, Goggins M. DNA methylation alterations in the pancreatic juice of patients with suspected pancreatic disease. *Cancer Res*. 2006 Jan 15;66(2):1208-17. doi: 10.1158/0008-5472.CAN-05-2664. PMID: 16424060.
157. Matsumura K, Hayashi H, Uemura N, Ogata Y, Zhao L, Sato H, Shiraishi Y, Kuroki H, Kitamura F, Kaida T, Higashi T, Nakagawa S, Mima K, Imai K, Yamashita YI, Baba H. Thrombospondin-1 overexpression stimulates loss of Smad4 and accelerates malignant behavior via TGF- β signal activation in pancreatic ductal adenocarcinoma. *Transl Oncol*. 2022 Dec;26:101533. doi: 10.1016/j.tranon.2022.101533. Epub 2022 Sep 15. PMID: 36115074; PMCID: PMC9483797.
158. McGuigan A, Kelly P, Turkington RC, Jones C, Coleman HG, McCain RS. Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes. *World J Gastroenterol*. 2018 Nov 21;24(43):4846-4861. doi: 10.3748/wjg.v24.i43.4846. PMID: 30487695; PMCID: PMC6250924.
159. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res*. 2005 Oct 13;33(18):5868-77. doi: 10.1093/nar/gki901. PMID: 16224102; PMCID: PMC1258174.
160. Melnikov AA, Scholtens D, Talamonti MS, Bentrem DJ, Levenson VV. Methylation profile of circulating plasma DNA in patients with pancreatic cancer. *J Surg Oncol*. 2009 Feb 1;99(2):119-22. doi: 10.1002/jso.21208. PMID: 19065635.
161. Menezes S, Okail MH, Jalil SMA, Kocher HM, Cameron AJM. Cancer-associated fibroblasts in pancreatic cancer: new subtypes, new markers, new targets. *J Pathol*. 2022 Jul;257(4):526-544. doi: 10.1002/path.5926. Epub 2022 Jun 7. PMID: 35533046; PMCID: PMC9327514.
162. Michaud DS, Vrieling A, Jiao L, Mendelsohn JB, Steplowski E, Lynch SM, Wactawski-Wende J, Arslan AA, Bas Bueno-de-Mesquita H, Fuchs CS, Gross M, Helzlsouer K, Jacobs EJ, Lacroix A, Petersen G, Zheng W, Allen N, Ammundadottir L, Bergmann MM, Boffetta P, Buring JE, Canzian F, Chanock SJ, Clavel-Chapelon F, Clipp S, Freiberg MS, Michael Gaziano J, Giovannucci EL, Hankinson S, Hartge P, Hoover RN, Allan Hubbell F, Hunter DJ, Hutchinson A, Jacobs K, Kooperberg C, Kraft P, Manjer J, Navarro C, Peeters PH, Shu XO, Stevens V, Thomas G, Tjønneland A, Tobias GS, Trichopoulos D, Tumino R, Vineis P, Virtamo J, Wallace R, Wolpin BM, Yu K, Zeleniuch-Jacquotte A, Stolzenberg-Solomon RZ.

- Alcohol intake and pancreatic cancer: a pooled analysis from the pancreatic cancer cohort consortium (PanScan). *Cancer Causes Control*. 2010 Aug;21(8):1213-25. doi: 10.1007/s10552-010-9548-z. Epub 2010 Apr 7. PMID: 20373013; PMCID: PMC3098295.
163. Midha S, Chawla S, Garg PK. Modifiable and non-modifiable risk factors for pancreatic cancer: A review. *Cancer Lett*. 2016 Oct 10;381(1):269-77. doi: 10.1016/j.canlet.2016.07.022. Epub 2016 Jul 25. PMID: 27461582.
164. Mishra NK, Guda C. Genome-wide DNA methylation analysis reveals molecular subtypes of pancreatic cancer. *Oncotarget*. 2017 Apr 25;8(17):28990-29012. doi: 10.18632/oncotarget.15993. PMID: 28423671; PMCID: PMC5438707.
165. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, Rashid NU, Williams LA, Eaton SC, Chung AH, Smyla JK, Anderson JM, Kim HJ, Bentrem DJ, Talamonti MS, Iacobuzio-Donahue CA, Hollingsworth MA, Yeh JJ. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet*. 2015 Oct;47(10):1168-78. doi: 10.1038/ng.3398. Epub 2015 Sep 7. PMID: 26343385; PMCID: PMC4912058.
166. Nagarajan RP, Zhang B, Bell RJ, Johnson BE, Olshen AB, Sundaram V, Li D, Graham AE, Diaz A, Fouse SD, Smirnov I, Song J, Paris PL, Wang T, Costello JF. Recurrent epimutations activate gene body promoters in primary glioblastoma. *Genome Res*. 2014 May;24(5):761-74. doi: 10.1101/gr.164707.113. Epub 2014 Apr 7. PMID: 24709822; PMCID: PMC4009606.
167. Nasca V, Chiaravalli M, Piro G, Esposito A, Salvatore L, Tortora G, Corbo V, Carbone C. Intraductal Pancreatic Mucinous Neoplasms: A Tumor-Biology Based Approach for Risk Stratification. *Int J Mol Sci*. 2020 Sep 2;21(17):6386. doi: 10.3390/ijms21176386. PMID: 32887490; PMCID: PMC7504137.
168. Natale F, Vivo M, Falco G, Angrisano T. Deciphering DNA methylation signatures of pancreatic cancer and pancreatitis. *Clin Epigenetics*. 2019 Sep 6;11(1):132. doi: 10.1186/s13148-019-0728-8. PMID: 31492175; PMCID: PMC6729090.
169. Neureiter D, Jäger T, Ocker M, Kiesslich T. Epigenetics and pancreatic cancer: pathophysiology and novel treatment aspects. *World J Gastroenterol*. 2014 Jun 28;20(24):7830-48. doi: 10.3748/wjg.v20.i24.7830. PMID: 24976721; PMCID: PMC4069312.
170. Nones K, Waddell N, Song S, Patch AM, Miller D, Johns A, Wu J, Kassahn KS, Wood D, Bailey P, Fink L, Manning S, Christ AN, Nourse C, Kazakoff S, Taylor D, Leonard C, Chang DK, Jones MD, Thomas M, Watson C, Pinese M, Cowley M, Rooman I, Pajic M; APCI; Butturini G, Malpaga A, Corbo V, Crippa S, Falconi M, Zamboni G, Castelli P, Lawlor RT, Gill AJ, Scarpa A, Pearson JV, Biankin AV, Grimmond SM. Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signalling. *Int J Cancer*. 2014 Sep 1;135(5):1110-8. doi: 10.1002/ijc.28765. Epub 2014 May 9. PMID: 24500968.

-
171. Ohuchida K, Mizumoto K, Egami T, Yamaguchi H, Fujii K, Konomi H, Nagai E, Yamaguchi K, Tsuneyoshi M, Tanaka M. S100P is an early developmental marker of pancreatic carcinogenesis. *Clin Cancer Res.* 2006 Sep 15;12(18):5411-6. doi: 10.1158/1078-0432.CCR-06-0298. PMID: 17000674.
172. Ojajärvi IA, Partanen TJ, Ahlbom A, Boffetta P, Hakulinen T, Jourenkova N, Kauppinen TP, Kogevinas M, Porta M, Vainio HU, Weiderpass E, Wesseling CH. Occupational exposures and pancreatic cancer: a meta-analysis. *Occup Environ Med.* 2000 May;57(5):316-24. doi: 10.1136/oem.57.5.316. PMID: 10769297; PMCID: PMC1739949.
173. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet.* 1998 Jul;19(3):219-20. doi: 10.1038/890. PMID: 9662389.
174. Omura N, Goggins M. Epigenetics and epigenetic alterations in pancreatic cancer. *Int J Clin Exp Pathol.* 2009;2(4):310-26. Epub 2008 Nov 15. PMID: 19158989; PMCID: PMC2615589.
175. Omura N, Li CP, Li A, Hong SM, Walter K, Jimeno A, Hidalgo M, Goggins M. Genome-wide profiling of methylated promoters in pancreatic adenocarcinoma. *Cancer Biol Ther.* 2008 Jul;7(7):1146-56. doi: 10.4161/cbt.7.7.6208. Epub 2008 Apr 29. PMID: 18535405; PMCID: PMC2763640.
176. O'Neill RS, Stoita A. Biomarkers in the diagnosis of pancreatic cancer: Are we closer to finding the golden ticket? *World J Gastroenterol.* 2021 Jul 14;27(26):4045-4087. doi: 10.3748/wjg.v27.i26.4045. PMID: 34326612; PMCID: PMC8311531.
177. Pang Y, Zhang Y, Zhang HY, Wang WH, Jin G, Liu JW, Zhu ZJ. MUC13 promotes lung cancer development and progression by activating ERK signalling. *Oncol Lett.* 2022 Jan;23(1):37. doi: 10.3892/ol.2021.13155. Epub 2021 Dec 1. PMID: 34966453; PMCID: PMC8669675.
178. Park M, Lee J, Kim Y, Yi KS, Cho BS, Choi CH, Park KS. Pancreatic serous cystic neoplasm mimicking intraductal papillary mucinous neoplasm: Two case reports and literature review. *Medicine (Baltimore).* 2023 Feb 3;102(5):e32820. doi: 10.1097/MD.00000000000032820. PMID: 36749262; PMCID: PMC9901981.
179. Park W, Chawla A, O'Reilly EM. Pancreatic Cancer: A Review. *JAMA.* 2021 Sep 7;326(9):851-862. doi: 10.1001/jama.2021.13027. Erratum in: *JAMA.* 2021 Nov 23;326(20):2081. doi: 10.1001/jama.2021.19984. PMID: 34547082; PMCID: PMC9363152.
180. Parkkila S, Pan PW, Ward A, Gibadulinova A, Oveckova I, Pastorekova S, Pastorek J, Martinez AR, Helin HO, Isola J. The calcium-binding protein S100P in normal and malignant human tissues. *BMC Clin Pathol.* 2008 Feb 18;8:2. doi: 10.1186/1472-6890-8-2. PMID: 18282279; PMCID: PMC2254630.
181. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci Rep.* 2015 Aug 17;5:13087. doi: 10.1038/srep13087. PMID: 26278466; PMCID: PMC4538374.
-

-
182. Passaro A, Al Bakir M, Hamilton EG, Diehn M, André F, Roy-Chowdhuri S, Mountzios G, Wistuba II, Swanton C, Peters S. Cancer biomarkers: Emerging trends and clinical implications for personalized treatment. *Cell*. 2024 Mar 28;187(7):1617-1635. doi: 10.1016/j.cell.2024.02.041. PMID: 38552610; PMCID: PMC7616034.
183. Patel H, Zanos T, Hewitt DB. Deep Learning Applications in Pancreatic Cancer. *Cancers (Basel)*. 2024 Jan 19;16(2):436. doi: 10.3390/cancers16020436. PMID: 38275877; PMCID: PMC10814475.
184. Patel SH, Edwards MJ, Ahmad SA. Intracellular Ion Channels in Pancreas Cancer. *Cell Physiol Biochem*. 2019;53(S1):44-51. doi: 10.33594/000000193. PMID: 31834994.
185. Patra KC, Bardeesy N, Mizukami Y. Diversity of Precursor Lesions For Pancreatic Cancer: The Genetics and Biology of Intraductal Papillary Mucinous Neoplasm. *Clin Transl Gastroenterol*. 2017 Apr 6;8(4):e86. doi: 10.1038/ctg.2017.3. PMID: 28383565; PMCID: PMC5415899.
186. Peng DF, Kanai Y, Sawada M, Ushijima S, Hiraoka N, Kitazawa S, Hirohashi S. DNA methylation of multiple tumor-related genes in association with overexpression of DNA methyltransferase 1 (DNMT1) during multistage carcinogenesis of the pancreas. *Carcinogenesis*. 2006 Jun;27(6):1160-8. doi: 10.1093/carcin/bgi361. Epub 2006 Mar 14. PMID: 16537562.
187. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Dijk S, Muhlhausler B, Stirzaker C, Clark SJ. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol*. 2016 Oct 7;17(1):208. doi: 10.1186/s13059-016-1066-1. PMID: 27717381; PMCID: PMC5055731.
188. Pogue-Geile KL, Chen R, Bronner MP, Crnogorac-Jurcevic T, Moyes KW, Dowen S, Otey CA, Crispin DA, George RD, Whitcomb DC, Brentnall TA. Palladin mutation causes familial pancreatic cancer and suggests a new cancer mechanism. *PLoS Med*. 2006 Dec;3(12):e516. doi: 10.1371/journal.pmed.0030516. PMID: 17194196; PMCID: PMC1751121.
189. Porta M, López T, Pumarega J, Jariod M, Crous-Bou M, Marco E, Rifà J, Grimalt JO, Malats N, Real FX; PANKRAS II Study Group. In pancreatic ductal adenocarcinoma blood concentrations of some organochlorine compounds and coffee intake are independently associated with KRAS mutations. *Mutagenesis*. 2009 Nov;24(6):513-21. doi: 10.1093/mutage/geb037. Epub 2009 Oct 1. PMID: 19797353.
190. Pradhan S, Talbot D, Sha M, Benner J, Hornstra L, Li E, Jaenisch R, Roberts RJ. Baculovirus-mediated expression and characterization of the full-length murine DNA methyltransferase. *Nucleic Acids Res*. 1997 Nov 15;25(22):4666-73. doi: 10.1093/nar/25.22.4666. PMID: 9358180; PMCID: PMC147102.
-

191. Puckett Y, Garfield K. Pancreatic Cancer. [Updated 2022 Sep 26]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK518996/>
192. Qasim A, Patel JB. ANCA-Associated Vasculitis. [Updated 2024 Aug 31]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK554372/>
193. Qi L, Zhou B, Chen J, Hu W, Bai R, Ye C, Weng X, Zheng S. Significant prognostic values of differentially expressed-aberrantly methylated hub genes in breast cancer. *J Cancer*. 2019 Oct 21;10(26):6618-6634. doi: 10.7150/jca.33433. PMID: 31777591; PMCID: PMC6856906.
194. Quist J, Taylor L, Staaf J, Grigoriadis A. Random Forest Modelling of High-Dimensional Mixed-Type Data for Breast Cancer Classification. *Cancers (Basel)*. 2021 Feb 27;13(5):991. doi: 10.3390/cancers13050991. PMID: 33673506; PMCID: PMC7956671.
195. Ragupathi A, Singh M, Perez AM, Zhang D. Targeting the *BRCA1/2* deficient cancer with PARP inhibitors: Clinical outcomes and mechanistic insights. *Front Cell Dev Biol*. 2023 Mar 22;11:1133472. doi: 10.3389/fcell.2023.1133472. PMID: 37035242; PMCID: PMC10073599.
196. Rahman F, Cotterchio M, Cleary SP, Gallinger S. Association between alcohol consumption and pancreatic cancer risk: a case-control study. *PLoS One*. 2015 Apr 9;10(4):e0124489. doi: 10.1371/journal.pone.0124489. PMID: 25856529; PMCID: PMC4391718.
197. Ramai D, Lanke G, Lai J, Barakat M, Chandan S, Ofosu A, Dhaliwal A, Adler DG. Early- and late-onset pancreatic adenocarcinoma: A population-based comparative study. *Pancreatology*. 2021 Jan;21(1):124-129. doi: 10.1016/j.pan.2020.12.007. Epub 2020 Dec 10. PMID: 33334692.
198. Rastogi T, Devesa S, Mangtani P, Mathew A, Cooper N, Kao R, Sinha R. Cancer incidence rates among South Asians in four geographic regions: India, Singapore, UK and US. *Int J Epidemiol*. 2008 Feb;37(1):147-60. doi: 10.1093/ije/dym219. Epub 2007 Dec 19. PMID: 18094016.
199. Rawla P, Sunkara T, Gaduputi V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World J Oncol*. 2019 Feb;10(1):10-27. doi: 10.14740/wjon1166. Epub 2019 Feb 26. PMID: 30834048; PMCID: PMC6396775.
200. Risch HA, Yu H, Lu L, Kidd MS. ABO blood group, *Helicobacter pylori* seropositivity, and risk of pancreatic cancer: a case-control study. *J Natl Cancer Inst*. 2010 Apr 7;102(7):502-5. doi: 10.1093/jnci/djq007. Epub 2010 Feb 24. PMID: 20181960; PMCID: PMC2902822.
201. Robertson KD, Uzvolgyi E, Liang G, Talmadge C, Sumegi J, Gonzales FA, Jones PA. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Res*. 1999 Jun 1;27(11):2291-8. doi: 10.1093/nar/27.11.2291. PMID: 10325416; PMCID: PMC148793.

-
202. Robertson KD, Wolffe AP. DNA methylation in health and disease. *Nat Rev Genet.* 2000 Oct;1(1):11-9. doi: 10.1038/35049533. PMID: 11262868.
203. Rodríguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. *Nat Med.* 2011 Mar;17(3):330-9. doi: 10.1038/nm.2305. PMID: 21386836.
204. Rosato V, Polesel J, Bosetti C, Serraino D, Negri E, La Vecchia C. Population attributable risk for pancreatic cancer in Northern Italy. *Pancreas.* 2015 Mar;44(2):216-20. doi: 10.1097/MPA.0000000000000251. PMID: 25479588.
205. Rosen RD, Sapra A. TNM Classification. [Updated 2023 Feb 13]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK553187/>
206. Rosty C, Ueki T, Argani P, Jansen M, Yeo CJ, Cameron JL, Hruban RH, Goggins M. Overexpression of S100A4 in pancreatic ductal adenocarcinomas is associated with poor differentiation and DNA hypomethylation. *Am J Pathol.* 2002 Jan;160(1):45-50. doi: 10.1016/S0002-9440(10)64347-7. PMID: 11786397; PMCID: PMC1867115.
207. Saha G, Singh R, Mandal A, Das S, Chattopadhyay E, Panja P, Roy P, DeSarkar N, Gulati S, Ghatak S, Ghosh S, Banerjee S, Roy B, Ghosh S, Chaudhuri D, Arora N, Biswas NK, Sikdar N. A novel hotspot and rare somatic mutation p.A138V, at TP53 is associated with poor survival of pancreatic ductal and periampullary adenocarcinoma patients. *Mol Med.* 2020 Jun 17;26(1):59. doi: 10.1186/s10020-020-00183-1. PMID: 32552660; PMCID: PMC7302128.
208. Saini A, Mastana S, Myers F, Lewis MP. 'From death, lead me to immortality' - mantra of ageing skeletal muscle. *Curr Genomics.* 2013 Jun;14(4):256-67. doi: 10.2174/1389202911314040004. PMID: 24294106; PMCID: PMC3731816.
209. Sasahira T, Kirita T, Nishiguchi Y, Kurihara M, Nakashima C, Bosserhoff AK, Kuniyasu H. A comprehensive expression analysis of the MIA gene family in malignancies: MIA gene family members are novel, useful markers of esophageal, lung, and cervical squamous cell carcinoma. *Oncotarget.* 2016 May 24;7(21):31137-52. doi: 10.18632/oncotarget.9082. PMID: 27145272; PMCID: PMC5058745.
210. Sato N, Maitra A, Fukushima N, van Heek NT, Matsubayashi H, Iacobuzio-Donahue CA, Rosty C, Goggins M. Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma. *Cancer Res.* 2003 Jul 15;63(14):4158-66. PMID: 12874021.
211. Sato N, Matsubayashi H, Fukushima N, Goggins M. The chemokine receptor CXCR4 is regulated by DNA methylation in pancreatic cancer. *Cancer Biol Ther.* 2005 Jan;4(1):70-6. doi: 10.4161/cbt.4.1.1378. Epub 2005 Jan 15. PMID: 15662133.
212. Sato N, Parker AR, Fukushima N, Miyagi Y, Iacobuzio-Donahue CA, Eshleman JR, Goggins M. Epigenetic inactivation of TFPI-2 as a common mechanism associated with growth and invasion of pancreatic ductal adenocarcinoma. *Oncogene.* 2005 Jan 27;24(5):850-8. doi: 10.1038/sj.onc.1208050. PMID: 15592528.
-

-
213. Schawkat K, Manning MA, Glickman JN, Morteale KJ. Pancreatic Ductal Adenocarcinoma and Its Variants: Pearls and Perils. *Radiographics*. 2020 Sep-Oct;40(5):1219-1239. doi: 10.1148/rg.2020190184. Epub 2020 Jul 17. PMID: 32678699.
214. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis*. 2010 Jan;31(1):27-36. doi: 10.1093/carcin/bgp220. Epub 2009 Sep 13. PMID: 19752007; PMCID: PMC2802667.
215. Shaw R. The epigenetics of oral cancer. *Int J Oral Maxillofac Surg*. 2006 Feb;35(2):101-8. doi: 10.1016/j.ijom.2005.06.014. Epub 2005 Sep 8. PMID: 16154320.
216. Shen CN, Goh KS, Huang CR, Chiang TC, Lee CY, Jeng YM, Peng SJ, Chien HJ, Chung MH, Chou YH, Hsieh CC, Kulkarni S, Pasricha PJ, Tien YW, Tang SC. Lymphatic vessel remodeling and invasion in pancreatic cancer progression. *EBioMedicine*. 2019 Sep;47:98-113. doi: 10.1016/j.ebiom.2019.08.044. Epub 2019 Sep 5. PMID: 31495721; PMCID: PMC6796580.
217. Sheng YH, He Y, Hasnain SZ, Wang R, Tong H, Clarke DT, Lourie R, Oancea I, Wong KY, Lumley JW, Florin TH, Sutton P, Hooper JD, McMillan NA, McGuckin MA. MUC13 protects colorectal cancer cells from death by activating the NF- κ B pathway and is a potential therapeutic target. *Oncogene*. 2017 Feb 2;36(5):700-713. doi: 10.1038/onc.2016.241. Epub 2016 Jul 11. PMID: 27399336; PMCID: PMC5541270.
218. Shi J, Kantoff PW, Wooster R, Farokhzad OC. Cancer nanomedicine: progress, challenges and opportunities. *Nat Rev Cancer*. 2017 Jan;17(1):20-37. doi: 10.1038/nrc.2016.108. Epub 2016 Nov 11. PMID: 27834398; PMCID: PMC5575742.
219. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. 2023 Jan;73(1):17-48. doi: 10.3322/caac.21763. PMID: 36633525.
220. Sikdar N, Saha G, Dutta A, Ghosh S, Shrikhande SV, Banerjee S. Genetic Alterations of Periapillary and Pancreatic Ductal Adenocarcinoma: An Overview. *Curr Genomics*. 2018 Sep;19(6):444-463. doi: 10.2174/1389202919666180221160753. PMID: 30258276; PMCID: PMC6128383.
221. Sim W, Lim WM, Hii LW, Leong CO, Mai CW. Targeting pancreatic cancer immune evasion by inhibiting histone deacetylases. *World J Gastroenterol*. 2022 May 14;28(18):1934-1945. doi: 10.3748/wjg.v28.i18.1934. PMID: 35664961; PMCID: PMC9150054.
222. Singh M, Jha RP, Shri N, Bhattacharyya K, Patel P, Dhamnetiya D. Secular trends in incidence and mortality of cervical cancer in India and its states, 1990-2019: data from the Global Burden of Disease 2019 Study. *BMC Cancer*. 2022 Feb 7;22(1):149. doi: 10.1186/s12885-022-09232-w. PMID: 35130853; PMCID: PMC8819855.
223. Sinha V, Shinde S, Saxena S, Thakur S, Walia T, Dixit V, Tiwari AK, Vishvakarma NK, Dwivedi M, Shukla D. A Comprehensive Review of Diagnostic and Therapeutic Strategies for the Management of Pancreatic Cancer. *Crit Rev Oncog*. 2020;25(4):381-404. doi: 10.1615/CritRevOncog.2020035971. PMID: 33639064.
-

-
224. Storz P, Crawford HC. Carcinogenesis of Pancreatic Ductal Adenocarcinoma. *Gastroenterology*. 2020 Jun;158(8):2072-2081. doi: 10.1053/j.gastro.2020.02.059. Epub 2020 Mar 19. PMID: 32199881; PMCID: PMC7282937.
225. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*. 2007 Jan 25;8:25. doi: 10.1186/1471-2105-8-25. PMID: 17254353; PMCID: PMC1796903.
226. Su Q, Zhu EC, Qu YL, Wang DY, Qu WW, Zhang CG, Wu T, Gao ZH. Serum level of co-expressed hub miRNAs as diagnostic and prognostic biomarkers for pancreatic ductal adenocarcinoma. *J Cancer*. 2018 Oct 11;9(21):3991-3999. doi: 10.7150/jca.27697. PMID: 30410604; PMCID: PMC6218787.
227. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021 May;71(3):209-249. doi: 10.3322/caac.21660. Epub 2021 Feb 4. PMID: 33538338.
228. Sung H, Siegel RL, Rosenberg PS, Jemal A. Emerging cancer trends among young adults in the USA: analysis of a population-based cancer registry. *Lancet Public Health*. 2019 Mar;4(3):e137-e147. doi: 10.1016/S2468-2667(18)30267-6. Epub 2019 Feb 4. PMID: 30733056.
229. Syren P, Andersson R, Bauden M, Ansari D. Epigenetic alterations as biomarkers in pancreatic ductal adenocarcinoma. *Scand J Gastroenterol*. 2017 Jun-Jul;52(6-7):668-673. doi: 10.1080/00365521.2017.1301989. Epub 2017 Mar 16. PMID: 28301276.
230. Tan AC, Jimeno A, Lin SH, Wheelhouse J, Chan F, Solomon A, Rajeshkumar NV, Rubio-Viqueira B, Hidalgo M. Characterizing DNA methylation patterns in pancreatic cancer genome. *Mol Oncol*. 2009 Dec;3(5-6):425-38. doi: 10.1016/j.molonc.2009.03.004. Epub 2009 Apr 22. PMID: 19497796; PMCID: PMC5527529.
231. Tanase C, Gheorghisan-Galateanu AA, Popescu ID, Mihai S, Codrici E, Albulescu R, Hinescu ME. CD36 and CD97 in Pancreatic Cancer versus Other Malignancies. *Int J Mol Sci*. 2020 Aug 6;21(16):5656. doi: 10.3390/ijms21165656. PMID: 32781778; PMCID: PMC7460590.
232. 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 Jul 2;47(W1):W556-W560. doi: 10.1093/nar/gkz430. PMID: 31114875; PMCID: PMC6602440.
233. Thingholm LB, Andersen L, Makalic E, Southey MC, Thomassen M, Hansen LL. Strategies for Integrated Analysis of Genetic, Epigenetic, and Gene Expression Variation in Cancer: Addressing the Challenges. *Front Genet*. 2016 Feb 1;7:2. doi: 10.3389/fgene.2016.00002. PMID: 26870081; PMCID: PMC4740898.
234. Toth R, Schiffmann H, Hube-Magg C, Büscheck F, Höflmayer D, Weidemann S, Lebok P, Fraune C, Minner S, Schlomm T, Sauter G, Plass C, Assenov Y, Simon R, Meiners J, Gerhäuser C. Random forest-based modelling to detect biomarkers for prostate cancer progression. *Clin*

