IN SILICO ANALYSIS FOR GENOMIC UNDERSTANDING TO ASCERTAIN IMPACT OF DIABETES RISK IN PROSTATE CANCER

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY
in
Molecular Biology and Genetic Engineering

By

Bhumandeep Kour

Registration Number: 11916445

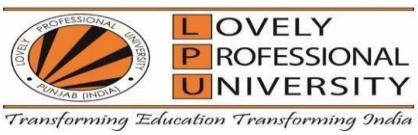
Supervised By

Dr. Anjuvan Singh (15950)
Department of Biotechnology
(Professor)
School of Bioengineering and Biosciences,
Lovely Professional University,
Punjab, India

Co-Supervised by

Dr. Prashanth N. Suravajhala
(Principal Scientist)
Amrita School of Biotechnology,
Amrita University, Kerala, India

Dr. Sugunakar Vuree
(Professor of Practice)
Vignan's Foundation for Science, Technology & Research,
Guntur, India



LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2024

DECLARATION

I, hereby declared that the presented work in the thesis entitled "*In Silico* Analysis for Genomic Understanding to Ascertain Impact of Diabetes Risk in Prostate Cancer" in fulfilment of degree of **Doctor of Philosophy (Ph. D.)** is the outcome of research work carried out by me under the supervision of Dr. Anjuvan Singh, working as Professor, in the Department of Biotechnology, of Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigators. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)

Name of the scholar: Bhumandeep Kour

Registration No.: 11916445

Department/school: Department of Molecular Biology and Genetic Engineering, School of Bioengineering and Biosciences,

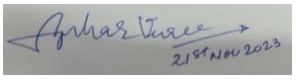
Lovely Professional University, Punjab, India

CERTIFICATE

This is to certify that the work reported in the Ph. D. thesis entitled "*In Silico* Analysis for Genomic Understanding to Ascertain Impact of Diabetes Risk in Prostate Cancer" submitted in fulfillment of the requirement for the award of degree of **Doctor of Philosophy (Ph.D.)** in the Molecular Biology and Genetic Engineering, is a research work carried out by Bhumandeep Kour, Registration No.11916445, is bonafide record of his/her original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

(Signature of Supervisor)
Name of supervisor: Dr. Anjuvan Singh
Designation: Professor
Department/school: Department of Biotechnology, School of Bioengineering and Biosciences,
University: Lovely Professional University, Punjab,India

(Signature of Co- Supervisor)
Dr. Prashanth N. Suravajhala
Designation: Principal Scientist
Department/school: Department of Biotechnology
University: Amrita School of Biotechnology, Amrita University, Kerala, India



(Signature of Co- Supervisor)
Dr. Sugunakar Vuree
Designation: Professor of Practice
Vignan's Foundation for Science, Technology & Research, Guntur, India

ABSTRACT

Prostate cancer (PCa) has become a major health concern being the third most prevalent cancer in India. On the other hand, type-2 diabetes mellitus (T2DM) and metabolic disorders contribute to a major chunk of the diseases associated with cancers. The link between Diabetes and Prostate Cancer (PCa) reveals shared risk factors, and notably, hyperglycemia increases prostate-specific antigen (PSA) level which is commonly used as a PCa biomarker. However, PSA's association with variables often turns out to be an erroneous positive result or even end up in needless biopsies of elderly people. Besides this, many antidiabetic drugs (like metformin) are found showing anticancer effects and increasing the survival rate of the PCa patients with diabetes. One of the prime objectives of the study was to explore the crosstalk of tumor metabolism that intersects between PCa and diabetes, with a focus on identifying variants and biomarkers beyond PSA for early detection.

With Genome Wide Association Studies (GWAS) and Next Generation Studies (NGS) proving to be boon for precision medicine, this thesis employed two different studies, first a meta-analysis study of already published datasets and then a preliminary study of 5 clinical cases and their genetic profiling with the help of whole exome sequencing (WES) from our consortium, viz. CAPCI. We conducted a comprehensive literature review focusing on meta-analysis to gather clinical parameters related to comorbidities such as diabetes, obesity, as well as PCa, recognizing their potential influence on complex processes of carcinogenesis, particularly in relation to PCa. The data was transformed into a semi-binary format for machine learning (ML) analysis. Publicly available datasets, our published RNA-seq datasets, and whole-exome sequencing data were cross-checked to identify common role players in PCa, diabetes, and obesity. Protein-Protein interaction (PPI) studies narrowed down to common genetic factors, revealing *BLM, TMPO* and *FOXP1*. Intriguingly, *BLM* gene was the only common gene to all three conditions, interacting with both *FOXP1* and *TMPO.*

The cases we have considered were prostate adenocarcinoma cancer from north-western region of India. We identified a total of 123,480 variants from our genomes of cases studied by WES and further delved into discerning pathogenic variants and selected mutations were validated using Sanger Sequencing. The studies through databanks, pathway enrichment analysis and literature search showed that most of the genes are enriched in the pathways of glucose metabolism, insulin resistance and other comorbidities. The prostate gland primarily produces prostate fluid under the regulation of the androgen receptor (AR), which mediates hormone signaling. AR stimulation is known to be involved in PCa. Additionally, AR

modulates insulin sensitivity, insulin secretion, obesity, and oxidative stress all of which impact the diabetes progression which marks AR pathway as a common target and needs to be explored with reference to both conditions. Since, lncRNAs are increasingly recognized as important players in cancer pathogenesis, including PCa, serving as signal transduction mediators. Hence, to bridge the gap between regulatory mechanisms of PCa and AR we aimed to identify lncRNAs and attempted to ask how AR signaling interacts with lncRNA networks in PCa biology and diabetes to uncover connections, shedding light on potential biomarkers and therapeutic targets.

Taken together, this study has allowed us to bring insights into ascertaining different phenotypes in the form of diabetes, PCa, and obesity. Also, the result of this study confirms that there is a plausible association between PCa and comorbidities like diabetes. The study moreover highlights the need to consider the comorbidities at the time of both diagnosis and treatment. Further investigation is warranted to understand the mechanism and early prediction of PCa.

Dedicated to my beloved father

Late S. Joginder Singh

I dedicate my thesis to the memory of my father, Late S. Joginder Singh, whose endless love, guidance, support and encouragement helped me start the Ph.D. journey and have been my guiding light throughout this academic journey. His wisdom, strength, and belief in my abilities have shaped the person I am today. Though he is no longer with us in the physical realm, his spirit lives on in every page of this work. This testament is to the values you instilled in me and the dreams you inspired. Your legacy continues to inspire me to reach new heights. I hope this achievement will fulfil the dream he envisioned for me. I miss you dearly father and wish you were here to share this achievement.

With eternal love,

Bhumandeep Kour

ACKNOWLEDGEMENT

Completing the Ph.D. thesis is a hard job. It was like a 'roller coaster' ride, but enormously rewarding. Now, I feel privileged to express my gratitude to those who contributed in one way or the other to accomplish my thesis. It is because of them that my Ph.D. experience has been one that I will cherish forever.

First and foremost, I would like to thank ***God*** *almighty for everything planned for me. I am forever grateful to God for everything in life. Whatever I do will never repay what he has provided me with.*

This work would not have been possible to complete without the encouragement, guidance, and constant support of my Supervisor ***Dr. Anjuvan Singh,*** *Professor, School of Bioengineering and Biosciences, Lovely Professional University (LPU), Punjab, India****.*** *With great respect, I acknowledge his contribution to bringing my Ph.D. to the final stage. I am really grateful to him for his suggestions during discussions, which helped me to grow more and more. I am speechless how you have believed in me and supported me whenever any problem arises and enlightened me with the right path. This work owes a lot to him. As a student, having a supervisor always motivating me to be happy and deal with situations calmly is helping me grow as a person also, I offer you a great respect. Within a short time I have learnt a lot from you.*

I express my deepest gratitude to my Co-supervisor, ***Dr. Prashanth N. Suravajhala****, Principal scientist, School of Biotechnology, Amrita University, Kerala. I am indebted for his support, kind words, and timely discussions whenever required. Recalling the day I met Sir first, I wished to work with him and working with him was like a dream come true. I can never thank you enough for your inputs and guidance during my Ph.D. as well as your encouragement, all the help and positive criticism has fuelled my productivity. You have always put your scholars at first priority; listened calmly to their problems; being always there for them and never let us wait for you. You always valued time for both yours and every other person, that's what I learned from you the most for life "value time". I will cherish working with you with potential future workings and meetings. I am inspired to see how you think differently and find new creative ways to approach scientific questions. Also, I thank you for being such a great and patient mentor on handling my difference of opinions and whenever I gave you disappointments with pleasant smiles and uplifting personalities. Thank you for everything!*

*I would also thank my second co-supervisor **Dr. Sugunakar Vuree,** professor of Practice, Vignan's Foundation for Science, technology and Research, Guntur, India, for showing direction when I started from scratch. From the beginning steps like the first steps of a small child, you were there to guide me. We have always looked upon you as a great strength of us not just as a supervisor but as an elder brother. Thank you for giving a channel to my Ph.D. and I reached here with your guidance. I will always appreciate the time we spent together discussing science to make every project more impactful and your words "Bhuman let your hard work speak" has given me immense motivation during my Ph.D. journey. Thank you for bearing all these years with so much patience.*

*I wish to convey my sincere thanks to **Dr. Neeta Raj Sharma**, Associate Dean, School of Bioengineering and Biosciences, for her kind support. I would also like to thank **Mr. Ashok Mittal**, Chancellor LPU, **Mrs. Rashmi Mittal** Pro-Chancellor LPU, **Dr. Ramesh Kanwar**, Vice-Chancellor LPU, and **Dr. Monica Gulati** Registrar, LPU for their vision ('Think Big') motivation and support.*

*I also wish to express my gratitude towards clinical collaborators, **Dr. Devendra Sharma** for providing samples. I would like to show my deepest gratitude and respect to **Dr. Jayaraman Valadi**, Distinguished Professor, FLAMES University, Pune, for making time and giving contributions in my Ph.D. work. I am extremely grateful for having **Dr. Nirmal Lohiya**, Professor University of Rajasthan for his valuable comments, knowledge, showing right directions and leading the CAPCI as an esteemed member. I am also thankful to **Dr. Renuka Suravajhala,** Assistant Professor**,** Amrita University, Kerala, for her valuable help for teaching me some in silico part of my Ph.D. work. Also, I would like to thank **Dr. Barkha Khilwani** for always being there for advising me as a friend and working together as colleagues in CAPCI projects.*

*Special thanks to my lab mates and co scholars **Dr**. **Nidhi Shukla, Shalini Rajagopal, Swati Arora, Sumit Sheoran, Satbir Kour and Dr. Santosh Kumari** from discussing science to solving problems, giving suggestions and also for their unconditional support and encouragement. I feel honoured to have Ph.D. friends **Dr**. **Navneet Kour, Dr. Ankit Awasthi, Dr. Leander Corrie, Dr. Dharavat Nagaraju, Dipatnu Banik, Ujwal Dahal, Sonawane Bhushan Laxman, Sourbh Suren Garg, Anjum Ayub, Bhupinder Singh, Dr. Agnibha Das Majumdar,** and others who have always been there every day to offer help, providing motivation, and care. Special thanks to my family members for their constant financial support, encouragement, care and hearing me out all this time. I also express my*

*acknowledgement to my lab Attendants **Mr. Fedrick Gill** for their cooperation and guidance. I also thank HOL **Dr. Khan, Mr. Kuldeep Singh & Mr. Sunny Gupta** for timely providing the lab facility.*

*Finally, I express my special and greatest acknowledgement to my mother **Mrs. Harbans Kour** who has dreamed this and built passion in me for science. Inexplicable thanks to my father **Late S. Joginder Singh** for providing me with warmth and nurturing me with all the best facilities. I always wanted to follow his path and be my ideal. I also express my sincere gratitude to my elder brother **S. Gurvinder Pal Singh** for giving me constant support, motivation to do better and providing me all the strength to stay strong, be patient and humble in the Ph.D. journey.*

Bhumandeep Kour

Table of Content:

S. No.	Title	Page No.
Chapter 1	**INTRODUCTION**	1-8
Chapter 2	**REVIEW OF LITERATURE**	9
2.1.	Prostate gland, anatomy, and region of cancer development	9-10
2.2.	Prostate cancer development	11-12
2.3.	Epidemiology	12-14
2.4	Causes	15
2. 4.1.	Age and ethnicity	15
2.4.2.	Genetic Disparities	15-16
2.4.3	Infection and Inflammations	16
2.4.4.	Androgens	16
2.4.5.	Diet, Fitness and Obesity	17
2.4.6.	Life style	18
2.5	Symptoms	18
2.6	Diagnosis	19
2.7	Treatment	20-21
2.8	Biomarkers of PCa	21-22
2.9	Comorbidities of PCa	22
2.9.1.	Diabetes Mellitus	23-25
2.9.2.	Obesity as a Phenotype	25
2.9.3.	Hypertension	26
2.9.4.	Urolithiasis	26-28
2.10.	Diabetes and Cancer	28-32
2.11.	Diabetes and Prostate cancer	32-35
2.12.	LncRNAs	35-36
2.13.	LncRNA and Prostate Cancer	36-39
2.14.	LncRNA and Diabetes	39-41
2.15.	Research Gap	41
2.16	Aim and Objective, Hypothesis	42-45
Chapter 3	**MATERIALS AND METHOD**	46
3.1.	**To screen variants in PCa patients affected with diabetes, thereby establishing a causal relationship between them**	46

	using integrated systems genomic approaches	
3.1.1.	Integrated machine learning approaches on publicly available dataset for identification of plausible candidate	46
3.1.1.1.	NCBI dbSNP	46
3.1.1.2.	ClinVar	46-47
3.1.1.3.	Dataset	47
3.1.1.4.	Clinical parameters of PCa	48
3.1.1.5.	Clinical parameters of diabetes associated with cancer/PCa	51
3.1.1.6.	Clinical parameters of obesity-associated with cancer/PCa	52
3.1.1.7.	Machine learning Algorithms	53-56
3.1.1.8.	Correlational Study	56
3.1.1.9.	Differential Analysis: GEPIA 2 (Gene Expression Profiling Interactive Analysis)	57
3.1.1.10.	CBioportal	58
3.1.1.11.	Survival Analysis	58
3.1.2.	**To establish a causal relationship between PCa (PCa) and type-2-diabetes through WES**	59
3.1.2.1.	Clinical Cases	59-60
3.1.2.2.	Next Generation Sequencing (NGS)- WES	60-65
3.1.2.3	Downstream analysis of VCF Files	65
3.1.2.3.1.	SNPnexus	66-67
3.1.2.3.2.	Batch Entrez	67
3.1.3.	Sanger Sequencing for Validation of WES results	67-71
3.1.4.	Study of Mutational Hotspots of PTEN	71
3.1.5.	Categorizing the variants or candidate genes for PCa in Urolithiasis as Comorbidity	72-73
3.2.	**To identify and explore the role of lncRNAs as potential targets using lncRNA-protein interactions**	73
3.2.1.	PCa Causal Proteins, LncRNAs and Androgen Receptors (AR)	73-75
3.2.2.	Molecular Docking studies:	76
3.2.3.	Work flow of molecular docking	76-79
3.2.4.	Correlation of Type-2-diabetes associated lncRNAs with PCa	79

	related lncRNAs identified through RNA-seq study	
3.3.	**To perform protein-protein interaction (PPI) network analyses and identify pathways associated with T2D and PCa *in silico* method**	80
3.3.1.	Interactome network	80
3.3.2.	Phenolyzer	78
3.3.3.	Gene/Pathway enrichment:	80-81
3.3.4.	Reactome	82
3.3.5.	KEGG pathway	82
3.3.6.	Panther Database	83
3.3.7.	Cytoscape	83
Chapter 4	**RESULTS**	84
4.1.	**To screen variants in PCa and diabetes using integrated systems genomic approaches**	84
4.1.1.	Result of parallel done Meta-analysis Study	84
4.1.1.1	Comparison of ClinVar unveiled *BLM* as a key player in Diabetes, PCa and Obesity	84
4.1.1.2.	Overlapping variants in PCa ClinVar and GWAS central	85
4.1.1.3.	More variant alliances between PRACTICAL Consortium GWAS Central and Clinvar of PCa	85
4.1.1.4.	Comparison of RNA-Seq data revealed *PP1MB* and *SFTPC* as common	87
4.1.1.5.	Regression algorithms on PCa, diabetes and obesity data sets suggest potential improvements for predictive models	87
4.1.1.6.	Mutational analysis *FOXP1*: a highly mutated in Prostate adenocarcinoma	90
4.1.1.7.	Gene expression patterns highlighted *PPP2CA* and *FOXP1* as highly expressed genes	91
4.1.1.8.	*BLM* showed significant impact on disease free survival in survival analysis	94
4.1.2.	**Causal relationship between PCa (PCa) and type-2-diabetes (T2D) through WES**	95

4.1.2.1.	Quality Check DNA QC report cDNA QC report	95-96
4.1.2.2.	Molarity of electronic ladder and sample	97-104
4.1.2.3.	Sample QC report	104
4.1.2.4.	Results of down-analysis on sequenced data	104
4.1.2.5.	MYO15A and BRCA2 mutations appeared common with the previous pilot study	105-121
4.1.2.6.	Correlation of sequencing results using meta-analysis	121
4.1.2.6.1.	Comparison with ClinVar data of PCa	121
4.1.2.6.2.	Comparison of CAPCI (our cohort) with PRACTICAL consortium yielded significant variants	122
4.1.2.6.3.	Comparison with GWAS Central PCa	124
4.1.2.6.4.	A suspicious mutation of SLC16A13 discovered on comparison with GWAS central Diabetes Datasets	125
4.1.2.6.5.	Comparison with ClinVar Diabetes	128
4.1.2.6.7.	Mutation in our Prostate cancer Cohort	129-136
4.1.2.3.	Results of Sanger validation	136-139
4.2.	**To identify and explore the role of lncRNAs as potential targets using lncRNA-protein interactions**	139
4.2.1.	Molecular Docking and Visualization results	139
4.2.1.1.	TLXINB and SCARNA10 LncRNA showed highest PCa related genes interactions	140
4.2.1.2.	ANKRD20A9P, SCARNA10 and NPBWR1 lncRNAs showed more AR interactions affinity	142
4.2.2.	Common lncRNAs identified through correlation of PCa related lncRNAs from RNA-seq study and publically available datasets of T2D lncRNA	144
4.3	**To perform protein-protein interaction (PPI) network analyses and identify pathways associated with T2D and PCa *in silico* method**	146
4.3.1.	Interactome network using GeneMANIA for Metadata analysis	146
4.3.1.1.	*BLM*, *TMPO* and *FOXP1* as common interacting partners	146
4.3.2.	Interactome networks using GeneMania for our cohort gene set	147