- Epigenetics. 2019 Oct 22;11(1):148. doi: 10.1186/s13148-019-0736-8. PMID: 31640781; PMCID: PMC6805338.
235. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* 2021 Sep 27;13(1):152. doi: 10.1186/s13073-021-00968-x. PMID: 34579788; PMCID: PMC8477474.
236. Tripathi S, Tabari A, Mansur A, Dabbara H, Bridge CP, Daye D. From Machine Learning to Patient Outcomes: A Comprehensive Review of AI in Pancreatic Cancer. *Diagnostics (Basel).* 2024 Jan 12;14(2):174. doi: 10.3390/diagnostics14020174. PMID: 38248051; PMCID: PMC10814554.
237. Tu J, Huang Z, Wang Y, Wang M, Yin Z, Mei X, Li M, Li L. Transcriptome analysis of the procession from chronic pancreatitis to pancreatic cancer and metastatic pancreatic cancer. *Sci Rep.* 2021 Feb 9;11(1):3409. doi: 10.1038/s41598-021-83015-4. PMID: 33564087; PMCID: PMC7873308.
238. Turanli B, Yildirim E, Gulfidan G, Arga KY, Sinha R. Current State of "Omics" Biomarkers in Pancreatic Cancer. *J Pers Med.* 2021 Feb 14;11(2):127. doi: 10.3390/jpm11020127. PMID: 33672926; PMCID: PMC7918884.
239. Ueki T, Toyota M, Skinner H, Walter KM, Yeo CJ, Issa JP, Hruban RH, Goggins M. Identification and characterization of differentially methylated CpG islands in pancreatic carcinoma. *Cancer Res.* 2001 Dec 1;61(23):8540-6. PMID: 11731440.
240. Umemoto K, Takahashi H, Morizane C, Yamada I, Shimizu S, Shioji K, Yoshida Y, Motoya M, Mizuno N, Kojima Y, Terashima T, Uesugi K, Ueno M, Furuse J, Akimoto T, Ikeda M. FOLFIRINOX in advanced pancreatic cancer patients with the double-variant type of UGT1A1 *28 and *6 polymorphism: a multicenter, retrospective study. *Cancer Chemother Pharmacol.* 2021 Mar;87(3):397-404. doi: 10.1007/s00280-020-04206-w. Epub 2021 Jan 2. PMID: 33386925; PMCID: PMC7889544.
241. Urrutia G, Salmonson A, Toro-Zapata J, de Assuncao TM, Mathison A, Dusetti N, Iovanna J, Urrutia R, Lomberg G. Combined Targeting of G9a and Checkpoint Kinase 1 Synergistically Inhibits Pancreatic Cancer Cell Growth by Replication Fork Collapse. *Mol Cancer Res.* 2020 Mar;18(3):448-462. doi: 10.1158/1541-7786.MCR-19-0490. Epub 2019 Dec 10. PMID: 31822519; PMCID: PMC7056528.
242. Van der Auwera I, Yu W, Suo L, Van Neste L, van Dam P, Van Marck EA, Pauwels P, Vermeulen PB, Dirix LY, Van Laere SJ. Array-based DNA methylation profiling for breast cancer subtype discrimination. *PLoS One.* 2010 Sep 7;5(9):e12616. doi: 10.1371/journal.pone.0012616. PMID: 20830311; PMCID: PMC2935385.
243. van der Put NM, Gabreëls F, Stevens EM, Smeitink JA, Trijbels FJ, Eskes TK, van den Heuvel LP, Blom HJ. A second common mutation in the methylenetetrahydrofolate reductase gene:

- an additional risk factor for neural-tube defects? *Am J Hum Genet.* 1998 May;62(5):1044-51. doi: 10.1086/301825. PMID: 9545395; PMCID: PMC1377082.
244. Van Wesenbeeck L, Janssens L, Meeuws H, Lagatie O, Stuyver L. Droplet digital PCR is an accurate method to assess methylation status on FFPE samples. *Epigenetics.* 2018;13(3):207-213. doi: 10.1080/15592294.2018.1448679. Epub 2018 Apr 18. PMID: 29527977; PMCID: PMC5997148.
245. Versemann L, Hessmann E, Ulisse M. Epigenetic Therapeutic Strategies to Target Molecular and Cellular Heterogeneity in Pancreatic Cancer. *Visc Med.* 2022 Feb;38(1):11-19. doi: 10.1159/000519859. Epub 2021 Nov 19. PMID: 35291698; PMCID: PMC8874235.
246. Vincent A, Omura N, Hong SM, Jaffe A, Eshleman J, Goggins M. Genome-wide analysis of promoter methylation associated with gene expression profile in pancreatic adenocarcinoma. *Clin Cancer Res.* 2011 Jul 1;17(13):4341-54. doi: 10.1158/1078-0432.CCR-10-3431. Epub 2011 May 24. PMID: 21610144; PMCID: PMC3131423.
247. Virnig BA, Shamliyan T, Tuttle TM, Kane RL, Wilt TJ. Diagnosis and management of ductal carcinoma in situ (DCIS). *Evid Rep Technol Assess (Full Rep).* 2009 Sep;(185):1-549. PMID: 20629475; PMCID: PMC4781639.
248. Vrba L, Futscher BW, Oshiro M, Watts GS, Menashi E, Hu C, Hammad H, Pennington DR, Golconda U, Gavini H, Roe DJ, Shroff RT, Nelson MA. Liquid biopsy, using a novel DNA methylation signature, distinguishes pancreatic adenocarcinoma from benign pancreatic disease. *Clin Epigenetics.* 2022 Feb 22;14(1):28. doi: 10.1186/s13148-022-01246-2. PMID: 35193708; PMCID: PMC8864826.
249. Vrieling A, Bueno-de-Mesquita HB, Boshuizen HC, Michaud DS, Severinsen MT, Overvad K, Olsen A, Tjønneland A, Clavel-Chapelon F, Boutron-Ruault MC, Kaaks R, Rohrmann S, Boeing H, Nöthlings U, Trichopoulou A, Moutsiou E, Dilis V, Palli D, Krogh V, Panico S, Tumino R, Vineis P, van Gils CH, Peeters PH, Lund E, Gram IT, Rodríguez L, Agudo A, Larrañaga N, Sánchez MJ, Navarro C, Barricarte A, Manjer J, Lindkvist B, Sund M, Ye W, Bingham S, Khaw KT, Roddam A, Key T, Boffetta P, Duell EJ, Jenab M, Gallo V, Riboli E. Cigarette smoking, environmental tobacco smoke exposure and pancreatic cancer risk in the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer.* 2010 May 15;126(10):2394-403. doi: 10.1002/ijc.24907. PMID: 19790196.
250. Wang Q, Gu L, Adey A, Radlwimmer B, Wang W, Hovestadt V, Bähr M, Wolf S, Shendure J, Eils R, Plass C, Weichenhan D. Tagmentation-based whole-genome bisulfite sequencing. *Nat Protoc.* 2013 Oct;8(10):2022-32. doi: 10.1038/nprot.2013.118. Epub 2013 Sep 26. PMID: 24071908.
251. Wang Y, Wang Y, Xu C, Liu Y, Huang Z. Identification of Novel Tumor-Microenvironment-Regulating Factor That Facilitates Tumor Immune Infiltration in Colon Cancer. *Mol Ther*

- Nucleic Acids. 2020 Aug 29;22:236-250. doi: 10.1016/j.omtn.2020.08.029. PMID: 33230430; PMCID: PMC7515980.
252. Wang YT, Gou YW, Jin WW, Xiao M, Fang HY. Association between alcohol intake and the risk of pancreatic cancer: a dose-response meta-analysis of cohort studies. *BMC Cancer*. 2016 Mar 12;16:212. doi: 10.1186/s12885-016-2241-1. PMID: 26968702; PMCID: PMC4788838.
253. Wang Z, Embaye KS, Yang Q, Qin L, Zhang C, Liu L, Zhan X, Zhang F, Wang X, Qin S. A Novel Metabolism-Related Signature as a Candidate Prognostic Biomarker for Hepatocellular Carcinoma. *J Hepatocell Carcinoma*. 2021 Mar 16;8:119-132. doi: 10.2147/JHC.S294108. PMID: 33758763; PMCID: PMC7981163.
254. Warshaw AL, Fernández-del Castillo C. Pancreatic carcinoma. *N Engl J Med*. 1992 Feb 13;326(7):455-65. doi: 10.1056/NEJM199202133260706. PMID: 1732772.
255. Waters AM, Der CJ. KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer. *Cold Spring Harb Perspect Med*. 2018 Sep 4;8(9):a031435. doi: 10.1101/cshperspect.a031435. PMID: 29229669; PMCID: PMC5995645.
256. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet*. 2007 Apr;39(4):457-66. doi: 10.1038/ng1990. Epub 2007 Mar 4. PMID: 17334365.
257. Weeks ME, Hariharan D, Petronijevic L, Radon TP, Whiteman HJ, Kocher HM, Timms JF, Lemoine NR, Crnogorac-Jurcevic T. Analysis of the urine proteome in patients with pancreatic ductal adenocarcinoma. *Proteomics Clin Appl*. 2008 Jul;2(7-8):1047-57. doi: 10.1002/prca.200780164. Epub 2008 Jul 10. PMID: 21136905.
258. Wei K, Hackert T. Surgical Treatment of Pancreatic Ductal Adenocarcinoma. *Cancers (Basel)*. 2021 Apr 20;13(8):1971. doi: 10.3390/cancers13081971. PMID: 33923884; PMCID: PMC8074119.
259. Whitcomb DC, Gorry MC, Preston RA, Furey W, Sossenheimer MJ, Ulrich CD, Martin SP, Gates LK Jr, Amann ST, Toskes PP, Liddle R, McGrath K, Uomo G, Post JC, Ehrlich GD. Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nat Genet*. 1996 Oct;14(2):141-5. doi: 10.1038/ng1096-141. PMID: 8841182.
260. Wilentz RE, Geradts J, Maynard R, Offerhaus GJ, Kang M, Goggins M, Yeo CJ, Kern SE, Hruban RH. Inactivation of the p16 (INK4A) tumor-suppressor gene in pancreatic duct lesions: loss of intranuclear expression. *Cancer Res*. 1998 Oct 15;58(20):4740-4. PMID: 9788631.
261. Williams JK, Schwarz JL, Keutgen XM. Surgery for metastatic pancreatic neuroendocrine tumors: a narrative review. *Hepatobiliary Surg Nutr*. 2023 Feb 28;12(1):69-83. doi: 10.21037/hbsn-22-238. Epub 2023 Jan 16. PMID: 36860243; PMCID: PMC9944533.
262. Wolpin BM, Chan AT, Hartge P, Chanock SJ, Kraft P, Hunter DJ, Giovannucci EL, Fuchs CS. ABO blood group and the risk of pancreatic cancer. *J Natl Cancer Inst*. 2009 Mar

- 18;101(6):424-31. doi: 10.1093/jnci/djp020. Epub 2009 Mar 10. PMID: 19276450; PMCID: PMC2657095.
263. Wood LD, Canto MI, Jaffee EM, Simeone DM. Pancreatic Cancer: Pathogenesis, Screening, Diagnosis, and Treatment. *Gastroenterology*. 2022 Aug;163(2):386-402.e1. doi: 10.1053/j.gastro.2022.03.056. Epub 2022 Apr 7. PMID: 35398344; PMCID: PMC9516440.
264. Worm J, Aggerholm A, Guldberg P. In-tube DNA methylation profiling by fluorescence melting curve analysis. *Clin Chem*. 2001;47(7):1183-9. PMID: 11427447.
265. Wu H, Ou S, Zhang H, Huang R, Yu S, Zhao M, Tai S. Advances in biomarkers and techniques for pancreatic cancer diagnosis. *Cancer Cell Int*. 2022 Jun 28;22(1):220. doi: 10.1186/s12935-022-02640-9. PMID: 35761336; PMCID: PMC9237966.
266. Xi T, Zhang G. Integrated analysis of tumor differentiation genes in pancreatic adenocarcinoma. *PLoS One*. 2018 Mar 29;13(3):e0193427. doi: 10.1371/journal.pone.0193427. PMID: 29596435; PMCID: PMC5875763.
267. Xiong Z, Laird PW. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res*. 1997 Jun 15;25(12):2532-4. doi: 10.1093/nar/25.12.2532. PMID: 9171110; PMCID: PMC146738.
268. Xu S, Furukawa T, Kanai N, Sunamura M, Horii A. Abrogation of DUSP6 by hypermethylation in human pancreatic cancer. *J Hum Genet*. 2005;50(4):159-167. doi: 10.1007/s10038-005-0235-y. Epub 2005 Apr 12. PMID: 15824892.
269. Yi JM, Guzzetta AA, Bailey VJ, Downing SR, Van Neste L, Chiappinelli KB, Keeley BP, Stark A, Herrera A, Wolfgang C, Pappou EP, Iacobuzio-Donahue CA, Goggins MG, Herman JG, Wang TH, Baylin SB, Ahuja N. Novel methylation biomarker panel for the early detection of pancreatic cancer. *Clin Cancer Res*. 2013 Dec 1;19(23):6544-6555. doi: 10.1158/1078-0432.CCR-12-3224. Epub 2013 Oct 2. PMID: 24088737; PMCID: PMC4310572.
270. Yokoyama S, Hamada T, Higashi M, Matsuo K, Maemura K, Kurahara H, Horinouchi M, Hiraki T, Sugimoto T, Akahane T, Yonezawa S, Kornmann M, Batra SK, Hollingsworth MA, Tanimoto A. Predicted Prognosis of Patients with Pancreatic Cancer by Machine Learning. *Clin Cancer Res*. 2020 May 15;26(10):2411-2421. doi: 10.1158/1078-0432.CCR-19-1247. Epub 2020 Jan 28. PMID: 31992588.
271. Yorifuji T, Tsuda T, Doi H, Grandjean P. Cancer excess after arsenic exposure from contaminated milk powder. *Environ Health Prev Med*. 2011 May;16(3):164-70. doi: 10.1007/s12199-010-0182-x. Epub 2010 Sep 29. PMID: 21431798; PMCID: PMC3078290.
272. Yousef A, Yousef M, Chowdhury S, Abdilleh K, Knafl M, Edelkamp P, Alfaro-Munoz K, Chacko R, Peterson J, Smaglo BG, Wolff RA, Pant S, Lee MS, Willis J, Overman M, Doss S, Matrisian L, Hurd MW, Snyder R, Katz MHG, Wang H, Maitra A, Shen JP, Zhao D. Impact of KRAS mutations and co-mutations on clinical outcomes in pancreatic ductal adenocarcinoma.

- NPJ Precis Oncol. 2024 Feb 3;8(1):27. doi: 10.1038/s41698-024-00505-0. PMID: 38310130; PMCID: PMC10838312.
273. Yuan J, He H, Chen C, Wu J, Rao J, Yan H. Combined high expression of CD47 and CD68 is a novel prognostic factor for breast cancer patients. *Cancer Cell Int*. 2019 Sep 11;19:238. doi: 10.1186/s12935-019-0957-0. PMID: 31528120; PMCID: PMC6737685.
274. Zhang BL, He N, Huang YB, Song FJ, Chen KX. ABO blood groups and risk of cancer: a systematic review and meta-analysis. *Asian Pac J Cancer Prev*. 2014;15(11):4643-50. doi: 10.7314/apjcp.2014.15.11.4643. PMID: 24969898.
275. Zhang CY, Liu S, Yang M. Clinical diagnosis and management of pancreatic cancer: Markers, molecular mechanisms, and treatment options. *World J Gastroenterol*. 2022 Dec 28;28(48):6827-6845. doi: 10.3748/wjg.v28.i48.6827. PMID: 36632312; PMCID: PMC9827589.
276. Zhang J, Li S, Liu F, Yang K. Role of CD68 in tumor immunity and prognosis prediction in pan-cancer. *Sci Rep*. 2022 May 12;12(1):7844. doi: 10.1038/s41598-022-11503-2. PMID: 35550532; PMCID: PMC9098459.
277. Zhang Y, Ma Y, Wang Y, Mukhopadhyay D, Bi Y, Ji B. Aurora kinase a inhibitor MLN8237 suppresses pancreatic cancer growth. *Pancreatol*. 2022 Jun;22(5):619-625. doi: 10.1016/j.pan.2022.03.019. Epub 2022 Apr 13. PMID: 35550115; PMCID: PMC9189053.
278. Zhao L, Cui Q, Lu Z, Chen J. Aberrant methylation of RASSF2A in human pancreatic ductal adenocarcinoma and its relation to clinicopathologic features. *Pancreas*. 2012 Mar;41(2):206-11. doi: 10.1097/MPA.0b013e318223d1a5. PMID: 21792082.
279. Zhao Z, Li X, Wang F, Xu Y, Liu S, Han Q, Yang Z, Huang W, Yin Z, Liu Q, Tan H, Ma T, Si S, Huang J, Yuan H, Li W, Liu R. Pathogenic genomic alterations in Chinese pancreatic cancer patients and their therapeutical implications. *Cancer Med*. 2023 May;12(10):11672-11685. doi: 10.1002/cam4.5871. Epub 2023 Mar 31. PMID: 36999792; PMCID: PMC10242355.
280. Zheng YT, Yang HY, Li T, Zhao B, Shao TF, Xiang XQ, Cai WM. Amiloride sensitizes human pancreatic cancer cells to erlotinib in vitro through inhibition of the PI3K/AKT signalling pathway. *Acta Pharmacol Sin*. 2015 May;36(5):614-26. doi: 10.1038/aps.2015.4. Epub 2015 Apr 13. PMID: 25864651; PMCID: PMC4422948.
281. Zhou F, Yao H, Ma Z, Hu X. Investigating small molecule compounds targeting psoriasis based on cMAP database and molecular dynamics simulation. *Skin Res Technol*. 2023 Apr;29(4):e13301. doi: 10.1111/srt.13301. PMID: 37113091; PMCID: PMC10234157.
282. Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, Laird PW, Berman BP. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet*. 2018 Apr;50(4):591-602. doi: 10.1038/s41588-018-0073-4. Epub 2018 Apr 2. PMID: 29610480; PMCID: PMC5893360.
283. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level

- datasets. *Nat Commun.* 2019 Apr 3;10(1):1523. doi: 10.1038/s41467-019-09234-6. PMID: 30944313; PMCID: PMC6447622.
284. Zhu Y, Zhang JJ, Zhu R, Zhu Y, Liang WB, Gao WT, Yu JB, Xu ZK, Miao Y. The increase in the expression and hypomethylation of MUC4 gene with the progression of pancreatic ductal adenocarcinoma. *Med Oncol.* 2011 Dec;28 Suppl 1:S175-84. doi: 10.1007/s12032-010-9683-0. Epub 2010 Oct 5. PMID: 20922503.



PUBLICATIONS



SEPTEMBER 16, 2024

PHD THESIS

Department of Life Sciences and Biotechnology, Jadavpur University

9.0: List of Publications

1. **Bararia A**, Dey S, Gulati S, Ghatak S, Ghosh S, Banerjee S, Sikdar N. Differential methylation landscape of pancreatic ductal adenocarcinoma and its precancerous lesions. *Hepatobiliary Pancreat Dis Int.* 2020 Jun;19(3):205-217. doi: 10.1016/j.hbpd.2020.03.010. Epub 2020 Apr 8. PMID: 32312637.
2. Sikdar, N., **Bararia, A.**, Dutta, A., Banerjee, S. (2022). Long Noncoding RNA Acting as Therapeutic Target for Oxidative Stress-Induced Pancreatic Cancer. In: Chakraborti, S. (eds) *Handbook of Oxidative Stress in Cancer: Therapeutic Aspects*. Springer, Singapore. https://doi.org/10.1007/978-981-16-5422-0_77
3. Chatterjee A*, **Bararia A***, Ganguly D, Mondal PK, Roy P, Banerjee S, Ghosh S, Gulati S, Ghatak S, Chattopadhyay BK, Basu P, Chatterjee A, Sikdar N. DNA methylome in pancreatic cancer identified novel promoter hypermethylation in *NPY* and *FAIM2* genes associated with poor prognosis in Indian patient cohort. *Cancer Cell Int.* 2022 Nov 3;22(1):334. doi: 10.1186/s12935-022-02737-1. PMID: 36329447; PMCID: PMC9635159. (*Joint first Co-author)
4. **Bararia A**, Chakraborty P, Roy P, Chattopadhyay BK, Das A, Chatterjee A, Sikdar N. Emerging role of non-invasive and liquid biopsy biomarkers in pancreatic cancer. *World J Gastroenterol.* 2023 Apr 21;29(15):2241-2260. doi: 10.3748/wjg.v29.i15.2241. PMID: 37124888; PMCID: PMC10134423.
5. **Bararia A**, Das A, Mitra S, Banerjee S, Chatterjee A, Sikdar N. Deoxyribonucleic acid methylation driven aberrations in pancreatic cancer-related pathways. *World J Gastrointest Oncol.* 2023 Sep 15;15(9):1505-1519. doi: 10.4251/wjgo.v15.i9.1505. PMID: 37746645; PMCID: PMC10514732.
6. Amlan Das, **Akash Bararia**, Sanghamitra Mukherjee and Nilabja Sikdar. Chronic pancreatitis as a driving factor for pancreatic cancer: An epidemiological understanding (*Accepted in World Journal of Clinical Oncology*)
7. Ranit Seal, **Akash Bararia**, Bitan kumar Chattopadhyay and Nilabja Sikdar. Irreversible electroporation for metastatic pancreatic carcinoma with liver metastasis: what does the evidence say? (*Accepted in World Journal of Clinical Case*)
8. **Akash Bararia**, Agniswar Chakraborty, Nicolas Keestra, Morgan Jones, Gourav Ghosh, Debabrata Ghosh Dastidar, Aniruddha Chatterjee, Sankhadeep Dutta, Sumit Mukherjee and Nilabja Sikdar. *MOGAT2*, *MIA*, and *CD68* as novel putative methylation-driven genes identified using network deconvolution approaches in pancreatic cancer with prognostic implications. (*Under Peer Review in Scientific Reports*)
9. **Akash Bararia**, Agniswar Chakraborty, Gourav Ghosh, Debabrata Ghosh Dastidar, Sumit Mukherjee and Nilabja Sikdar, A multiface approach using supervised algorithms and clinical models to generate high accuracy signatures for pancreatic cancer. (*Under Peer Review in Computers in Biology and Medicine*)
10. **Akash Bararia**, Arunima Maiti, Gourav Ghosh, Debabrata Ghosh Dastidar, Sumit Mukherjee and Nilabja Sikdar, Identification of an Adverse Prognosis Associated Novel Onco-channel gene, *KCNJ5* in Indian Cohort of Pancreatic Cancer (*Under Peer Review in Heliyon*)

10.0: Conferences and Workshops attended.

- Topic: Novel putative methylation-driven genes identified using network deconvolution approaches in pancreatic cancer.
Venue: 13th East Zonal Oncology Symposium at Novotel Kolkata on 10th February 2024.
(Poster Presentation)
- Attended NIOS 2024 organized by Netaji Subhas Chandra Bose Cancer Hospital.
Venue: The Global Conference for Oncology at Westin Hotels and Resorts on 6th and 7th January 2024.
- Topic: DNA methylome in pancreatic cancer identified novel promoter hyper-methylation in *NPY* and *FAIM2* genes associated with poor prognosis in Indian patient cohort
Venue: Society of Biological Chemists, Kolkata Chapter (2022) at Sister Nivedita University, Kolkata. **(Oral presentation)**
- Topic: R in Research Methodology (2022); Venue: Online Mode: Organized by COMMCAD.
- Topic: International Training Workshop on Single Cell Transcriptomics, Organized by NIBMG (INDIA) and MIT (USA) as part of Human Cell Atlas during 17th – 22nd January 2022.



APPENDIX



SEPTEMBER 16, 2024

PHD THESIS

Department of Life Sciences and Biotechnology, Jadavpur University

Table 1: PanIN Classification

Pancreatic Intraepithelial Neoplasia Classification		
Oldest Terminology	WHO 2010 Classification	Baltimore Consensus meeting 2014
Mucinous (goblet cell) metaplasia, flat duct lesion without atypia, mucinous ductal hyperplasia, simple hyperplasia, mucinous cell hyperplasia, flat duct hyperplasia, nonpapillary epithelial hypertrophy	PanIN 1A	Low grade PanIN
Papillary hyperplasia, papillary duct lesion without atypia, ductal hyperplasia, adenomatoid hyperplasia	PanIN 1B	
Atypical hyperplasia, papillary duct lesion with atypia, low grade dysplasia, moderate dysplasia	PanIN 2	
Carcinoma in situ, intraductal carcinoma, high grade dysplasia, severe dysplasia	PanIN 3	High grade PanIN

Table 2: TNM Classification

Tumour staging based on TNM classification Stage	T	N	M
Stage IA	T1	N0	M0
Stage IB	T2	N0	M0
Stage IIA	T3	N0	M0
Stage IIB	T1-T3	N1	M0
Stage III	T4	any N	M0
Stage IV	any T	any N	M1

Table 3: Syndromes and Genes associated with Pancreatic Ductal Adenocarcinoma

Syndrome	Genes	PDAC Risk
Peutz-Jeghers Syndrome	<i>LKB1</i>	35% lifetime
FAMMM syndrome	<i>CDKN2A, CDK4</i>	13-22%
HBOC	<i>BRCA1, BRCA2</i>	2-3.5%
Li-Fraumeni Syndrome	<i>TP53</i>	7.3%
HNCFAP	<i>MLH1, MSH2APC</i>	3.7% 4.5%
Hereditary Pancreatitis	<i>PRSS1, SPINK1</i>	30-40%
Cystic Fibrosis	<i>CFTR</i>	<5%
FPC syndrome	<i>PALB2, BRCA1, ATM, FANCC</i>	2FDR 8-12%, >3FDR 16-38%

Table 4: Primer Sequences

Primers for MSP			
Gene	Methylated forward	Methylated reverse	Amplicon size(bp)
<i>FAIM2</i>	TTATTGGGGTTGTTTCGTAGC	ATAACCGCAACCTACCTACGTC	278
<i>NPY</i>	CGGGTGGGAGTTATTTAAGC	CTCCACAAAACGAAAACTACGT	264
<i>FOXE1</i>	TTTTTGGTAGAGGTTTGAAAAGTTC	CCTTAAAAAAACCACGAAACGTA	275
<i>KCNA6</i>	GGAATTGGGAAGAGTTAGTTTCG	CTTACTAAAATACCCGAAACCG	145
<i>RASSF1</i>	CGAAGTACGGGTTTAATCGG	GCGAACTAAAAACGATAACCAC	193
<i>SIGIRR</i>	TTTTTTAAGTGAGGAGGGCG	ACCTAAATCCCAACTCGAAA	192
	Unmethylated forward	Unmethylated reverse	
<i>FAIM2</i>	TTTATTTATTGGGGTTGTTTTGTAGT	TAAATAACCACAACCTACCTACATC	286
<i>NPY</i>	GTGGGTGGGAGTTATTTAAGTGT	ACTCCACAAAACAAAAAATACATC	266
<i>FOXE1</i>	TTTTGGTAGAGGTTTGAAAAGTTTG	CTCCTTAAAAAAACCACAAAACATA	277
<i>KCNA6</i>	AATTGGGAAGAGTTAGTTTGGT	AAAACCTTACTAAAATACCCAAAACCAC	147
<i>RASSF1</i>	GTGAAGTATGGGTTTAATTGGG	CCCACAACTAAAAACAATAACCAC	197
<i>SIGIRR</i>	TTTTTAAGTGAGGAGGGTGGAG	AACCTAAATCCCAACTCAAAA	192
Primers for gene expression			
Gene	Forward	Reverse	Annealing temperatures
<i>SIGIRR</i>	TCTGTGGCGACTTCATCTGG	ACCGTCCTTGTTCAAGCTCA	59°C
<i>RASSF1</i>	TCTGTGGCGACTTCATCTGG	ACCGTCCTTGTTCAAGCTCA	58°C
<i>KCNA6</i>	AGGATGAAGACGATTCCTACACA	AGGGGAAGATAGCCACCAAGT	58°C
<i>FAIM2</i>	CTAGAACCCACGCGGAGGAA	CCTCCTTCTTCTCGCCATGC	57°C
<i>FOXE1</i>	GACCACGGTGGACTTCTACG	GCGGACACGAACCGATCTAT	57.5°C
<i>NPY</i>	GGAAAACGATCCAGCCCAGA	CAGGGTCTTCAAGCCGAGTT	58°C
<i>SLITRK3</i>	TTCCATAGCTGAGATGCTTCACA	GGAATCGGGGTAGTCCATCC	58°C
<i>IRF4</i>	ACAGCAGTTCTTGTCAGAG	GAGGTTCTACGTGAGCTG	58°C
<i>MX2</i>	CCAGTTCCTCTCAGTCCCAAGATT	TACTGGATGATCAAGGGAACGTGG	59°C
<i>GALR1</i>	CCCCATCATGTCATCCACCT	ATGGGGTTCACCGAGGAGTT	57°C

Table 5: Details of the PAAD Samples in the TCGA Cohort (n=185). The Red Marked Samples Were Excluded from the Analysis. Highlighted Samples were used in Discovery Cohort

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender	CLI_race
1	TCGA-2J-AAB1	stage iib	t3	n1	m0	male	Eu
2	TCGA-2J-AAB4	stage iib	t2	n1	m0	male	Eu
3	TCGA-2J-AAB6	stage iia	t3	n0	m0	male	Eu
4	TCGA-2J-AAB8	stage iib	t3	n1	m0	male	Eu
5	TCGA-2J-AAB9	stage iib	t3	n1	m0	female	Eu
6	TCGA-2J-AABA	stage iib	t3	n1	m0	male	Eu
7	TCGA-2J-AABE	stage iia	t3	n0	m0	male	Eu
8	TCGA-2J-AABF	stage iib	t3	n1	m0	male	Eu
9	TCGA-2J-AABH	stage iia	t3	n0	m0	male	Eu
10	TCGA-2J-AABI	stage iia	t3	n0	m0	female	Eu
11	TCGA-2J-AABK	stage iib	t3	n1	m0	male	Eu
12	TCGA-2J-AABO	stage iib	t3	n1	m0	male	Eu
13	TCGA-2J-AABP	stage iib	t3	n1	m0	female	Eu
14	TCGA-2J-AABR	stage iia	t3	n0	m0	female	Eu
15	TCGA-2J-AABT	stage iib	t3	n1	m0	female	Eu
16	TCGA-2J-AABU	stage iib	t3	n1	m0	male	Eu
17	TCGA-2J-AABV	stage iib	t3	n1	m0	male	Eu
18	TCGA-2L-AAQA	stage iib	t3	n1	mx	male	Eu
19	TCGA-2L-AAQE	stage iib	t3	n1	m0	male	Eu
20	TCGA-2L-AAQI	stage iib	t3	n1	mx	male	Eu
21	TCGA-2L-AAQJ	stage iii	t4	n0	mx	female	Eu
22	TCGA-2L-AAQL	stage iib	t3	n1	mx	male	Eu
23	TCGA-2L-AAQM	stage iib	t3	n1	mx	male	Eu
24	TCGA-3A-A9I5	NA	NA	n0	mx	male	Eu
25	TCGA-3A-A9I7	stage iib	t3	n1	mx	male	AA
26	TCGA-3A-A9I9	stage iia	t3	n0	mx	male	Eu
27	TCGA-3A-A9IB	stage iib	t3	n1	mx	female	Eu
28	TCGA-3A-A9IC	stage iib	t3	n1	mx	female	Eu
29	TCGA-3A-A9IH	stage ia	t1	n0	mx	female	Eu
30	TCGA-3A-A9IJ	stage ib	t2	n0	mx	male	Eu

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender	CLI_race
31	TCGA-3A-A9IL	stage i	t1	NA	mx	female	Eu
32	TCGA-3A-A9IN	stage ib	t2	nx	mx	female	Eu
33	TCGA-3A-A9IO	NA	tx	nx	mx	male	Eu
34	TCGA-3A-A9IR	stage ib	t2	n0	mx	female	Eu
35	TCGA-3A-A9IS	stage ib	t2	n0	mx	male	Eu
36	TCGA-3A-A9IU	stage iib	t3	n1	mx	male	Eu
37	TCGA-3A-A9IV	stage ib	t2	n0	mx	female	Eu
38	TCGA-3A-A9IX	stage ia	t1	n0	mx	male	Eu
39	TCGA-3A-A9IZ	stage iib	t3	n1	mx	male	Eu
40	TCGA-3A-A9J0	stage iib	t3	n1	mx	male	Eu
41	TCGA-3E-AAAY	stage iib	t3	n1	mx	male	Eu
42	TCGA-3E-AAAZ	stage iia	t3	n0	mx	male	Eu
43	TCGA-F2-6879	stage iib	t3	n1	m0	male	Eu
44	TCGA-F2-6880	stage iib	t3	n1	m0	male	Eu
45	TCGA-F2-7273	stage iib	t3	n1	m0	male	AA
46	TCGA-F2-7276	stage iib	t3	n1	m0	male	Eu
47	TCGA-F2-A44G	stage iib	t3	n1	m0	female	Eu
48	TCGA-F2-A44H	stage iia	t3	n0	m0	male	Eu
49	TCGA-F2-A7TX	stage iib	t3	n1	m0	male	Eu
50	TCGA-F2-A8YN	stage iia	t3	n0	m0	male	Eu
51	TCGA-FB-A4P5	stage iib	t3	n1	mx	female	Eu
52	TCGA-FB-A4P6	stage iib	t3	n1	mx	male	Eu
53	TCGA-FB-A545	stage iib	t3	n1	m0	female	Eu
54	TCGA-FB-A5VM	stage ib	t2	n0	m0	male	Eu
55	TCGA-FB-A78T	stage iib	t3	n1	m0	female	Eu
56	TCGA-FB-A7DR	stage iia	t3	nx	m0	male	As
57	TCGA-FB-AAPP	stage iib	t3	n1	mx	male	Eu
58	TCGA-FB-AAPQ	stage iib	t3	n1	mx	male	Eu
59	TCGA-FB-AAPS	stage iib	t2	n1	mx	female	Eu
60	TCGA-FB-AAPU	stage iia	t3	n0	mx	female	As
61	TCGA-FB-AAPY	stage iib	t2	n1	mx	male	Eu
62	TCGA-FB-AAPZ	stage iib	t2	n1	mx	male	Eu
63	TCGA-FB-AAQ0	stage iia	t3	n0	mx	male	Eu

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender	CLI_race
64	TCGA-FB-AAQ1	stage iib	t3	n1	mx	male	Eu
65	TCGA-FB-AAQ2	stage iib	t3	n1	mx	female	Eu
66	TCGA-FB-AAQ3	stage iib	t3	n1	mx	female	Eu
67	TCGA-FB-AAQ6	stage ia	t1	n0	mx	male	Eu
68	TCGA-FZ-5919	stage iib	t3	n1	mx	female	Eu
69	TCGA-FZ-5920	stage iib	t3	n1	m0	male	AA
70	TCGA-FZ-5921	stage iib	t3	n1	m0	female	NA
71	TCGA-FZ-5922	stage iia	t3	n0	m0	male	Eu
72	TCGA-FZ-5923	stage iv	t4	n1	m1	male	Eu
73	TCGA-FZ-5924	stage iia	t3	n1	m0	male	Eu
74	TCGA-FZ-5926	stage iii	t3	n1	m0	female	Eu
75	TCGA-H6-8124	stage iib	t3	n1	m0	female	NA
76	TCGA-H6-A45N	stage iib	t3	n1	mx	female	Eu
77	TCGA-H8-A6C1	stage iia	t3	n0	mx	male	Eu
78	TCGA-HV-A5A3	stage iia	t3	nx	mx	male	As
79	TCGA-HV-A5A4	stage iia	t3	n0	mx	female	As
80	TCGA-HV-A5A5	stage iib	t3	n1	m0	female	As
81	TCGA-HV-A5A6	stage iib	t3	n1b	m0	female	As
82	TCGA-HV-A7OL	stage iia	t3	n0	mx	male	As
83	TCGA-HV-A7OP	stage iib	t3	n1	m0	male	As
84	TCGA-HV-AA8V	stage iib	t3	n1	mx	male	As
85	TCGA-HV-AA8X	stage iib	t2	n1	mx	female	As
86	TCGA-HZ-7289	stage iib	t3	n1	mx	male	Eu
87	TCGA-HZ-7918	stage iib	t3	n1	m0	male	Eu
88	TCGA-HZ-7919	stage iib	t3	n1	m0	female	Eu
89	TCGA-HZ-7920	stage ib	t2	n0	mx	male	Eu
90	TCGA-HZ-7922	stage iib	t3	n1	mx	female	Eu
91	TCGA-HZ-7923	stage iia	t3	n0	mx	male	Eu
92	TCGA-HZ-7924	stage iia	t3	n0	mx	female	Eu
93	TCGA-HZ-7925	stage iib	t3	n1	mx	male	Eu
94	TCGA-HZ-7926	stage iib	t3	n1	mx	male	AA
95	TCGA-HZ-8001	stage iii	t4	n0	mx	male	Eu
96	TCGA-HZ-8002	stage iib	t3	n1	mx	male	Eu

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender	CLI_race
97	TCGA-HZ-8003	stage iib	t3	n1	mx	female	Eu
98	TCGA-HZ-8005	stage iib	t3	n1	mx	male	Eu
99	TCGA-HZ-8315	stage iia	t3	n0	mx	female	Eu
100	TCGA-HZ-8317	stage iib	t3	n1	mx	female	Eu
101	TCGA-HZ-8519	stage ia	t1	n0	mx	male	Eu
102	TCGA-HZ-8636	stage iv	t3	n0	m1	female	Eu
103	TCGA-HZ-8637	stage iib	t3	n1	mx	female	Eu
104	TCGA-HZ-8638	stage ia	t1	n0	mx	female	Eu
105	TCGA-HZ-A49G	stage iib	t2	n1	mx	female	Eu
106	TCGA-HZ-A49H	stage iib	t3	n1	mx	female	Eu
107	TCGA-HZ-A49I	stage iib	t3	n1	mx	male	Eu
108	TCGA-HZ-A4BH	stage iib	t3	n1b	mx	male	Eu
109	TCGA-HZ-A4BK	stage iib	t3	n1	mx	male	Eu
110	TCGA-HZ-A77O	stage iib	t2	n1	mx	female	Eu
111	TCGA-HZ-A77P	stage iib	t3	n1	mx	male	Eu
112	TCGA-HZ-A77Q	stage iib	t3	n1	mx	female	Eu
113	TCGA-HZ-A8P0	stage iib	t3	n1	mx	male	Eu
114	TCGA-HZ-A8P1	stage ib	t2	n0	mx	male	Eu
115	TCGA-HZ-A9TJ	stage iv	t3	n0	m1	male	NA
116	TCGA-IB-7644	stage iv	t3	n1	m1	female	Eu
117	TCGA-IB-7645	stage iib	t3	n1	m0	female	NA
118	TCGA-IB-7646	stage iib	t3	n1	m0	male	Eu
119	TCGA-IB-7647	stage iib	t3	n1	m0	male	Eu
120	TCGA-IB-7649	stage iib	t3	n1	m0	female	NA
121	TCGA-IB-7651	stage iib	t3	n1	m0	female	Eu
122	TCGA-IB-7652	stage iib	t3	n1	m0	female	As
123	TCGA-IB-7654	stage iib	t2	n1	m0	male	Eu
124	TCGA-IB-7885	stage iib	t3	n1	m0	female	Eu
125	TCGA-IB-7886	stage iib	t3	n1	m0	male	Eu
126	TCGA-IB-7887	stage iib	t3	n1	m0	female	Eu
127	TCGA-IB-7888	stage iia	t3	n0	m0	female	AA
128	TCGA-IB-7889	stage iib	t3	n1	m0	female	Eu
129	TCGA-IB-7890	stage ib	t2	n0	m0	male	NA

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender	CLI_race
130	TCGA-IB-7891	stage iib	t3	n1	m0	female	Eu
131	TCGA-IB-7893	stage iia	t3	n0	m0	male	Eu
132	TCGA-IB-7897	stage iib	t3	n1	m0	female	Eu
133	TCGA-IB-8126	stage iii	t4	n1	m0	female	Eu
134	TCGA-IB-8127	stage iib	t3	n1	m0	male	Eu
135	TCGA-IB-A5SO	stage iib	t3	n1	m0	male	Eu
136	TCGA-IB-A5SP	stage iia	t3	n0	m0	male	Eu
137	TCGA-IB-A5SQ	stage ib	t2	n0	m0	female	Eu
138	TCGA-IB-A5SS	stage iib	t3	n1	m0	female	Eu
139	TCGA-IB-A5ST	stage iib	t3	n1	m0	female	Eu
140	TCGA-IB-A6UF	stage iib	t3	n1	m0	male	Eu
141	TCGA-IB-A6UG	stage iib	t3	n1	m0	male	Eu
142	TCGA-IB-A7LX	stage iib	t3	n1	mx	male	Eu
143	TCGA-IB-A7M4	stage iib	t3	n1	mx	male	Eu
144	TCGA-IB-AAUM	stage iib	t2	n1	m0	male	Eu
145	TCGA-IB-AAUN	stage ib	t2	n0	m0	female	Eu
146	TCGA-IB-AAUO	stage iib	t3	n1	m0	female	Eu
147	TCGA-IB-AAUP	stage iib	t3	n1	m0	male	Eu
148	TCGA-IB-AAUQ	stage iib	t3	n1	m0	male	Eu
149	TCGA-IB-AAUR	stage iib	t3	n1	m0	male	Eu
150	TCGA-IB-AAUS	stage iib	t3	n1	m0	female	Eu
151	TCGA-IB-AAUT	stage iib	t2	n1	m0	male	Eu
152	TCGA-IB-AAUU	stage iib	t3	n1	m0	male	Eu
153	TCGA-IB-AAUV	stage iib	t3	n1	m0	male	Eu
154	TCGA-IB-AAUW	stage iib	t3	n1	m0	female	Eu
155	TCGA-L1-A7W4	stage iib	t3	n1	m0	female	Eu
156	TCGA-LB-A7SX	stage iib	t3	n1	mx	female	AA
157	TCGA-LB-A8F3	stage iia	t3	n0	mx	female	AA
158	TCGA-LB-A9Q5	stage iib	t3	n1	mx	female	Eu
159	TCGA-M8-A5N4	stage iia	t3	n0	m0	female	NA
160	TCGA-OE-A75W	stage iia	t3	n0	m0	male	Eu
161	TCGA-PZ-A5RE	stage iib	t3	n1	mx	female	Eu
162	TCGA-Q3-A5QY	stage iib	t3	n1	mx	male	Eu

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender	CLI_race
163	TCGA-Q3-AA2A	stage ib	t3	n1	mx	female	Eu
164	TCGA-RB-A7B8	stage iib	t3	n1	mx	female	Eu
165	TCGA-RB-AA9M	stage iib	t1	n1	mx	male	Eu
166	TCGA-RL-AAAS	stage ib	t2	n0	m0	female	Eu
167	TCGA-S4-A8RM	stage iib	t3	n1	mx	male	Eu
168	TCGA-S4-A8RO	stage iib	t3	n1	mx	female	Eu
169	TCGA-S4-A8RP	stage iib	t3	n1	mx	female	Eu
170	TCGA-US-A774	stage iib	t3	n1	mx	female	Eu
171	TCGA-US-A776	stage iia	t3	n0	mx	male	Eu
172	TCGA-US-A779	stage iib	t3	n1b	mx	female	Eu
173	TCGA-US-A77E	stage iib	t3	n1	mx	male	Eu
174	TCGA-US-A77G	stage iib	t3	n1	mx	male	Eu
175	TCGA-US-A77J	stage iib	t3	n1b	mx	female	Eu
176	TCGA-XD-AAUG	stage iv	t3	n1	m1	female	Eu
177	TCGA-XD-AAUH	stage iib	t3	n1	m0	female	Eu
178	TCGA-XD-AAUI	stage iib	t3	n1	mx	female	Eu
179	TCGA-XD-AAUL	stage iia	t3	n0	mx	male	Eu
180	TCGA-XN-A8T3	stage ib	t2	n0	m0	male	Eu
181	TCGA-XN-A8T5	stage ib	t2	n0	m0	female	Eu
182	TCGA-YB-A89D	stage iib	t3	n1	mx	male	Eu
183	TCGA-YH-A8SY	stage iii	t3	n1	mx	female	Eu
184	TCGA-YY-A8LH	stage iib	t3	n1	mx	female	Eu
185	TCGA-Z5-AAPL	stage iia	t3	n0	m0	female	Eu

Table 6: Details of the PAAD Samples in the TCGA Cohort (n=185). The Red Marked Samples Were Excluded from the Analysis.

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender
1	TCGA-2J-AAB1	stage iib	t3	n1	m0	male
2	TCGA-2J-AAB4	stage iib	t2	n1	m0	male
3	TCGA-2J-AAB6	stage iia	t3	n0	m0	male
4	TCGA-2J-AAB8	stage iib	t3	n1	m0	male
5	TCGA-2J-AAB9	stage iib	t3	n1	m0	female
6	TCGA-2J-AABA	stage iib	t3	n1	m0	male
7	TCGA-2J-AABE	stage iia	t3	n0	m0	male
8	TCGA-2J-AABF	stage iib	t3	n1	m0	male
9	TCGA-2J-AABH	stage iia	t3	n0	m0	male
10	TCGA-2J-AABI	stage iia	t3	n0	m0	female
11	TCGA-2J-AABK	stage iib	t3	n1	m0	male
12	TCGA-2J-AABO	stage iib	t3	n1	m0	male
13	TCGA-2J-AABP	stage iib	t3	n1	m0	female
14	TCGA-2J-AABR	stage iia	t3	n0	m0	female
15	TCGA-2J-AABT	stage iib	t3	n1	m0	female
16	TCGA-2J-AABU	stage iib	t3	n1	m0	male
17	TCGA-2J-AABV	stage iib	t3	n1	m0	male
18	TCGA-2L-AAQA	stage iib	t3	n1	mx	male
19	TCGA-2L-AAQE	stage iib	t3	n1	m0	male
20	TCGA-2L-AAQI	stage iib	t3	n1	mx	male
21	TCGA-2L-AAQJ	stage iii	t4	n0	mx	female
22	TCGA-2L-AAQL	stage iib	t3	n1	mx	male
23	TCGA-2L-AAQM	stage iib	t3	n1	mx	male
24	TCGA-3A-A9I5	NA	NA	n0	mx	male
25	TCGA-3A-A9I7	stage iib	t3	n1	mx	male
26	TCGA-3A-A9I9	stage iia	t3	n0	mx	male
27	TCGA-3A-A9IB	stage iib	t3	n1	mx	female

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender
28	TCGA-3A-A9IC	stage iib	t3	n1	mx	female
29	TCGA-3A-A9IH	stage ia	t1	n0	mx	female
30	TCGA-3A-A9IJ	stage ib	t2	n0	mx	male
31	TCGA-3A-A9IL	stage i	t1	NA	mx	female
32	TCGA-3A-A9IN	stage ib	t2	nx	mx	female
33	TCGA-3A-A9IO	NA	tx	nx	mx	male
34	TCGA-3A-A9IR	stage ib	t2	n0	mx	female
35	TCGA-3A-A9IS	stage ib	t2	n0	mx	male
36	TCGA-3A-A9IU	stage iib	t3	n1	mx	male
37	TCGA-3A-A9IV	stage ib	t2	n0	mx	female
38	TCGA-3A-A9IX	stage ia	t1	n0	mx	male
39	TCGA-3A-A9IZ	stage iib	t3	n1	mx	male
40	TCGA-3A-A9J0	stage iib	t3	n1	mx	male
41	TCGA-3E-AAAY	stage iib	t3	n1	mx	male
42	TCGA-3E-AAAZ	stage iia	t3	n0	mx	male
43	TCGA-F2-6879	stage iib	t3	n1	m0	male
44	TCGA-F2-6880	stage iib	t3	n1	m0	male
45	TCGA-F2-7273	stage iib	t3	n1	m0	male
46	TCGA-F2-7276	stage iib	t3	n1	m0	male
47	TCGA-F2-A44G	stage iib	t3	n1	m0	female
48	TCGA-F2-A44H	stage iia	t3	n0	m0	male
49	TCGA-F2-A7TX	stage iib	t3	n1	m0	male
50	TCGA-F2-A8YN	stage iia	t3	n0	m0	male
51	TCGA-FB-A4P5	stage iib	t3	n1	mx	female
52	TCGA-FB-A4P6	stage iib	t3	n1	mx	male
53	TCGA-FB-A545	stage iib	t3	n1	m0	female
54	TCGA-FB-A5VM	stage ib	t2	n0	m0	male
55	TCGA-FB-A78T	stage iib	t3	n1	m0	female

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender
56	TCGA-FB-A7DR	stage iia	t3	nx	m0	male
57	TCGA-FB-AAPP	stage iib	t3	n1	mx	male
58	TCGA-FB-AAPQ	stage iib	t3	n1	mx	male
59	TCGA-FB-AAPS	stage iib	t2	n1	mx	female
60	TCGA-FB-AAPU	stage iia	t3	n0	mx	female
61	TCGA-FB-AAPY	stage iib	t2	n1	mx	male
62	TCGA-FB-AAPZ	stage iib	t2	n1	mx	male
63	TCGA-FB-AAQ0	stage iia	t3	n0	mx	male
64	TCGA-FB-AAQ1	stage iib	t3	n1	mx	male
65	TCGA-FB-AAQ2	stage iib	t3	n1	mx	female
66	TCGA-FB-AAQ3	stage iib	t3	n1	mx	female
67	TCGA-FB-AAQ6	stage ia	t1	n0	mx	male
68	TCGA-FZ-5919	stage iib	t3	n1	mx	female
69	TCGA-FZ-5920	stage iib	t3	n1	m0	male
70	TCGA-FZ-5921	stage iib	t3	n1	m0	female
71	TCGA-FZ-5922	stage iia	t3	n0	m0	male
72	TCGA-FZ-5923	stage iv	t4	n1	m1	male
73	TCGA-FZ-5924	stage iia	t3	n1	m0	male
74	TCGA-FZ-5926	stage iii	t3	n1	m0	female
75	TCGA-H6-8124	stage iib	t3	n1	m0	female
76	TCGA-H6-A45N	stage iib	t3	n1	mx	female
77	TCGA-H8-A6C1	stage iia	t3	n0	mx	male
78	TCGA-HV-A5A3	stage iia	t3	nx	mx	male
79	TCGA-HV-A5A4	stage iia	t3	n0	mx	female
80	TCGA-HV-A5A5	stage iib	t3	n1	m0	female
81	TCGA-HV-A5A6	stage iib	t3	n1b	m0	female
82	TCGA-HV-A7OL	stage iia	t3	n0	mx	male
83	TCGA-HV-A7OP	stage iib	t3	n1	m0	male