4.3.3.	Pathway Enrichment of Gene sets from our PCa Cohort	148
4.3.3.1.	Kegg Mapper and Kegg Pathway	148
4.3.3.2.	Reactome identified top six significant pathways with focus on DNA repair genes *Tp53, PALB2, BRCA1, BRCA2* and *ATM*	153
4.3.3.3.	Panther DB Analysis of pathways indicating molecular and biological processes	157-160
4.3.3.4	Cytoscape-CytoHubba	160-165
4.3.4.	Results of urolithiasis	165-167
Chapter 5	**DISCUSSIONS**	168-180
Chapter 6	**CONCLUSIONS AND FUTURE PERSPECTIVES**	181-183
	BIBLIOGRAPHY	184-217
	APPENDICES-Supplementary images	218-232
	LIST OF PUBLICATIONS	233
	LIST OF CONFERENCES AND WORKSHOPS	224

List of Tables:

S. No.	Table Title	Page No
3.1	List of different regression algorithms used for regression analysis	53
3.2	List of different regression algorithms applied in Weka software applied on datasets of PCa, diabetes mellitus, and obesity and their Root mean square error (RMSE) value	56
3.3	Summary of sequenced data including total number of reads both forward and backward and the size of total sequenced data	63
3.4	Quantification by Qubit® 4.0 Fluorometer	69
3.5	Master mix for Amplification	69
3.6	PCR cycling conditions	69
3.7	Primer details	70
3.8	List of PCa Proteins, Androgen Receptor and PDB Ids used in the molecular docking study	74
4.1	Functions and pathways involved in the common genes among ClinVar PCa and ClinVar diabetes from Genecards	88
4.2	Details of common genes retrieved using genecards	89
4.3	Table representing the DNA quality at both ratio A260/280 (1.9) and A260/230 (2.0-2.2) by using Nanodrop 1000	95
4.4	Table showing details of quality check of cDNA	96
4.5	Sample table for Electronic ladder	98
4.6	Peak Table for ladder	98
4.7	Sample table for sample 1	99
4.8	Peak table for sample 1	99
4.9	Sample table for sample 2	100
4.10	Peak Table for sample 2	100
4.11	Sample Table for Sample 3	101
4.12	Peak Table for sample 3	101
4.13	Sample table for Sample 4	102
4.14	Peak table for Sample 4	102
4.15	Sample table for sample 5	103
4.16	Peak table for sample 5	103

4.17	Significant common variants and genes with previous cases	105
4.18	Other 37 significant pathogenic variants from our cases	108
4.19	Tabular summary of the general associations of identified mutated gene with PCa and diabetes	113
4.20	ClinVar and common of 5 cases	121
4.21	Variants common with PRACTICAL Consortium	123
4.22	Variants common with GWAS central data of PCa	124
4.23	Variants common with GWAS central Diabetes	126
4.24	Variants common with ClinVar Diabetes	128
4.25	Analysis and identification of Genotype	136
4.26	The potential candidate 5-lncRNA's interactions with 3-PCa proteins (PDB id) and binding energies, RMSD and Interacting residues were summarized	141
4.27	The potential candidate 5-lncRNA's interactions with 7-AR proteins (PDB id) and binding energies, RMSD and Interacting residues were summarized	143
4.28	List of common PCa related LncRNAs with T2D related lncRNA database	144
4.29	List of pathways in which our input gene set playing role obtained by KEGG Mapper and pathway	148
4.30	List of prostate cancer related pathways from KEGG pathway database	153
4.31	List of pathways related to the gene set using the reactome database	154
4.32	List of overall pathways and genes playing role in	157
4.33	List of Biological processes associated with genes set	158
4.34	List of Molecular processes associated with genes set	159

List of Figures:

S. No.	Figure Title	Page No
2.1	Anatomy of Prostate gland. A. Posteriorly placed at the neck of bladder, B. Cross-sectional representation of the different lobes in prostate gland (anterior, posterior, lateral and middle, C. Histological representation of the prostate duct	10
2.2	Development of cancer in prostate gland from stage 1 to stage 4. In stage 1 a neoplastic growth starts in one lobe of gland, stage 2 it reaches to second later lobe of the gland, spreading to seminal vesicle defines the third stage and in stage 4 when it reaches the lymph nodes and urethra becomes narrower it starts spreading to other organs and called marks the beginning of metastatic stage	11
2.3	Estimated number of new cases, deaths and prevalence in 2020, males of above 45 years	13-14
2.4	Predicted incidence and mortality from 2020 to 2040, males, age 45-85+, Asia	14
2.5	Predicted incidence and mortality from 2020 to 2040, males, age 45-85+, India	14
2.6:	Risk factors of Prostate cancer	18
2.7	Symptoms of Prostate Cancer	19
2.8	Showing the interplay of hyperglycemia, insulin resistance, hyperinsulinemia, AR signaling, and anticancer Metformin mode of action	44
3.1	Flow chart representing identification, filtering, mapping (to SNP database, ClinVar, and SNPnexus), selecting, and finally, semi-binary data conversion of (1, 0.5, 0, -0.5, -1) with respective significance.	48
3.2	Different clinical parameters of **A**. PCa (Clinical Significance, PSA Level, Gleason Grading, Minor Allele Frequency, DRE Value), **B**. Different clinical parameters of Diabetes (Clinical Significance, Minor Allele Frequency, Glycated Hemoglobin: HbA1c, WBC Count Test, Fasting Blood Glucose, Body Mass Index); **C**. Different clinical parameters of Obesity (Clinical Significance, Minor Allele Frequency, Body	49-50

	Mass Index, LDL/HDL) for PCa, and their different ranges based on risk, and their conversion into semi-binary data (-1, -0.5, 0, 0.5, 1)	
3.3	Representation of Matrices we prepared by semi binary conversion of data	55
3.4	Flow chart for the correlational study of ClinVar variants (PCa, diabetes, and obesity) and published GWAS data from GWAS Central, PRACTICAL Consortium, Exome data, and RNA Seq data of both prostate cancer and diabetes mellitus for identifying common variants	57
3.5	Interface of the GEPIA 2 tool used for expression and survival analysis	58
3.6	Diagrammatic illustration of Whole Exome Sequencing approach	61
3.7	Workflow of library quality check using Agilent 4150 Tapestation	62
3.8	Work flow of WES benched marked pipeline. After obtaining the raw reads the sequenced data downstream analysis performed in 3 phases; Pre-processing, Variant discovery and Variant Prioritization	64
3.9	Set of commands used for WES pipelines	64
3.10	Workflow of vcf files analysis	65
3.11	Three steps to run SNPnexus: First information, providing the query format (genomic coordinates/chromosome regions/dbSNP/vcf file) and selection of all the annotation categories and finally submitting the query.	66
3.12	Image showing the interface of Batch Entrez on which we selected database on SNP to provide the data type and uploaded .text file for analysis	67
3.13	Steps involved in Sanger sequencing using capillary gel electrophoresis	68
3.14	Overview of cancer variant prioritization based on coding variants, conservational screening, and functional prediction.	72
3.15	A flowchart for the steps of gnomAD to identify the variant types and their rarity of Urolithiasis related genes. We mapped Re -seq	73

	proteins (19) of urolithiasis to the SNP database which resulted in the identification of 50896 different SNP's, which were subjected to ClinVar/dbVar. Later from the given categorized list (pathogenic, likely pathogenic, conflicting predictions, uncertain significant, benign and likely benign) we filtered out those genes which are pathogenic (61) and are having uncertain significance (100). These SNPs can be verified on the basis of type of variation (missense or nonsense) or can also be analysed for their associations with other diseases	
3.16	Showing interface of HDOCK server and the uploaded FASTA format sequence of the both Receptor (PCa protein) and ligand (LncRNA)	77
3.17	Systematic work flow of the 11 lncRNAs and the 36 targeted PCa and AR Proteins	78
3.18	Work flow for Gene Enrichment Analysis using Kegg Mapper and Kegg Pathway database, Reactome, Panther-Db and Cytoscape software	81
4.1	Venn plot for identifying commonalities between Clinvar datasets of prostate cancer, diabetes, and obesity.	85
4.2	**A, B**: Venn Plot to identify common variants among ClinVar prostate cancer, diabetes, GWAS central, PRACTICAL consortium for prostate cancer, and prostate cancer exome data; **C, D**: Venn Plot to identify common variants among ClinVar prostate cancer, Diabetes, Obesity, GWAS central, PRACTICAL consortium for prostate cancer and prostate cancer exome data	86
4.3	**A:** Venn plot showing no commonality between exome data and obesity ClinVar data; **B:** Venn plot for GWAS central data of both type-2-diabetes and prostate cancer, PRACTICAL Consortium, and ClinVar Prostate cancer; **C:** Venn plot showing some commonalities between RNAseq results of three prostate cancer samples and ClinVar data of Diabetes	86
4.4	Linear plots for RMSE standard deviations of PCa, DM and Obesity	87
4.5	Overall percent and type of genetic mutations in *FOXP1*,	91

	SFTPC, PPP2CA, PPARG, PPM1B, TMEM67, TMPO, and *BLM* related to prostate adenocarcinoma.	
4.6	**a:** Box plots for expression of common genes compared to TCGA tumor-normal datasets of Prostate adenocarcinomas (PRAD) **b:** Box plots for expression of common genes in compared to TCGA tumor-normal datasets of Prostate adenocarcinomas (PRAD) and GTEx data (multiple genes based on Z scores) **A**. Only tumor tissue expression; **B**. Match TCGA normal data; **C**. Match TCGA normal and GTEx data **c:** Expression in comparison to TCGA tumor-normal dataset of Prostate Adenocarcinomas (PRAD) dataset and GTEx data for TMPO gene	92, 93
4.7	Comparison of overall survival (OS) and disease-free survival (DFS) among *FOXP1, TMPO, TMEM67, TMPO, BLM, PPP2CA*, PPM1B, and PPARG *using* GEPIA 2 tool with PRAD dataset for both TCGA normalized data and GTEx data.	94
4.8	Image of agarose gel electrophoresis of all tumor samples and Ladder	96
4.9	The TapeStation Electropherogram report of electronic ladder. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values as per the different size of fragments seen ranging from 50-1500 bps.	97
4.10	The TapeStation Electropherogram report of sample (gDNA) H20/6591. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values of lower mark, a major peak in the region of 275-305 bps	99
4.11	The TapeStation Electropherogram report of sample (gDNA) H20/8329. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values of lower mark, a major peak in the region of 360-380 bps	100
4.12	The TapeStation Electropherogram report of sample (gDNA)	101

	H20/8030. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values of lower mark, a major peak in the region of 275-530 bps	
4.13	The TapeStation Electropherogram report of sample (gDNA) H20/7111. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values of lower mark, a major peak in the region of 305-540 bps	102
4.14	The TapeStation Electropherogram report of sample (gDNA) H20/7255. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values of lower mark, a major peak in the region of 210-320bps	103
4.15	Venn plot showing the common number of variants among the cases	105
4.16	Venn plots of comparison of each sample variants with the ClinVar data of PCa, diabetes and obesity	122
4.17	V Venn plots for cohort and PRACTICAL consortium	123
4.18	Commonalities with GWAS data sets of PCa	125
4.19	ɔmmonalities with GWAS data sets of diabetes	126
4.20	All the results images **(a-as)** of cBioportal mutational analysis done mapping the identified genes to the PCa datasets. Upright lollipops indicating the position of the different protein changes caused by mutations related to PCa in that particular region. *PTEN, BRCA2, ATM, MYO15A, CACNA1C, ABCC9, TSC2, SZT2* and *PLCE1* are high in protein changes.	129-136
4.21	**a.** Chromatograph of Sanger validated gene **GJB2** **b.** Chromatograph of Sanger validated gene **MPO** **c.** Chromatograph of Sanger validated gene **ITGB4** **d.** Chromatograph of Sanger validated gene **MYRF** **e.** Chromatograph of Sanger validated gene **ATM** **f.** Chromatograph of Sanger validated gene **MYRF** **g.** Chromatograph of Sanger validated gene **HNF1A** **h.** Chromatograph of Sanger validated gene **GJB2** Chromatograph of Sanger validated gene **MYRF**	137-139
4.22	Represent the 3D visualization using Pymol the complexes of	140

	potential 5-lncRNAs and interacting residues with 3-PCa proteins **(a-e).**	
4.23	**i.** Represent the 3D visualization using Pymol the complexes of potential 5-lncRNAs and interacting residues with 7 targeted AR proteins **(a-f).**	142
4.24	Statistical overview of 11 lncRNAs and 3 lncRNAs those are common of which SCARNA10 as potential biomarker using Venny 2.1.	144
4.25	**A**. Interactome Network of commonly associated genes (WRN, DNAJC6, RET, NF1AND TMEM67) of prostate cancer and diabetes from ClinVar data, TMPO is shown as a common interacting partner; **B**. Interactome network of TMEM67, SFTPC, and PPMIB showing mainly genetic interactions (green) with one common FOXP1; **C**. Interactome network of BLM gene found as the only one common gene between prostate cancer, diabetes, and obesity that is interacting with both the earlier identified FOXP1(physical interactions shown in red color) and TMPO (co-expression shown in purple color) which can be seen in **D**.	146
4.26	PPI map of selected Sanger validated gene set. Different color of edges indicates different kind sof interactions among the proteins *viz*. Co-expression (purple), pathway (sky blue), genetic interactions (green), co-localization (blue), physical interactions (red) and predicted with orange color.	147
4.27	Prostate cancer pathway retrieved from KEGG pathway	153
4.28	Pie chart distribution of the genes and the related pathways by different colours. Major portion represented that the genes are not well explored for the pathway associations (purple portion).	157
4.29	Pie chart distribution of the genes and the biological processes represented by different colours	158
4.30	Pie chart distribution of the genes and the molecular processes represented by different colours.	159
4.31	Clustering coefficient networks showing top 10 genes for *FOXP1, PPM1B, PPARG, TMEM67, TMPO, SFTPC, PPP2CA*	160

	and *BLM* genes identified in Meta-Analysis	
4.32	Clustering coefficient networks showing top 10 genes for all 42 mutated genes and *FOXP1*, *TMPO* and *BLM* identified from the meta-analysis study. *DLD, PDHX, NDUFB4, NDUFB7, BLM, UQCRC1, NDUFV1, PARP1, CHEK1* and *PTEN* are the top 10 genes that are functioning together.	161
4.33	Pictorial representation to illustrate the difference between the normal metabolism and hyperglycaemic (diabetes) situation in *PTEN* pathway.	164
4.34	Figure representing the signaling pathway involving few mutated genes from our study playing a role in cancer development and diabetes.	165
4.35	A schematic representation of an interaction map of all the 22 urolithiasis and other 20 associated genes using GeneMania, forming a network of Urolithiasis causal genes with other genes.	166
4.36	Disease gene based network analysis using Phenolyzer. All the edges indicate the propensity of interactions.	166
Supplementary Figure 1	Supplementary Figures of FastQC files of Clinical samples	218-232

Abbreviations:

WHO	World Health Organization
PCa	Prostate Cancer
GLOBOCAN	Global Cancer Observatory
PBCR	Population based cancer Registry
PSA	Prostate Specific Antigen
DRE	Digital Rectal Exam
DM	Diabetes Mellitus
T2DM	Type-2-diabetes Mellitus
AR	Androgen Receptor
ATP	Adenosine Tyro phosphine
TF	Transcription Factor
IGF	Insulin Growth Factor
PI3K	Phosphoinositide 3-kinase
mTOR	Mammalian Target of Rapamycin
lncRNA	Long Non-coding RNA
NGS	Next Generation Sequencing
WES	Whole Exome Sequencing
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
TCGA	Cancer Genome Atlas Program
CAPCI	Cancer Prostate Consortium of India
RNA-Seq	RNA Sequencing
ADT	Androgen Deprivation Therapy
CRPC	Castration Resistant Cancer
PRAD	Prostate Adenocarcinoma
BMI	Body Mass Index
BPH	Benign Hyperplasia
PCA3	Prostate Cancer Antigen 3
GDM	Gestational Diabetes Mellitus
T1DM	Type 1 Diabetes Mellitus
GWAS	Genome Wide Association Studies
UGT1A1	Hepatic Bilirubin UDP-glucuronosyltransferase 1A1

AMPK	Adenosine Monophosphate Kinase
ATM	Ataxia Melanogaster
MF	Metformin
DHT	Dihydrotestosterone
GSIS	Glucose Stimulated Insulin Secretion
miRNA	Micro Ribonucleic Acid
ncRNA	Non-coding Ribose Nucleic Acid
CirRNA	Circular Ribose Nucleic Acid
RBP	RNA-binding protein
PTEN	Phosphatase and Tensin Homolog
TNM	Tumor Node Metastasis
GEO	Gene Expression Omnibus
QC	Quality Check
SAM	Sequence Alignment Map
BAM	Binary Alignment Map
VCF	Variant Calling Format
SNP	Single Nucleotide Polymorphism
CNV	Copy Number Variant
NCBI	National Centre of Biotechnology Information
ClinVar	Clinically Verified
MYO15A	Myosin Factor IV A
BRCA2	BReast CAncer Gene 2
BRCA1	BReast CAncer Gene 1
APOA4	Apolipoprotein A4
NTRK1	Neurotrophic Receptor Tyrosine Kinase 1
HNF1A	HNF1 Homeobox A
ERCC5	Excision Repair 5 Endonuclease
SEMA4A	Semaphorin 4A
TSC2	TSC Complex Subunit 2
ABCA4	ATP Binding Cassette Subfamily A Member 4
SDHB	Succinate Dehydrogenase Complex Iron Sulfur Subunit B
NR2E3	Nuclear Receptor Subfamily 2 Group E Member 3
GJB2	Gap Junction Protein Beta 2

SMPD1	Sphingomyelin Phosphodiesterase
MPZ	Myelin Protein Zero
PKP2	Plakophilin 2
CACNA2D4	Calcium Voltage-Gated Channel Auxiliary Subunit Alpha2delta 4
TOR1AIP1	Torsin 1A Interacting Protein
FGD4	FYVE, RhoGEF And PH Domain Containing 4
RNPC3	RNA Binding (RNP1, RRM) Containing 3
POLR3B	RNA Polymerase III Subunit B
TMX2	Thioredoxin Related Transmembrane Protein 2
ABCC9	ATP Binding Cassette Subfamily C Member 9
PDHX	Pyruvate dehydrogenase Complex Component X
UBR1	Ubiquitin Protein Ligase E3 Component N-Recognin 1
CLN5	CLN5 Intracellular Trafficking Protein
PGM1	Phosphoglucomutase 1
COQ8A	Coenzyme Q8A
AGL	Amylo-Alpha-1, 6-Glucosidase, 4-Alpha-Glucanotransferase
CACNA1C	Calcium Voltage-Gated Channel Subunit-Alpha 1C
PDE2A	Phosphodiesterase
DOCK7	Dedicator Of Cytokinesis 7
SZT2	SZT2 Subunit Of KICSTOR Complex
NDUFV1	NADH: Ubiquinone Oxidoreductase Core Subunit VI
THSD1	Thrombospondin Type 1 Domain Contacting 1
PLCE1	Phospholipase C Epsilon 1
ITGB4	Integrin Subunit Beta 4
CHEK1	Checkpoint Kinase1
TUBGCP4	Tubulin Gamma Complex Component 4
COL2A1	Collagen Type II Alpha 1 Chain
PARP1	Poly(ADP-Ribose) Polymerase 1
CTPS1	CTP Synthase 1
SLC16A13	Solute Carrier Family 16 Member 13
MYRF	Myelin Regulatory Factor
MPO	Myeloperoxidase
BLM	BLM RecQ Like Helicase

TMPO	Thymopoietin
FOXP1	Forkhead Box P1
TMEM67	Transmembrane Protein 67
LNCaP	Lymph Node Carcinoma of the Prostate
PPI	Protein-Protein Interactions
ML	Machine Learning

CHAPTER 1: INTRODUCTION

A greater part of deaths worldwide are due to non-communicable diseases, and cancer is likely to score the top most position. It is also considered the only significant barricade for the rising life span of all countries in the present world (Bochen Cao et al., 2018) (Ramesh & Kosalram, 2023) (Witts et al., 2024). In 2015, WHO claimed cancer as the first or second main reason for mortality earlier than the age of 70 years among 91 over 172 countries; it also lays third or fourth position in 22 other countries (Bray et al., 2018) (R. D. Smith & Mallath, 2019). Among all, PCa (PCa) is specific and unique to male reproductive health, a neoplasia of the outer peripheral region of the prostate gland and associated with unique reprogramming of glucose metabolism (Bader & McGuire, 2020). Among all male malignancies, PCa, an adenocarcinoma, is one of the most prevalent in men worldwide, affecting the prostate gland, testis and androgen receptor pathways (Tyagi et al., 2022). As PCa arises at the periphery of the gland, it has a peripheral epithelium with exocrine glands made up of secretory luminal cells with PCa that is ultimately tissue origin-specific. PCa is entirely tissue-origin-specific, and therefore, it is categorized into prostate epithelial carcinoma and prostate adenocarcinomas (Bader & McGuire, 2020).

Global Cancer Observatory (GLOBOCAN) data for 2020 predicts that by 2040, a total of 63100 more cases will be added and 256,000 deaths will happen across Asia. Recent Population Based Registries (PBCR) statistics 2020 for India and that across the globe by the International Agency for Research on Cancer reported data for incidence rates for five Asian nations (China, India, Japan, Republic of Korea, and Thailand), three North American nations (the USA, Canada, and Brazil), two Oceanian nations (Australia and New Zealand), and four European nations (the UK, France, Germany, and Italy) for the years 2008 to 2012. In all the mentioned locations, the incidence of men below 40 was extremely low and rose sharply beyond that age. Prevalence peaked at ages 75–84, 70–79, and 70–79 years, respectively, across Asia (apart from China and India), America, Oceania, and Europe. In general, Asia's incidence was typically lower than that of other western continents, even as several Asian countries have shown an expanding rate of incidence and mortality (R. Sharma et al., 2024). A piece of precise information regarding the weight of Asian PCa is shortening, which is ethnically, economically, and geographically heterogeneous (Ha Chung et al., 2019) The low occurrence rate in Asian men compared to western men might be because of a shortage of prostate-specific antigen systemic screening approaches (Ha Chung et al., 2019) (Kimura & Egawa, 2018).

Japan and South Korea have higher incidence rates of PCa compared to other Asian nations, with Japan having a prevalence of 520 per 100,000 and South Korea 370 per 100,000 (Hori & Palmer, 2021) (Ha Chung et al., 2019). PCa is the most commonly diagnosed malignancy among males in western nations, with increasing incidence rates (Mittal et al., 2012) (Wang et al., 2022) (Jiang et al., 2024). The etiology of PCa is not fully understood but is associated with age, family history (42% genetic background association), first-degree relatives being more at risk of developing it, and dietary factors such as red meat, fat, diary, and egg intake in Asian men (Ha Chung et al., 2019) (Khan et al., 2019). However, due to its complex and polygenic character, its pathophysiology is not yet fully understood, and PCa has become such a typical malady (Khan et al., 2019) (Mittal et al., 2012). Genetic variations contribute to differences in PCa incidence between western and Asian populations (Ha Chung et al., 2019). Acquired and congenital risk factors also affect PCa development. Latent PCa means men who show no undeniable signs during their life expectancy, and the cancer-causing tumor is unexpectedly discovered while examining the dead body (autopsy). On the other hand, accidental PCa is, by chance, identified in prostatic tissue extracted as a non-malignant ailment (Kimura & Egawa, 2018). In the pathogenesis of PCa, genetic factors are crucial. There are few Indian reports on hereditary PCa susceptibility (Mittal et al., 2012) (Vietri et al., 2021) (Berenguer et al., 2023). Only 25-30% of the patients are accurately staged clinically, based on a physical examination (Ravi et al., 2021). Comparatively fewer PCa instances are reported in India than in other Western nations with common symptoms like frequent occurrences can contribute to raising PCa risk (Patel & Klein, 2009). Smoking increases the risk by 1.6 times (Kenfield et al., 2011) (Mittal et al., 2012). Genetic variations contribute to differences in PCa incidence between Western and Asian populations (Ha Chung et al., 2019). Acquired and congenital risk factors also affect PCa development. Latent PCa means men who show no undeniable signs during their life expectancy, and the cancer-causing tumor is unexpectedly discovered while examining the dead body (autopsy). On the other hand, accidental PCa is, by chance, identified in prostatic tissue extracted as a non-malignant ailment(Kimura & Egawa, 2018). In the pathogenesis of PCa, genetic factors are crucial. There are few Indian reports on hereditary PCa susceptibility (Kimura & Egawa, 2018) (Mittal et al., 2012) (Vietri et al., 2021) (Berenguer et al., 2023). Only 25-30% of the patients are accurately staged clinically, based on a physical examination (Ravi et al., 2021). Comparatively fewer PCa instances are reported in India than in other Western nations with common symptoms like frequent urination difficulties, a worn-out urinary flow, bloody pee or semen, and severe lower back, hip, or thigh discomfort are some prevalent PCa symptoms. Similar symptoms, though, might also be caused by benign prostatic hyperplasia or prostatitis,

and as a result, it is extremely challenging to identify PCa from typical symptoms (Sarkar et al., 2022). Therefore, an improved awareness of epidemiology in Asia, including its modifiable and non-modifiable conditions of risk, might make it possible to make better healthcare decisions and develop policies (Ha Chung et al., 2019).

PCa screening in past years has significantly been geared toward Prostate Serum Antigen (PSA) testing, Digital Rectal Examination (DRE) or prostatectomy or at autopsy to identify latent cancers in asymptomatic individuals (Bray et al., 2018). Prostate Specific Antigen (PSA) is a 33kD proenzyme that belongs to the Kallikrein group of proteases and is also known as human Kallikrein 3. It is the most often used biomarker for identifying PCa. Its concentration in seminal fluid ranges from 0.5-2 g/L in men's urine and saliva (Khan et al., 2019) (Sarkar et al., 2022). The two common techniques for diagnosing PCa are Gleason scoring of tissue images from microarrays and the microscopic analysis of stained specimens by pathologists. However, scoring PCa tissue images requires much effort, is prone to arbitrary decisions from different observers, and displays low reliability. Pathologists use needle biopsy in the biography during the conventional PCa diagnosis, relying on Gleason grading system, which assesses glandular structures. While glandular structures eventually disappear in high-grade tumours, epithelial cells do not show in low-grade tumours. The tissue for PCa is separated into five growth patterns, numbered from 1 to 5, each corresponding to a distinct cell tissue shape and a better prognosis. And yet, the distinction between normal tissue and high risk tissue is little and challenging (Tyagi et al., 2022). In geriatric medical care (medical care for aged people), the prostate-specific antigen (PSA) titer is a standard clinical procedure frequently used to detect early occurrences of PCa. Numerous international studies have shown population-specific normative PSA readings with different thresholds for PCa. The universal use of normalized standard values causes many false-positive screening findings, which causes undue stress and frequently unneeded invasive follow-ups. Regular PSA testing of serum boosts PCa (PCa) detection over digital rectal examination (DRE), which enhances the DRE's ability to predict cancer, which results in the early identification of prostate malignancies. The most effective method for early detection of PCa is the combined use of DRE and serum PSA testing. This combination can change the likelihood that PCa is present. This combination has shown a higher detection rate when screening tests with serum PSA measurements than with DRE alone (Rahul Unnikrishnan et al., 2021). However, relying solely on PSA levels above 20ng/mL for high-risk classification may overlook other pathological indicators, especially in certain populations like Indian men. Therefore, clinical symptoms and pathological characteristics should also be considered for accurate risk assessment (Garg et al., 2022).

Comorbidities are frequently overlooked while making PCa treatment decisions, leaving patients in a quandary about whether to pursue aggressive therapy. The explanation for this disparity might be doctors and patients making incorrect assumptions about extended life periods with single comorbidities, believing that they will have less influence on males, and ignoring good mortality estimation when making treatment decisions (Chamie et al., 2012). According to a study, individuals who have regular contact with healthcare facilities for comorbidities receive early diagnosis due to routine testing. This shows that comorbidities may be associated with cancer aetiology to promote aggressiveness and could help in early prognosis (Fleming et al., 2006) (Tiruye et al., 2024). For example, diabetes, chronic pain, anxiety, depression, hypertension, obesity and urolithiasis are known to be associated with PCa. Obesity is known to play a role in the recurrence of PCa along with other comorbidities like diabetes, hormonal imbalance and heart problems (Wilson et al., 2022) (Tiruye et al., 2024). Hypertension is generally found as an after-effect of treatment therapy (Xiaolei Zhu & Wu, 2019). Another condition observed known as urolithiasis, which could be considered a comorbidity, is mainly located at the metastatic stage and seen during diagnosis and treatment (J. M. Liu et al., 2013). Urolithiasis has been linked to PCa via calcium homeostasis, ROS, androgen receptors (X. Sun et al., 2021).

The comorbidity of diabetes and cancer is that both are multifactorial, severe, heterogeneous, and life-threatening diseases associated with acute, chronic, and fatal illnesses. Several epidemiological studies have suggested that the chances of developing many cancer types, *viz*., urinary tract, pancreatic, breast, liver, kidney, female reproductive tract and colorectal increase with diabetes (particularly DM) (Vigneri et al., 2009) (Pearson-Stuttard et al., 2021). It was shown that there is an increase in malignancies in Korean and Japanese people who had DM from 4.7% up to 21.9% after a survey conducted for 10 years (Suh & Kim, 2011). Despite numerous investigations, the association between DM and cancer remains unclear due to the complex nature of type 2 diabetes mellitus, which is characterized by hyperglycemia and hormonal imbalances. Various factors such as obesity, medication, metabolic control, and diet in diabetic patients may influence this association (Vigneri et al., 2009). Since diabetes is underdiagnosed in adults (3-5%), even if we believe that DM increases cancer risk, this association is still underrated (Vigneri et al., 2009) (R. K. Shahid et al., 2021).

In the prostate gland, the role of glucose metabolism is unique as compared to others. The prostate gland is more glycolytic with a tremendous collection of citrate (Beier et al., 2023). The primary role of a prostate gland is to produce prostate fluid. This function depends on the hormone signalling mediated through the hormone receptor transcription factor AR (androgen

receptor). Androgens bind the androgen receptor in the cytoplasm, which moves to the nucleus, where they activate genes for PCa. Since, PCa originates from the periphery region of the gland, there is high glycolysis and production of pyruvate, which is converted into lactate to maintain the redox nature for producing ATP molecules and citrate, which accumulate in prostate fluid. There seems to be no Krebs cycle or oxidative phosphorylation (Bader & McGuire, 2020). As per "the bioenergetics theory of prostate malignancy" proposed by Franklin and Costello, in prostate cells, the Kreb's cycle is inhibited, leading to inefficient energy production. To compensate, these cells increase glycolysis to survive and sustain citrate production. However, PCa cells undergo metabolic reprogramming, oxidizing citrate to produce ATP and restoring energy efficiency. Unlike many other cancers, PCa cells do not rely heavily on increased glucose consumption early in tumor development (Cutruzzolà et al., 2017). Another essential feature of the prostate gland is the highest deposition of zinc inside it, which is a critical regulator of various enzymes and TF's (transcription factors). This Zn inhibits the aconitase enzyme (ACO2), which truncates the TCA cycle. On the other hand, prostate adenocarcinomas have an intact TCA cycle, oxidative phosphorylation and lipogenesis, leading to cell proliferation in the prostate. They have low or limited glycolysis (Bader & McGuire, 2020). Zinc has antitumor effects (induce apoptosis, inhibit anti-apoptotic protein NF-κB, reduce the invasive and proangiogenic capabilities in metastasis) however its level falls in PCa. Instead, they switch to the Warburg effect only in metastatic stages. This delay in the metabolic shift limits the effectiveness of early diagnosis methods (Cutruzzolà et al., 2017).

Some *in vitro* studies have shown that increased exposure of PCa cells to glucose increases the AR expression, suggesting that the elevated glucose promotes tumor aggressiveness. In some diabetic cases the volume of prostate has increased with no clear indication of oncological staging (Antunes et al., 2018). Although the exact involvement of glucose metabolism remains unclear, evidence has suggested its link to PCa progression, as diabetes mellitus patients tend to exhibit more progressive PCa phenotypes. This suggests a relationship between both diseases particularly in later stages (Eidelman et al., 2017).

Increased insulin due to insulin promotion promotes the IGF-1 secretion that causes several impacts on cancer development (through powerful mitogenic effects) (Szablewski, 2024). *IGF-1* and insulin bind to their particular receptors on the cell's surface, activating the *PI3K/Akt/mTOR* and *Ras/Raf/MAPK* signalling pathways. This concludes that the PCa cell exhibits a shift towards oxidative phosphorylation and lipogenesis. Further, metformin, an antidiabetic drug, has shown anticancer effects by lowering the insulin in the serum and

blocking *PI3K/Akt/mTOR* while activating *AMPK*, which inhibits *mTOR*. While metformin is effective in improving survival rates in aggressive PCa, its impact on lowering cancer PCa incidence remains uncertain. It is known that enhanced metabolism in diabetic patient's leads to increased PCa cell proliferation and tumor formation (Baxi et al., 2024). Thus, studying the underlying mechanisms behind PCa metabolic functions wherein phosphorylases and kinases such as PI3K activate glucose uptake and control glycolytic enzymes directly would hold importance (Jang et al., 2013) (M. Shahid et al., 2019) and could suggest a relationship between diabetes and PCa. Even a study conducted on 6,403 Americans found that higher dietary sugar intake correlates with elevated serum PSA levels. Every 1gm intake increases PSA level by 0.003ng/m, with *P* less than 0.05. This link may be due to increased activation of inflammatory cytokines and serum uric acid, leading to chronic inflammation. However, both natural and added sugars have the same chemical structure but may show a wide range of physiological effects to regulate inflammation (Z. Liu et al., 2021). Further, intake of dietary sugars is found to be associated with metabolic syndromes like high blood pressure, LDL cholesterol, triglycerides, uric acid and inflammation (Z. Liu et al., 2021). Also, the glucose level during diagnosis and treatment of PCa is an independent predictor of PCa recurrence (50% higher risk) compared to patients with normal glucose levels. Therefore, this evidence suggests glucose is an influencer and controls the modifiable risk factors that lead to recurrences of PCa and its progression (Wright et al., 2013). However, PSA works as a protease to cleave the insulin-like growth factor binding protein-3 (IGFBP-3), which could modulate the IGFs (Insulin-like growth) and lead to the proliferation of both normal and abnormal prostatic cells (Cohen et al., 1992).

Furthermore, obesity which contributes to insulin resistance is also a critical factor in PCa development. Many meta-analyses indicate a slight but significant increase in prostate cancer risk associated with obesity, particularly for more aggressive forms. Insulin and IGF-1 levels, which tend to rise with obesity, may produce tumor growth and contribute to the transition of PCa cells to an androgen-resistant state. Men with obesity frequently have decreased testosterone levels, which can stimulate the production of adipokines including adiponectin and leptin. Obese people have greater leptin levels, which promotes cancer cells to proliferate and resist dying, particularly in aggressive stages of cancer. On the other hand the adiponectin, which is decreased in obese people, appears to aid in cancer prevention. Several biological processes may be involved in this protective effect. All in all, this indicates that adipokines are significant in risk of aggressive PCa. Nevertheless, further studies are required to clarify the ways in which these variables interact with insulin resistance to promote PCa.

Also, PCa risk is frequently assessed in obese male using BMI, which is a rough indicator of obesity. It is important to look at the interaction between muscle and fat, specifically its metabolic status, in order to comprehend how obesity contributes to insulin resistance in PCa (Di Sebastiano et al., 2018).

Since PCa is heterogeneous clinically and pathologically, most of the glands with cancer have more than one individual tumor focus with diverse pathological features and no clear shared somatic commonality (Lindberg et al., 2013) (Segura-Moreno et al., 2021) (X. Yu et al., 2023). NGS development, such as DNA and RNA sequencing, enables the discovery of the genomic, transcriptomic and epigenomic landscapes of individual malignant expansions. Various genomic abnormalities, frequent and rare auxiliary variations (deletions and insertions), fusion transcripts and copy-number variations could be examined concurrently. In combination with applications of the bioinformatic approach, NGS advancements are being increasingly utilized to study many genes affordably and have been employed in assessing cancer clinical samples, offering NGS-centred molecular diagnostic services (Nair et al, 2021)

However, there have been no previous attempts at genotype-phenotype associations using next-generation sequencing (NGS) investigations in India. NGS breakthroughs have hailed the finding of biomarkers from Chinese and Caucasian/European heritage, but nothing is known about the Indian phenotype/variant of PCa (Gupta et al., 2020). Till today, oncologists in India are in a dilemma over routine use of NGS because most of them are unsure of the report interpretations. Oncologists believe that we are evolving continuously as are our genetic profiles, too (Pathak et al., 2024).

Though few studies employing whole transcriptomic and exome analyses from our lab have recently been done to identify genetic variants related to PCa, there are still a large number of variants from samples to be screened. Our CAPCI (bioclues.org/capci), last accessed on April 12, 2024) group has shown that over 30 causal genes were known to be associated with PCa risk in the Indian sub-population and besides that novel lncRNAs (LINCO1440, ENSG00000234855, SOX2OT, ENST00000647843.1 and ENSG00000287903) and differentially expressed genes (DEGs) were identified using RNA sequencing (Gupta et al., 2020) (N. Shukla et al., 2023).

Through an intense literature review and considering the shortcomings in proper diagnosis and treatment, we aimed to identify the known potential biomarkers for early detection with the help of NGS. We hope that this study will help to identify new variants or signatures for prostate cancer in association with comorbidities like diabetes. Also, common pathways will

be explored for them with respect to PCa development and diabetes. Investigation of the underlying mechanisms of such a complex disease and its other disease associations will help to improve treatment. The study may underlie some evidence elucidating the association of phenotype traits with distinct variants of Indian origin.

Therefore, the Cancer Prostate Consortium of India (D. Sharma et al., 2023), a pilot study from the Systems Genomics Lab, reported mutations in a small north-western cohort of Indians. We believe this was the first research in India performed using WES to screen mutations and infer the genetic associations (polygenetic) between PCa and diabetes. It is likely that obesity can also contribute to an increased risk of PCa through the imbalance of circulating molecules, the promotion of cancer cell proliferation by adipokines, and the complex interplay of age, diabetes and obesity. However, the debate on such associations and the exact nature of relationships is going on (Mistry et al., 2007) (Freedland & Aronson, 2004).