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender
84	TCGA-HV-AA8V	stage iib	t3	n1	mx	male
85	TCGA-HV-AA8X	stage iib	t2	n1	mx	female
86	TCGA-HZ-7289	stage iib	t3	n1	mx	male
87	TCGA-HZ-7918	stage iib	t3	n1	m0	male
88	TCGA-HZ-7919	stage iib	t3	n1	m0	female
89	TCGA-HZ-7920	stage ib	t2	n0	mx	male
90	TCGA-HZ-7922	stage iib	t3	n1	mx	female
91	TCGA-HZ-7923	stage iia	t3	n0	mx	male
92	TCGA-HZ-7924	stage iia	t3	n0	mx	female
93	TCGA-HZ-7925	stage iib	t3	n1	mx	male
94	TCGA-HZ-7926	stage iib	t3	n1	mx	male
95	TCGA-HZ-8001	stage iii	t4	n0	mx	male
96	TCGA-HZ-8002	stage iib	t3	n1	mx	male
97	TCGA-HZ-8003	stage iib	t3	n1	mx	female
98	TCGA-HZ-8005	stage iib	t3	n1	mx	male
99	TCGA-HZ-8315	stage iia	t3	n0	mx	female
100	TCGA-HZ-8317	stage iib	t3	n1	mx	female
101	TCGA-HZ-8519	stage ia	t1	n0	mx	male
102	TCGA-HZ-8636	stage iv	t3	n0	m1	female
103	TCGA-HZ-8637	stage iib	t3	n1	mx	female
104	TCGA-HZ-8638	stage ia	t1	n0	mx	female
105	TCGA-HZ-A49G	stage iib	t2	n1	mx	female
106	TCGA-HZ-A49H	stage iib	t3	n1	mx	female
107	TCGA-HZ-A49I	stage iib	t3	n1	mx	male
108	TCGA-HZ-A4BH	stage iib	t3	n1b	mx	male
109	TCGA-HZ-A4BK	stage iib	t3	n1	mx	male
110	TCGA-HZ-A77O	stage iib	t2	n1	mx	female
111	TCGA-HZ-A77P	stage iib	t3	n1	mx	male

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender
112	TCGA-HZ-A77Q	stage iib	t3	n1	mx	female
113	TCGA-HZ-A8P0	stage iib	t3	n1	mx	male
114	TCGA-HZ-A8P1	stage ib	t2	n0	mx	male
115	TCGA-HZ-A9TJ	stage iv	t3	n0	m1	male
116	TCGA-IB-7644	stage iv	t3	n1	m1	female
117	TCGA-IB-7645	stage iib	t3	n1	m0	female
118	TCGA-IB-7646	stage iib	t3	n1	m0	male
119	TCGA-IB-7647	stage iib	t3	n1	m0	male
120	TCGA-IB-7649	stage iib	t3	n1	m0	female
121	TCGA-IB-7651	stage iib	t3	n1	m0	female
122	TCGA-IB-7652	stage iib	t3	n1	m0	female
123	TCGA-IB-7654	stage iib	t2	n1	m0	male
124	TCGA-IB-7885	stage iib	t3	n1	m0	female
125	TCGA-IB-7886	stage iib	t3	n1	m0	male
126	TCGA-IB-7887	stage iib	t3	n1	m0	female
127	TCGA-IB-7888	stage iia	t3	n0	m0	female
128	TCGA-IB-7889	stage iib	t3	n1	m0	female
129	TCGA-IB-7890	stage ib	t2	n0	m0	male
130	TCGA-IB-7891	stage iib	t3	n1	m0	female
131	TCGA-IB-7893	stage iia	t3	n0	m0	male
132	TCGA-IB-7897	stage iib	t3	n1	m0	female
133	TCGA-IB-8126	stage iii	t4	n1	m0	female
134	TCGA-IB-8127	stage iib	t3	n1	m0	male
135	TCGA-IB-A5SO	stage iib	t3	n1	m0	male
136	TCGA-IB-A5SP	stage iia	t3	n0	m0	male
137	TCGA-IB-A5SQ	stage ib	t2	n0	m0	female
138	TCGA-IB-A5SS	stage iib	t3	n1	m0	female
139	TCGA-IB-A5ST	stage iib	t3	n1	m0	female

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender
140	TCGA-IB-A6UF	stage iib	t3	n1	m0	male
141	TCGA-IB-A6UG	stage iib	t3	n1	m0	male
142	TCGA-IB-A7LX	stage iib	t3	n1	mx	male
143	TCGA-IB-A7M4	stage iib	t3	n1	mx	male
144	TCGA-IB-AAUM	stage iib	t2	n1	m0	male
145	TCGA-IB-AAUN	stage ib	t2	n0	m0	female
146	TCGA-IB-AAUO	stage iib	t3	n1	m0	female
147	TCGA-IB-AAUP	stage iib	t3	n1	m0	male
148	TCGA-IB-AAUQ	stage iib	t3	n1	m0	male
149	TCGA-IB-AAUR	stage iib	t3	n1	m0	male
150	TCGA-IB-AAUS	stage iib	t3	n1	m0	female
151	TCGA-IB-AAUT	stage iib	t2	n1	m0	male
152	TCGA-IB-AAUU	stage iib	t3	n1	m0	male
153	TCGA-IB-AAUV	stage iib	t3	n1	m0	male
154	TCGA-IB-AAUW	stage iib	t3	n1	m0	female
155	TCGA-L1-A7W4	stage iib	t3	n1	m0	female
156	TCGA-LB-A7SX	stage iib	t3	n1	mx	female
157	TCGA-LB-A8F3	stage iia	t3	n0	mx	female
158	TCGA-LB-A9Q5	stage iib	t3	n1	mx	female
159	TCGA-M8-A5N4	stage iia	t3	n0	m0	female
160	TCGA-OE-A75W	stage iia	t3	n0	m0	male
161	TCGA-PZ-A5RE	stage iib	t3	n1	mx	female
162	TCGA-Q3-A5QY	stage iib	t3	n1	mx	male
163	TCGA-Q3-AA2A	stage ib	t3	n1	mx	female
164	TCGA-RB-A7B8	stage iib	t3	n1	mx	female
165	TCGA-RB-AA9M	stage iib	t1	n1	mx	male
166	TCGA-RL-AAAS	stage ib	t2	n0	m0	female
167	TCGA-S4-A8RM	stage iib	t3	n1	mx	male

Srl	tcga_participant_barcode	CLI_pathologic_stage	CLI_pathology_T_stage	CLI_pathology_N_stage	CLI_pathology_M_stage	CLI_gender
168	TCGA-S4-A8RO	stage iib	t3	n1	mx	female
169	TCGA-S4-A8RP	stage iib	t3	n1	mx	female
170	TCGA-US-A774	stage iib	t3	n1	mx	female
171	TCGA-US-A776	stage iia	t3	n0	mx	male
172	TCGA-US-A779	stage iib	t3	n1b	mx	female
173	TCGA-US-A77E	stage iib	t3	n1	mx	male
174	TCGA-US-A77G	stage iib	t3	n1	mx	male
175	TCGA-US-A77J	stage iib	t3	n1b	mx	female
176	TCGA-XD-AAUG	stage iv	t3	n1	m1	female
177	TCGA-XD-AAUH	stage iib	t3	n1	m0	female
178	TCGA-XD-AAUI	stage iib	t3	n1	mx	female
179	TCGA-XD-AAUL	stage iia	t3	n0	mx	male
180	TCGA-XN-A8T3	stage ib	t2	n0	m0	male
181	TCGA-XN-A8T5	stage ib	t2	n0	m0	female
182	TCGA-YB-A89D	stage iib	t3	n1	mx	male
183	TCGA-YH-A8SY	stage iii	t3	n1	mx	female
184	TCGA-YY-A8LH	stage iib	t3	n1	mx	female
185	TCGA-Z5-AAPL	stage iia	t3	n0	m0	female

Table 7: Characteristics of the Samples Included in Validation Cohorts

Validation cohort 1					
Sample N0	Sex	Age (Y)	Tobacco intake	TNM Grade	Histopathological Report
PDAC_1	F	48	No	T2N1M0	Poorly differentiated adenocarcinoma
PDAC_2	F	74	No	pT3N0	Well differentiated adenocarcinoma
PDAC_3	M	38	No	T3N1M0	Poorly differentiated adenocarcinoma
PDAC_4	M	65	Yes	pT3N1	Ductal adenocarcinoma, Moderately differentiated
PDAC_5	M	64	Yes	pT3N1	Ductal adenocarcinoma, Moderately differentiated
PDAC_6	M	43	No	PT3N0	Well differentiated pancreatic ductal adenocarcinoma, extends to the ampulla of vater and duodenam
PDAC_7	M	40	Yes	pT3N1	Poorly differentiated adenocarcinoma
Validation cohort 2					
CMC_PANC_001	M	45	No	T2N1M0	Well differentiated adenocarcinoma
CMC_PANC_018	M	48	Yes	T2N1M0	Well differentiated adenocarcinoma
CMRI_PANC_020	F	74	No	pT3N0	Well differentiated adenocarcinoma
CMC_PANC_023	M	60	Yes	T3N1M0	Poorly differentiated adenocarcinoma
CMRI_PANC_024	F	60	No	pT3pN0	Well differentiated adenocarcinoma
CMC_PANC_025	M	44	No	T2N1M0	Poorly differentiated adenocarcinoma
CMRI_PANC_026	F	35	No	pT3pN0	Well differentiated adenocarcinoma
CMRI_PANC_053	M	43	No	PT3N0	Well differentiated adenocarcinoma
CMC_PANC_056	F	50	No	pT2N0Mx	Well differentiated adenocarcinoma
CMC_PANC_060	M	65	Yes	T2N1M0	Moderate Differentiated adenocarcinoma
CMC_PANC_064	M	60	No	T2N1M0	Moderate Differentiated adenocarcinoma
CMC_PANC_067	M	58	No	T2N0Mx	Well differentiated adenocarcinoma
TMC_PANC_076	M	40	Yes	pT3N1	Poorly differentiated adenocarcinoma
CMC_PANC_097	F	60	No	T3N1Mx	Well differentiated adenocarcinoma
CMRI_PANC_098	M	70	No	pT3N1	Well differentiated adenocarcinoma
TMC_PANC_120	M	53	No	T3N1M0	Well differentiated adenocarcinoma
CMC_PANC_122	M	68	No	pT3N2MX	Poorly differentiated adenocarcinoma

SSKM_PANC_124	F	54	No	T2N1M0	Moderate Differentiated adenocarcinoma
CMRI_PANC_135	M	62	Yes	T3N1M0	Well differentiated adenocarcinoma
SSKM_PANC_139	M	53	Yes	T3N1M0	Moderate Differentiated adenocarcinoma
TMC_PANC_217	M	54	No	pT2N1	Moderate Differentiated adenocarcinoma
SSKM_PANC_224	F	53	No	T2N1M0	Well differentiated adenocarcinoma
Validation cohort 3					
CMC_PANC_001	M	45	Yes	T2N1M0	Well differentiated adenocarcinoma
RGK_PANC_005	F	48	No	T2N1M0	Poorly differentiated adenocarcinoma
CMRI_PANC_010	M	46	Yes	PT3N0	Moderate Differentiated adenocarcinoma
CMC_PANC_018	M	48	Yes	T2N1M0	Well differentiated adenocarcinoma
RGK_PANC_022	M	38	No	T3N1M0	Poorly differentiated adenocarcinoma
CMC_PANC_023	M	60	Yes	T3N1M0	Poorly differentiated adenocarcinoma
CMC_PANC_025	M	44	No	T2N1M0	Poorly differentiated adenocarcinoma
TMC_PANC_031	M	64	No	pT3N1	Moderate Differentiated adenocarcinoma
TMC_PANC_042	M	65	Yes	pT3N1	Moderate Differentiated adenocarcinoma
TMC_PANC_048	M	64	Yes	pT3N1	Moderate Differentiated adenocarcinoma
CMC_PANC_064	M	60	No	T2N1M0	Moderate Differentiated adenocarcinoma
CMC_PANC_067	M	58	No	T2N0Mx	Well differentiated adenocarcinoma
TMC_PANC_076	M	40	Yes	pT3N1	Poorly differentiated adenocarcinoma
TMC_PANC_084	M	39	Yes	pT3N1	Moderate Differentiated adenocarcinoma
SSKM_PANC_089	M	57	Yes	T2N1M0	Poorly differentiated adenocarcinoma
CMC_PANC_099	M	60	No	T2N0Mx	Well differentiated adenocarcinoma
SSKM_PANC_107	F	50	No	T2N1M0	Moderate Differentiated adenocarcinoma
TMC_PANC_120	M	53	No	T3N1M0	Moderate Differentiated adenocarcinoma
CMC_PANC_122	M	68	No	pT3N2MX	Poorly differentiated adenocarcinoma
TMC_PANC_147	F	70	No	pT3N1	Moderate Differentiated adenocarcinoma
CMRI_PANC_027	F	49	No	PT4N1	Well Differentiated Periapillary adenocarcinoma
TMC_PANC_055	M	52	Yes	PT3N0	Poorly differentiated adenocarcinoma
CMC_PANC_060	M	65	Yes	T2N1M0	Moderately differentiated adenocarcinoma of pancreas
SSKM_PANC_065	M	48	Yes	T2N0M0	Moderately Differentiated Periapillary Adenocarcinoma
TMC_PANC_079	M	56	No	pT2N0	Moderate Differentiated adenocarcinoma

TMC_PANC_094	M	46	Yes	pT1N0	Well differentiated adenocarcinoma
Validation cohort 4					
SSKM_PANC_089	M	57	Yes	T2N1M0	Poorly differentiated adenocarcinoma
TMC_PANC_031	M	64	No	pT3N1	Moderate Differentiated adenocarcinoma
CMC_PANC_099	M	60	No	T2N0Mx	Well differentiated adenocarcinoma
CMC_PANC_097	F	60	No	T3N1Mx	Well differentiated adenocarcinoma
CMRI_PANC_098	M	70	No	pT3N1	Well differentiated adenocarcinoma
TMC_PANC_042	M	65	Yes	pT3N1	Moderate Differentiated adenocarcinoma
CMC_PANC_122	M	68	NA	pT3N2MX	Poorly differentiated adenocarcinoma
TMC_PANC_076	M	40	Yes	pT3N1	Poorly differentiated adenocarcinoma
TMC_PANC_084	M	39	Yes	pT3N1	Moderate Differentiated adenocarcinoma
CMC_PANC_023	M	60	Yes	T3N1M0	Poorly differentiated adenocarcinoma
TMC_PANC_260	F	60	No	T3N0	Moderate Differentiated adenocarcinoma
TMC_PANC_268	M	67	NA	T2N1	Moderate Differentiated adenocarcinoma
TMC_PANC_255	M	64	No	T2N0	Moderate Differentiated adenocarcinoma
TMC_PANC_226	M	69	No	T3N1M1	Moderate Differentiated adenocarcinoma
TMC_PANC_258	F	67	No	T3N1M0	Moderate Differentiated adenocarcinoma
TMC_PANC_266	M	54	NA	T2N0	Moderate Differentiated adenocarcinoma
TMC_PANC_245	F	68	No	T2N1	Poorly differentiated adenocarcinoma
TMC_PANC_267	F	62	NA	T2N1	Moderate Differentiated adenocarcinoma
TMC_PANC_242	M	69	NA	T2N1	Moderate Differentiated adenocarcinoma
SSKM_PANC_218	F	56	No	T2N0M0	Well differentiated adenocarcinoma

Table 8: Details of the Probes (n=7832) that were Differentially Methylated between Pancreatic Cancer (n=9) and Solid Tissue Normal (n=9) Samples in the TCGA Cohort. Normalized Beta Values for Each Probe is Mentioned. "_C" Denotes Control and "_D" Denotes Cancer Tissue Samples.(Available in soft copy)

Table 9: Details of 156 Differentially Methylated Positions (DMPs), Identified in Validation Cohort 1 Including PDAC Tumors (n=7) Compared to Adjacent Normal Tissue Samples (n=7). The Chromosomal Locations of the DMPs are Mentioned According to hg19 Reference Genome. Annotation of CpG Islands was Mentioned as Mentioned in the Annotation Provided by the Beadchip.

Differentially_methylated	ProbeID	chr	pos	pval	qval	Delta_beta	Gene	UCSC_RefGene_Group	Relation_to_Island	Regulatory_Feature_Group	TCGA Status
HYPO	cg02297063	chr19	1132389	2.32E-07	0.004	-0.285	<i>SBNO2</i>	Body	OpenSea	Promoter_Associated	uncommon
HYPO	cg02791973	chr19	1132256	3.10E-07	0.004	-0.272	<i>SBNO2</i>	Body	OpenSea	Promoter_Associated	common
HYPO	cg03140412	chr11	415111	2.41E-06	0.004	-0.252	<i>SIGIRR</i>	5'UTR	N_Shore	Promoter_Associated	uncommon
HYPO	cg19864130	chr19	1132420	1.42E-06	0.004	-0.249	<i>SBNO2</i>	Body	OpenSea	Promoter_Associated	common
HYPO	cg21435684	chr17	80255457	5.92E-06	0.004	-0.248			S_Shelf	Promoter_Associated	common
HYPO	cg06752054	chr21	42733747	2.64E-06	0.004	-0.248	<i>MX2</i>	TSS1500	OpenSea	Promoter_Associated	uncommon
HYPO	cg04029159	chr11	415080	3.60E-06	0.004	-0.242	<i>SIGIRR</i>	5'UTR	N_Shore		common
HYPO	cg16389901	chr14	54423807	3.41E-06	0.004	-0.240	<i>BMP4</i>	TSS1500	S_Shore		uncommon
HYPO	cg21011133	chr2	25142198	1.93E-05	0.004	-0.238	<i>ADCY3</i>	TSS200	N_Shore		common
HYPO	cg21130374	chr21	42734266	3.87E-06	0.004	-0.236	<i>MX2</i>	5'UTR	OpenSea	Promoter_Associated	uncommon
HYPO	cg03178838	chr19	1132339	3.11E-06	0.004	-0.230	<i>SBNO2</i>	Body	OpenSea	Promoter_Associated	uncommon
HYPO	cg19664267	chr6	33245163	4.42E-06	0.004	-0.229	<i>B3GALT4</i>	1stExon	Island	Promoter_Associated	common
HYPO	cg04786142	chr1	234908381	1.41E-05	0.004	-0.229			OpenSea	Promoter_Associated	uncommon
HYPO	cg11496630	chr6	3024178	2.56E-06	0.004	-0.229			OpenSea	Promoter_Associated	common

Differentially_methylated	ProbeID	chr	pos	pval	qval	Delta_beta	Gene	UCSC_RefGene_Group	Relation_to_Island	Regulatory_Feature_Group	TCGA Status
HYPO	cg14341177	chr9	95475787	7.00E-05	0.005	-0.228	<i>BICD2</i>	Body	N_Shore		uncommon
HYPO	cg16238993	chr19	1132372	1.75E-06	0.004	-0.228	<i>SBNO2</i>	Body	OpenSea	Promoter_Associated	common
HYPO	cg24965479	chr17	59216714	1.63E-05	0.004	-0.226	<i>BCAS3</i>	Body	OpenSea		uncommon
HYPO	cg20928945	chr7	960411	8.19E-06	0.004	-0.226	<i>ADAP1</i>	Body	S_Shore		uncommon
HYPO	cg00768487	chr1	26611311	4.83E-05	0.005	-0.226	<i>UBXN11</i>	Body	OpenSea	Promoter_Associated	uncommon
HYPO	cg14931884	chr10	369977	3.68E-06	0.004	-0.224	<i>DIP2C</i>	Body	S_Shore		common
HYPO	cg23885932	chr19	1132294	1.42E-06	0.004	-0.223	<i>SBNO2</i>	Body	OpenSea	Promoter_Associated	common
HYPO	cg12907959	chr22	36552828	4.06E-06	0.004	-0.223	<i>APOL3</i>	5'UTR	OpenSea		common
HYPO	cg05792169	chr15	85874227	1.04E-05	0.004	-0.221			OpenSea	Promoter_Associated	uncommon
HYPO	cg12255995	chr19	1132537	1.67E-06	0.004	-0.221	<i>SBNO2</i>	Body	OpenSea	Promoter_Associated	uncommon
HYPO	cg21571658	chr15	70785399	1.29E-05	0.004	-0.221			OpenSea		uncommon
HYPO	cg17534029	chr20	62318426	2.12E-06	0.004	-0.220	<i>RTEL1</i>	Body	Island		uncommon
HYPO	cg25416067	chr17	75315667	4.83E-06	0.004	-0.220	<i>SEPT9</i>	5'UTR	OpenSea	Promoter_Associated	uncommon
HYPO	cg24039081	chr3	197155143	3.37E-06	0.004	-0.219			OpenSea		common
HYPO	cg13905238	chr3	197155273	5.05E-06	0.004	-0.219			OpenSea		uncommon
HYPO	cg02633924	chr17	75315563	7.61E-06	0.004	-0.218	<i>SEPT9</i>	5'UTR	OpenSea	Promoter_Associated	common

Differentially_methylated	ProbeID	chr	pos	pval	qval	Delta_beta	Gene	UCSC_RefGene_Group	Relation_to_Island	Regulatory_Feature_Group	TCGA Status
HYPO	cg11575912	chr19	5828299	1.94E-07	0.004	-0.217	<i>NRTN</i>	3'UTR	Island		uncommon
HYPO	cg21151432	chr2	25142229	1.40E-05	0.004	-0.217	<i>ADCY3</i>	TSS200	N_Shore		uncommon
HYPO	cg18082788	chr6	149806339	8.01E-06	0.004	-0.216	<i>ZC3H12D</i>	TSS200	OpenSea	Promoter_Associated	common
HYPO	cg26348487	chr19	2488560	3.84E-06	0.004	-0.216			Island		uncommon
HYPO	cg01649611	chr2	43521066	1.89E-05	0.004	-0.216	<i>THADA</i>	Body	OpenSea		uncommon
HYPO	cg12560128	chr12	113416061	1.05E-05	0.004	-0.215	<i>OAS2</i>	TSS1500	OpenSea	Promoter_Associated	common
HYPO	cg23261919	chr10	135072960	1.03E-05	0.004	-0.215			N_Shelf		common
HYPO	cg21333861	chr6	33244976	1.19E-05	0.004	-0.214	<i>B3GALT4</i>	1stExon	Island	Promoter_Associated	uncommon
HYPO	cg18442362	chr7	44677772	1.90E-05	0.004	-0.214	<i>OGDH</i>	Body	OpenSea	Promoter_Associated	common
HYPO	cg23066280	chr7	158032496	2.35E-05	0.004	-0.214	<i>PTPRN2</i>	Body	OpenSea		common
HYPO	cg01947066	chr15	74289586	2.05E-05	0.004	-0.213	<i>PML</i>	Body	N_Shore	Promoter_Associated	uncommon
HYPO	cg20765408	chr13	25085669	2.34E-05	0.004	-0.213	<i>PARP4</i>	5'UTR	N_Shore		common
HYPO	cg07009284	chr8	76318515	9.27E-06	0.004	-0.213			N_Shore		common
HYPO	cg24227481	chr7	73417185	6.52E-06	0.004	-0.212			OpenSea		uncommon
HYPO	cg16474725	chr6	45979349	1.71E-05	0.004	-0.212	<i>CLIC5</i>	Body	N_Shelf	Promoter_Associated	uncommon
HYPO	cg09410512	chr2	224564400	1.96E-05	0.004	-0.212			OpenSea		common

Differentially_methylated	ProbeID	chr	pos	pval	qval	Delta_beta	Gene	UCSC_RefGene_Group	Relation_to_Island	Regulatory_Feature_Group	TCGA Status
HYPO	cg05267204	chr2	55360999	8.96E-06	0.004	-0.211			OpenSea		uncommon
HYPO	cg13786863	chr11	68610743	1.05E-05	0.004	-0.211	<i>CPT1A</i>	TSS1500	N_Shore		common
HYPO	cg01973676	chr7	101596404	2.83E-05	0.005	-0.211	<i>CUX1</i>	Body	OpenSea		common
HYPO	cg12247222	chr19	17403313	1.39E-06	0.004	-0.211	<i>ABHD8</i>	3'UTR	Island	Promoter_Associated	uncommon
HYPO	cg21526750	chr14	23291749	4.40E-06	0.004	-0.211			Island		uncommon
HYPO	cg04263436	chr6	33245467	3.69E-06	0.004	-0.210	<i>B3GALT4</i>	1stExon	Island	Promoter_Associated	common
HYPO	cg10536999	chr7	26193109	1.77E-05	0.004	-0.210	<i>NFE2L3</i>	Body	S_Shore		uncommon
HYPO	cg03663556	chr7	50900556	1.69E-05	0.004	-0.210			OpenSea		uncommon
HYPO	cg26733301	chr20	61585924	5.06E-06	0.004	-0.209	<i>SLC17A9</i>	Body	S_Shore	Promoter_Associated	uncommon
HYPO	cg06936290	chr7	13889871	8.90E-06	0.004	-0.209			OpenSea		uncommon
HYPO	cg25508319	chr11	128778506	1.38E-05	0.004	-0.209	<i>KCNJ5</i>	5'UTR	OpenSea		common
HYPO	cg25637655	chr6	29911542	4.01E-05	0.005	-0.209	<i>HLA-A</i>	Body	S_Shore		uncommon
HYPO	cg12666727	chr1	42128487	6.03E-06	0.004	-0.208	<i>HIVEP3</i>	5'UTR	Island		uncommon
HYPO	cg20834178	chr4	6944071	3.16E-05	0.005	-0.208	<i>TBC1D14</i>	Body	OpenSea		uncommon
HYPO	cg09495643	chr19	1132313	6.48E-06	0.004	-0.208	<i>SBNO2</i>	Body	OpenSea	Promoter_Associated	common
HYPO	cg06021088	chr2	127822551	4.85E-06	0.004	-0.208	<i>BIN1</i>	Body	OpenSea		uncommon