AR stimulation is known to be involved in PCa, and AR also modulates insulin sensitivity, insulin secretion, obesity, and oxidative stress, all of which impact diabetes progression, which marks the AR pathway as a common target and needs to be explored with reference to both conditions. Since then, lncRNAs are increasingly recognized as important players in cancer pathogenesis, including PCa, serving as signal transduction mediators. Hence, to bridge the gap between the regulatory mechanisms of PCa and AR, we have also conducted a collaborative RNA-seq study on a small cohort from the north-western region of India. We used the identified lncRNAs in the current study, where we attempted to ask how AR signaling interacts with lncRNA networks in PCa biology and diabetes to uncover connections, shedding light on potential biomarkers and therapeutic targets.

CHAPTER 2: REVIEW OF LITERATURE

2.1. Prostate gland, anatomy, and region of cancer development:

The prostate gland is an upturned pyramid looking, exocrine in nature, surrounding the urethra and placed in the middle of the genitourinary lining and bladder neck. Ejaculatory ducts traverse obliquely into the gland and enter the prostatic urethra at verumontanum **(Figure 2.1.A)**. On an average, mature gland weighs 15 to 20g, roughly the size of a walnut, and measures 4 (transverse) × 3 (anteroposterior) × 3 (craniocaudal) cm. It performs the role of an additional sex gland (Coakley & Hricak, 2000). The prostate gland begins to form during the third month of pregnancy as invaginations of epithelium from the urogenital sinus of the posterior region, driven by the mesenchyme that lies under it. The enzyme 5α-reductase converts foetal testosterone into the hormone 5α-dihydrotestosterone, which is necessary for the prostate gland to develop normally. Deficiency of the 5α-reductase can therefore result in significant deformities related to external genitalia as well as a minor or barely detectable PCa (Hammerich et al., 2008) (Gogola et al., 2024).

Earlier prostate was internally described in terms of five lobes (anterior, middle, posterior and two laterals) **(Figure 2.1.B)**. However, according to modern descriptions, the prostate comprises acinar (glandular) and some non-glandular components like urethra and fibrous tissue with stromal at the anterior region and prostatic urethra. The glandular part of the prostate has both outer and inner elements characterized by their location, histology and ductal anatomy. Further, the inner part includes transition and periurethral zones, whereas the outer portion comprises peripheral and central zones (Coakley & Hricak, 2000) (Gogola et al., 2024).

According to McNeal's (1981) description of anatomical zones in the prostate gland (McNeal, 1981), the transition zone forms the 10-15% of glandular tissues of an adult prostate gland, divided into two equal regions on the lateral sides of the urethra. This region is usually resistant to cancer development but is susceptible to age-related benign hyperplasia (BPH) or nodal hyperplasia and rarely adenocarcinomas. A cone-shaped central zone forms the gland's base, with the cone's tip where ejaculatory ducts converge. The transition and central zones give rise to minor PCa foci but are at very low risk for cancer progression. The peripheral region mostly constitutes glandular tissues and contributes to 70% of the young prostate. Carcinomas, post-inflammatory atrophy and chronic prostatitis often originate from this region compared to others (McNeal, 1981).

Histologically, gland ducts are branched, lined with a secretory layer of columnar cells and an underneath layer of basal cells, where the lumen of ducts is generally filled with corpora amylacea **(Figure 2.1.C)**. The gland's surrounding fibrous tissue comprises the prostatic capsule **(Figure 2.1.B)**. On the superior region of the base of the gland, seminal vesicles (SV) are placed and form the ejaculatory duct complex. These vesicles are generally resilient to any ailment affecting the prostate gland. Still, its involvement (seminal vesicle involvement, SVI) because of PCa could be among the most crucial indicators of PCa (Hammerich et al., 2008).

PCa is typically multifocal and most prevalent in the posterior region, near the prostatic capsule, known as the periphery region, compared to the other areas. Since the PCa and Gleason grading in patients is often heterogeneous, both primary with high prevalence and secondary with lower prevalence grading is preferred. Hence, both patterns are added to form the Gleason scoring (Hammerich et al., 2008).

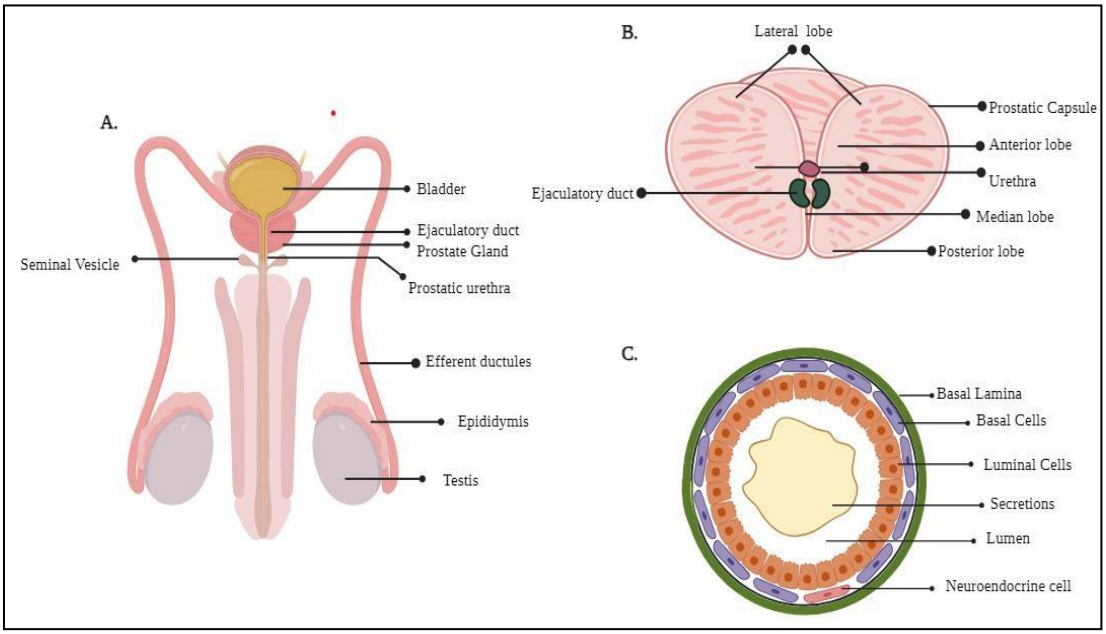


Figure 2.1: Anatomy of Prostate gland. A. Posteriorly placed at the neck of the bladder, B. Cross-sectional representation of the different lobes in the prostate gland (anterior, posterior, lateral and middle, C. Histological representation of the prostate duct (Figure created through BioRender https://www.biorender.com/)

2.2. PCa development:

The transformation from a normal prostate to a neoplastic or malignant prostate undergoes several steps. It starts with prostatic intraepithelial neoplasias (PIN) by proliferation in luminal cells, then becomes localized, reaches an advanced stage called prostate adenocarcinomas (PRAD), and finally, locally invasive cancer stage by degrading the basal cell layer and basal lamina. The sensitivity towards the hormone plays an essential role in PCa, as several androgen deprivation therapies (ADT) have been reported to block the androgen pathways; however, it can develop resistance to ADT, thus inviting (metastatic) castration-resistant PCa (CRPC). The first spot to start the metastasis is the lymph node, which is adjacent to the prostate gland, where the primary tumor would have already formed. Later, it spreads to distant organs like bones, lungs, and liver. Metastasis which forms bone as osteoblastic and osteolytic lesions is often found leading to frequent bone fractures and body pain. **Figure 2.2** represents the four stages of PCa development and the beginning of metastasis. Like every other cancer type, epithelial-mesenchymal transition (EMT) also plays a role in PCa metastasis (G. Wang et al., 2018). As per the "cell origin model", the columnar luminal cells, which produce secretory proteins and the basal cells, represent the main two epithelial cells, and a rare neuroendocrine cell plays a role in the origin of PCa. These three cells are found to arise from the basal progenitors (multipotent) during the organogenesis of the prostate gland (S. H. Lee & Shen, 2015).

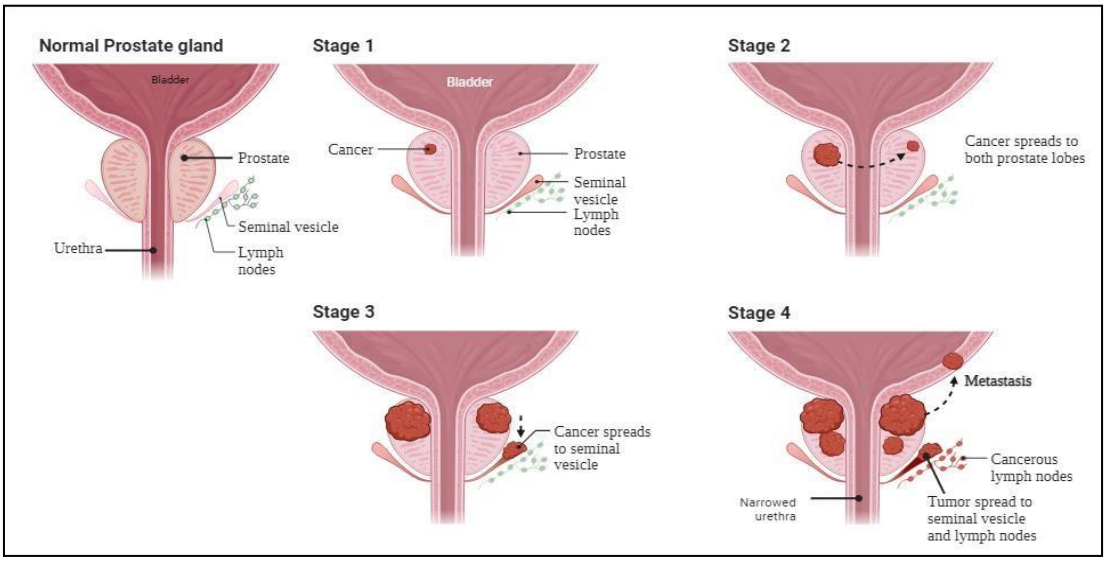


Figure 2.2: Cancer development in the prostate gland from stage 1 to stage 4. In stage 1, a neoplastic growth starts in one lobe of the gland, in stage 2, it reaches the sector lobe of the gland, spreading to the seminal vesicle, which defines the third stage; and in stage 4, when it

comes to the lymph nodes, and the urethra becomes narrower, it starts spreading to other organs and marks the beginning of the metastatic stage (Figure created with BioRender, https://www.biorender.com/, last accessed on April 1, 2024)

Several growth factors are secreted by the PCa cells, like adrenomedullin, endothelin 1, fibroblast growth factors, bone morphogenetic proteins and platelet-derived growth factors, which show paracrine signalling during metastasis in bone. Some proteases like urokinase-type plasminogen, prostate-specific antigen (PSA), insulin-like growth factor, transforming growth factor *β,* and matrix al metalloproteinase (G. Wang et al., 2018). However, the exact mechanism of PCa initiation is still unclear (S. H. Lee & Shen, 2015) (Testa et al., 2019).

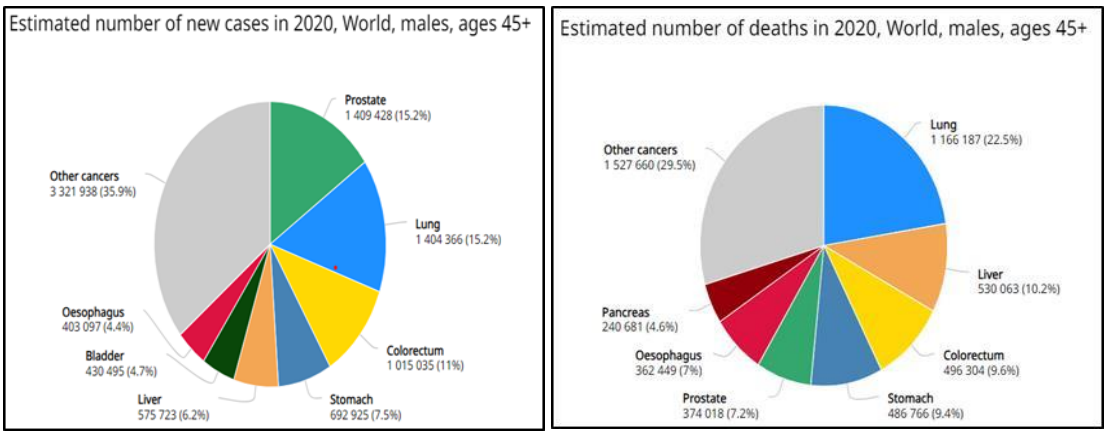
2.3. Epidemiology:

PCa was expected to be the second most common cancer and would be ranked as the fifth most significant reason for mortality in men caused by cancer by 2020 (Sung et al., 2021) (Oderda et al., 2023) (Hartley & Ahmad, 2023). It is projected to be diagnosed in one in seven American men and in one in every 25 men worldwide (Barsouk et al., 2020). It was estimated to have roughly 14 million fresh cases and global deaths of about 37,000 by 2020 **(Figure 2.3)**. Death rates aren't changing much, with 8.1 cases in transitioned countries and 5.9 cases in transitioning countries for every 100,000 men, respectively. However, the incidence of transitioned countries was three times higher than that of transitioning countries (Sung et al., 2021).

According to the latest Population Based Cancer Registry (PBCR) statistics for 2020 India, the total number of PCa cases is 41,532, with a crude rate of 5.7 among men (Mathur et al., 2020). In particular, the incidence rate is 37.5 cases per every 1 00,000 men in transitioning countries, compared to 11.3 cases per every 1 00,000 men in transitioning countries. In particular, PCa is most often diagnosed in more than half (112 countries) of the 185 countries globally. It was observed that the incidence of PCa varies regionally from 6.3 to 83.4 per 100,000 males, with maximum prevalence seen in North America, New Zealand/Australia, Western Europe, Northern Europe, South Africa and the Caribbean. However, the lowest prevalence was seen in North Africa and the Middle East. In the case of mortality, the regional pattern differed from the incidence as the highest deaths were seen in Polynesia/Micronesia, Sub-Saharan Africa and the Caribbean. There are about 48 countries, including South America (Chile, Venezuela and Ecuador), Central America, Africa, Sweden and the Caribbean, in which the major region of deaths in men is PCa (Sung et al., 2021).

Further, it has been reported that black men who are residents of both the United States (US) and the Caribbean regions and who are connected to West African ancestry have a higher incidence. This suggests that Western African genetic inheritance is a modulating factor in increasing the risk of PCa (Rebbeck et al., 2013).

In the late 1980s and early 1990s, the PCa incident rate spiked quickly in the US, Australia and Canada because of the widespread use of PSA screening broadly, but the rate fell sharply a few years later by 2000 (C. K. Zhou et al., 2016) (Sung et al., 2021) with countries in northern Europe, western Europe, central Asia, and south Asia also showing similar patterns. However, eastern countries like Bulgaria, Slovakia, Belarus and China continued to have a rise in PCa incidence rates (Culp et al., 2020) (Barsouk et al., 2020). Also, the sub-Saharan African region has reported an annual rise of 2% to 10% over time. With rising cases in Asia and Africa, their mortality trend also increased, which could be because of limited proper treatment and PSA screening. On the other hand, from 2009-2013, those countries with significant resources for health facilities that could manage to control the mortality rate, like high-income regions of the Americas, Europe and Oceania with highly advanced treatments, could conduct early screening at a high rate. Besides this, since 2010, the US has reported an increased incidence of last-stage PC, leading to the highest mortality rates from 2012 to 2017 (Sung et al., 2021). The most significant factor influencing the discrepancies in PCa incidence rates globally is probably different PCa detection procedures worldwide, for instance, the utilization of PSA tests and healthcare systems (C. K. Zhou et al., 2016). According to the Global Cancer Observatory (GLOBOCAN) data for 2020, it has been estimated that by 2040, 631000 more cases will be added in Asia and 256000 more deaths will be reported (**Figure 2.4)**. Also, by 2040, 64800 more cases will be added and 32200 deaths will occur in India (**Figure 2.5)**.



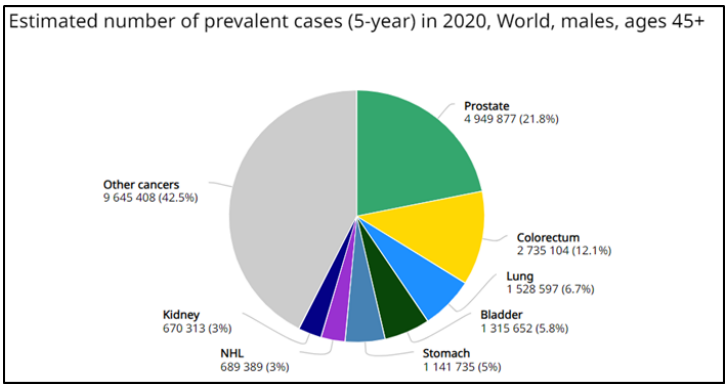


Figure 2.3: Estimated number of new cases (1409 428), deaths (734 018) and prevalence (4949 877) in 2020 among males over 45 years, GLOBOCAN 2020 https://gco.iarc.fr/

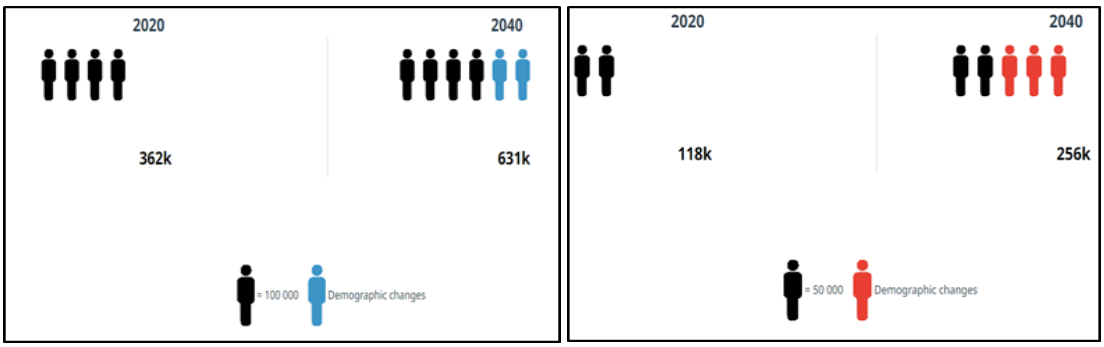


Figure 2.4: Predicted incidence (approx. 362000 to 631000 cases) and mortality (approximately 118000-256000 cases) from 2020 to 2040, males, age 45-85+, Asia, GLOBOCAN 2020 https://gco.iarc.fr/

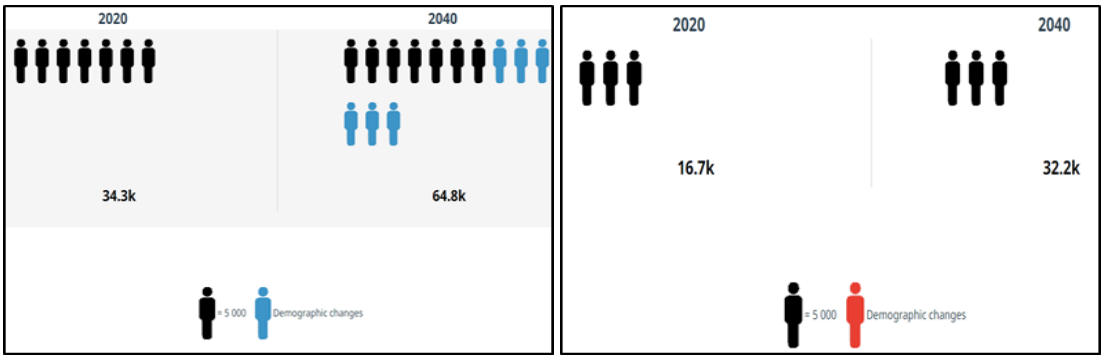


Figure 2.5: Predicted incidence (34300 to 64800 cases) and mortality (16700 to 32200) from 2020 to 2040, males, age 45-85+, India, GLOBOCAN 2020, https://gco.iarc.fr/

2.4. Causes:

Risk factors of heterogeneous cancer, PCa can be age, ethnicity, genetic mutation, lifestyle habits, infections and inflammations, changes in hormone levels and food habits (Oderda et al., 2023) **(Figure 2.6).**

2.4.1. Age and ethnicity

PCa is often found among older men, and incidence has shown a positive relationship with age (Barsouk et al., 2020) (Kimura & Egawa, 2018). Men under 50 years old have a very low overall incidence of PCa; they make up a mere 0.1% of all patients with the disease. Around 85% of cases are identified in patients 65 years old, whereas on crossing the 85 years of age, an aggregated global PCa threat lays up to 20% from 0.2%. Autopsies have shown that more than 75% of minor lesions are seen in men who are years old, 50% in sixth-decade and fourth-decades study claimed that the middle age single test of PSA might predict the chancesof PCa till the next 25 years. Men with PSA levels above 0.5 1.0 ng/ml have 2.5 fold higher risks, while those who have levels above 2.0 up to 3.0 ng/ml have a higher than 19 fold risk (Patel & Klein, 2009).