Differentially_methylated	ProbeID	chr	pos	pval	qval	Delta_beta	Gene	UCSC_RefGene_Group	Relation_to_Island	Regulatory_Feature_Group	TCGA Status
HYPO	cg24366211	chr11	68610706	7.80E-06	0.004	-0.207	<i>CPT1A</i>	TSS1500	N_Shore		uncommon
HYPO	cg03301058	chr6	90007856	2.99E-05	0.005	-0.207	<i>GABRR2</i>	Body	OpenSea		uncommon
HYPO	cg14418802	chr1	9299979	2.84E-05	0.005	-0.207	<i>H6PD</i>	5'UTR	OpenSea		uncommon
HYPO	cg26996569	chr12	121829743	2.47E-05	0.005	-0.207			OpenSea		uncommon
HYPO	cg04659689	chr7	2903377	6.86E-06	0.004	-0.207			OpenSea		common
HYPO	cg02853152	chr4	6690702	1.58E-05	0.004	-0.206			OpenSea	Promoter_Associated	common
HYPO	cg06172942	chr3	50378529	2.26E-06	0.004	-0.206	<i>RASSF1</i>	TSS200	Island		common
HYPO	cg20003368	chr6	44026475	3.83E-06	0.004	-0.206			OpenSea		uncommon
HYPO	cg03851835	chr11	107711742	1.79E-05	0.004	-0.206	<i>SLC35F2</i>	Body	OpenSea		common
HYPO	cg07805542	chr1	9779309	9.36E-06	0.004	-0.205	<i>PIK3CD</i>	Body	S_Shelf	Promoter_Associated	common
HYPO	cg13393476	chr19	39894666	4.35E-06	0.004	-0.205			Island	Promoter_Associated	uncommon
HYPO	cg04650676	chr17	75315486	1.29E-05	0.004	-0.205	<i>SEPT9</i>	5'UTR	OpenSea	Promoter_Associated	common
HYPO	cg12273284	chr10	12490844	4.33E-05	0.005	-0.205	<i>CAMK1D</i>	Body	OpenSea	Promoter_Associated	uncommon
HYPO	cg10142520	chr11	65344604	1.43E-05	0.004	-0.205	<i>EHBP1L1</i>	Body	S_Shore	Promoter_Associated	uncommon
HYPO	cg04470054	chr17	78830472	5.39E-05	0.005	-0.205	<i>RPTOR</i>	Body	N_Shore		common
HYPO	cg16476991	chr13	114831404	6.66E-06	0.004	-0.205	<i>RASA3</i>	Body	OpenSea	Promoter_Associated	common

Differentially_methylated	ProbeID	chr	pos	pval	qval	Delta_beta	Gene	UCSC_RefGene_Group	Relation_to_Island	Regulatory_Feature_Group	TCGA Status
HYPO	cg03108070	chr6	33245490	6.78E-06	0.004	-0.204	<i>B3GALT4</i>	1stExon	Island	Promoter_Associated	common
HYPO	cg18517266	chr19	5827811	7.63E-07	0.004	-0.204	<i>NRTN</i>	Body	Island		common
HYPO	cg08047457	chr3	50378413	7.82E-07	0.004	-0.204	<i>RASSF1</i>	TSS200	Island		uncommon
HYPO	cg09126279	chr4	40675171	2.89E-05	0.005	-0.203			OpenSea	Promoter_Associated	uncommon
HYPO	cg20970886	chr3	192568714	7.36E-06	0.004	-0.203	<i>C3orf59</i>	Body	OpenSea		common
HYPO	cg09884146	chr2	65593908	1.26E-05	0.004	-0.203	<i>SPRED2</i>	5'UTR	OpenSea		uncommon
HYPO	cg26994377	chr4	140579069	1.75E-05	0.004	-0.203			OpenSea		uncommon
HYPO	cg18582260	chr13	25085301	4.60E-05	0.005	-0.203	<i>PARP4</i>	5'UTR	N_Shore		uncommon
HYPO	cg21390801	chr21	42733729	9.10E-06	0.004	-0.203	<i>MX2</i>	TSS1500	OpenSea	Promoter_Associated	common
HYPO	cg23987789	chr1	7842406	1.78E-05	0.004	-0.202			N_Shore		common
HYPO	cg06879394	chr6	13274151	6.34E-05	0.005	-0.202	<i>PHACTR1</i>	Body	OpenSea		uncommon
HYPO	cg25690715	chr17	75315567	1.61E-05	0.004	-0.202	<i>SEPT9</i>	5'UTR	OpenSea	Promoter_Associated	common
HYPO	cg17184704	chr10	11727286	1.79E-05	0.004	-0.202			OpenSea		uncommon
HYPO	cg16066354	chr22	39417451	7.77E-06	0.004	-0.201	<i>APOBEC3D</i>	1stExon	OpenSea		common
HYPO	cg22191981	chr6	41472382	7.11E-06	0.004	-0.201			N_Shore		uncommon
HYPO	cg21853021	chr11	2920146	2.22E-05	0.004	-0.201	<i>SLC22A18AS</i>	Body	N_Shelf		common

Differentially_methylated	ProbeID	chr	pos	pval	qval	Delta_beta	Gene	UCSC_RefGene_Group	Relation_to_Island	Regulatory_Feature_Group	TCGA Status
HYPO	cg12359904	chr3	156848454	6.20E-06	0.004	-0.201			OpenSea		common
HYPO	cg10893986	chr1	3229739	1.19E-05	0.004	-0.200	<i>PRDM16</i>	Body	OpenSea		uncommon
HYPO	cg22027433	chr17	62208049	1.04E-05	0.004	-0.200	<i>ERN1</i>	TSS1500	Island		uncommon
HYPO	cg21912556	chr8	18605482	2.13E-05	0.004	-0.200	<i>PSD3</i>	Body	OpenSea		common
HYPER	cg14780632	chr11	65816521	4.25E-06	0.004	0.200	<i>GAL3ST3</i>	5'UTR	Island		uncommon
HYPER	cg21179088	chr7	54612333	9.19E-06	0.004	0.201	<i>VSTM2A</i>	Body	Island		common
HYPER	cg18474072	chr10	664762	4.17E-05	0.005	0.201	<i>DIP2C</i>	Body	OpenSea		common
HYPER	cg26460092	chr15	83776420	1.41E-05	0.004	0.201	<i>TM6SF1</i>	5'UTR	Island		common
HYPER	cg27501878	chr20	61638588	2.55E-07	0.004	0.201	<i>BHLHE23</i>	TSS1500	Island		common
HYPER	cg14502484	chr5	1446208	6.64E-06	0.004	0.201	<i>SLC6A3</i>	TSS1500	Island		uncommon
HYPER	cg15672437	chr13	113764555	4.78E-06	0.004	0.201	<i>F7</i>	Body	Island		uncommon
HYPER	cg12071328	chr11	20690930	2.10E-06	0.004	0.202	<i>NELL1</i>	TSS200	Island		uncommon
HYPER	cg01360618	chr3	69591659	3.06E-07	0.004	0.202			Island		common
HYPER	cg03078269	chr12	54354974	8.81E-06	0.004	0.202			Island		uncommon
HYPER	cg03061682	chr15	28352098	7.43E-05	0.005	0.202			Island		common
HYPER	cg10300684	chr14	29236323	4.06E-07	0.004	0.202	<i>FOXP1</i>	1stExon	N_Shore		common

Differentially_methylated	ProbeID	chr	pos	pval	qval	Delta_beta	Gene	UCSC_RefGene_Group	Relation_to_Island	Regulatory_Feature_Group	TCGA Status
HYPER	cg16620382	chr6	43211213	1.96E-06	0.004	0.202	<i>TTBK1</i>	TSS200	Island		uncommon
HYPER	cg01805540	chr10	18429749	5.76E-07	0.004	0.203	<i>CACNB2</i>	1stExon	Island		common
HYPER	cg09619624	chr22	18985632	6.25E-05	0.005	0.203			Island		common
HYPER	cg16364121	chr17	35291924	4.50E-06	0.004	0.204			Island		uncommon
HYPER	cg23774356	chr22	19137874	7.79E-06	0.004	0.204	<i>GSC2</i>	TSS200	Island		uncommon
HYPER	cg11701471	chr8	54164051	3.96E-06	0.004	0.204	<i>OPRK1</i>	5'UTR	Island		uncommon
HYPER	cg25884711	chr7	24323840	4.07E-07	0.004	0.205	<i>NPY</i>	5'UTR	Island		common
HYPER	cg18770350	chr1	236849966	1.09E-05	0.004	0.205	<i>ACTN2</i>	1stExon	Island		uncommon
HYPER	cg11258943	chr9	100615357	3.70E-05	0.005	0.205	<i>FOXE1</i>	TSS200	Island		common
HYPER	cg16703956	chr5	1446217	2.29E-06	0.004	0.205	<i>SLC6A3</i>	TSS1500	Island		uncommon
HYPER	cg23189410	chr3	147125712	1.20E-06	0.004	0.205	<i>ZIC4</i>	TSS1500	N_Shore		uncommon
HYPER	cg01873886	chr14	54419825	5.48E-07	0.004	0.206	<i>BMP4</i>	5'UTR	N_Shore		uncommon
HYPER	cg18863595	chr3	6902845	2.65E-06	0.004	0.206	<i>GRM7</i>	1stExon	Island		common
HYPER	cg26115633	chr10	129535893	7.39E-06	0.004	0.207	<i>FOXI2</i>	1stExon	Island		uncommon
HYPER	cg07819944	chr10	124897153	1.11E-06	0.004	0.207	<i>HMX3</i>	Body	Island		uncommon
HYPER	cg14394550	chr8	22548400	5.53E-06	0.004	0.207	<i>EGR3</i>	Body	Island		uncommon

Differentially_methylated	ProbeID	chr	pos	pval	qval	Delta_beta	Gene	UCSC_RefGene_Group	Relation_to_Island	Regulatory_Feature_Group	TCGA Status
HYPER	cg21277995	chr6	393239	1.85E-07	0.004	0.207	<i>IRF4</i>	Body	Island		uncommon
HYPER	cg10356613	chr17	35294491	1.02E-05	0.004	0.207	<i>LHX1</i>	TSS200	Island		uncommon
HYPER	cg03109827	chr10	133110349	5.26E-07	0.004	0.208	<i>TCERG1L</i>	TSS1500	Island		common
HYPER	cg03502002	chr18	74962133	3.55E-06	0.004	0.209	<i>GALR1</i>	1stExon	Island		common
HYPER	cg13703049	chr7	27213610	5.34E-06	0.004	0.209	<i>HOXA10</i>	Body	Island		common
HYPER	cg27363829	chr3	164914614	1.95E-06	0.004	0.209	<i>SLITRK3</i>	TSS200	OpenSea		uncommon
HYPER	cg18486102	chr12	50297777	2.46E-06	0.004	0.209	<i>FAIM2</i>	TSS200	Island		common
HYPER	cg11254700	chr19	53561386	1.64E-05	0.004	0.209			Island		uncommon
HYPER	cg17839237	chr7	19157193	2.11E-06	0.004	0.210	<i>TWIST1</i>	1stExon	Island		uncommon
HYPER	cg26809635	chr12	54355087	1.23E-05	0.004	0.210			Island		uncommon
HYPER	cg19870512	chr12	4919081	9.89E-06	0.004	0.210	<i>KCNA6</i>	5'UTR	Island		common
HYPER	cg18483908	chr22	19017420	8.30E-07	0.004	0.211			Island		uncommon
HYPER	cg27578811	chr15	83316223	7.90E-06	0.004	0.213	<i>CPEB1</i>	Body	Island		uncommon
HYPER	cg13601435	chr9	120507431	5.97E-07	0.004	0.213			Island		uncommon
HYPER	cg14500486	chr6	159655392	3.06E-06	0.004	0.214	<i>FNDC1</i>	Body	Island		uncommon
HYPER	cg11940285	chr3	129693385	1.09E-06	0.004	0.215	<i>TRH</i>	5'UTR	Island		common

Differentially_methylated	ProbeID	chr	pos	pval	qval	Delta_beta	Gene	UCSC_RefGene_Group	Relation_to_Island	Regulatory_Feature_Group	TCGA Status
HYPER	cg21270847	chr1	164604806	1.05E-06	0.004	0.219	PBX1	Body	OpenSea		uncommon
HYPER	cg06947913	chr12	50297793	1.33E-06	0.004	0.220	FAIM2	TSS200	Island		uncommon
HYPER	cg22399133	chr2	208989248	4.44E-06	0.004	0.222	CRYGD	1stExon	Island		common
HYPER	cg02604524	chr3	128720884	4.16E-06	0.004	0.222	CCDC48	1stExon	Island		common
HYPER	cg24221648	chr13	79170283	8.09E-06	0.004	0.222			Island		uncommon
HYPER	cg18989174	chr7	149470820	0.00546998	0.048	0.223	ZNF467	TSS1500	Island		common
HYPER	cg01227537	chr3	147127012	3.12E-06	0.004	0.224	ZIC1	TSS200	Island		uncommon
HYPER	cg17495912	chr13	37006340	1.80E-05	0.004	0.224	CCNA1	5'UTR	Island		common
HYPER	cg18862481	chr3	129693370	3.91E-06	0.004	0.227	TRH	5'UTR	Island		uncommon
HYPER	cg13791254	chr9	100616607	4.57E-06	0.004	0.234	FOXE1	1stExon	Island		uncommon
HYPER	cg20302133	chr1	111217194	6.85E-06	0.004	0.236	KCNA3	1stExon	Island	Promoter_Associated	uncommon
HYPER	cg13405887	chr9	132382812	1.35E-06	0.004	0.236	C9orf50	1stExon	Island		common
HYPER	cg18318649	chr5	87974369	0.004556668	0.043	0.240	LOC645323	Body	N_Shore		common
HYPER	cg06304097	chr10	133109867	1.08E-06	0.004	0.265	TCERG1L	1stExon	Island		common

Table 10: Details of the Probes that were Differentially Methylated both in the TCGA Cohort and in our Study (n=68).

ProbeID	diffmeth	chr	pos	Gene_Name	Delta_beta_ourstudy	qval (our study)	Delta_beta_TCGA	qval (TCGA)
cg04786142	HYPO	chr1	234908381		-0.2292	0.0043	-0.4093	0.0227
cg20928945	HYPO	chr7	960411	<i>ADAP1</i>	-0.2262	0.0042	-0.4294	0.0085
cg14931884	HYPO	chr10	369977	<i>DIP2C</i>	-0.2235	0.0042	-0.4133	0.0080
cg05792169	HYPO	chr15	85874227		-0.2212	0.0042	-0.3108	0.0392
cg13905238	HYPO	chr3	197155273		-0.2187	0.0042	-0.3530	0.0214
cg18442362	HYPO	chr7	44677772	<i>OGDH</i>	-0.2142	0.0044	-0.3856	0.0213
cg23066280	HYPO	chr7	158032496	<i>PTPRN2</i>	-0.2141	0.0045	-0.3740	0.0247
cg20765408	HYPO	chr13	25085669	<i>PARP4</i>	-0.2131	0.0045	-0.3671	0.0447
cg09410512	HYPO	chr2	224564400		-0.2116	0.0044	-0.3420	0.0193
cg01973676	HYPO	chr7	101596404	<i>CUX1</i>	-0.2109	0.0046	-0.4235	0.0204
cg21526750	HYPO	chr14	23291749		-0.2105	0.0042	-0.3276	0.0455
cg03663556	HYPO	chr7	50900556		-0.2100	0.0043	-0.3519	0.0121
cg06936290	HYPO	chr7	13889871		-0.2089	0.0042	-0.3578	0.0432
cg25508319	HYPO	chr11	128778506	<i>KCNJ5</i>	-0.2088	0.0042	-0.3084	0.0322
cg02853152	HYPO	chr4	6690702		-0.2063	0.0043	-0.3475	0.0347
cg03851835	HYPO	chr11	107711742	<i>SLC35F2</i>	-0.2057	0.0043	-0.3338	0.0251
cg07805542	HYPO	chr1	9779309	<i>PIK3CD</i>	-0.2055	0.0042	-0.3770	0.0196
cg12273284	HYPO	chr10	12490844	<i>CAMK1D</i>	-0.2050	0.0048	-0.3698	0.0473
cg09126279	HYPO	chr4	40675171		-0.2034	0.0046	-0.3326	0.0477
cg20970886	HYPO	chr3	192568714	<i>C3orf59</i>	-0.2033	0.0042	-0.3966	0.0030
cg09884146	HYPO	chr2	65593908	<i>SPRED2</i>	-0.2033	0.0042	-0.3199	0.0255
cg18582260	HYPO	chr13	25085301	<i>PARP4</i>	-0.2029	0.0048	-0.3757	0.0234
cg14780632	HYPER	chr11	65816521	<i>GAL3ST3</i>	0.2000	0.0042	0.3037	0.0014
cg21179088	HYPER	chr7	54612333	<i>VSTM2A</i>	0.2005	0.0042	0.3048	0.0102
cg26460092	HYPER	chr15	83776420	<i>TM6SF1</i>	0.2009	0.0043	0.3256	0.0044

ProbeID	diffmeth	chr	pos	Gene_Name	Delta_beta_ourstudy	qval (our study)	Delta_beta_TCGA	qval (TCGA)
cg27501878	HYPER	chr20	61638588	<i>BHLHE23</i>	0.2010	0.0042	0.3679	0.0006
cg14502484	HYPER	chr5	1446208	<i>SLC6A3</i>	0.2012	0.0042	0.2401	0.0079
cg15672437	HYPER	chr13	113764555	<i>F7</i>	0.2013	0.0042	0.2791	0.0010
cg01360618	HYPER	chr3	69591659		0.2018	0.0042	0.2874	0.0061
cg03078269	HYPER	chr12	54354974		0.2019	0.0042	0.3822	0.0056
cg10300684	HYPER	chr14	29236323	<i>FOXG1</i>	0.2022	0.0042	0.2450	0.0019
cg16620382	HYPER	chr6	43211213	<i>TTBK1</i>	0.2024	0.0042	0.2718	0.0105
cg01805540	HYPER	chr10	18429749	<i>CACNB2</i>	0.2025	0.0042	0.2587	0.0041
cg16364121	HYPER	chr17	35291924		0.2040	0.0042	0.3168	0.0015
cg23774356	HYPER	chr22	19137874	<i>GSC2</i>	0.2042	0.0042	0.2877	0.0251
cg25884711	HYPER	chr7	24323840	<i>NPY</i>	0.2046	0.0042	0.2793	0.0017
cg18770350	HYPER	chr1	236849966	<i>ACTN2</i>	0.2048	0.0042	0.4078	0.0013
cg16703956	HYPER	chr5	1446217	<i>SLC6A3</i>	0.2050	0.0042	0.2497	0.0024
cg23189410	HYPER	chr3	147125712	<i>ZIC4</i>	0.2054	0.0042	0.3372	0.0029
cg01873886	HYPER	chr14	54419825	<i>BMP4</i>	0.2059	0.0042	0.2458	0.0297
cg18863595	HYPER	chr3	6902845	<i>GRM7</i>	0.2062	0.0042	0.2770	0.0251
cg26115633	HYPER	chr10	129535893	<i>FOXI2</i>	0.2066	0.0042	0.3680	0.0060
cg07819944	HYPER	chr10	124897153	<i>HMX3</i>	0.2069	0.0042	0.2893	0.0309
cg14394550	HYPER	chr8	22548400	<i>EGR3</i>	0.2070	0.0042	0.3252	0.0114
cg21277995	HYPER	chr6	393239	<i>IRF4</i>	0.2073	0.0042	0.3110	0.0013
cg10356613	HYPER	chr17	35294491	<i>LHX1</i>	0.2074	0.0042	0.3099	0.0059
cg03109827	HYPER	chr10	133110349	<i>TCERG1L</i>	0.2081	0.0042	0.3490	0.0004
cg03502002	HYPER	chr18	74962133	<i>GALR1</i>	0.2085	0.0042	0.4111	0.0015
cg13703049	HYPER	chr7	27213610	<i>HOXA10</i>	0.2087	0.0042	0.3581	0.0048
cg18486102	HYPER	chr12	50297777	<i>FAIM2</i>	0.2089	0.0042	0.3265	0.0005
cg11254700	HYPER	chr19	53561386		0.2093	0.0043	0.3303	0.0206

ProbeID	diffmeth	chr	pos	Gene_Name	Delta_beta_ourstudy	qval (our study)	Delta_beta_TCGA	qval (TCGA)
cg17839237	HYPER	chr7	19157193	<i>TWIST1</i>	0.2100	0.0042	0.3595	0.0011
cg26809635	HYPER	chr12	54355087		0.2100	0.0042	0.3138	0.0036
cg19870512	HYPER	chr12	4919081	<i>KC6</i>	0.2103	0.0042	0.3313	0.0495
cg27578811	HYPER	chr15	83316223	<i>CPEB1</i>	0.2128	0.0042	0.3008	0.0021
cg13601435	HYPER	chr9	120507431		0.2132	0.0042	0.3373	0.0005
cg11940285	HYPER	chr3	129693385	<i>TRH</i>	0.2146	0.0042	0.3543	0.0001
cg06947913	HYPER	chr12	50297793	<i>FAIM2</i>	0.2204	0.0042	0.3949	0.0002
cg22399133	HYPER	chr2	208989248	<i>CRYGD</i>	0.2216	0.0042	0.3783	0.0009
cg02604524	HYPER	chr3	128720884	<i>CCDC48</i>	0.2220	0.0042	0.3074	0.0143
cg24221648	HYPER	chr13	79170283		0.2221	0.0042	0.4542	0.0008
cg17495912	HYPER	chr13	37006340	<i>CC1</i>	0.2241	0.0044	0.4753	0.0009
cg18862481	HYPER	chr3	129693370	<i>TRH</i>	0.2268	0.0042	0.3330	0.0001
cg13791254	HYPER	chr9	100616607	<i>FOXE1</i>	0.2342	0.0042	0.3373	0.0007
cg20302133	HYPER	chr1	111217194	<i>KC3</i>	0.2358	0.0042	0.4826	0.0004
cg13405887	HYPER	chr9	132382812	<i>C9orf50</i>	0.2362	0.0042	0.3349	0.0036
cg18318649	HYPER	chr5	87974369	<i>LOC645323</i>	0.2398	0.0429	0.1558	0.0011
cg06304097	HYPER	chr10	133109867	<i>TCERG1L</i>	0.2647	0.0042	0.5046	0.0006

Table 11: Details of the Probes that were not Differentially Methylated Commonly in the TCGA Cohort and in our Study (n=88).