Ethnically, Asians have fewer incidents, particularly in India, China and Japan, where annually, 1.9 fresh cases are rising per every 100,000 men in the population. However, in the US, 161 fresh cases are added for every 100,000 men in the population (Patel & Klein 2009). African Americans have a higher incidence in the US (Barsouk et al., 2020). Interestingly, there has been a rise in cases when the origin of Japan moves to developing countries like the US, indicating that there must be external unknown factors that might be playing a role in increasing risk (Patel & Klein, 2009).

2.4.2. Genetic Disparities

It suggested that having a first-degree relative with PCa (familial) enhanced a man's likelihood of developing the disease. The chance of having PCa is inversely correlated with the age at which family members were affected and the number of affected family members (Xu et al., 2021). PCa can be categorized as hereditary, sporadic, or familial with early-onset cases (below or till the age of 55) found to be largely hereditary (43%), although sporadic instances make up the majority of PCa patients (85%). Hereditary cases can be considered as a subtype of the familial type, which exhibits a consistent genetic susceptibility pattern for PCa in the Mendelian pattern of inheritance. Twin studies say the probability of two twins

developing PCa could be because of the same environment and genetic makeup (concordance), especially for the identical and non-identical**.** Segregation analysis indicated that the tendency to cluster the PCa in the family could be because of its inheritance in an autosomal dominant fashion. Other than single nuclear polymorphism hypermethylation in promoter regions, gene fusions (fusion-oncogenes) like *TMPRSS2-ETV1* and *TMPRSS2-ERG* and deletions like in *PTEN* are also the genetic reasons for PCa development. Some of the susceptible genes for PCa are *MSR1, BRCA2, ATM, CHEK2, MSR1, OGG1,* and *RNASEL* (Patel & Klein, 2009) (Barsouk et al., 2020) (Lin & Jin, 2023). *BRCA1* has reported low risk in comparison to *BRCA2* (Barsouk et al., 2020).

2.4.3. Infection and Inflammation

PCa development can develop from infection related inflammations (prostatitis) caused by sexual intercourse or urine and the hyperproliferative. According to some studies, many viruses, like herpesvirus type-8, herpes simplex type-1, hepatitis B and C, human papillomavirus type-16, and cytomegalovirus are found to be associated with PCa development by turning on oncogenes and neoplastic transformations. In African populations, both prostatitis and gonorrhoea are seen as associated. Observations from several studies are inconsistent in declaring a clear relationship between infection and PCa (Patel & Klein, 2009).

In epithelial atrophy, a range of lesions that are probably caused by oxidative damage, infection, and hypoxia can cause mutations in cells that divide quickly, which could aid in the growth of cancer. This condition is also known as proliferative inflammatory atrophy and commonly coexists with high-grade prostatic intraepithelial neoplasias (Patel & Klein, 2009) (Oseni et al., 2023).

2.4.4. Androgens

Dihydrotestosterone (in tissues) and testosterone (circulating) are the two principal androgens found inside the male adult body of a male. These androgens are crucial in the development, maturity and maintenance of a prostate gland and require AR for their mode of action. However, few studies have shown that the high level of testosterone in serum increases the risk for PCa. Recently, a polymorphism in the gene (*SRD5A2*) coding for 5α reductase isozyme type 2 has been associated with PCa development. Even the conflicting observations are noted in the associations of mutations in genes related to testosterone biosynthesis and its

degradation with PCa, like Cytochrome P450 17 (CYP 17), HSD3B1 and 2 (Patel & Klein, 2009).

2.4.5. Diet, Fitness and Obesity

The risk of PCa is also related to the average amount of fat intake as well as polyunsaturated fat. Obesity has recently been considered a risk factor for inducing PCa by the crosstalk of oxidative stress caused by insulin resistance and inflammation caused by an increase in circulating inflammatory factors (IL-6, TNF and leptin) as a consequence (Patel & Klein, 2009) (Szablewski, 2024). Studies have shown that with a 5% increase in 5 kg/m^2 BMI, the risk for PCa is raised in an obese person (Patel & Klein, 2009). A trial study comparing BMI showed a 29% increase in risk with a BMI above 30 kg/m^2 and an 18% decrease in risk with BMI below 25 kg/m^2. Obesity is also found to increase the risk of developing leptin resistance. Leptin is known to play a role in stimulating proliferation and PCa progression in cell lines. Besides this, the intake of red meat prepared at high temperatures causes the formation of carcinogenic heterocyclic amines and less vegetable and fruit intake. Further, Vitamin D has been claimed in several studies to have a protective role against PCa by increasing apoptosis and (Patel & Klein, 2009) (Dovey et al., 2023).

2.4.6. Life style

Smoking is considered mostly a lifestyle factor as it causes a rise in the level of circulating androgens on exposure to cadmium, which brings high oxidative stress. Other than smoking, drinking alcohol and sexually transmitted infections during intercourse can contribute to raising PCa risk (Patel & Klein, 2009), while smoking increases the risk by 1.6 times (Kenfield et al., 2011).

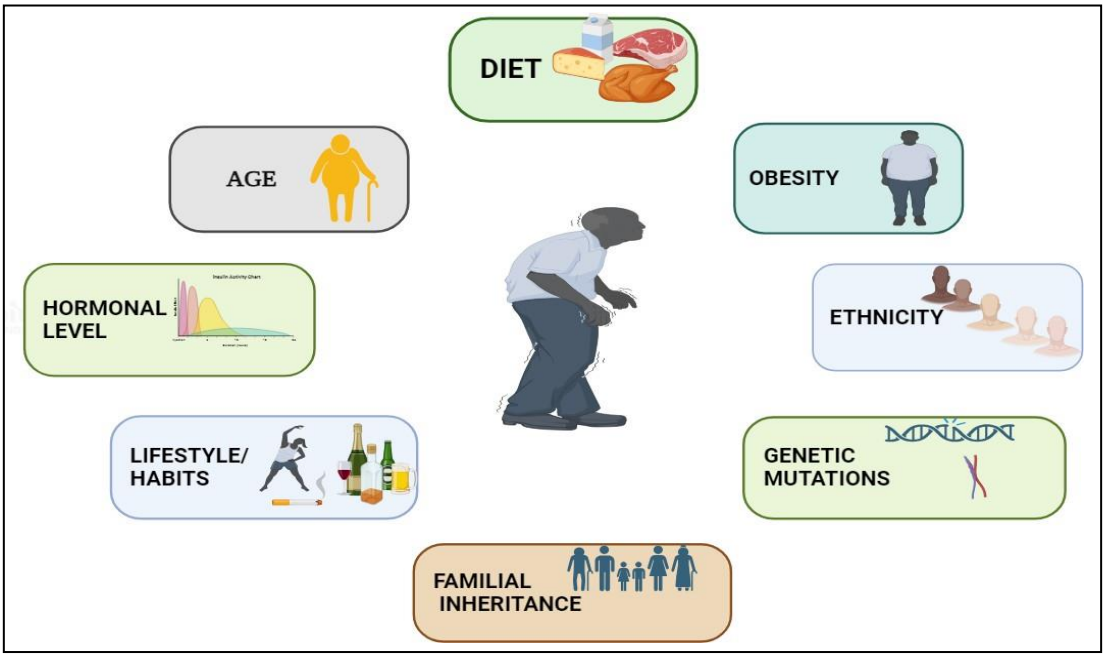


Figure 2.6: Risk factors of PCa includes diet, age, obesity, ethnicity, lifestyle, habits, familial inheritance, genetic mutations and hormonal level (Figure created with BioRender https://www.biorender.com/)

2.5. Symptoms:

General symptoms of PCa include prostate pain, sexual problems, erectile dysfunction, urinary problems (micturition and pain), loss of appetite, constipation, fatigue, diarrhoea, dyspnea, nausea, insomnia, cognitive functioning, problems with partner and financial difficulties (Bestmann et al., 2007). Bone pain, loss of weight, oedema in the lower end, intravascular coagulant anemia and impediment in the upper region of the urinary duct are seen in metastatic patients (Hernandez & Thompson, 2004). Treatment may also cause adverse effects on the body, like urine inconsistency, erectile dysfunction, insomnia and bowel issues. Pain, fatigue, and depression exist simultaneously as "symptom clusters" in PCa survivor patients (Baden et al., 2020). Depression and suicidal thoughts are frequently seen in PCa patients (one in every six cases). Selecting the right treatment is the reason for depression (Fervaha et al., 2019) **(Figure 2.7).**

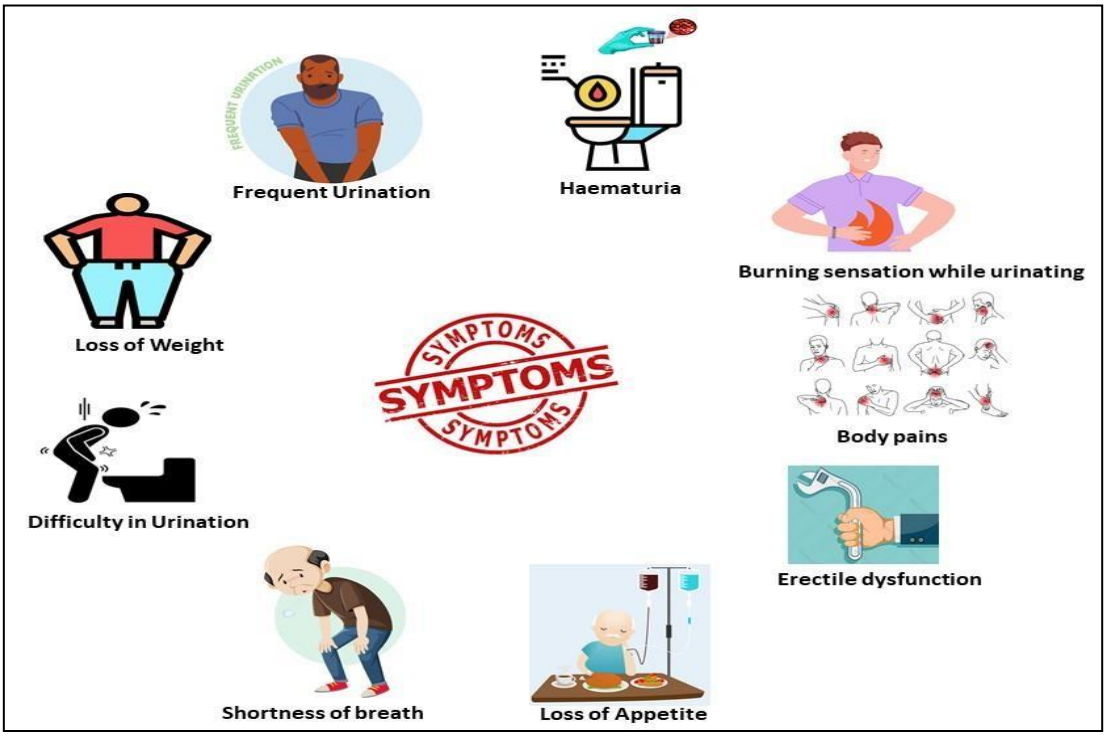


Figure 2.7: Symptoms of PCa includes loss of weight, frequent urination, haematuria, burning sensation while urinating, shortness of breath, loss of appetite and difficulty in urination

2.6. Diagnosis

Since, the PCa is heterogeneous in regards to the spectrum of tumors, starting from minor and indolent up to one that is a threat to life, the goal of primary identification is the evaluation of local and farthest tumor aggressiveness via staging. The conventional way of diagnosing PCa is first, with the help of DRE and PSA levels in the blood and then biopsy guided by TRUS (Descotes, 2019).

Prostate Specific Antigen (PSA):

The serum level of a glycoprotein produced by the prostate epithelium in men called PSA gets elevated in PCa and prostatitis. A cut-off value of 0.4ng/ml is set with a 46% accuracy rate. With the increasing age of a patient, accuracy can be up to 91% (Barry, 2001) (Kachuri et al., 2023).

Unfortunately, the drawback in diagnosis with the level of PSA is that PSA is specific to the prostate organ but not the PCa, which explains why there is an overlap in the levels of PCa and BPH. Therefore, it can be detected even if the level of PSA is low, indicating that there is no particular threshold value for PSA. Consequently, the decision regarding the threshold of a

PSA level at which a doctor would advise a biopsy is still debatable. However, the high levels can surely testify to PCa. According to a study, age-related ranges for PSA testing can improve the diagnosis specificity rate. A study has proposed cut-off values for age-related ranges of 2.5, 3.5, 4.5 and 6.5, respectively, for the ages 40s, 50s, 60s, and 70s. The level of free PSA (fPSA) is checked in those men whose DRE test is normal and whose PSA level falls in the range of 4ng/ml to 10ng/ml. Other than this, PSA velocity (>0.75 ng/ml/year) could be associated with PCa and could help in the stratification of risk before checking the total PSA (tPSA) level (Descotes, 2019).

Recently, phi test known as the prostate health index done from blood was used to combine tPSA, fPSA and proPSA blood tests. When compared to tPSA or fPSA/tPSA alone, phi results can increase the accuracy of PCa diagnosis. It has shown a greater specificity of 95%. Another test is called a 4K score, which is a combination of four Kallikrein (intact PSA, tPSA, fPSA, and a peptidase called hK2). Studies have suggested its high specificity in the diagnosis of high-grade cancers, even of 7 Gleason grade and could help in reducing unnecessary indolent biopsies by 41% (Descotes, 2019). Clinical staging is mainly based on whether a prostatic nodule is present or not in the DRE test. Gleason grading of biopsies is used to estimate the chances of organ restricted disease, invasion of seminal vesicles and involvement of lymph nodes of the pelvis (Hernandez & Thompson, 2004).

The urinary test includes PCa gene 3 mRNA testing (PCA3), which gets upregulated in PCa patients (Descotes, 2019). Multiparametric MRI (mpMRI) is also used for local staging. Metastasis in lymph nodes can be detected through CT scans, MRI and monogram's (Descotes, 2019).

2.7. Treatment

2.7.1. Radical prostatectomy:

Radical prostatectomy, considered the primary approach to treatment, in general, has been reported to increase the survival rate ranging from 90-94%, metastasis free survival from 82-87%, and PSA free 10 year survival rate ranging from 52-68% (Hernandez & Thompson, 2004) (Bejrananda & Pliensiri, 2023).

2.7.2. Radiation therapy:

In this approach, a beam is administered externally to treat the local tumors. Out of the different types of radiation therapy, the choice for a particular patient is made based on the range of risk inferred by several pre-treatment tests like Gleason grading score, level of PSA and clinical stage. To improve the outcomes of the therapy, hormonal therapy is provided in

combination with the patients who are at higher risk. Those who have undergone radiotherapy are less likely to have bowel issues in comparison to those who have undergone radical prostatectomy (Hernandez & Thompson, 2004).

2.7.3. Cryosurgery:

This kind of surgery is done to kill the tissues or cells on direct exposure to cold and followed by hot treatment, which delays the blood supply and damages the structure of tissues (Hernandez & Thompson, 2004).

2.7.4. Hormonal therapy/Androgen Deprived Therapy:

ADT is also known as surgical castration and is the most preferred treatment for the treatment of advanced stage adenocarcinoma PCa (Sharifi et al., 2005) (Y. H. A. Lee et al., 2023). The positive effects of ADY were first observed by Huggins and Hodges in metastatic patients (Huggins 1941). Androgen receptors (ARs) belong to the family of nuclear receptors and regulate the expression of target genes as ligand induced transcription factors. These include sex steroids (progestin, androgens and estrogen), steroids of the adrenal gland (mineralocorticoids and glucocorticoids), vitamin D, fatty acids, retinoid and thyroid hormones (Miyamoto *et al.*, 2004). It basically involves the inhibition of male hormones, specifically testosterone and ultimately the reduction and inhibition of ARs. There are two ways for hormonal therapy: Luteinizing Hormone-Releasing Hormone (LHRH) analogue therapy (suppresses LH-RH secretion) and LHRH antagonist therapy (which antagonizes the LH-RH receptor) (Sekhoacha et al., 2022). Others include estrogen Diethylstilbestrol (suppresses LH-RH secretion), steroidal antiandrogens (antagonize AR in a particular tissue and suppress LH-RH), non-steroidal antiandrogens (antagonize AR), 5α-reductase inhibitors and adrenal androgen inhibitors (to suppress adrenal corticoids and testicular steroid genesis) (Miyamoto et al., 2004).

2.8. Biomarkers of PCa

Recently, urine has been considered as important as blood as a biofluid for diagnosis because of its easy collection and the presence of materials from all the body tissues directly. Therefore, cells, DNA, RNA and proteins from the prostate in the urine could be potential biomarkers. However, the Food and Drug Administration (FDA) has only approved PCA3 (PCa Antigen 3) as a urinary marker. Its level is estimated by taking the ratio of PCA3 and PSA mRNA. Since it is, the cut-off value is also not clear; therefore, PCA3 testing is also considered for those who have had a negative biopsy previously. Thus, genetic testing along

with the results of family history PSA and DRE is considered for calculating risk. The genetic testing includes the Mi Prostate (the combined result of PSA of serum and TMPRSS2-ERG with PCA3) or SelectMDx (measure of DLX1 and HOX6 gene mRNA) score. Another score called ExoDx Prostate IntelliScore is for ERG transcript expression in exosomes and PCA3 (Salciccia et al., 2021).