ProbeID	pval	qval	delbeta	Diff Meth	chr	pos	UCSC_RefGene_Name	UCSC_RefGene_Group	_TCGA
cg02297063	2.32E-07	0.004166748	-0.285033714	HYPO	chr19	1132389	<i>SBNO2</i>	Body	uncommon
cg02791973	3.10E-07	0.004166748	-0.271503766	HYPO	chr19	1132256	<i>SBNO2</i>	Body	uncommon
cg03140412	2.41E-06	0.004166748	-0.251576886	HYPO	chr11	415111	<i>SIGIRR</i>	5'UTR	uncommon
cg19864130	1.42E-06	0.004166748	-0.248782292	HYPO	chr19	1132420	<i>SBNO2</i>	Body	uncommon
cg21435684	5.92E-06	0.004166748	-0.248466091	HYPO	chr17	80255457	NA	NA	uncommon
cg06752054	2.64E-06	0.004166748	-0.248003978	HYPO	chr21	42733747	<i>MX2</i>	TSS1500	uncommon
cg04029159	3.60E-06	0.004166748	-0.242021526	HYPO	chr11	415080	<i>SIGIRR</i>	5'UTR	uncommon
cg16389901	3.41E-06	0.004166748	-0.239843339	HYPO	chr14	54423807	<i>BMP4</i>	TSS1500	uncommon
cg21011133	1.93E-05	0.004358099	-0.238099024	HYPO	chr2	25142198	<i>ADCY3</i>	TSS200	uncommon
cg21130374	3.87E-06	0.004166748	-0.236160564	HYPO	chr21	42734266	<i>MX2</i>	5'UTR	uncommon
cg03178838	3.11E-06	0.004166748	-0.229749493	HYPO	chr19	1132339	<i>SBNO2</i>	Body	uncommon
cg19664267	4.42E-06	0.004166748	-0.229345331	HYPO	chr6	33245163	<i>B3GALT4</i>	1stExon	uncommon
cg11496630	2.56E-06	0.004166748	-0.228546255	HYPO	chr6	3024178	NA	NA	uncommon
cg14341177	7.00E-05	0.005272077	-0.228204074	HYPO	chr9	95475787	<i>BICD2</i>	Body	uncommon
cg16238993	1.75E-06	0.004166748	-0.227504899	HYPO	chr19	1132372	<i>SBNO2</i>	Body	uncommon
cg24965479	1.63E-05	0.004326151	-0.226492002	HYPO	chr17	59216714	<i>BCAS3</i>	Body	uncommon
cg23885932	1.42E-06	0.004166748	-0.222921987	HYPO	chr19	1132294	<i>SBNO2</i>	Body	uncommon
cg12907959	4.06E-06	0.004166748	-0.222865701	HYPO	chr22	36552828	<i>APOL3</i>	5'UTR	uncommon
cg12255995	1.67E-06	0.004166748	-0.221182183	HYPO	chr19	1132537	<i>SBNO2</i>	Body	uncommon
cg21571658	1.29E-05	0.00423714	-0.220923508	HYPO	chr15	70785399	NA	NA	uncommon
cg17534029	2.12E-06	0.004166748	-0.220387729	HYPO	chr20	62318426	<i>RTEL1</i>	Body	uncommon
cg25416067	4.83E-06	0.004166748	-0.21971255	HYPO	chr17	75315667		5'UTR	uncommon
cg24039081	3.37E-06	0.004166748	-0.219381074	HYPO	chr3	197155143	NA	NA	uncommon
cg02633924	7.61E-06	0.004166748	-0.217685511	HYPO	chr17	75315563		5'UTR	uncommon
cg11575912	1.94E-07	0.004166748	-0.217060608	HYPO	chr19	5828299	<i>NRTN</i>	3'UTR	uncommon
cg21151432	1.40E-05	0.004246296	-0.216898657	HYPO	chr2	25142229	<i>ADCY3</i>	TSS200	uncommon
cg18082788	8.01E-06	0.004166748	-0.216475551	HYPO	chr6	149806339	<i>ZC3H12D</i>	TSS200	uncommon
cg26348487	3.84E-06	0.004166748	-0.215982196	HYPO	chr19	2488560	NA	NA	uncommon
cg01649611	1.89E-05	0.004358099	-0.21552631	HYPO	chr2	43521066	<i>THADA</i>	Body	uncommon
cg12560128	1.05E-05	0.004166748	-0.214886778	HYPO	chr12	113416061	<i>OAS2</i>	TSS1500	uncommon

ProbeID	pval	qval	delbeta	Diff Meth	chr	pos	UCSC_RefGene_Name	UCSC_RefGene_Group	_TCGA
cg23261919	1.03E-05	0.004166748	-0.214811587	HYPO	chr10	135072960	NA	NA	uncommon
cg21333861	1.19E-05	0.004206736	-0.214388997	HYPO	chr6	33244976	B3GALT4	1stExon	uncommon
cg01947066	2.05E-05	0.004383302	-0.213283565	HYPO	chr15	74289586	PML	Body	uncommon
cg07009284	9.27E-06	0.004166748	-0.21254764	HYPO	chr8	76318515	NA	NA	uncommon
cg24227481	6.52E-06	0.004166748	-0.211782109	HYPO	chr7	73417185	NA	NA	uncommon
cg16474725	1.71E-05	0.004326151	-0.211689105	HYPO	chr6	45979349	CLIC5	Body	uncommon
cg05267204	8.96E-06	0.004166748	-0.211433725	HYPO	chr2	55360999	NA	NA	uncommon
cg13786863	1.05E-05	0.004166748	-0.211144068	HYPO	chr11	68610743	CPT1A	TSS1500	uncommon
cg12247222	1.39E-06	0.004166748	-0.210917134	HYPO	chr19	17403313	ABHD8	3'UTR	uncommon
cg04263436	3.69E-06	0.004166748	-0.210131691	HYPO	chr6	33245467	B3GALT4	1stExon	uncommon
cg10536999	1.77E-05	0.004339511	-0.210039283	HYPO	chr7	26193109	NFE2L3	Body	uncommon
cg26733301	5.06E-06	0.004166748	-0.209463136	HYPO	chr20	61585924	SLC17A9	Body	uncommon
cg25637655	4.01E-05	0.004755949	-0.208798757	HYPO	chr6	29911542	HLA-A	Body	uncommon
cg12666727	6.03E-06	0.004166748	-0.208234078	HYPO	chr1	42128487	HIVEP3	5'UTR	uncommon
cg20834178	3.16E-05	0.004641472	-0.20807154	HYPO	chr4	6944071	TBC1D14	Body	uncommon
cg09495643	6.48E-06	0.004166748	-0.207743319	HYPO	chr19	1132313	SBNO2	Body	uncommon
cg06021088	4.85E-06	0.004166748	-0.207737024	HYPO	chr2	127822551	BIN1	Body	uncommon
cg24366211	7.80E-06	0.004166748	-0.207395094	HYPO	chr11	68610706	CPT1A	TSS1500	uncommon
cg03301058	2.99E-05	0.004607748	-0.206841022	HYPO	chr6	90007856	GABRR2	Body	uncommon
cg14418802	2.84E-05	0.004593272	-0.206814758	HYPO	chr1	9299979	H6PD	5'UTR	uncommon
cg26996569	2.47E-05	0.004501806	-0.206684769	HYPO	chr12	121829743	NA	NA	uncommon
cg04659689	6.86E-06	0.004166748	-0.206589532	HYPO	chr7	2903377	NA	NA	uncommon
cg06172942	2.26E-06	0.004166748	-0.206266992	HYPO	chr3	50378529	RASSF1	TSS200	uncommon
cg20003368	3.83E-06	0.004166748	-0.205836796	HYPO	chr6	44026475	NA	NA	uncommon
cg13393476	4.35E-06	0.004166748	-0.20517207	HYPO	chr19	39894666	NA	NA	uncommon
cg04650676	1.29E-05	0.00423714	-0.205126947	HYPO	chr17	75315486		5'UTR	uncommon
cg10142520	1.43E-05	0.004262603	-0.205020266	HYPO	chr11	65344604	EHBP1L1	Body	uncommon
cg04470054	5.39E-05	0.004968058	-0.204885137	HYPO	chr17	78830472	RPTOR	Body	uncommon
cg16476991	6.66E-06	0.004166748	-0.204680315	HYPO	chr13	114831404	RASA3	Body	uncommon
cg03108070	6.78E-06	0.004166748	-0.20443028	HYPO	chr6	33245490	B3GALT4	1stExon	uncommon
cg18517266	7.63E-07	0.004166748	-0.203525351	HYPO	chr19	5827811	NRTN	Body	uncommon

ProbeID	pval	qval	delbeta	Diff Meth	chr	pos	UCSC_RefGene_Name	UCSC_RefGene_Group	_TCGA
cg08047457	7.82E-07	0.004166748	-0.203501321	HYPO	chr3	50378413	<i>RASSF1</i>	TSS200	uncommon
cg26994377	1.75E-05	0.004326151	-0.203064941	HYPO	chr4	140579069	NA	NA	uncommon
cg21390801	9.10E-06	0.004166748	-0.202502728	HYPO	chr21	42733729	<i>MX2</i>	TSS1500	uncommon
cg23987789	1.78E-05	0.004346823	-0.202457896	HYPO	chr1	7842406	NA	NA	uncommon
cg06879394	6.34E-05	0.005145955	-0.201945359	HYPO	chr6	13274151	<i>PHACTR1</i>	Body	uncommon
cg25690715	1.61E-05	0.004323287	-0.201745344	HYPO	chr17	75315567		5'UTR	uncommon
cg17184704	1.79E-05	0.004346823	-0.201612803	HYPO	chr10	11727286	NA	NA	uncommon
cg16066354	7.77E-06	0.004166748	-0.201004744	HYPO	chr22	39417451	<i>APOBEC3D</i>	1stExon	uncommon
cg22191981	7.11E-06	0.004166748	-0.200930679	HYPO	chr6	41472382	NA	NA	uncommon
cg21853021	2.22E-05	0.004436858	-0.200679113	HYPO	chr11	2920146	<i>SLC22A18AS</i>	Body	uncommon
cg12359904	6.20E-06	0.004166748	-0.20055932	HYPO	chr3	156848454	NA	NA	uncommon
cg10893986	1.19E-05	0.004206736	-0.200419204	HYPO	chr1	3229739	<i>PRDM16</i>	Body	uncommon
cg22027433	1.04E-05	0.004166748	-0.200278337	HYPO	chr17	62208049	<i>ERN1</i>	TSS1500	uncommon
cg21912556	2.13E-05	0.004408225	-0.200271623	HYPO	chr8	18605482	<i>PSD3</i>	Body	uncommon
cg18474072	4.17E-05	0.004791797	0.200852091	HYPER	chr10	664762	<i>DIP2C</i>	Body	uncommon
cg12071328	2.10E-06	0.004166748	0.201590033	HYPER	chr11	20690930	<i>NELL1</i>	TSS200	uncommon
cg03061682	7.43E-05	0.00534141	0.201951187	HYPER	chr15	28352098	NA	NA	uncommon
cg09619624	6.25E-05	0.005140634	0.202960226	HYPER	chr22	18985632	NA	NA	uncommon
cg11701471	3.96E-06	0.004166748	0.204376107	HYPER	chr8	54164051	<i>OPRK1</i>	5'UTR	uncommon
cg11258943	3.70E-05	0.004716354	0.204838082	HYPER	chr9	100615357	<i>FOXE1</i>	TSS200	uncommon
cg27363829	1.95E-06	0.004166748	0.20880463	HYPER	chr3	164914614	<i>SLITRK3</i>	TSS200	uncommon
cg18483908	8.30E-07	0.004166748	0.211233972	HYPER	chr22	19017420	NA	NA	uncommon
cg14500486	3.06E-06	0.004166748	0.213971134	HYPER	chr6	159655392	<i>FNDC1</i>	Body	uncommon
cg21270847	1.05E-06	0.004166748	0.219146448	HYPER	chr1	164604806	<i>PBX1</i>	Body	uncommon
cg18989174	0.00546998	0.048315118	0.222834363	HYPER	chr7	149470820	<i>ZNF467</i>	TSS1500	uncommon
cg01227537	3.12E-06	0.004166748	0.223612005	HYPER	chr3	147127012	<i>ZIC1</i>	TSS200	uncommon

Table 12: Details of the 91 Genes with Differential Methylation (gDMs), Red= Hypomethylated; Green=Hypermethylated; Blue=Significantly Different

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
MX2	chr21:41,361,944-41,409,393	MX Dynamin Like GTPase 2	potent role against viral infection, involved in interferon signaling, tumor suppressor; regulated by IFN beta	15.49	1.92	<0.01	<0.01
APOL3	chr22:36,140,323-36,166,177	Apolipoprotein L3	lipid transport, over expressed in cancer; regulated by IFN alfa	20.14	3.66	<0.01	
OAS2	chr12:112,978,395-113,011,723	2'-5'-Oligoadenylate Synthetase 2	upregulated in pancreatic cancer, works like anti-viral immune response, induced by interferons; regulated by IFN beta	16.55	0.48	<0.01	<0.01
RASSF1	chr3:50,329,782-50,340,980	Ras Association Domain Family Member 1	highly expressed in PAAD; tumour suppressor gene	23.25	13.75		
SPRED2	chr2:65,307,425-65,432,637	Sprouty Related EVH1 Domain Containing 2	control ERK signaling; negatively regulate Ras and MAPK signaling; tumour suppressor	14.78	4.95	<0.01	<0.01
KCNJ5	chr11:128,891,356-128,921,163	Potassium Inwardly Rectifying Channel Subfamily J Member 5	aldosterone-producing adenomas.	0.91	3.67	<0.05	**
NRTN	chr19:5,805,387-5,828,324	Neurturin	promotes survival of neuronal populations.	1.29	5.53	<0.05	0.012
ABHD8	chr19:17,292,131-17,310,236	Abhydrolase Domain Containing 8	hydrolase activity	7.12	4.03		

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
RTEL1	chr20:63,657,810-63,696,253	Regulator Of Telomere Elongation Helicase 1	stability, protection and elongation of telomere	7.52	4.28		
SBNO2	chr19:1,107,637-1,174,268	Strawberry Notch Homolog 2	regulator in the proinflammatory cascade, transcriptional corepression of NF-kappaB in macrophages	39.73	9.65	<0.05	0.03
SIGIRR	chr11:405,716-417,455	Single Ig And TIR Domain Containing	IL-1R8, negatively regulate anti-inflammatory functions of IL37 and IL-1; regulated by IFN beta	59.01	26.35	<0.05	***
ADAP1	chr7:897,900-955,407	ArfGAP With Dual PH Domains 1	GTPase activator activity	43.76	5.76	<0.05	0.02
EHBP1L1	chr11:65,576,034-65,592,650	EH Domain Binding Protein 1 Like 1	May act as Rab effector protein and play a role in vesicle trafficking	47.65	6.68	<0.05	
PRDM16	chr1:3,069,168-3,438,621	PR/SET Domain 16	Binds DNA and functions as a transcriptional regulator	2.35	2.33		
RASA3	chr13:113,977,783-114,132,623	RAS P21 Protein Activator 3	Inhibitory regulator of the Ras-cyclic AMP pathway.	16.1	3.93	<0.05	
RPTOR	chr17:80,544,819-80,966,371	Regulatory Associated Protein Of MTOR Complex 1	Involved in the control of the mammalian target of rapamycin complex 1	8.04	2.88	<0.05	0.05
SLC17A9	chr20:62,952,690-62,969,585	Solute Carrier Family 17 Member 9	Involved in vesicular storage and exocytosis of ATP	14.67	2.15	<0.05	<0.01

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
ACTN2	chr1:236,664,141-236,764,631	Actinin Alpha 2	involved in binding actin to the membrane.	0.26	1.37		
GAL3ST3	chr11:66,040,765-66,049,161	Galactose-3-O-Sulfotransferase 3	distinct tissue distribution and substrate specificities	0.02	0.19		
BHLHE23	chr20:63,005,927-63,007,035	Basic Helix-Loop-Helix Family Member E23	previously reported to be hypermethylated,transcriptional repressor	0	0		
NELL1	chr11:20,669,551-21,575,686	Neural EGFL Like 1	hypermethylated in renal cancer cell line	0.01	0		
FOXG1	chr14:28,766,787-28,770,277	Forkhead Box G1	associated with Rett syndrome	0	0		
TTBK1	chr6:43,243,481-43,288,259	Tau Tubulin Kinase 1	tumour suppressor gene; hypermethylated in many breast and prostate cancer	0	0		
CACNB2	chr10:18,140,424-18,545,881	Calcium Voltage-Gated Channel Auxiliary Subunit Beta 2	Calcium channel, voltage-dependent; hypermethylated in esophageal adenocarcinoma	1.06	2.96		
GSC2	chr22:19,146,993-19,150,292	Goosecoid Homeobox 2	transcriptional regulation; hypermethylated in cancers	0	0		
NPY	chr7:24,284,188-24,291,862	Neuropeptide Y	hypermethylated in gastric cancer; related to pancreatic cancer risk;	1.36	1.87		

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
FOXE1	chr9:97,853,226-97,856,717	Forkhead Box E1	one of the most frequently hypermethylated TSG in pancreatic cancer	0.02	0.01		
GRM7	chr3:6,770,001-7,743,038	Glutamate Metabotropic Receptor 7	Glutamate Metabotropic Receptor 7	0.07	0.01		
SLITRK3	chr3:165,186,720-165,197,109	SLIT And NTRK Like Family Member 3	involved in controlling neurite outgrowth. Enhanced expression of this gene was found in gastrointestinal stromal tumor	0.18	0.01		
TWIST1	chr7:19,020,991-19,117,636	Twist Family BHLH Transcription Factor 1	hypermethylated in pancreatic cancer and many other cancers; promote EMT and chemoresistance; transcription factor	5	0.2	<0.05	0.06
TRH	chr3:129,974,688-129,977,935	Thyrotropin Releasing Hormone		0.04	0		
FAIM2	chr12:49,866,896-49,904,217	Fas Apoptotic Inhibitory Molecule 2	hypermethylated in pancreatic cancer	2.67	0.51	<0.05	0.02
CRYGD	chr2:208,121,607-208,124,524	Crystallin Gamma D	involved in cataract formation.	0.09	0		
ZIC4	chr3:147,386,046-147,407,476	Zic Family Member 4	involved in Dandy-Walker malformation	0	0.01		
ZIC1	chr3:147,393,422-147,510,293	Zic Family Member 1	over expressed in gastric and pancreatic cancer; promoter hypermethylation	0	0.01		

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
			called reduced expression				
KCNA6	chr12:4,806,877-4,851,927	Potassium Voltage-Gated Channel Subfamily A Member 6	Involved in insulin secretion, neuronal excitability, epithelial electrolyte transport	0	0		
KCNA3	chr1:110,653,560-110,674,940	Potassium Voltage-Gated Channel Subfamily A Member 3	potassium gated ion channel; deregulated in cancers; generally overexpressed in cancer	0.28	0.05		
C9orf50	chr9:129,612,225-129,622,275	Chromosome 9 Open Reading Frame 50	Involved in Spastic Paraplegia 20	0.14	0		
TCERG1L	chr10:131,092,391-131,311,721	Transcription Elongation Regulator 1 Like	tumour suppressor; hypermethylated in pancreatic cancers	0.13	0.1		
CPEB1	chr15:82,543,201-82,648,861	Cytoplasmic Polyadenylation Element Binding Protein 1	play a role in cell proliferation and tumorigenesis.	0.96	0.57		
HMX3	chr10:123,135,970-123,139,423	H6 Family Homeobox 3	Diseases associated with HMX3 include Superior Semicircular Canal Dehiscence	0.07	0		
HOXA10	chr7:27,170,592-27,180,261	Homeobox A10	regulate gene expression, morphogenesis, and differentiation.	2.78	0.03	<0.01	<0.01
IRF4	chr6:391,752-411,443	Interferon Regulatory Factor 4	activation of innate and adaptive immune systems.	0.74	0.05		

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
SLC6A3	chr5:1,392,794-1,445,440	Solute Carrier Family 6 Member 3	dependence on alcohol and cocaine, protection against nicotine dependence.	0.03	0.01		
VSTM2A	chr7:54,542,064-54,571,080	V-Set And Transmembrane Domain Containing 2A	Diseases include Phosphoserine Phosphatase Deficiency and Serine Deficiency associated with pre-B-cell acute lymphoblastic leukemia.	0.13	0.03		
PBX1	chr1:164,555,584-164,899,296	PBX Homeobox 1	Plays a role in regulating insulin levels and body fat accumulation in response to a high fat diet	16	13.31		
ADCY3	chr2:24,819,169-24,920,237	Adenylate Cyclase 3	Shows activity against Vif-deficient HIV-1	26.39	7.14	<0.05	
APOBEC3D	chr22:39,021,113-39,033,277	Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3D	galactosyltransferase activity and ganglioside	3.13	0.26	<0.05	
B3GALT4	chr6:33,277,123-33,284,832	Beta-1,3-Galactosyltransferase 4	Stimulates histone acetyltransferase activity	13.73	1.42	<0.05	
BCAS3	chr17:60,677,453-61,392,838	BCAS3 Microtubule Associated Cell Migration Factor	Facilitates the binding of RAB6A to the Golgi by stabilizing its GTP-bound form	9.98	3.3	<0.05	
BICD2	chr9:92,711,363-92,764,841	BICD Cargo Adaptor 2		9.46	2.63	<0.05	

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
BIN1	chr2:127,048,023-127,107,288	Bridging Integrator 1	Is a key player in the control of plasma membrane curvature, membrane shaping and membrane remodeling	32.53	8.91	<0.05	
BMP4	chr14:53,949,736-53,958,761	Bone Morphogenetic Protein 4	Induces SRC phosphorylation which, in turn, activates VEGFR2, leading to an angiogenic response	26.34	1.22	<0.05	<0.05
C3orf59	chr3:192,796,815-192,917,856	Mab-21 Domain Containing 2	enables protein-containing complex binding	0	0		
CAMK1D	chr10:12,349,514-12,835,545	Calcium/Calmodulin Dependent Protein Kinase ID	differentiation and activation of neutrophil cells, and apoptosis of erythroleukemia cells	4.07	5.38		
CCDC48	chr3:129,001,304-129,040,742	EF-Hand And Coiled-Coil Domain Containing 1	enables calcium ion binding	0	0		
CCNA1	chr13:36,430,488-36,442,870	Cyclin A1	primarily function in the control of the germline meiotic cell cycle	0.24	0.02		
CLIC5	chr6:45,880,827-46,129,819	Chloride Intracellular Channel 5	Can insert into membranes and form poorly selective ion channels	2.99	0.4	<0.05	
CPT1A	chr11:68,754,620-68,844,410	Carnitine Palmitoyltransferase 1A	Plays an important role in hepatic triglyceride metabolism	18.04	7.26	<0.05	

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
CUX1	chr7:101,815,904-102,283,958	Cut Like Homeobox 1	Transcription factor involved in the control of neuronal differentiation in the brain	54.69	16.34	<0.05	
DIP2C	chr10:274,190-689,710	Disco Interacting Protein 2 Homolog C	Diseases associated with DIP2C include Spastic Hemiplegia and Fraser Syndrome 1	5.95	2.03	<0.05	
EGR3	chr8:22,687,659-22,693,480	Early Growth Response 3	immediate early growth response, gene 3	3.8	0.41	<0.05	
ERN1	chr17:64,039,080-64,132,469	Endoplasmic Reticulum To Nucleus Signaling 1	acts as a key sensor for the endoplasmic reticulum unfolded protein response	6.86	12.75		
F7	chr13:113,105,773-113,120,685	Coagulation Factor VII	Initiates the extrinsic pathway of blood coagulation	0.16	0.06		
FNDC1	chr6:159,169,400-159,272,108	Fibronectin Type III Domain Containing 1	May be an activator of G protein signaling	12.98	0.34	<0.05	0.02
FOXI2	chr10:127,737,185-127,741,183	Forkhead Box I2	Possible transcriptional activator.	0.02	0.02		
GABRR2	chr6:89,254,464-89,315,299	Gamma-Aminobutyric Acid Type A Receptor Subunit Rho2	plays a role in retinal neurotransmission	0.13	0.21		
GALR1	chr18:77,249,848-77,277,900	Galanin Receptor 1	Receptor for the hormone galanin	0.01	0.01		
H6PD	chr1:9,234,767-9,271,337	Hexose-6-Phosphate Dehydrogenase/Glucose 1-Dehydrogenase	provide reducing equivalents such as NADPH	16.08	6.38	<0.05	0.01

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
HIVEP3	chr1:41,506,365-42,035,934	HIVEP Zinc Finger 3	Essential regulator of adult bone formation	4.12	1.04		
HLA-A	chr6:29,941,260-29,945,884	Major Histocompatibility Complex, Class I, A	In complex with B2M/beta 2 microglobulin displays primarily viral and tumor-derived peptides on antigen-presenting cells	1174.56	61.31	<0.01	
LHX1	chr17:36,936,785-36,944,612	LIM Homeobox 1	May play a role in early mesoderm formation	0.03	0		
LOC645323	chr5:88,499,582-88,691,057	Long Intergenic Non-Protein Coding RNA 461	may be involved in tumorigenesis, as depletion by siRNA suppressed glioma cell division	0	0		
NFE2L3	chr7:26,152,198-26,187,137	Nuclear Factor, Erythroid 2 Like 3	Activates erythroid-specific, globin gene expression	19.41	5.43	<0.05	<0.01
OGDH	chr7:44,606,572-44,709,066	Oxoglutarate Dehydrogenase	required for lysine succinylation of histones:	41.13	10.64	<0.05	
OPRK1	chr8:53,225,724-53,251,637	Opioid Receptor Kappa 1	functions as receptor for various synthetic opioids	0.01	0		
PARP4	chr13:24,420,930-24,512,993	Poly(ADP-Ribose) Polymerase Family Member 4	enables NAD ⁺ ADP-ribosyltransferase activity	30.55	5.95	<0.05	<0.01
PHACTR1	chr6:12,716,554-13,290,462	Phosphatase And Actin Regulator 1	Involved in the regulation of cortical neuron migration and	2.15	2.12		

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
			dendrite arborization				
PIK3CD	chr1:9,629,889-9,729,114	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Delta	activating signaling cascades involved in cell growth, survival, proliferation, motility and morphology.	4.57	1.12	<0.05	
PML	chr15:73,994,673-74,047,827	PML Nuclear Body Scaffold	including tumor suppression, transcriptional regulation, apoptosis, senescence, DNA damage response, and viral defense mechanisms	52.4	7.26	<0.05	
PSD3	chr8:18,527,303-19,086,909	Pleckstrin And Sec7 Domain Containing 3	Guanine nucleotide exchange factor for ARF6	4.44	2.11		
PTPRN2	chr7:157,539,056-158,587,823	Protein Tyrosine Phosphatase Receptor Type N2	Plays a role in vesicle-mediated secretory processes.	14.6	9.96		
40057	chr17:77,280,569-77,500,596	Septin 9	May play a role in the internalization of 2 intracellular microbial pathogens	144.79	14.77		
SLC22A18AS	chr11:2,887,344-2,903,575	Solute Carrier Family 22 Member 18 Antisense	Diseases associated with SLC22A18AS include Beckwith-Wiedemann Syndrome.	3.94	0.07	<0.05	
SLC35F2	chr11:107,790,991-107,928,293	Solute Carrier Family 35 Member F2	Putative solute transporter	10.83	2.01	<0.05	

GENE NAME	GENOMIC LOCATION (GRCh38/hg38)	FULL NAME	FUNCTION	Expression in cancer (TPM)	Expression in normal (TPM)	P-value (diff in expression)	Stage-wise change (p-value)
TBC1D14	chr4:6,909,242-7,033,118	TBC1 Domain Family Member 14	Plays a role in the regulation of starvation-induced autophagosome formation	16.46	7.08	<0.05	0.01
THADA	chr2:43,230,851-43,596,046	THADA Armadillo Repeat Containing	functions in megakaryopoiesis and platelet formation.	17.8	6.09	<0.05	
TM6SF1	chr15:83,107,572-83,145,403	Transmembrane 6 Superfamily Member 1	May function as sterol isomerase	3.02	0.39	<0.05	
UBXN11	chr1:26,281,328-26,318,363	UBX Domain Protein 11	Promotes RHOA activation mediated by GNA12 and GNA13	17.26	10.09		
ZC3H12D	chr6:149,446,795-149,485,014	Zinc Finger CCCH-Type Containing 12D	regulate cell growth likely by suppressing RB1 phosphorylation	0.77	0.11		
ZNF467	chr7:149,764,182-149,776,323	Zinc Finger Protein 467	promotes adipocyte differentiation and suppresses osteoblast differentiation in the bone marrow	5.13	4.81		

Table 13: Summary of the Result Obtained from PCMDB. The Number of Studies that had Reported for Hypo- (Red) or Hyper-Methylation (Green) of gDMs, Either in Pancreatic Cancer Tissue or Cell Lines are listed. Methylation of the highlighted genes were validated using MSP.

GENE	Tissue			Tissue			Cell Lines			Cell Lines		
	Hypo	Hyper	Notknown	% of studies showing hypo-methylation	% of studies showing hyper-methylation	Total	Hypo	Hyper	Notknown	% of studies showing hypo-methylation	% of studies showing hyper-methylation	Total
<i>MX2</i>												0
<i>APOL3</i>												0
<i>OAS2</i>												0
<i>RASSF1</i>								5	15		25	20
<i>SPRED2</i>		1			100	1						0
<i>KCNJ5</i>								9			100	9
<i>NRTN</i>												0
<i>ABHD8</i>												0
<i>RTEL1</i>												0
<i>SBNO2</i>							9			100		9
<i>SIGIRR</i>												0
<i>EHBP1L1</i>		1			100	1						
<i>ADAP1</i>												
<i>SLC17A9</i>		1			100	1						
<i>RPTOR</i>		1			100	1						
<i>RASA3</i>		1			100	1	9			100		9
<i>PRDM16</i>		1			100	1	9	9		50	50	18
<i>ACTN2</i>								21	19		190.5	40
<i>GAL3ST3</i>												0
<i>BHLHE23</i>		1			100	1		9			100	9
<i>NELL1</i>		1			100	1		24	25		49	49

GENE	Tissue			Tissue			Cell Lines			Cell Lines		
	Hypo	Hyper	Notknown	% of studies showing hypo-methylation	% of studies showing hyper-methylation	Total	Hypo	Hyper	Notknown	% of studies showing hypo-methylation	% of studies showing hyper-methylation	Total
<i>FOXG1</i>		1			100	1	9	9		50	50	18
<i>TTBK1</i>		1			100	1		9			100	9
<i>CACNB2</i>		1			100	1		27	23		54	50
<i>GSC2</i>		1			100	1		9			100	9
<i>NPY</i>		2			100	2		32	9		78	41
<i>FOXE1</i>	3	1	8	25	8.3	12	1	18	8	3.7	66.7	27
<i>SLC6A3</i>		1			100	1		24	25		49	49
<i>GRM7</i>												0
<i>SLITRK3</i>								11	29		27.5	40
<i>TWIST1</i>		2			100	2		87	63		58	150
<i>TRH</i>								20	8		71.4	28
<i>FAIM2</i>		1			100	1		21	8		72.4	29
<i>CRYGD</i>												0
<i>ZIC4</i>		1			100	1		9			100	9
<i>ZIC1</i>		1			100	1		15	14		51.7	29
<i>KCNA6</i>		1			100	1		30	20		60	50
<i>KCNA3</i>		1			100	1		28	22		56	50
<i>C9orf50</i>		1			100	1		9			100	9
<i>TCERG1L</i>								9	1		90	10
<i>CPEB1</i>		1			100	1		9			100	9
<i>HMX3</i>		1			100	1		9			100	9
<i>HOXA10</i>		1			100	1		24	5		82.8	29
<i>IRF4</i>		1			100	1		20	9		69	29
<i>VSTM2A</i>		1			100	1		9			100	9
<i>PBX1</i>												0

GENE	Tissue			Tissue			Cell Lines			Cell Lines		
	Hypo	Hyper	Notknown	% of studies showing hypo-methylation	% of studies showing hyper-methylation	Total	Hypo	Hyper	Notknown	% of studies showing hypo-methylation	% of studies showing hyper-methylation	Total
GALR1		2			100	2		78	44		63.9	122
FOXI2		1			100	1		9			100	9
FNDC1								9			100	9
F7		1			100	1		9			100	9
EGR3		1			100	1		9			100	9
CCNA1		2			100	2		58	23		71.6	81
CCDC48								9			100	9
BMP4	1	1		50	50	2		12			100	12
BICD2		1			100	1						
TM6SF1		1			100	1		10	11		47.61	21
ZNF467		1			100	1						
OPRK1								15	25		37.5	40
LHX1	36	1	11	75	2.08	48	4	25	18	8.51	53.19	47

Table 14: Fold Changes of the Expressions of gDMs in the Validation Cohort.

SAMPLE_ID	FOLD/SI_GIRR_G	FOLD/SI_GIRR_A	FOLD/KC_NA6_G	FOLD/K_CNA6_A	FOLD/RAS_SF1_G	FOLD/RA_SSF1_A	FOLD/FA_IM2_G	FOLD/F_AIM2_A	FOLD/NPY_G	FOLD/NPY_A	FOLD/FO_XE1_G	FOLD/FO_XE1_A	FOLD/SLI_TRK3_G	FOLD/SLI_TRK3_A
1T	221.99	34.72	55.32	4.08	71033.65	5240.66	620.88	45.81	21.43	1.58	268.81	19.83	0.02	0.01
5T	36.98	7.82	442.44	757.55	88664.96	151812.49	0.34	0.58	43.64	74.72	0.00	0.03	ND	ND
10T	2.00	2.70	0.35	0.40	0.65	0.76	ND	ND	2.58	2.97	1.12	1.30	0.66	0.76
18T	0.51	1.02	0.19	0.14	2714.67	1988.68	1.38	1.01	1.83	1.34	8.34	6.11	0.14	0.10
22T	1.87	1.62	6.31	3.91	0.04	0.03	4.04	2.50	26.68	16.53	14.50	8.99	0.35	0.21
25T	5.75	1.85	ND	ND	4778.74	8566.86	0.03	0.06	0.12	0.21	0.19	0.35	ND	ND
31T	2.74	2.28	0.46	0.62	1.05	1.42	0.90	1.22	1.98	2.69	0.95	1.29	ND	ND
42T	0.45	0.09	0.08	0.01	6511.78	860.97	0.41	0.05	1.16	0.15	0.00	0.00	0.09	0.01
48T	2.58	0.15	3.63	0.51	6.48	0.92	2.01	0.28	1.39	0.20	3.75	0.53	0.00	0.00
64T	9.03	0.05	ND	ND	5.09	0.23	ND	ND	ND	ND	ND	ND	ND	ND
67T	3.53	2.74	0.71	1.42	8.30	16.65	4.92	9.87	ND	ND	0.00	0.00	5.14	10.32
76T	4696.99	3046.31	1198.67	486.09	1850467.03	750413.89	22.12	8.97	1.00	0.40	20.37	8.26	ND	ND
84T	11.58	3.59	ND	ND	36.50	8.83	1.97	0.48	ND	ND	55.91	13.52	6.51	1.58
99T	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.10	0.00	0.02	0.00	0.00	0.00	0.02
107T	0.22	0.15	ND	ND	2440.58	4193.51	0.04	0.08	0.09	0.16	0.00	0.00	0.06	0.11
120T	6.82	3.32	ND	ND	0.06	0.01	4.36	0.89	18.50	3.76	40.16	8.17	7.23	1.47
122T	176.44	20.54	214.97	12.43	1.34	0.08	ND	ND	103.21	5.97	1467.29	84.83	117.27	6.78
147T	13.27	8.82	ND	ND	0.23	0.13	3.22	1.79	0.37	0.21	91.50	51.04	8.30	4.63
SAMPLE_ID	FOLD/SI_GIRR_G	FOLD/SI_GIRR_A	FOLD/KC_NA6_G	FOLD/K_CNA6_A	FOLD/RAS_SF1_G	FOLD/RA_SSF1_A	FOLD/FA_IM2_G	FOLD/F_AIM2_A	FOLD/NPY_G	FOLD/NPY_A	FOLD/FO_XE1_G	FOLD/FO_XE1_A	FOLD/SLI_TRK3_G	FOLD/SLI_TRK3_A
1T	UP	UP	UP	UP	UP	UP	UP	UP	UP	Down	UP	UP	Down	Down
5T	UP	UP	UP	UP	UP	UP	Down	Down	UP	UP	Down	Down	ND	ND
10T	UP	UP	Down	Down	Down	Down	ND	ND	UP	UP	Down	Down	Down	Down
18T	Down	Down	Down	Down	UP	UP	Down	Down	Down	Down	UP	UP	Down	Down
22T	UP	Down	UP	UP	Down	Down	UP	UP	UP	UP	UP	UP	Down	Down
25T	UP	UP	ND	ND	UP	UP	Down	Down	Down	Down	Down	Down	ND	ND
31T	UP	UP	Down	Down	Down	Down	Down	Down	UP	UP	Down	Down	ND	ND
42T	Down	Down	Down	Down	UP	UP	Down	Down	Down	Down	Down	Down	Down	Down
48T	UP	Down	UP	Down	UP	Down	UP	Down	Down	Down	UP	Down	Down	Down

64T	UP	Down	ND	ND	UP	Down	ND	ND	ND	ND	ND	ND	ND	ND
67T	UP	UP	Down	Down	UP	UP	UP	UP	ND	ND	Down	Down	UP	UP
76T	UP	UP	UP	UP	UP	UP	UP	UP	Down	Down	UP	UP	ND	ND
84T	UP	UP	ND	ND	UP	UP	UP	Down	ND	ND	UP	UP	UP	Down
99T	Down	Down	Down	Down	Down	Down	Down	Down	Down	Down	Down	Down	Down	Down
107T	Down	Down	ND	ND	UP	UP	Down	Down	Down	Down	Down	Down	Down	Down
120T	UP	UP	ND	ND	Down	Down	UP	Down	UP	UP	UP	UP	UP	Down
122T	UP	UP	UP	UP	Down	Down	ND	ND	UP	UP	UP	UP	UP	UP
147T	UP	UP	ND	ND	Down	Down	UP	Down	Down	Down	UP	UP	UP	UP
	HYPO	HYPO	HYPER	HYPER	HYPO	HYPO	HYPER	HYPER	HYPER	HYPER	HYPER	HYPER	HYPER	HYPER
	77.70%	61.11%	33.33%	38.88%	61.11%	50%	38.88%	61.11%	44.44%	50%	44.44%	50%	44.44%	55.50%
ND= Not Derived														

Table 15: Comparison of Methylation Status and Gene Expression Status of Selected Genes Mapped to Enriched Pathway in Pancreatic Cancer.

Gene	Function	Methylation status (our data)	TCGA VS GTEx (RNA SEQ)	TCGA 9T vs 9N (METH ARRAY)	TCGA 185T vs 10N (METH ARRAY)
RASSF1A	highly expressed in PAAD; tumour suppressor gene	Hypo	>0.05	Hyper	0.051
KCNJ5		Hypo	<0.05	Hypo	0.139
SIGIRR	IL-1R8, negatively regulate anti-inflammatory functions of IL37 and IL-1; regulated by IFN beta	Hypo	<0.05	Hypo	0.000109
OAS2	upregulated in pancreatic cancer, works like anti-viral immune response, induced by interferons; regulated by IFN beta	Hypo	<0.05	Hypo	0.0000188
FOXP1	associated with Rett syndrome	Hyper	>0.05	Hyper	0.00000225
TWIST1	hypermethylated in pancreatic cancer and many other cancers; promote EMT and chemoresistance; transcription factor	Hyper	<0.05	Hyper	0.000452
KCNA6	Involved in insulin secretion, neuronal excitability	Hyper		Hyper	0.000000894
KCNA3	potassium gated ion channel	Hyper	>0.05	Hyper	0.00000806
IRF4	activation of innate and adaptive immune systems.	Hyper	>0.05	Hyper	0.000115

Table 16: Functional Analysis of gDMs with Metascape

Term	Description	LogP	Log(q-value)	Symbols
GO:0043269	regulation of ion transport	-7.833380217	-3.472	ACTN2,BIN1,CACNB2,GRM7,KCNA3,KCNA6,KCNJ5,OPRK1,PML,TRH,TWIST1,CLIC5,SLC6A3,RASA3,THADA
GO:0007423	sensory organ development	-6.739708485	-2.855	BMP4,CRYGD,FOGX1,LHX1,PBX1,SLC6A3,TWIST1,ZIC1,SPRED2,HMX3,FOXE1,NELL1,HOXA10,NRTN,PML,SBNO2
R-HSA-909733	Interferon alpha/beta signaling	-6.370927133	-2.854	HLA-A,IRF4,MX2,OAS2,PML,PIK3CD,TWIST1,OPRK1,APOBEC3D
GO:0097305	response to alcohol	-6.064617431	-2.744	ADCY3,CPT1A,F7,OPRK1,SLC6A3,TRH,H6PD,NPY,GRM7,CACNB2,OAS2,GABRR2,FOGX1,PIK3CD,FOXE1,GALR1,PTPRN2,ERN1
GO:0048562	embryonic organ morphogenesis	-5.65903039	-2.528	BMP4,FOGX1,FOXE1,LHX1,TWIST1,ZIC1,HMX3,PBX1,HOXA10,PRDM16,ACTN2,MX2,BCAS3
GO:0002009	morphogenesis of an epithelium	-5.348879479	-2.349	FOXE1,LHX1,PBX1,PIK3CD,PML,TWIST1,BMP4,EGR3,ERN1,PRDM16,SPRED2,OPRK1,RTEL1,NRTN
WP1403	AMP-activated protein kinase (AMPK) signaling	-4.922386212	-2.134	CPT1A,PIK3CD,CCNA1,RPTOR,PML,ADCY3,NPY,CPEB1,EGR3,ERN1,HLA-A,KCNJ5
GO:0043086	negative regulation of catalytic activity	-3.690969969	-1.472	BIN1,BMP4,GALR1,GRM7,BCAS3,RPTOR,SPRED2,PHACTR1,CPT1A,RASA3
GO:0043627	response to estrogen	-3.297624453	-1.258	F7,OPRK1,BCAS3
ko05202	Transcriptional misregulation in cancer	-3.281271615	-1.247	HOXA10,PBX1,PML,CCNA1
GO:0045664	regulation of neuron differentiation	-3.151030299	-1.182	BIN1,BMP4,FOGX1,PBX1,CUX1,SLITRK3,TTBK1

Table 17: Top 4 Enriched Pathways Observed from our Dataset.

Category	Molecular function	Functional phenotype in PDA	Reference
Regulation of ion transport	massive fibrosis of the tumor stroma (desmoplasia) /immune evasion. / involved in gemcitabine resistance	PDAC microenvironment population / uncontrolled proliferation, evasion of apoptosis, sustained angiogenesis and promotion of invasion and migration /low overall survival.	https://www.frontiersin.org/articles/10.3389/fphar.2020.568993/full ; https://www.frontiersin.org/articles/10.3389/fphar.2020.586599/full ;
Interferon signalling	involved in the development of PDAC. / affect innate /e viability and migration of human pancreatic cancer cell immune signaling pathways	PDAC development / treatment targeting IFN and TLR3 /PDAC cell proliferation and invasion signaling pathways	https://pubmed.ncbi.nlm.nih.gov/32727774/ ; https://www.pnas.org/content/118/8/e2012469118
Morphogenesis and development	collagen-rich dense extracellular matrix (ECM) that promotes malignancy of cancer cells and presents a barrier for drug delivery. /secretes proteinase, promotes fibrillar collagen deposition from both cancer cells	pro-tumorigenic roles /program of gene expression /progenitor cell expansion and migration known as branching morphogenesis, /treatment-refractory malignancy. / pancreatic cancer invasiveness	https://www.nature.com/articles/s41467-021-22490-9#Sec9 / ; https://www.molbiolcell.org/doi/10.1091/mbc.e09-03-0203 ; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3117430/
Transcriptional misregulation	grade-specific gene expression programs in human PDAC /controlling differentiation in PDACs	poor outcomes and early progression of pancreatic cancer / prognosis in pancreatic cancer. / correlation with lymphatic and perineural invasion / poor prognosis.	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6337016/ ; https://pubmed.ncbi.nlm.nih.gov/26769127/

Table 18: Results of GO enrichment analysis on each MCODE network (Log (p-value) is mentioned with each annotation)

Network	Annotation	Genes
MCODE_1	R-HSA-913531 Interferon Signalling -10.8;R-HSA-909733 Interferon alpha/beta signalling -9.8;R-HSA-877300 Interferon gamma signalling -9.3	IRF4, PML, MX2, OAS2, HLA-A
MCODE_2	R-HSA-418594 G alpha (i) signalling events -9.7;R-HSA-388396 GPCR downstream signalling -8.3;R-HSA-372790 Signaling by GPCR -8.0	ADCY3, NPY, GALR1, OPRK1

Table 19: Summary of the Output Obtained from the Pearson Corelation Test

		Correlations																						
		Sex	Alcohol	Smoking	Chewing	DiabeticHistory	JaundiceHistory	PancreaticLesion	TumorStage	LymphNode	Survivalstatus	Days	TP53	KRAS	Her2	FAIM2NPY	FOLDFAIM2_A	FOLDNPY_A	FOLDLITRK3_A	FOLDSIGIRR_A	FOLDKCN6_A	FOLDRASSF1_A	FoldMX2	FoldGALR1
Sex	Pearson Correlation			-.419*																				
	Sig. (2-tailed)			.037																				
Alcohol	Pearson Correlation			.577*																				
	Sig. (2-tailed)			.003																				
Smoking	Pearson Correlation	-.419*	.577*																					
	Sig. (2-tailed)	.037	.003																					
Chewing	Pearson Correlation													.552*										
	Sig. (2-tailed)													.005										
Diabetic History	Pearson Correlation															-.484*								
	Sig. (2-tailed)															.026								

		Correlations																						
		Sex	Alcohol	Smoking	Chewing	DiabeticHistory	JaundiceHistory	PancreaticLesion	TumorStage	LymphNode	Survivalstatus	Days	TP53	KRAS	Her2	FAIM2NPY	FOLDFAIM2_A	FOLDNPY_A	FOLDSLITRK3_A	FOLDSIGIRR_A	FOLDKCNA6_A	FOLDRASSF1_A	FoldMX2	FoldGALR1
Jaundice History	Pearson Correlation								.559*	.559*														
	Sig. (2-tailed)								.038	.038														
Major Illness	Pearson Correlation			-.730																				
	Sig. (2-tailed)			.062																				
Pancreatic Lesion	Pearson Correlation																					1.000*		
	Sig. (2-tailed)																					0.000		
Tumor Stage	Pearson Correlation						.559*			1.000*	-.545*													-.538**
	Sig. (2-tailed)						.038			0.000	.013													.008
Lymph Node	Pearson Correlation						.559*		1.000*		-.545*													-.538**
	Sig. (2-tailed)						.038		0.000		.013													.008
Survivalstatus	Pearson Correlation								-.545*	-.545*		.536*	-.453*									-.683**		

Correlations																								
		Sex	Alcohol	Smoking	Chewing	DiabeticHistory	JaundiceHistory	PancreaticLesion	TumorStage	LymphNode	Survivalstatus	Days	TP53	KRAS	Her2	FAIM2NPY	FOLDFAIM2_A	FOLDNPY_A	FOLDSLITRK3_A	FOLDSIGIRR_A	FOLDKCNA6_A	FOLDRASSF1_A	FoldMX2	FoldGALR1
	Sig. (2-tailed)								.013	.013		.018	.045									.004		
Days	Pearson Correlation										.536*					-.517*					-.866**			
	Sig. (2-tailed)										.018					.040					.001			
TP53	Pearson Correlation										-.453*													
	Sig. (2-tailed)										.045													
KRAS	Pearson Correlation				.552*																			
	Sig. (2-tailed)				.005																			
FAIM2NPY	Pearson Correlation					-.484*						-.517*					.669*	.583*			.592*		.478*	
	Sig. (2-tailed)					.026						.040					.001	.004			.016		.045	
FOLDFAIM2_A	Pearson Correlation															.669*			.545*					
	Sig. (2-tailed)															.001			.019					

		Correlations																						
		Sex	Alcohol	Smoking	Chewing	DiabeticHistory	JaundiceHistory	PancreaticLesion	TumorStage	LymphNode	Survivalstatus	Days	TP53	KRAS	Her2	FAIM2NPY	FOLDFAIM2_A	FOLDNPY_A	FOLDSLITRK3_A	FOLDSIGIRR_A	FOLDKCNA6_A	FOLDRASSF1_A	FoldMX2	FoldGALR1
FOLDNPY_A	Pearson Correlation															.583*								
	Sig. (2-tailed)															.004								
FOLDSLITRK3_A	Pearson Correlation																.545*							.658**
	Sig. (2-tailed)																.019							.002
FOLDSIGIRR_A	Pearson Correlation																						.685*	.586**
	Sig. (2-tailed)																						.001	.003
FOLDKCNA6_A	Pearson Correlation											-.866**				.592*							.625*	
	Sig. (2-tailed)											.001				.016							.022	
FOLDRASSF1_A	Pearson Correlation							1.000*			-.683**													
	Sig. (2-tailed)							0.000			.004													
FoldMX2	Pearson Correlation															.478*				.685*	.625*			

		Correlations																						
		Sex	Alcohol	Smoking	Chewing	DiabeticHistory	JaundiceHistory	PancreaticLesion	TumorStage	LymphNode	Survivalstatus	Days	TP53	KRAS	Her2	FAIM2NPY	FOLDFAIM2_A	FOLDNPY_A	FOLDLITRK3_A	FOLDSIGIRR_A	FOLDKCNA6_A	FOLDRASSF1_A	FoldMX2	FoldGALR1
	Sig. (2-tailed)															.045				.001	.022			
FoldGALR1	Pearson Correlation								-.538**	-.538**									.658*	.586*				
	Sig. (2-tailed)								.008	.008									.002	.003				
*. Correlation is significant at the 0.05 level (2-tailed).																								
**. Correlation is significant at the 0.01 level (2-tailed).																								
a. Cannot be computed because at least one of the variables is constant.																								

Table 20: Summary of the Output obtained from the Kendell Tau Corelation Test

Correlations																														
Ken dall' s tau_ b	Sex	Corre lation Coeff icient				- .41 9*																					.4 57 *			
		Sig. (2- tailed)				.04 0																						.0 33		
	Alcohol	Corre lation Coeff icient				.57 7**																						- .0 14		
		Sig. (2- tailed)				.00 6																						.9 50		
	Smoking	Corre lation Coeff icient	- .41 9*		.57 7**																									
		Sig. (2- tailed)	.04 0		.00 6																									
	Chewing	Corre lation Coeff icient														.55 2**														
		Sig. (2- tailed)														.00 8														
	DiabeticHi story	Corre lation Coeff icient																- .48 4*												
		Sig. (2- tailed)																												

Correlations																												
		Sig. (2- tailed)																	.03 0									
JaundiceHistory	Correlation Coefficient									.55 9*	.55 9*																	
	Sig. (2- tailed)									.04 4	.04 4																	
TumorStage	Correlation Coefficient						.5 59*			1.0 00			-. .54 5*														-. .53 8*	
	Sig. (2- tailed)						.0 44						.01 8														.01 2	
LymphNode	Correlation Coefficient						.5 59*			1.0 00*			-. .54 5*														-. .53 8*	
	Sig. (2- tailed)						.0 44						.01 8														.01 2	
TumorHistologyType	Correlation Coefficient																				.49 4*							
	Sig. (2- tailed)																				.03 0							
SurvivalStatus	Correlation Coefficient									-	-			.46	-											-		

Correlations																												
	tus	lation Coeff icient									.54 5*	.54 5*			8*	.45 3*									.68 3**			
		Sig. (2- tailed)									.01 8	.01 8			.01 8	.04 8									.00 8			
	Days	Corre lation Coeff icient												.46 8*					- .45 3*						- .73 0*			
		Sig. (2- tailed)												.01 8					.03 7						.01 1			
	TP53	Corre lation Coeff icient												- .45 3*														
		Sig. (2- tailed)												.04 8														
	FAIM2NP Y	Corre lation Coeff icient					- .48 4*								- .45 3*					.66 9**	.58 3**				.59 2*			.47 8*
		Sig. (2- tailed)					.03 0								.03 7					.00 3	.00 8				.02 2			.04 9
	FOLDFAI M2_A	Corre lation Coeff icient																	.66 9**				.51 4*					
		Sig. (2- tailed)																	.00 3				.02 8					

Correlations																												
		tailed)																										
	FOLDNPY_A	Correlation Coefficient																.583**										
		Sig. (2-tailed)																.008										
	FOLDSLI TRK3_A	Correlation Coefficient										.494*							.514*								.682**	
		Sig. (2-tailed)										.030							.028								.003	
	FOLDSIGI RR_A	Correlation Coefficient																								.685**	.586**	
		Sig. (2-tailed)																								.004	.006	
	FOLDKC NA6_A	Correlation Coefficient												-.730*				.592*								.625*		
		Sig. (2-tailed)												.011				.022								.030		
	FOLDRAS SF1_A	Correlation Coefficient								1.000*				-.683**														

Correlations																													
		icient																											
		Sig. (2-tailed)											.008																
	FoldMX2	Corre lation Coeff icient																.478*					.685**	.625*					
		Sig. (2-tailed)																.049					.004	.030					
	FoldGALR 1	Corre lation Coeff icient										-.538*	-.538*										.682**	.586**					
		Sig. (2-tailed)										.012	.012										.003	.006					
*. Correlation is significant at the 0.05 level (2-tailed).																													
**. Correlation is significant at the 0.01 level (2-tailed).																													