Earlier, PSA was the only FDA-approved serum biomarker. Blood biomarkers are less specific markers that are explored, though many are studied like fPSA. Lately, because of limitations with PSA, the FDA has approved a 4K score (Four-kallikrein) and prostate health index (PHI) serum assay. PHI is the combined test of tPSA, p2PSA and fPSA (Salciccia et al., 2021).

Metabolomics, a study of metabolites playing a role in any biological process, is an effective method for clinical research that could help to identify biomarkers and would provide better knowledge for disease pathophysiology by bridging the gap between genotype and phenotype. Nuclear Magnetic Resonance (NMR) and Mass spectrophotometry (MS) with targeted and untargeted approaches offer various perspectives on metabolite profiles. Later, for the identification of biomarkers and interpreting data, chemometric is applied. For instance, in the case of pancreatic cancer cysteine, cystathionine and homocysteine are found to be helpful in the detection of recurrence, while lysophosphatidylcholines are found to be associated with pancreatic cancer risk (Salciccia et al., 2021).

Exosomes are found in different biofluids like urine and plasma, generated by endosomes. These are considered potential biomarkers for diseases and cancers because they have specific markers (e.g., carbonic anhydrase IX, CA-IX, CD63, CD81, CF9, HSP70, etc.) obtained during the fusion of endosomes and plasma membranes. These are filled with DNA, micro-RNAs and mRNAs. Cancer cells increase the release of exosomes because of extracellular acidosis and hypoxia. A study found that the level of exosomes PSA was significantly higher in plasma as compared to BPH patients, suggesting high sensitivity (96%) and specificity (100%). Further, it could help in distinguishing BPH from PCa (Salciccia et al., 2021).

2.9. Comorbidities of PCa

While making decisions for PCa, comorbidities are very poorly considered, due to which the patient faces a dilemma for definite treatment as to whether they should go for aggressive treatment or not. As a consequence, patients with comorbidity have post-treatment traumas. The Charlson Comorbidity Index is a method which can assess mortality risk and survival. A study has explained that those who have single comorbidities like obesity, alcoholism,

smoking and mobility problems are receiving aggressive treatments as compared to those who are without comorbidities. The reason for such discrepancy could be because of false assumptions by both doctors and patients of long survival periods with single comorbidity as they are not going to impact much on men and hence, neglecting proper estimation of mortality during treatment. However, for men who have peripheral vascular problems, diabetes with organ damage and chronic obstructive pulmonary disease (COPD) and have a 10 year life expectancy over the age of 75 years, then aggressive treatment (surgery or radiation therapy) may not be the right option for them (Chamie et al., 2012).

A study in 1999 by Post PN *et al.* estimated the chance of having a minimum of single comorbidity for patients between 60 and 69 ages is 38%, 48% for 70-74% ages and 53 percent for patients who are 75 years old or above (Post et al., 1999). Among all diseases, COPD and cardiovascular diseases are very common (Post et al., 1999). Previous studies have examined the effect and outcomes of ADT on patients with co-existing major cardiovascular disease and inferred that therapy could be fatal or have other adverse effects. Hence, careful consideration of the risk-benefit ratio of each health condition is required (Wu et al., 2020).

To investigate the association between co-existing medical conditions and the PCa stage at which it is diagnosed, it was shown that patients who have coronary heart disease, dyslipidemia or benign hypertension have a lower chance of advanced PCa than those who have no co-existing medical condition (Fleming et al., 2006). Conversely, a higher risk of advanced stage is among those who have renal disorders, peripheral vascular diseases and substance addiction. Considering racial disparities, comorbidities of renal disease are increasing of last stage PCa with the same frequency in both races (black and white). However, it is seen that substance adductors among white men are at greater risk as compared to black men (Fleming et al., 2006). Further, it has been observed, according to the study, those patients who are in regular touch with healthcare centres for comorbidities are getting early diagnosis because of regular testing. This suggests that comorbidities interact with the pathogenesis of cancer to increase aggressiveness (Fleming et al., 2006).

2.9.1. Diabetes Mellitus

Diabetes mellitus (DM) occurs as a result of insufficient insulin production or impaired insulin sensitivity, and it has become a serious threat to people's health. It is a heterogeneous problem with numerous aetiologies comprising three main types, viz., type 1 diabetes mellitus (T1DM), type-2-diabetes mellitus (T2DM), and gestational diabetes mellitus (GDM). Understanding the biological mechanisms associated with them would allow us to identify candidate proteins and genes (X. Tang et al., 2016). The emergence of genome-wide

association studies (GWAS) has substantially enhanced our understanding of the genetic basis of disease risk in the past few years. Prior to the introduction of GWASs in 2006, very little information was available about the genes that influence common complicated or multifactorial diseases and quantitative traits. These research findings imply that susceptibility to prevalent diseases is influenced by a variety of genetic topologies, including common genetic variants with minimal effects and uncommon variants with substantial impact sizes. Nevertheless, the combination of candidate T2DM genes discovered using GWASs does not fully confirm established features of disease pathogenesis. Several system-level approaches have been used to bridge the gap between genome and phenome correlation (Jain et al., 2013). Computational analyses of disease-linked genes using interactome and toxic genomic data help us to connect T2DM candidate genes found in GWAS with disease pathophysiology, including abnormal pancreatic cell formation and function and insulin sensitivity. On the other hand, computational predictions of potential proteins/genes are less expensive and time-saving than experimental methods. In order to unravel the genetic roots of common disorders, it is necessary to understand the complexity of the gene–phenotype connection. Recent research employing the human interactome and phenome has uncovered not just common phenotypic and genetic overlap between diseases but also a modular architecture of the genetic landscape of human diseases, opening up new avenues for reducing the complexity of humans. Because diseases are rarely caused by the malfunction of a single protein, a more comprehensive and robust interactome is essential for identifying groups of interconnected proteins associated with disease aetiology. GDM is categorised as insulin resistance leading to hyperglycemia during pregnancy, which mostly retracts after parturition. According to the WHO, the prevalence rate is 15.8%, accounting for about 20.4 million live births, with the majority of cases in pregnant women above the age of 35 years (Who.int, last accessed on April 2, 2024). The International Diabetes Federation in 2019 estimated a prevalence of 28.5% in India, with incidence varying in each state due to challenges in screening strategies and paucity of consensus among physicians and healthcare providers in prepartum and postpartum management of GDM. Pregnant women with GDM have an inherent risk of developing T2DM post-delivery or later in life. The offspring is also susceptible to any form of diabetes postnatally or in the long term. The genetic factors responsible for GDM and the future risk of developing T2DM through epidemiological and physiological studies reveal commonality in susceptibility loci, which implies that most of the diabetes genes are involved in causing GDM. The few key genes that share common variants are *KCNJ11, GCK, HNF1A, TCF7L2, CDKAL1, KCNQ1, CDKN2A, MTNR1B, SRR, HHEX, TCF2, SLC30A8, and IGF2BP2*. Genetic similarities between T1DM and GDM are less

studied, and a study among Asian Indian women with GDM showed the presence of pancreatic autoantibodies like GAD, which is a biomarker for T1DM (Ranjit Unnikrishnan et al., 2016) (Rout et al., 2022).

2.9.2. Obesity as a phenotype

Men who have both obesity and PCa have higher chances of recurrence, last stages and PCa related deaths. Moreover, obese PCa would face adverse during or post-treatment effects and would more likely lead to CRPC. Besides this, obesity can also cause other comorbidities like diabetes and cardiovascular diseases. It can also affect men psychologically, socially, mobility and overall life quality (Wilson et al., 2022) (Tiruye et al., 2024). It can be *vice versa*, that PCa patients who are not obese can get obese later out of stress, change in eating, or less physical action because of side effects from treatment and ADT (Wilson et al., 2022). Although the precise physiological processes linking obesity with PCa are not completely known, it is assumed that the metabolic modulations and inflammatory environment associated with high body fat play an important role (Wilson et al., 2022) (Mistry et al., 2007).

Hormonal imbalances associated with obesity can also promote PCa progression instead of becoming a risk aspect directly. It is assumed that the adipose tissue actively functions as an endocrine organ and secretes adipokines (Mistry et al., 2007). In obese patients, the level of circulating ratio of several molecules like steroid hormones (testosterone), insulin-like growth factor, insulin, and adipokines also imbalances (Mistry et al., 2007) (Freedland & Aronson, 2004) (Szablewski, 2024). As a result, it may cause prostate cells to get exposed to elevated circulating adipokines or help in retropubic space invasions and promote tumor progression. Furthermore, the interaction of leptin (adipokines) and adiponectin was also reported to promote PCa cell line proliferation. However, only adipokines can't justify this association, and more study is required (Mistry et al., 2007). Body Mass Index (BMI) above 30 kg/m^2 is supposed to raise the chances of having PCa in adults of age above 21 by 9%. Men aged 50 or above 59 are at risk 58% for PCa, suggesting a complex relationship of age and obesity with respect to PCa. A controversy related to high BMI is also there, even as few studies showed a protective effect of high BMI and BMI estimation at adult ages is ambiguous results in prediction (Freedland & Aronson, 2004).

2.9.3. Hypertension

Arterial hypertension increases the risk of PCa by about 15% (Radišauskas et al., 2016). A drug used to treat PCa called enzalutamide is found to increase the risk of developing hypertension and high grade cardiovascular problems. It is suggested that enzalutamide is an inhibitor of androgen signalling, causing androgen deprivation. Androgen deprivation is associated with hypertension because of endothelial dysfunction, causing the proliferation of smooth muscles and vasoconstriction. A fall in testosterone is also an indicator of endothelial dysfunction. Recently, hypertension as a side effect of ADT has frequently been seen. Other ADT drugs that increase the incidence of hypertension are *leuprolide, darolutamide, apalutamide, abiraterone, firmagon and bicalutamide.* Anti-VEGF is also known to increase hypertension in any cancer patient (Xiaolei Zhu & Wu, 2019).

2.9.4. Urolithiasis

When stones develop in any portion of the renal system-the kidney, bladder, or urethra, it is known as urolithiasis, commonly referred to as urinary tract stone disease. It is the third most prevalent urinary illness globally, with different incidence rates in different countries. It has been discovered to be between 1% and 5% in Asian nations, 5% to 9% in Europe, and 7% to 13% in North America. Globally, the incidence rate was 62 per 100,000 people per year; however, it was 197 per 100,000 people annually for men and women. Although there was a very high rate of reversion and frequency, the risk factors have not yet been classified, despite the fact that the danger of familial inheritance in higher possibilities of stone formation is similar to 2.57 times that in men. Although the genetic tendencies to manifest urolithiasis are not known, the intricate interactions between genetic and external environmental factors would perhaps allow us to have a deeper understanding of the disease. Within the first five years following the initial incident, over 50% of patients with kidney stones are at risk of another incident (X. Sun et al., 2021). The complicated pathophysiology of urolithiasis involves the production of tiny stones that often do not cause any symptoms; nevertheless, if the stones become larger than 5mm, it becomes an obstruction to the ureter and causes excruciating agony. The kidney stone excretes itself from the body through the urine. While bladder stones can grow naturally or can be incorporated into kidney-derived stones through the urine stream, certain stones can also enter the pelvic area and start to harm the kidneys. Some microorganisms are also known to play a role, i.e., stone formation, like urea splitting (Proteus species), forms the struvite stones (infection stones) in the kidney. Not only urinary, but also intestinal microbes may also be responsible as a consequence of their associations with other diseases, such as hypertension, diabetes mellitus, obesity, and other metabolic

syndromes. Interestingly, 70% of calcium stones are composed of oxalate as gut microbiota like oxalate metabolizing bacterial species (OMBS) specifically O*xalobacter formigenes*, ingests oxalate and activates its secretion to form oxalate stones (X. Sun et al., 2021). Furthermore, a correlation between increased body mass index (BMI) and the formation of kidney stones was also seen on the basis of some evident studies, which have shown that there is more mineral excretion (Uric acid, Calcium, Sodium, Oxalate, Phosphorous, Sulphate, Magnesium, Cysteine and Citrate) and a fall in the pH of urine (X. Sun et al., 2021). The prevalence of calcium oxalate is higher, irrespective of obesity. More specifically, in obese patients, uric acid stones (63%) are more common when compared to non-obese (11%) patients. Since obesity, along with age, is more susceptible to stone formation, both of these are associated risk factors for urolithiasis. Low pH endures phosphate or calcium stones, but high pH favours uric acid stones. On the other hand, inflammatory agents (eicosanoids) having poly fatty acids lower the excretion of calcium, thereby further reducing kidney stone formation, and these fatty acids can change vitamin D synthesis, which controls the calcium excretion (X. Sun et al., 2021). Stone formation is still an unclear process; however, the mass spectrometric proteomic approach for matrix proteins might be helpful in understanding stone formation. Urolithiasis has posed itself as a universal health issue. This new era has already given approaches which are far less invasive, but the rate of recurrence has proved the advancement inadequate for urolithiasis prevention. Despite other lifestyle reasons, it may be possible that some urolithiasis genes and their polymorphisms are associated with developing PCa (X. Sun et al., 2021). By using the bioinformatics approach, we have also reviewed the role of Urolithiasis and PCa in relation to genetic association and associated risk factors. We have identified their pathogenic and likely pathogenic polymorphisms by co-relational studies (X. Sun et al., 2021). Efforts are on the way to identify and screen urolithiasis related mutations using NGS approaches, *viz*. WES, RNA-Seq etc.

Recently, more detailed work on PCa has been included using WES and identifying some differentially expressed genes (DEGs). These DEGs are also subjected to analysis in relation to urolithiasis.

Association of Urolithiasis and PCa:

Urolithiasis and PCa can be seen at the same time and are treated simultaneously in many cases. By applying conditional logistic regression, a case-control study has confirmed the associations of PCa with previous urinary calculi (kidney, bladder and other unspecified sites). For example, Urinary calculi can be classified into two subcategories; primarily

because of metabolic issues and secondary as a result of several diseases (comorbidity), medicines, dietary habits (fat intake), anatomy, infections or more likely in those patients with diabetes or glucose intolerance. Thus, dietary changes like more fat consumption on one side promote stone formation, and, on the other side, the hormonal imbalance caused by AR ultimately leads to PCa progression. While treating a PCa patient with luteinizing hormone-releasing hormone analogues, they produce androgen deficiency along with osteoporosis through certain cytokines, mainly IL-1, 6, 10 and TGF-1. This state of bone loss also increases calcium and phosphorus minerals in the serum and urine, resulting in renal calcium deposition in the kidneys (nephrolithiasis). Androgen therapy also lowers ammonium excretion, resulting in an increase of free ions inside the tubular lumen, which increases the risk of uric acid lithiasis. Therefore, oxidative stress and decreased urinary pH because of androgen deprivation therapy (ADT) also cause urolithiasis. According to some follow-up cases, people undergoing testosterone replacement therapy have also shown associations by inhibiting an active calcium transporter (TRPV5) and increasing calcium excretion and oxalate levels in urine. A study says doing brachytherapy for PCa has a high chance of lithiasis because of iodine seeds acting as nidus. Also, next-generation (NGS)-based exome analysis of common genes between Asians and Caucasians for PCa found 9 genes, and *UGT1A*1 was one of them. However, *UGT1A*1, already mentioned, is responsible for urolithiasis by forming Atazanavir related to kidney stones (X. Sun et al., 2021).

UGTIA1: Protease inhibitors like atazanavir are also well-known to be associated with kidney stones. A case study of 74 year old HIV patients reported kidney stones with a composition of atazanavir and renal colic. Atazanavir linked lithiasis has a prevalence rate of 7%. This type of nephrolithiasis is considered rare in comparison to other types of lithiasis. Genotype TC of rs10929303, GC of rs1042640, and rs8330 allele of UGT1A-30-UTR independently contribute to risk in these HIV patients. Depending upon the types of protease inhibitor drugs, the composition of the stone will change accordingly (*emtricitabine, tenofovir, ritonavir, indinavir, nelfinavir*). Therefore, urolithiasis has shown its impact on PCa through calcium homeostasis, ROS or androgen receptors (X. Sun et al., 2021) (Taguchi et al., 2017).

2.10. Diabetes and cancer

Some meta-analysis studies have suggested that the highest relative risk observed for increased chances of cancer in diabetes is with liver and pancreatic cancer. In comparison to any other tissue type, liver cells are more exposed to high concentrations of insulin during portal circulation, as seen in insulin-resistant hyperinsulinemia type-2-diabetes patients. Therefore, the mitogenic impact of insulin is particularly involved in raising liver cancer due

to the higher physiological exposure of normal liver cells to insulin. Besides this, exogenous insulin supply also makes liver cells exposed to the same level of insulin. Hepatitis B and C virus infections, and other diabetes related disorders, including steatosis and cirrhosis, are also assumed to indirectly expedite the development of cancer in diabetic patients. The risk of pancreatic cancer starts with the pre-diabetic stage. Studies have observed that it increases relative risk by increasing glucose tolerance impairments. High exposure of exocrine pancreatic cells to high insulin (hyperinsulinemia) has also been observed in pancreatic cancer development because of the same blood flow with insulin secretory cells (islets). In the case of kidney cancer, the risk is generally referred to as hypertension, obesity and hyperinsulinemia. Patients with DM frequently have urinary tract infections, increasing the chances of cancer in reproductive organs. Diabetes increases the risk of both endometrial and breast cancer by unbalancing the sex hormones and delaying menarche, specifically in those women who have type-1-diabetes and fertility disorders. In the case of PCa, although diagnosis at an early stage (neoplasia forms) in diabetic patients reduces the cancer incidence, once the patient becomes overweight and insulin resistant, the chances of the patient ending his life with PCa will increase. However, diabetes has not only increased the relative risk of incidence but has also increased mortality more drastically in breast cancer and colorectal cancer (Vigneri et al., 2009). DM and cancer have common risk factors like obesity, physical inactivity, age and smoking associated with hyperglycemia, insulin resistance to mention a few (Smith and Gale 2010).

Being an autoimmune disorder, type 1 diabetes requires an exogenous supply of insulin when the Beta-cells of the pancreas are unable to produce endogenous insulin. Exogenous insulin cannot mimic the compartment distribution pattern of endogenous insulin, which is first supplied to the liver where it retains the required level of insulin (80%), and the left hormone flows through the systemic circulatory system to reach peripheral tissues (liver/peripheral ratio (9:1). In contrast to endogenous insulin, exogenous insulin, reaches both liver and peripheral tissue at the same time and at the same concentrations. Hence, exogenous insulin leads to a two-fold to five-fold higher hyperinsulinemia in peripheral tissues. While in T2DM, there is endogenous hyperinsulinemia because of insulin resistance and this insulin can't replace the fuel storage of the body (Vigneri et al., 2009) (Szablewski, 2024). Hence, the unused material, i.e., glucose, is found in excess. The excess of both glucose and insulin results in several other hormonal abnormalities leptin, glucagon, etc.). With passing years of DM, the beta cells decrease insulin secretion because of an increase in the rate of apoptosis, which is not balanced by neogenesis. When this stage arises, the T2DM patient's body starts working metabolically like a type-1-diabetes patient with both endogenous hyperinsulinemia

and exogenous insulin requirement. As a consequence, the duration of diabetes and the requirement of insulin differently affect the exposure of tissue to insulin. It is well known that through the mitogenic action of insulin, hyperinsulinemia enhances cancer initiation in multiple ways For instance, high insulin binds and activates the insulin-like growth factor 1(IGF-1) receptor and decreases the IGF-1 binding proteins, setting IGf-1 free as an active growth factor (Noto et al., 2013) (Vigneri et al., 2009). Secondly, high insulin may increase the insulin receptor IR level, having two isoforms, IR-A and IR-B. IRs after alternative splicing increases the mitogenic effects (Vigneri et al., 2009) (J. Li & Huang, 2024).

Over the years, several antidiabetic drugs, *viz.* biguanides (metformin), thiazolidinedione, sulphonylureas, meglitinides and α- glycosidase inhibitors are in use (Huo et al., 2023). A patient is prescribed a variety of drugs over the years of diabetes duration. These drugs help to decrease the hyperinsulinemia (Vigneri et al., 2009). The biguanide metformin had been recently found to reduce the cancer risk in DM patients. Apart from reducing the insulin, these antidiabetic drugs are showing anticancer effects by stimulating AMPK, known to induce glucose consumption by muscle cells and also stimulating its upstream regulator LKB1 (tumor suppressor gene). AMPK works as an anti-proliferative by reducing insulin and IGF-1 signalling (DeCensi et al., 2010) (Vigneri et al., 2009).

It can inhibit the *MAPK*, *AKT* and *mTOR* to arrest the cell at the cell cycle (Vigneri et al., 2009). The sulphonylureas are known to increase insulin and increase the cancer risk. Vice-versa, many anticancer therapies can induce hyperglycemia. For example, glucocorticoid use has shown diabetogenic action by causing insulin resistance when used at high doses. Corticosteroids used as anti-androgens have also shown a negative impact on the glucose metabolism. Androgen therapy decreases insulin sensitivity, alters the lipid profile, and raises the risk of diabetes and heart problems (Vigneri et al., 2009).

Metformin mainly works in two ways. First, it reduces the production of glucose by blocking one step of an aerobic cellular energy production molecule called ATP via activating the *AMPK* enzyme. The other way metformin works independently is *AMPK*, where *ATM* is assumed to modulate the action of metformin by some other parallel pathway. It has been observed in some patients with ataxia telangiectasia who were having insulin resistance (Birnbaum & Shaw, 2011). It was indicated that the *ATM* could regulate *AMPK* by phosphorylating its upstream activator, LKB, to affect the cellular process, or it can regulate the *AMPK* independently of *LKB1*. It regulates the other targets of metformin independent of AMPK (Birnbaum & Shaw, 2011).

Hyperglycemia favours malignancies as it fulfils those cellular processes which consume the highest level of energy, like cell proliferation; thereby, glucose serves as a fuel to cancer

progression. This was provided by many studies wherein starved tumor cells reduced their malignant growth. Hyperglycemia also supports angiogenesis through microRNA-467 upregulation. It produces oxidative stress and inflammation by the interaction of advanced glycation end products and their receptors, favouring cancer development (Sciacca et al., 2013).

Biological connection between diabetes and cancer

The glucose level in the body is regulated by insulin, a hormone (peptide) which increases the glucose up taking and its assimilation. However, insulin resistance is stated when it becomes unable to perform this function in a diabetic patient. On the other hand, the beta cell continuously secretes insulin to make up and maintain balance, but it results in hyperinsulinemia (Wilcox, 2005). This increased level will trigger the production of IGF-1 from liver cells. IGF-1 will then binds to its tyrosine kinase receptor IGF-1Ra and stimulate various metabolic and mitogenic signalling pathways to control processes like cancer cell proliferation, differentiation, and apoptosis. Later, some downstream targets like PI3KB and rat sarcoma-mitogen-activated protein kinase/extracellular signal regulated kinase signalling pathways get stimulated. *PI3KB* signalling has a role in cancer cell survival and migration. While rat sarcoma mitogen-activated protein kinase, extracellular signal regulated kinase signalling pathway to control cancer cell proliferation and metabolism (Poloz & Stambolic, 2015). Hence, patients who have diabetes show increased levels of IGF-1, bringing them more susceptibility towards higher risk of different cancer development like breast, prostate and colorectal cancer (Szablewski, 2024). In general, cancer cells show the Warburg Effect, in which high glucose uptake is there for more glycolysis to provide energy to cancer cells. Therefore, this state of hyperglycemia in diabetic patients will provide cancer cells with good conditions for the survival and proliferation of a cancer cell. At the same time, tumor DNA and protein synthesis are linked to the metabolism of glucose. Consequently, a higher blood glucose level impacts the growth of a tumor and metastasis (M. Wang et al., 2020) both conditions, hyperinsulinemia and hyperglycemia, co-exist in nearly all patients of diabetes, it becomes hard to identify the individual part played by each of these abnormalities. A clear idea about whether or not hyperglycemia acts as an autonomous factor in the promotion of tumor growth and metastasis was taken up by (M. Wang et al., 2020). In the wake of assessing the different cell and molecular pathways, metformin (MF) might be considered a perfect specialist to be utilized either alone or in assistance as standard medications for PCa. MF has demonstrated to be a perfect remedial operator, being reasonable and non-harmful for cell targets. MF assumes a job as an enemy of proliferations and hostile to cancer causing

operators. MF can: (1) bring down the degrees of insulin, which therefore brings down cell expansion, (2) increment the initiation of the *AMPK* pathway that brings about development restraint legitimately to PCa cells, and (3) in a roundabout way follow up on *AMPK* enactment and reduce cell multiplication in malignant growth cells. MF smothers the androgen signalling pathway, and modifications of insulin-like development factor-1 (IGF-1) flagging pathways cause the development and multiplication of PCa, hindrance of the mTOR pathway, and lipogenesis. In light of epidemiological factors, patients with DM may be protectively affected by PCa (Zaidi et al., 2019). So, metformin diminishes hyperglycemia hyperinsulinemia, and the results of these conditions are dominated by diminishing hepatic gluconeogenesis. This medication may likewise act straightforwardly on diseases or cells in danger of change by instigating vitality stress, which eases back cell multiplication through initiation of *AMPK* or potentially other mechanisms (Klil-Drori et al., 2017). Patients treated with Metformin for T2DM indicated lower tumor mortality, showing a relationship between metformin and tumorigenesis (Zi et al., 2018).

2.11. Diabetes and PCa

Androgen receptor and PCa:

AR is a member of the steroid and nuclear receptor superfamily of nuclear transcription factor (NR3C4, nuclear receptor subfamily 3, group C, gene 4) and is involved in the regulation of normal growth and development of various target organs (W. Gao et al., 2005). The AR is a ligand-dependent transcription factor upon binding with androgens, *i.e.* native ligands 5α-dihydrotestosterone (DHT) and testosterone. AR initiates male sexual development and differentiation. In a nutshell, AR functions in response to androgens and regulates the transcription of genes via nuclear translocation (Heinlein & Chang, 2004) (Davey & Grossmann, 2016).

PCa is also dependent on androgen stimulation mediated by the AR, even as the latter plays a significant role in the growth and differentiation of the healthy prostate (Koochekpour, 2010). For example, it is well known that the AR complexes move to the nucleus and dimerizes before modulating the transcription of targeted genes, wherein, many genes become regulatory and invite candidate genes to interact with them. Furthermore, it also allows the drugs to be targeted on their AR grooves, thus making the active functional motifs (Koochekpour, 2010). The AR gene is located on the X-chromosome at the locus Xq11-12 and is composed of 8 exons coding for a ~2757 bp open reading frame and ~919 amino acids within a 10.6 kb mRNA (Brinkmann et al., 1989). AR signalling pathway involves various

mediators regulated by long non-coding RNA (lncRNA) by various mechanisms (Yongyong Yang et al., 2021). PCA3 modulates PCa cell survival via modulating AR signalling and is now used in PCa diagnosis (Lemos et al., 2019).

The testis of a fetus produces first testosterone and the 5α reductase, which later reduces the level of testosterone to undergo the process of prostate morphogenesis. Those foetuses lacking 5α reductase enzyme due to mutation or deficiency will develop an underdeveloped prostate. Once the prostate gland forms the androgens help to maintain the survival of secretory epithelial cells, which are known to transform into adenocarcinomas. However, the mere level of androgen is not the reason for the cancer progression, but the AR is equally important as mediators in cancer development. For instance, a fall in testosterone could be because of hyper expression of ARs. It is observed that the AR expression is highly heterogeneous in PCa, which usually correlates with ADT response. Besides this, the reason behind the cancer development in some cells that have cells that have lost the AR expression/gene is still unclear. Therefore, an attempt to complete AR ablation could give selective benefit to cancer growth or survival. Moreover, PSA expression is also observed to be primarily regulated by AR (Heinlein & Chang, 2004). Moreover, dysregulated lipid metabolism related pathways also contribute to escaping of tumors from eradication in PCa (Guerrero-Ochoa et al., 2024).

Role of Androgen Receptor in Diabetes:

A study showed that 43% of DM had a lower level of testosterone, and 57% of DM patients had a lower level of free testosterone, whereas only 7% of patients who had type 1 diabetes showed lower testosterone levels. The lower level of testosterone is generally linked to high insulin resistance in both type 1 and type-2-diabetes (inverse relationship). This inverse relation is seen majorly in type-2-diabetes (Grossmann et al., 2008). This relation can be seen vice versa as lower levels of androgen (testosterone) can ensure acquiring DM in future and hypogonadism. A fall in free testosterone level is also reported in obese people and is inversely related to the obesity degree. Obesity also causes insulin resistance, glucose intolerance, an increase in the level of leptin, and further decreases the androgen levels. After the castration, it has been observed in many studies that the level of insulin and improvement in glycemic control. However, the mechanism underlying insulin resistance is yet more to understand (Kapoor et al., 2006). It is suggested that when an androgen like testosterone binds its AR in liver cells, it activates the AR signal, which leads to glucose stimulated insulin

secretion (GSIS), and the deficiency of GSIS is associated with DM. These androgens might be involved differently in DM associated disorders (Sakkiah et al., 2018).

A study has reported that the elevated AR signalling is because of mutated IGF-1/insulin receptors in PCa with DM patients (J. Li & Huang, 2024). The authors concluded that this could be the reason for the poor prognosis of diabetes and PCa; another reason could be reduced estrogen receptor ligands causing the disinhibition of androgen pathway signalling (Lutz *et al.,* 2018). However, the growth factor IGF-I, which shares locus with lncH19 (IGF-II/H19), forms an imprinted gene. This silencing is found disrupted in different cancers including PCa. The association of adipose tissue and obesity is a known risk factor for both T2DM and PCa by disturbing cellular environments. As a result, hyperglycaemia or inflammatory, metabolic situations are hypothesized to be the cause of this loss of imprinting (LOI) (Kingshott et al., 2021).

Antidiabetic drug and PCa:

Metformin, an antidiabetic drug from several studies, has been proven to not only have an effect on glucose metabolism but also show interactions with androgen receptors. It plays a role in stabilizing PSA levels (Rout et al., 2022). In certain therapies, another commonly used method for T2DM, it is reported that glucagon-like peptide-1 receptor expression plays an anti-PCa effect. It helps in attenuating cell cycle progression. So, its forceful activation to express can be a potential therapeutic approach (Rout et al., 2022). Therefore, both metformin and certain therapies help in blocking cell cycle progression by reducing mTOR activity (Rout et al., 2022). Hypogonadism (decrease in level of testosterone) is also found associated with both diabetes and PCa. A fall in its serum level is capable of causing high graded PCa. Hence, T2DM is suggested to be a crucial predictor of high graded PCa, especially with benign prostatic hyperplasia (Rout et al., 2022). For early possible detection, PSA levels are broadly used, but its concentration shows variation due to several other comorbidities, age, and lifestyle, which makes it demand more precise analysis of test results. Based on a linear regression analysis, there is a fall in PSA in patients who are taking antidiabetics and obese people on haemodilution. This establishes an inverse relationship between diabetes obesity and PSA level. Such a study suggests deliberately checking the PSA level, especially in diabetic and obese patients (Rout et al., 2022). Both PCa and DM incidence is rising parallel with age. Despite the fact diabetes mellitus reduces the risk of PCa; DM can also increase its mortality (Rout et al., 2022). The understanding of the association between DM and PCa is still insufficient. Moreover, obesity makes its pathophysiology a more complex situation

(Rout et al., 2022). Besides this the glucagon-like-peptide-1 (GLP-1) used as external analogues in treating diabetes was found in lowering the risk of PCa to half (Stein et al., 2023).

2.12. LncRNAs

RNA molecules have one of a type known as non-coding RNA (ncRNA) which cannot code any protein. Studies on several molecular processes of a cell have reported their role in regulating them. Basically these ncRNAs are categorized further into small ncRNAs, which involve Circular RNAs (CirRNAs), microRNAs (MmRNAs), and long noncoding RNAs (lncRNAs) and long intergenic noncoding RNAs (lincRNAs) (S. Yang et al., 2018) It is suggested that all those ailments caused by corrupted transcriptome are not always limited to errors in making of protein-coding RNAs; however errors in several ncRNA expressions are also responsible. Despite the fact that the proof for connection between various human medical issues and *lncRNA* are increasing (Suravajhala et al., 2015) (Huang et al., 2021), ncRNA–protein connections have drawn huge interest for further studies and have become a new area for exploring causes of various diseases, especially cancers. These ncRNAs have also shown their associations in cancer (Statello et al., 2021). Recent studies on lncRNAs employed NGS for identifying novel targets (H. L. V. Wang & Chekanova, 2019). A systematic analysis is done on genes involved in subcutaneous adipose tissue from the thigh of AIT2DM patients followed by phenome-interactome map preparation. Several new pathways and genes along with those of non-coding RNAs involved in AIT2DM were found. They have been explained to be in relation with some pathways like NOTCH, tyrosine kinase JAK2, thus recruitment of *Src* and *STAT5* family kinases (signalling molecules) may assist as biomarkers on cell surface (Saxena et al., 2021). As already mentioned, among men PCa has become the most frequent cancer type, but still the approach of NGS studies were not in use before times. Lately, this approach has reported the identification of genetic biomarkers in Caucasians, but such studies are very rarely done on Indian PCa genetic variants (Gupta et al., 2020).

Numerous cellular processes depend on lncRNAs interactions with more than one RNA-binding protein (RBPs), which are frequently able to bind a variety of RNAs. Very few lncRNA are characterized fully functionally. Some of them remain attached to sites of transcription and regulate *cis* gene expression while the remaining acts as decoys as they prevent the transcription factors (TF) from interacting with DNA by binding with TF themselves. LncRNAs can even show interaction with chromatin modifying complexes to direct them to *transgenic* targets. Further, these can work like sponges for microRNAs

(miRNAs), help enhancers by binding them, can regulate antisense mRNAs after transcription or could be a scaffold for the formation of macromolecular complexes. Many RBPs can be cancer related proteins like MSL, DNA methyltransferase, trithorax-group proteins etc. and could interact with lncRNAs. The RBPs expression level in tumors fluctuates significantly which indicates their potential for diagnosis of PCa or any other cancer type. Metastasis associated lung adenocarcinoma transcript 1, *MALAT1*, is the first lncRNA to be upregulated and associated with PCa and other cancer types, as its interaction with the SR family of proteins (Ferrè et al., 2016). A near 2% of RNAs are coding, the rest of them are non-coding, called long non-coding RNAs, lncRNAs. These are RNA polymerase II synthesized, caped, polyadenylated and less than 200 nt. in size. Lately, these are well known for playing a role in several cellular processes like cell differentiation, metabolism, cell cycle, signal transduction, transcription, translation and post-translations, epigenetics, receptor interactions, DNA modifications, and many diseases including infections (Bridges et al., 2021).

2.13. LncRNA and PCa

2.13.1. LncRNA and Cancer:

To understand the role of lncRNAs in cancer, first, their expression, functions and structures need to be understood. Both epigenetic and transcriptional factors regulate the lncRNA expressions. Genes of lncRNA are highly methylated in comparison to any other protein coding gene which shows epigenetic control. LncRNAs are tissue specific even though they are expressed in lower levels. Genetic or epigenetic mutation in their genes causes deregulations and transforms the cell to initiate cancer progression (G. Yang et al., 2014). Besides this, the more the underlying roles of lncRNAs are being exposed, the more their importance and potential in cancer treatment is realized since they be involved in cell cycle regulation, epigenetic regulation, miRNA regulation, DNA damages, chromosomal instability, and tumor induction related signaling pathways. These usually bind or interact with DNAs, Proteins and other RNAs; and their location, sequence length and secondary structures also matter to show their effect in tumor development. LncRNAs binds as decoy to different transcriptional regulatory sites (showing *cis* and *trans* function) and other translational factors for example a novel cancer associated lncRNA SNHG9 interacts with LATS1 C-terminal facilitating liquid-liquid phase separation. Since, lncRNAs have also known to interfere in chemotherapy resistance for drugs by regulating expressions of target gene involved in drug efflux, autophagy, apoptosis, cancer stem cells and EMT, suggesting a new potential way to treat the cancers and as potential biomarkers of cancers. However, lncRNAs and their therapeutic importance is still less explored with respect to cancers despite

their high stability in body fluids which can make both treatment and diagnosis feasible (X. Yin et al., 2022) (X. Zhou et al., 2022).