Table 21: Validation of KCNJ5 expression and methylation

GSE211398	Gene	P value	Tumor Mean Value	Non-Tumor Mean Value	Difference	SE of difference	df	FC,Log2
	KCNJ5	0.000374101	2.58229	9.94222	-7.35993	1.80154	26	-1.9
GSE16515	ID	adj.P.Val	logFC	Gene.symbol	Gene.title			
	237186_at	3.37E-04	-1.1319655	KCNJ5	potassium voltage-gated channel subfamily J member 5			

Validation of KCNJ5 methylation status				
GSE74071	ID	adj.P.Val	P.Value	F
	cg25508319	2.94E-02	1.85E-03	8.07
GSE49149	ID	adj.P.Val	P.Value	logFC
	cg25508319	4.68E-20	1.14E-21	3.18E-01

Table 22: CMap derived *KCNJ5* associated small molecule targeting

Gene	Drug	Regulatory approval	Indication	Interaction score
KCNJ5	AMILORIDE	Approved	antihypertensive agent	1.310667864
KCNJ5	VERNAKALANT HYDROCHLORIDE	Not Approved		1.966001797
KCNJ5	PHORBOL 12-MYRISTATE 13-ACETATE	Not Approved		0.786400719
KCNJ5	FLUPIRTINE	Not Approved	analgesic	0.786400719
KCNJ5	VERAPAMIL	Approved	antihypertensive agent	0.310421336

Table 23: 47 Probes

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
cg13295878	9.02	8266.08	4.89	-0.57	18.61	0.57	1.20E+08	0	1.84	0.07	3.94
cg15849060	-6.48	0	4.26	-14.84	1.88	0	6.53	0	-1.52	0.13	2.96
cg07473553	0.49	1.64	2.69	-4.78	5.77	0.01	321.07	0	0.18	0.85	0.23
cg08210297	2.77	16	2.29	-1.72	7.27	0.18	1434.27	0	1.21	0.23	2.14
cg14659346	0.58	1.79	1.34	-2.05	3.21	0.13	24.9	0	0.44	0.66	0.59
cg19654061	-0.13	0.88	3.52	-7.02	6.77	0	867.7	0	-0.04	0.97	0.04
cg09291474	0.12	1.13	5.34	-10.34	10.59	0	39540.92	0	0.02	0.98	0.03
cg08196512	4.34	76.85	5.56	-6.56	15.24	0	4.16E+06	0	0.78	0.43	1.2
cg15944459	0.43	1.54	3.48	-6.38	7.24	0	1396.77	0	0.12	0.9	0.15
cg27505627	1.34	3.82	3.67	-5.86	8.54	0	5091.55	0	0.36	0.72	0.48
cg08006182	-0.41	0.66	2.67	-5.65	4.83	0	125.38	0	-0.15	0.88	0.19
cg20140201	-4.79	0.01	2.24	-9.18	-0.4	0	0.67	0	-2.14	0.03	4.94
cg04804772	-3.39	0.03	2.79	-8.85	2.07	0	7.91	0	-1.22	0.22	2.16
cg07010146	10.84	50822.58	6.5	-1.9	23.57	0.15	1.73E+10	0	1.67	0.1	3.39
cg11957248	-5.13	0.01	9.17	-23.11	12.85	0	3.80E+05	0	-0.56	0.58	0.8
cg12184624	5.03	152.66	5.65	-6.05	16.1	0	9.84E+06	0	0.89	0.37	1.42
cg27658026	1.76	5.8	8.22	-14.36	17.87	0	5.78E+07	0	0.21	0.83	0.27
cg09060908	-2.2	0.11	1.38	-4.91	0.5	0.01	1.66	0	-1.59	0.11	3.17
cg19249749	-3.54	0.03	1.96	-7.38	0.29	0	1.34	0	-1.81	0.07	3.83
cg27235315	2.01	7.47	3	-3.86	7.88	0.02	2649.55	0	0.67	0.5	0.99
cg06601993	-1.19	0.3	1.71	-4.55	2.16	0.01	8.69	0	-0.7	0.49	1.04
cg05209483	6.74	845.3	2.76	1.33	12.15	3.79	1.89E+05	0	2.44	0.01	6.1
cg14754494	-1.1	0.33	1.76	-4.54	2.35	0.01	10.51	0	-0.62	0.53	0.91
cg03531754	0.18	1.19	2.59	-4.9	5.25	0.01	191.14	0	0.07	0.95	0.08
cg08823209	-8.84	0	5.4	-19.42	1.73	0	5.65	0	-1.64	0.1	3.3
cg21965980	0.29	1.34	2.95	-5.5	6.08	0	437.68	0	0.1	0.92	0.12

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
cg12183875	-8.73	0	2.81	-14.23	-3.23	0	0.04	0	-3.11	<0.005	9.05
cg25152942	11.09	65451.92	2.99	5.23	16.95	186.72	2.29E+07	0	3.71	<0.005	12.23
cg27051686	2.45	11.6	3.34	-4.1	9	0.02	8104.49	0	0.73	0.46	1.11
cg12531542	-5.49	0	2.41	-10.21	-0.77	0	0.46	0	-2.28	0.02	5.47
cg15955277	0.76	2.13	2.28	-3.72	5.23	0.02	187.18	0	0.33	0.74	0.43
cg20938170	-23.88	0	6.66	-36.93	-10.83	0	0	0	-3.59	<0.005	11.55
cg25255988	9.9	19963.31	4.76	0.57	19.24	1.76	2.27E+08	0	2.08	0.04	4.73
cg06196689	-0.45	0.64	2.3	-4.96	4.06	0.01	57.78	0	-0.2	0.84	0.24
cg07497941	-0.78	0.46	2.64	-5.96	4.39	0	80.55	0	-0.3	0.77	0.38
cg20340346	0.69	2	1.38	-2.01	3.4	0.13	29.94	0	0.5	0.62	0.7
cg20911165	-1.13	0.32	1.99	-5.03	2.78	0.01	16.1	0	-0.56	0.57	0.81
cg24537237	1.41	4.09	1.68	-1.89	4.71	0.15	110.9	0	0.84	0.4	1.31
cg08836861	6.76	861.23	4.46	-1.98	15.49	0.14	5.36E+06	0	1.52	0.13	2.95
cg02733351	5.57	261.69	3.79	-1.86	12.99	0.16	4.38E+05	0	1.47	0.14	2.82
cg05626664	2.52	12.4	2.89	-3.14	8.17	0.04	3544.32	0	0.87	0.38	1.38
cg07359183	-8.52	0	5.03	-18.39	1.34	0	3.83	0	-1.69	0.09	3.47
cg04968426	0.46	1.58	2.52	-4.47	5.39	0.01	220.08	0	0.18	0.85	0.23
cg23382741	-4.26	0.01	2.89	-9.92	1.4	0	4.06	0	-1.47	0.14	2.83
cg26474043	-3.73	0.02	4.16	-11.88	4.42	0	83.25	0	-0.9	0.37	1.44
cg07402003	-6.22	0	3.94	-13.93	1.5	0	4.48	0	-1.58	0.11	3.13
cg03607648	2.64	14.01	2.17	-1.61	6.89	0.2	981.22	0	1.22	0.22	2.16

Table 24: Top 8 probes obtained using Nomogram

cg02643667	<i>TFF1</i>
cg02965295	<i>ANKRD44</i>
cg03607648	<i>UGT1A1</i>
cg04804772	<i>CALHM3</i>
cg06196689	<i>MUC13</i>
cg06601993	<i>CD36</i>
cg07210669	<i>S100P</i>
cg14323984	<i>S100P</i>

Table 25: 450K based methylation data validation

GSE74071		
Probe ID	adj.P.Val	P.Value
cg06601993	1.41E-01	2.36E-02
cg02643667	5.57E-01	3.08E-01
cg02965295	0.0784	0.00866
cg03607648	7.88E-01	6.37E-01
cg04804772	8.45E-01	7.31E-01
cg06196689	2.97E-01	9.11E-02
cg07210669	9.98E-03	3.81E-04
cg14323984	6.94E-01	4.90E-01
GSE 49149		
Probe ID	adj.P.Val	P.Value
cg06601993	5.25E-12	3.38E-13
cg02643667	9.70E-19	2.72E-20

GSE74071		
cg02965295	4.16E-15	1.80E-16
cg03607648	1.97E-08	2.15E-09
cg04804772	2.36E-21	4.97E-23
cg06196689	2.58E-06	4.03E-07
cg07210669	9.68E-26	1.32E-27
cg14323984	1.68E-33	1.01E-35

Table 26: RNA seq based data validation

Gene Name	P value	Tumor Mean Value	Non-Tumor Mean Value	Difference	SE of difference	t ratio	df	FC, Log2		
<i>S100P</i>	0.00231	263.557	2.37323	261.183	77.3391	3.37712	26	6.8		
<i>MUC13</i>	0.00211	57.1438	3.63091	53.5129	15.6747	3.41398	26	4.0		
Transcriptome analysis of the procession from chronic pancreatitis to pancreatic cancer and metastatic pancreatic cancer										
Name of dataset: DEGs in chronic pancreatitis					<i>TFF1</i>					
Name of dataset: DEGs in metastatic pancreatic cancer					<i>TFF1</i>	<i>S100P</i>				
Name of dataset: DEGs in pancreatic cancer					<i>TFF1</i>	<i>S100P</i>				

Table 27: MicroArray based expression data validation

GSE183795							
ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
8093950	3.22E-20	1.88E-22	10.7873	40.3282	1.38956	<i>S100P</i>	S100 calcium binding protein P
8090180	8.82E-14	3.29E-15	8.41517	23.9521	1.58256	<i>MUC13</i>	mucin 13, cell surface associated
8133876	1.18E-08	1.51E-09	-6.282	11.2276	-1.0641	<i>CD36</i>	CD36 molecule
8070579	5.80E-15	1.58E-16	8.86811	26.9301	1.82984	<i>TFF1</i>	trefoil factor 1
7936111	5.94E-02	4.11E-02	-2.0538	-4.8903	-0.0845	<i>CALHM3</i>	calcium homeostasis modulator 3
GSE28735							
ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
8093950	6.89E-14	1.43E-17	10.6013	29.4074	1.22608	<i>S100P</i>	S100 calcium binding protein P
8070579	4.87E-07	9.91E-09	6.31103	9.72389	1.94329	<i>TFF1</i>	trefoil factor 1
8057990	9.91E-01	9.87E-01	0.01687	-6.6364	0.00248	<i>ANKRD44</i>	ankyrin repeat domain 44
7936111	7.21E-01	6.27E-01	-0.4871	-6.5188	-0.0291	<i>CALHM3</i>	calcium homeostasis modulator 3
8133876	3.50E-06	1.18E-07	-5.7495	7.33643	-1.2905	<i>CD36</i>	CD36 molecule
8049349	1.12E-08	8.66E-11	7.34145	14.3025	1.50244	<i>UGT1A3///UGT1A1///UGT1A4///UGT1A9///UGT1A5///UGT1A6///UGT1A7///UGT1A8///UGT1A10</i>	UDP glucuronosyltransferase family 1 member A3///UDP glucuronosyltransferase family 1 member A1///UDP glucuronosyltransferase family 1 member A4///UDP glucuronosyltransferase family 1 member A9///UDP glucuronosyltransferase family 1 member A5///UDP glucuronosyltransferase family 1 member A6///UDP glucuronosyltransferase family 1 member A7///UDP glucuronosyltransferase family 1 member A8///UDP

GSE183795							
							glucuronosyltransferase family 1 member A10
GSE15471							
ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
232655_at	3.56E-04	1.60E-04	-3.9654	0.34046	-0.2465	<i>UGT1A1</i>	UDP glucuronosyltransferase family 1 member A1
228766_at	2.87E-02	1.98E-02	-2.3794	-4.092	-0.796	<i>CD36</i>	CD36 molecule
204351_at	9.23E-11	8.74E-13	8.5169	18.7866	3.66658	<i>S100P</i>	S100 calcium binding protein P
1554711_at	1.22E-03	6.23E-04	-3.5644	-0.9374	-0.2895	<i>CALHM3</i>	calcium homeostasis modulator 3
233361_at	3.81E-02	2.69E-02	-2.2552	-4.3614	-0.1119	<i>ANKRD44</i>	ankyrin repeat domain 44
205009_at	1.06E-04	4.14E-05	4.34313	1.62525	2.3991	<i>TFF1</i>	trefoil factor 1

RESEARCH

Open Access



DNA methylome in pancreatic cancer identified novel promoter hyper-methylation in *NPY* and *FAIM2* genes associated with poor prognosis in Indian patient cohort

Ankita Chatterjee^{2†}, Akash Bararia^{1†}, Debopriyo Ganguly², Pronoy Kanti Mondal¹, Paromita Roy³, Sudeep Banerjee³, Shibajyoti Ghosh⁴, Sumit Gulati⁵, Supriyo Ghatak⁵, Bitan Kumar Chattopadhyay⁶, Priyadarshi Basu², Aniruddha Chatterjee⁷ and Nilabja Sikdar^{1*}

Abstract

Background: Pancreatic ductal adenocarcinoma (PDAC) is one of the leading cancers worldwide and has a poor survival, with a 5-year survival rate of only 8.5%. In this study we investigated altered DNA methylation associated with PDAC severity and prognosis.

Methods: Methylome data, generated using 450 K bead array, was compared between paired PDAC and normal samples in the TCGA cohort (n = 9) and our Indian cohort (n = 7). The total Indian Cohort (n = 75) was split into cohort 1 (n = 7), cohort 2 (n = 22), cohort 3 (n = 26) and cohort 4 (n = 20). Validation of differential methylation (6 selected CpG loci) and associated gene expression for differentially methylated genes (10 selected gDMs) were carried out in separate validation cohorts, using MSP, RT-PCR and IHC correlations between methylation and gene expression were observed in TCGA, GTEx cohorts and in validation cohorts. Kaplan–Meier survival analysis was done to study differential prognosis, during 2–5 years of follow-up.

Results: We identified 156 DMPs, mapped to 91 genes (gDMs), in PDAC; 68 (43.5%) DMPs were found to be differentially methylated both in TCGA cohort and our cohort, with significant concordance at hypo- and hyper-methylated loci. Enrichments of “regulation of ion transport”, “Interferon alpha/beta signalling”, “morphogenesis and development” and “transcriptional dysregulation” pathways were observed among 91 gDMs. Hyper-methylation of *NPY* and *FAIM2* genes with down-regulated expression in PDAC, were significantly associated with poor prognosis in the Indian patient cohort.

Conclusions: Ethnic variations among populations may determine the altered epigenetic landscape in the PDAC patients of the Indian cohort. Our study identified novel differentially methylated genes (mainly *NPY* and *FAIM2*) and also validated the previously identified differentially methylated CpG sites associated with PDAC cancer patient's

[†]AnkC and AB both authors contributed equally in this manuscript work and preparation.

*Correspondence: snilabja@isical.ac.in; snilabja@gmail.com

¹ Biological Sciences Division, Human Genetics Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata, West Bengal 700108, India
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



Emerging role of non-invasive and liquid biopsy biomarkers in pancreatic cancer

Akash Bararia, Prosenjeet Chakraborty, Paromita Roy, Bitan Kumar Chattopadhyay, Amlan Das, Aniruddha Chatterjee, Nilabja Sikdar

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): A
Grade B (Very good): B
Grade C (Good): 0
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Bialek Ł, Poland;
Wang XB, China

Received: November 28, 2022

Peer-review started: November 28, 2022

First decision: January 11, 2023

Revised: February 2, 2023

Accepted: March 15, 2023

Article in press: March 15, 2023

Published online: April 21, 2023



Akash Bararia, Nilabja Sikdar, Human Genetics Unit, Indian Statistical Institute, Kolkata 700108, India

Prosenjeet Chakraborty, Department of Molecular Biosciences, SVYASA School of Yoga and Naturopathy, Bangalore 560105, India

Paromita Roy, Department of Pathology, Tata Medical Center, Kolkata 700160, India

Bitan Kumar Chattopadhyay, Department of General Surgery, IPGIMER & SSKM Hospital, Kolkata 700020, India

Amlan Das, Department of Biochemistry, Royal Global University, Assam 781035, India

Aniruddha Chatterjee, Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin 9061, New Zealand

Aniruddha Chatterjee, School of Health Sciences and Technology, University of Petroleum and Energy Studies, Dehradun 248007, India

Corresponding author: Nilabja Sikdar, PhD, Research Scientist, Doctor, Human Genetics Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata 700108, India. snilabja@isical.ac.in

Abstract

A global increase in the incidence of pancreatic cancer (PanCa) presents a major concern and health burden. The traditional tissue-based diagnostic techniques provided a major way forward for molecular diagnostics; however, they face limitations based on diagnosis-associated difficulties and concerns surrounding tissue availability in the clinical setting. Late disease development with asymptomatic behavior is a drawback in the case of existing diagnostic procedures. The capability of cell free markers in discriminating PanCa from autoimmune pancreatitis and chronic pancreatitis along with other precancerous lesions can be a boon to clinicians. Early-stage diagnosis of PanCa can be achieved only if these biomarkers specifically discriminate the non-carcinogenic disease stage from malignancy with respect to tumor stages. In this review, we comprehensively described the non-invasive disease detection approaches and why these approaches are gaining popularity for their early-stage diagnostic capability and associated clinical feasibility.



Deoxyribonucleic acid methylation driven aberrations in pancreatic cancer-related pathways

Akash Bararia, Amlan Das, Sangeeta Mitra, Sudeep Banerjee, Aniruddha Chatterjee, Nilabja Sikdar

Specialty type: Gastroenterology and hepatology

Provenance and peer review:

Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): A

Grade B (Very good): 0

Grade C (Good): C

Grade D (Fair): 0

Grade E (Poor): 0

P-Reviewer: Niu ZS, China; Jin S, China

Received: April 18, 2023

Peer-review started: April 18, 2023

First decision: May 12, 2023

Revised: May 29, 2023

Accepted: August 1, 2023

Article in press: August 1, 2023

Published online: September 15, 2023



Akash Bararia, Nilabja Sikdar, Human Genetics Unit, Indian Statistical Institute, Kolkata 700108, India

Amlan Das, Department of Biochemistry, Royal Global University, Assam 781035, India

Sangeeta Mitra, Department of Biochemistry and Biophysics, University of Kalyani, West Bengal 741235, India

Sudeep Banerjee, Department of Gastrointestinal Surgery, Tata Medical Center, Kolkata 700160, India

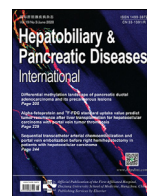
Aniruddha Chatterjee, Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin 9054, New Zealand

Aniruddha Chatterjee, School of Health Sciences and Technology, University of Petroleum and Energy Studies, Dehradun 248007, India

Corresponding author: Nilabja Sikdar, PhD, Research Scientist, Doctor, Human Genetics Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata 700108, India. snilabja@isical.ac.in

Abstract

Pancreatic cancer (PanCa) presents a catastrophic disease with poor overall survival at advanced stages, with immediate requirement of new and effective treatment options. Besides genetic mutations, epigenetic dysregulation of signaling pathway-associated enriched genes are considered as novel therapeutic target. Mechanisms beneath the deoxyribonucleic acid methylation and its utility in developing of epi-drugs in PanCa are under trails. Combinations of epigenetic medicines with conventional cytotoxic treatments or targeted therapy are promising options to improving the dismal response and survival rate of PanCa patients. Recent studies have identified potentially valid pathways that support the prediction that future PanCa clinical trials will include vigorous testing of epigenomic therapies. Epigenetics thus promises to generate a significant amount of new knowledge of biological and medical importance. Our review could identify various components of epigenetic mechanisms known to be involved in the initiation and development of pancreatic ductal adenocarcinoma and related precancerous lesions, and novel pharmacological strategies that target these components could potentially lead to breakthroughs. We aim to highlight the possibilities that exist and the potential therapeutic interventions.



Review Article

Differential methylation landscape of pancreatic ductal adenocarcinoma and its precancerous lesions

Akash Bararia^{a, #}, Subhankar Dey^{b, #}, Sumit Gulati^c, Supriyo Ghatak^c, Shibajyoti Ghosh^d, Sudeep Banerjee^e, Nilabja Sikdar^{a, *}^a Human Genetics Unit, Indian Statistical Institute, Kolkata, India^b Department of Zoology, New Alipore College, University of Calcutta, Kolkata, India^c Department of Gastroenterological Surgery, Calcutta Medical Research Institute, Kolkata, India^d Department of General Surgery, Medical College and Hospital, Kolkata, India^e Department of Gastrointestinal Surgery, Tata Medical Center, Rajarhat, Kolkata, India

ARTICLE INFO

Article history:

Received 3 June 2019

Accepted 18 March 2020

Available online 8 April 2020

Keywords:

Pancreatic cancer

Methylation

CpG islands

Epigenetic

Biomarkers

Precancerous lesions

ABSTRACT

Background: Pancreatic cancer is one of the most lethal diseases with an incidence almost equal to the mortality. In addition to having genetic causes, cancer can also be considered an epigenetic disease. DNA methylation is the premier epigenetic modification and patterns of aberrant DNA methylation are recognized to be a common hallmark of human tumor. In the multistage carcinogenesis of pancreas starting from precancerous lesions to pancreatic ductal adenocarcinoma (PDAC), the epigenetic changes play a significant role.

Data sources: Relevant studies for this review were derived via an extensive literature search in PubMed via using various keywords such as pancreatic ductal adenocarcinoma, precancerous lesions, methylation profile, epigenetic biomarkers that are relevant directly or closely associated with the concerned area of our interest. The literature search was intensively done considering a time frame of 20 years (1998–2018).

Result: In this review we have highlighted the hypermethylation and hypomethylation of the precancerous PDAC lesions (pancreatic intra-epithelial neoplasia, intraductal papillary mucinous neoplasm, mucinous cystic neoplasm and chronic pancreatitis) and PDAC along with the potential biomarkers. We have also achieved the early epigenetic driver that leads to progression from precancerous lesions to PDAC. A bunch of epigenetic driver genes leads to progression of precancerous lesions to PDAC (*ppENK*, *APC*, *p14/5/16/17*, *hMLH1* and *MGMT*) are also documented. We summarized the importance of these observations in therapeutics and diagnosis of PDAC hence identifying the potential use of epigenetic biomarkers in epigenetic targeted therapy. Epigenetic inactivation occurs by hypermethylation of CpG islands in the promoter regions of tumor suppressor genes. We listed all hyper- and hypomethylation of CpG islands of several genes in PDAC including its precancerous lesions.

Conclusions: The concept of the review would help to understand their biological effects, and to determine whether they may be successfully combined with other epigenetic drugs. However, we need to continue our research to develop more specific DNA-demethylating agents, which are the targets for hypermethylated CpG methylation sites.

© 2020 First Affiliated Hospital, Zhejiang University School of Medicine in China. Published by Elsevier B.V. All rights reserved.

Introduction

Pancreatic cancer (PanCa) is one of the most fatal malignant neoplasm around the world. Nowadays, PanCa is leading the 4th for cause of cancer mortality in USA and is estimated to be the 2nd leading cause for cancer related death by 2030 [1] (<https://seer.cancer.gov/>). Majority of PanCa is exocrine in origin that de-

velops from ductal cell and is known as pancreatic ductal adenocarcinomas (PDAC) [2]. PDAC is the most common form of PanCa that fills up more than 80% of cases. The aggressiveness of this malignant tumor is due to its asymptomatic feature and is usually detected only when it has already spread to surrounding tissues [3]. It is presumed that PDAC originated from the epithelium of ductal system, because of their ductal phenotype. It has been observed that hyperplastic and metaplastic changes of duct epithelium are commonly associated with ductal adenocarcinoma [3]. The clinical outcome of PDAC is high short-term mortality than any other neoplasm. As per the updated epidemiological scenario, PDAC remains

* Corresponding author.

E-mail address: snilabja@isical.ac.in (N. Sikdar).

Contributed equally.



Long Noncoding RNA Acting as Therapeutic Target for Oxidative Stress-Induced Pancreatic Cancer 78

Nilabja Sikdar, Akash Bararia, Ashmita Dutta, and Sudeep Banerjee

Contents

Introduction	1527
Treatment Strategy of Pancreatic Cancer	1529
Overview of Long Noncoding RNA (lncRNA) in Cancer	1530
Biological and Functional Roles of lncRNAs in Pancreatic Cancer	1532
Reactive Oxygen Species and Long Noncoding RNA in Cancer	1542
Long Noncoding RNA in Cancer Therapy Mediated via Oxidative Stress	1544
Long Noncoding RNA as Diagnostic and Prognostic Biomarkers for Pancreatic Cancer	1546
Long Noncoding RNAs in the Directive of Oxidative Stress, Reactive Oxygen Species and Pancreatic Cancer Progression	1549
Mechanistic Role of Hypoxia-Induced Long Noncoding RNA in Pancreatic Cancer	1551
Conclusion and Future Perspective	1553
References	1554

Abstract

Pancreatic cancer (PanCa) is a systemic disease with both locally advanced and malignant neoplastic phenotypes and has a poor prognosis. Among all sorts of PanCa, exocrine cancers account for approximately 95%, with Pancreatic Ductal Adenocarcinoma (PDAC) ranking fourth in terms of cancer mortality in the United States. Due to its increasing incidence, it is appraised to become the second leading cause of cancer by 2030. PDAC has a 5-year survival rate of 8%. PanCa, rightly known to be the “King of Cancer,” has mortality/incidence ratio of 98% due to lack of therapeutic advancement. PDAC and its associated hypoxic microenvironment play a major role in the formation of hypoxia-inducible factor-1 (HIF-1), which is highly expressed in 88% of PanCa tissues.

N. Sikdar (✉) · A. Bararia · A. Dutta
Human Genetics Unit, Indian Statistical Institute, Kolkata, West Bengal, India
e-mail: snilabja@isical.ac.in

S. Banerjee
Department of Gastrointestinal Surgery, Tata Medical Center, Kolkata, West Bengal, India