Studies have also suggested that the lncRNA can maintain the stemness of the cancer cells, for instance, the *LUNAR1* which is an lncRNA and regulated by notch signalling which directs the stem cells of leukemia and T-cells of acute lymphoblastic leukemia. Recently, the PRC complexes have been identified as promoting stemness and interacting with lncRNAs like ANRIL and *HOTAIR*. In aggressive PCa, PCAT-1 is reported as a highly upregulated lncRNA. Since lncRNA can be easily detected from body fluids and can regulate DNA, RNA, protein function and their synthesis, they are promising markers for diagnosis such as *PCA3*. Further, several studies have proved their importance in treatment for instance, lncRNAs are regulators of epigenetics, thus shutting down such lncRNAs which epigenetically regulate any cancerous gene expression (X. Yin et al., 2022) (Zhou et al., 2022). Recently a study used variational graph auto-encoder (VGAELDA) algorithms to find associations between lncRNAs and metastatic events and showed few novel lncRNAs are specifically regulating the metastatic events (cell invasion and cell migration) in different cancers. For instance lncRNA-MEG3 (breast, lung, cervical, colorectal), SNHG16 (breast, bladder, renal, gastric, ovarian, retinoblastoma, prostate), MALAT1 (lung), and MT1JP (cervical, colorectal, gallbladder, melanoma, thyroid, ovarian, prostate, esophageal, gallbladder) are some of them (Zhu et al., 2023)

2.13.2. PCa and LncRNA Biomarkers:

Recently, a multitude of novel lncRNAs dysregulated in PCa, have been identified by large scale RNA profiling projects. RNA sequencing identified a set of 121 PCa-associated intergenic non-coding RNA transcripts termed the PCAT family (Khilwani et al., 2023 Unpublished). Functional analyses of lncRNAs have revealed significant contributions to PCa by targeting relevant pathways and gene regulation mechanisms including PTEN/AKT and androgen receptor signalling as well as chromatin remodelling complexes (Aird et al., 2018). Among all the identified biomarkers of PCa, PCA3 showed the highest specificity in diagnosis as a biomarker and it expresses 60-100 times more in over 95% of the PCa cases. PCA3 is known to regulate apoptotic genes, angiogenesis, cell-cell adhesions, and mitogen-activated kinase kinase-1 and signal transductions. PCA3 shows more specificity in diagnosis when combined with a gene fusion testing of TMPRSS2-ERG (Misawa et al., 2017). PCa antigen 3 (PCA3) modulates PCa cell survival via modulating AR signalling and is now used in PCa diagnosis (Lemos et al., 2019). SChLAP1 (Second chromosome locus associated with Prostate-1) lncRNA expresses in CRPCs and is associated with biochemical recurrence risk,

metastasis, lethal progression and PCa related mortality. It is expressed in 25% of the PCa cases, predicts recurrence after radical prostatectomy and is known to interact with a chromatin remodelling complex called Switch-Sucrose Non-Fermentable. SPRY4-IT1 (SPRY4 intronic transcript 1) was markedly elevated in both patients and PC3 cell lines. SPRY4-IT1 plays a role in invasion, proliferation and apoptosis. MALAT1 is known as a poor prognostic lncRNA biomarker as it is found expressed in other types of cancers like lung, breast, colon, liver and pancreas, suggesting its role in promoting tumor growth and metastasis. Further, its high expression is associated with poor prognostic indicators like a high Gleason score suggesting an aggressive tumor stage, high PSA and high TNM. However, it is expressed mostly in CRCP patients and binds with a gene regulation protein called Enhancer of Zeste Homolog 2 (EZH2) to enhance the migration, invasions, and aggressiveness to bring metastasis turning out to be a poor prognostic factor. Another poor prognostic biomarker for PCa is TRPM2-AS (Transient receptor potential cation, subfamily M, member2-antisense transcript). It is associated with PCa cell apoptosis and is a known therapeutic target. *NEAT1* (Nuclear enriched abundant transcript) has shown resistance to antagonists of AR or androgen deprivation therapy. NEAT1 changes the epigenetic pattern by binding to the promoter region of the gene so that transcription initiates and also starts oncogenic growth. Its high expression is an indicator of early recurrence by raising biochemicals like PSA and metastasis. Around 50% of PCa tumors have elevated levels of *PCGEM1* lncRNA (PCa gene expression marker1). It has a role in the proliferation of cancer cells by regulating c-Myc and the formation of colonies. Another c-*Myc* interacting lncRNA is *PCAT1* which has a repressive impact on several genes like *BRCA2* and expresses highly in localized high grade tumors and metastatic PCa (Misawa et al., 2017).

Previous studies have also identified *GAS5*, *HULC* and *UCA1* as three lncRNAs linked with PCa radiosensitivity. A study compared TCGA datasets and lncRNAs expressions of both complete and incomplete radioresponse patients. They found that LINC01600 was highly upregulated in incomplete radio response patients. LINC01600 was also earlier reported overexpressed in adenocarcinomas of the lung and about 558 coding genes (involved in DNA repair, cell cycle, metabolism and radiosensitivity related) were showing co-expression with it (Rout et al., 2022). Another study by Iris et al., 2020 reported upregulated LINC00665 and LINC0026 in patients after radiotherapy. LINC00665 and LINC0026 play a role in survival and DNA repair mechanisms, suggesting them as potential markers of PCa (Eke et al., 2021) (Iris et al., 2020). Hence, the important role of lncRNAs has been identified in the development of PCa, promotion of castration-resistant PCa (CRPC), cell proliferation,

invasion, metastatic spread along with modulation of AR-mediated signalling (Yang et al., 2021). Lately, LincRNA-p21 expression, a biomarker present in urine, can be used to differentiate between a BPH and PCa. It was observed that the circulating lncRNAs can be more helpful in diagnosis of PCa (Beylerli et al., 2022).

2.14. LncRNA and Diabetes

Considering the dynamic role of lncRNAs as novel prognostic, diagnostic and predictive markers in PCa, lncRNAs may also serve as therapeutic targets aiding in the prevention, development and treatment of CRPC and metastasis of the disease. Functional analysis of lncRNAs could be done by deciphering lncRNA-protein interaction, as the function of most lncRNAs is dependent on interaction with protein-coding genes. In contrast to the limited experimental approaches available, lncRNA-protein interactions can be more effectively studied by employing different computational tools. It will be interesting to predict the lncRNA that is common in PCa and AR which can enable us to develop potential lncRNAs that are prognostic biomarkers. Epigenetic mechanisms that control intricate interactions between different genes and the environmental variables while retaining the sequence of a DNA include methylation, posttranslational modification of histones and lncRNAs (Sathishkumar et al., 2018). As it is well known that lncRNAs play a role in several molecular processes and lately an increasing amount of research has indicated that lncRNAs are important regulators of insulin secretion and resistance, functioning of beta cells, apoptosis, and metabolism of glucose. As a result, many investigations have observed the shifts in the expression of lncRNAs among DM patients and animal models of both type-1 and type-2-diabetes. LncRNAs are therefore believed to offer novel biomarkers with intriguing potential for the diagnosis of both type1 and type-2-diabetes. For instance, a study found the level of GAS5 lncRNA falls in T2DM patients of the US cohort and vice versa. Having low serum levels of S5 was at more risk of acquiring T2DM in future. Another study on cohorts from testified upregulation of ENST00000550337.1 in the blood could be highly promising in early diagnosis of early T2DM (Dieter et al., 2021).

Dieter et al. (2021), in a review explained that the levels of some lncRNAs (*Hotair, Malat1, Kncq1ot1, Miat and Anril*) were upregulated consistently, whereas the level of *Meg3 was downregulated.* Malat1, also a predictor of metastatic lung cancer, is upregulated in macrophages, serum and liver of D2TM animal models (J. Y. Liu et al., 2014) (S. X. Liu et al., 2019). Further, Malat1 has been seen to interact with *PI3K/Akt, MAPK/ERK, Wnt/β*-catenin, *NF-κB* insulin, *AMPK* and *FoxO* signalling related to DM (Dong et al., 2014).

Hotair is hyper expressed in blood and liver of T2DM and diabetic kidney disease. It interacts with *MAPK, TNF, FoxO, Akt and HIF-1* signalling (Shaker et al., 2019).

*Anril i*s a known predictor of many cancer types like liver, thyroid, colon, breast, lung and gliomas. It also interacts with TGFβ, MAPK, FoxO, AGE/RACE and PI3K/Akt signalling (Dieter et al., 2021). An antisense lncRNA, *Kncq1ot1,* regulates the development of beta cells and stress related protein processing in the endoplasmic reticulum (Dieter et al., 2021) (Kassem et al., 2001). High expression of *Miat* was found in rats with diabetic myocardium. However, the down regulation of *Meg3* was seen in the blood, serum and Islets of DM mice models. It is found to be involved in insulin synthesis by suppressing insulin gene transcription (D. Zhang et al., 2018). Sathiskumar *et al.,* also reported *XIST* (diabetic neuropathy), *LET* (upregulated in DM), *RNCR3I* (increases vasculature of the retina and upregulated), *THRIL* (controls TNF-alpha), *GM4419* (diabetic neuropathy), *SALRNA1* (downregulated in DM and negatively related to hyperglycemia, inflammation and senescence), *NBR2* and *PANDA* (AMPK signalling and liver fibrosis), *PLUTO* (upregulated in pre-diabetic and DM) and *LincRNA-p21* with varying expression levels because of mutation in DM patients (Sathishkumar et al., 2018) (Goyal et al., 2018). In high glucose induced inflammatory response in the case of diabetic retinopathy, lncRNA *H19* is also found to play an important role through the H19/miR-19b/SIRT1 axis (Luo et al., 2021). Differentially expressed lncRNA (LINC01128) is already known to increase the rate of cervical cancer progression and is also predicted as a biomarker of gestational hypertension. Similarly, Saxena et al. (2021), suggested that LINC01128 could serve as a biomarker for diabetes diagnosis and prognosis (Rout et al., 2022).

LINC01128 Thigh Adipose Tissue and identified LINC01128 to be relatively expressed (Tiwari et al., 2019). Further studies on GEO datasets of osteosarcoma have shown that LINC01128 expresses its oncogenic role. It revealed that increased expression of LINC01128 in OS patients is accompanied by their shorter survival. However, its knockdown turned down the proliferation, migration, and invasion. In OS, LINC01128 is identified to work as a sponge in triggering Wnt/β-Catenin signalling by promoting *MMP2* expression through miR-299-3p (Yao & Chen, 2020). In cervical cancer tissues, the expression of LINC01128 is found to be significantly high, and its fall suggests that it might lower the SFN (stratifin) at both the mRNA and protein levels. SFN, a known potential biomarker in cervical cancer, is also majorly expressed in the early stage of lung adenocarcinomas. It clearly explains how LINC01128 could accelerate cell processes like cell proliferation, migration, and invasion and even can inhibit apoptosis through SFN upregulation and release by binding miR-383-5p and also working as its antagonist (Rout et al., 2022). MiR-383 is under the regulation of

LINC01128. However, overexpression of miR-383 in T2DM serum reverses the cell apoptosis under high glucose in mouse β cells by *TLR4* and *APOC3* suppression (X. Cheng et al., 2020). Also, high LINC01128 was seen in stage III-IV CRC and mediated PRMT5 function, which is a mediator of methylation of proteins. In pancreatic cancer, it was found as an *EMT-LPS* (epithelial mesenchymal transition related lncRNA prognostic signature) molecule (Bo Cao et al., 2021) (Deng et al., 2022).

2.15. Research Gaps

Scientists have yet to identify the reasons behind variations in outcomes linking type-2 diabetes and PCa, which appears complex across different stages. Glucose transporters are crucial for PCa development since glucose assimilation in the prostate gland is unusual as compared to other tissues. Therefore, exploring glucose metabolism elements, including enzymes, metabolites, or other transporters has become an important area for more in-depth studies. Also, the behavioural understandings of the metabolic interchanging occurring between immune cells, tumor cells and stromal cells of the microenvironment in tumors are still blurry. The two conditions, neoplastic transformation and hyperglycaemia, often co-occur and their interaction complicates treatment by limiting drug selection. Thus, a shortage of understanding of these co-occurring conditions adds to greater death rates. While some pathways linking glucose and insulin metabolism to PCa have been proposed, the impact of varying extracellular degrees of glucose on malignant cells is weakly stated till now. However, the matter of concern is will the answers for metabolism-related medical problems decrease the chance of having PCa before it is diagnosed? Current research is focused on finding if a molecular subtype interaction with the metabolic environment can help in affecting the outcome of PCa mortality. Although the role of insulin is contributing, how exactly it is affecting is still unclear. Further, age significantly influences PCa risk, with the highest risk is seen in diabetic patients with 1–3 years of diabetes, after which progressively it drops. This suggests an undefined cause for this non-linearity. The reason behind the variation in the impact of geographical and socioeconomic is also demanding further detection. Though the associations between both of them are evident, more research is demanding to come out with some constant inferences and potential mechanisms. Future analyses should include the detection of more biomarkers for prostate neoplasm, especially since NGS has already identified some genetic variants in Caucasians but Asian India has not been explored yet.

2.16. Aim and Objectives

2.16.1. Hypothesis

Several confounders like obesity, drugs for treatment, metabolic control, diet, etc. in diabetic patients are likely to influence the link between cancer and diabetes. The exact incidence of association is still lacking except for the PCa case, wherein a decrease in incidence rate has been noted among diabetic men. Diabetes itself is underdiagnosed in adults (3-5%), so even if we believe that DM is increasing the chances of cancer, this association is still underrated. Moreover, whether diabetes via a different mechanism leads to cancer development aggressively or is it the host body that is not able to show resistance towards cancer development is not answered yet (Vigneri et al., 2009). Obesity is also recently considered as a risk factor for inducing PCa by crosstalk of oxidative stress caused by insulin resistance and inflammation caused by increase in circulating inflammatory factors (*IL-6, TNF* and leptin) as a consequence. Some studies say 5% per every increase of 5 kg/m^2 BMI; risk for PCa is raised in an obese person. In case of PCa, although diagnosis at early stage (neoplasias forms) in diabetic patients reduces the cancer incidence, but once the patient becomes overweight and insulin resistant, the chances of the patient ending his life with PCa will increase. However, diabetes has not only increased the relative risk of incidence but has also increased the mortality more drastically in breast cancer and colorectal cancer (Vigneri et al., 2009).

From studying literature intensively, it can be assumed that the AR mediates the connection between the PCa and Diabetes. In PCa it plays an important role in growth and differentiation of healthy prostate. Hence, PCa is dependent on androgen stimulation and AR mediates this stimulation. Even the fall in the testosterone could be related to hyper expression of ARs ultimately ending in tumor development. These testosterone (androgens) are also found low in in many DM patients and on other side fall in testosterone being an androgen when binds to AR, it stimulates the AR signaling in liver cells which leads to GSIS. This suggests its link with high insulin resistance. Further, this explains on AR therapy to reduce AR and GSIS, chances of having the diabetes is high indicating a risk for post AR therapy diabetes development. This elevating signaling is also supposed to be the result of a mutated IGF/insulin receptor and its contribution in poor prognosis of diabetic PCa patients. Few studies have also reported the elevated AR signaling in diabetic PCa patients. This highlights a potential interplay of the complex hormonal signaling pathway of two conditions which motivates this to explore more explicit associations at genetic level.

Metformin has shown that the beyond glucose metabolism but interferes with AR. In few cases even it is observed to stabilize the PSA level presenting PCa management property. Analysis has also inferred that the fall in PSA level diabetic patients who are taking

antidiabetic and obese patients, has an inverse relationship between diabetes, obesity and PSA levels. Therefore, while screening for diabetes and obesity PSA level should also be considered. Metformin mainly works two ways first; it reduces the production of glucose by blocking one step of an aerobic cellular energy production molecule called ATP via activating AMPK enzyme. The other way of metformin works independent of *AMPK*, where *ATM* is assumed to modulate the action of metformin by some other parallel pathway. It has been observed in some patients with ataxia telangiectasia who were having insulin resistance. It was indicated that the *ATM* could regulate *AMPK* by phosphorylating its upstream activator, LKB1 to affect the cellular process or it can regulate the *AMPK* independent of LKB1. It regulates the other targets of metformin independent of *AMPK* (Birnbaum & Shaw 2011) (**Figure: 2.7)**.

Antidiabetic therapies or metformin can certainly block cell progression by halting *mTOR*. The *mTOR* can stimulate glycolysis on getting phosphorylated from *AMPK* under AR dependent hyperglycemia and *PI3K/Akt* signaling pathway. In the *PI3K/Akt* pathway active mTOR can bring angiogenesis and cell survival. MF can inhibit *PI3K*, Akt directly and mTOR (**Figure: 2.7)**. Further, mTOR regulates the insulin secretion from Beta cells and leads to hyperinsulinemia and insulin resistance. This signaling activates mitogens *RAS-MAPK* and *PI3KB* through *IGF*-1 and which can induce cell proliferation, invasion and metastasis. Here, Metformin in many studies is found to inhibit hyperinsulinemia, *IGF-1* and *PI3KB* directly.

DM is not just a single disease but it is a cluster of metabolic ailments because of either missing insulin or low levels of insulin causing abnormal glucose level and its tolerance (Deepthi et al., 2017). In general, lack of any enzymes which are required to turn one metabolite into another results in a metabolic affliction in the body. Other ways, metabolite over accumulation can also manifest a metabolic disorder and hence diabetes fits a better example which is lack of enzyme and over accumulation of glucose (Nagamani et al., 2019). This indicates more has to be explored in connection of metabolic disorders and PCa.

On one side we could see it is lowering the chances of PCa but on the opposite side it is supposed to be responsible for PCa related deaths. So it is confirmed that the associations are very complex since the understanding of this association can be considered insufficient and obesity is assumed to complicate more pathophysiology.

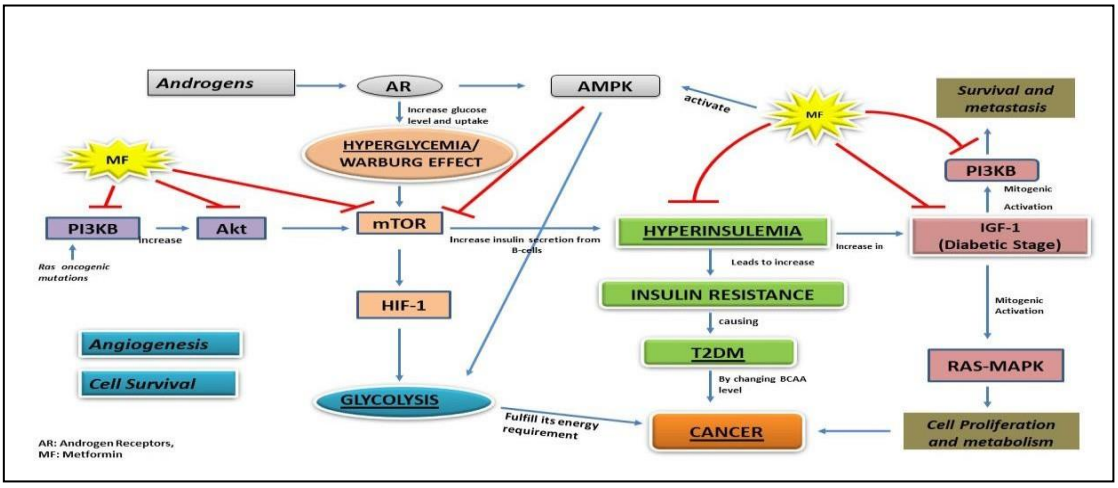


Figure: 2.7: Showing the interplay of hyperglycemia, insulin resistance, hyperinsulemia, AR signaling, and anticancer Metformin mode of action. Apart from the antidiabetic effect metformin contributes and anticancer effect through interfering in glucose metabolism pathway as it directly blocks PI3K, Akt, AMPK and mTOR to reduce glucose production (glycolysis) and energy production for cell progression, angiogenesis and cell survival. Metformin also puts checks on hyperinsulinemia and insulin resistance by blocking PI3K and IGF-1 to halt the RAS-MAPK activation. Hyperinsulinemia and insulin resistance are observed to play role in cancer development, cell proliferation and metastasis. **AR** (Androgen receptor), **PI3KB** (Phosphoinositide 3-kinase beta protein), **AMPK** (Adenosine monophosphate-activated protein kinase), **Akt** (Ak strain transforming), **mTOR** (Mammalian target of rapamycin), **HIF**-1 (Hypoxia inducing factor-1), **IGF1** (Insulin growth factor),**RAS-MAPK** (Ras/mitogen-activated protein kinase), **MF** (Metformin), **T2DM** (Type-2- diabetes mellitus)

Further, glucose transporters are taken as appropriate for PCa development since glucose assimilation in the prostate gland is unusual as compared to other differentiated tissues. So therefore, utilization of elements of glucose metabolism, together with enzymes, metabolites, or other transporters for typifying carcinomas has become an important area which requires more attention which must be further exploited and thoroughly considered. Additional knowledge of the metabolic dependencies that enable tumor development in the local surroundings will be significant for new findings. The two conditions neoplastic transformation and hyperglycemia are repeatedly present in co-morbidity; moreover their interaction has made treatment system further difficult by limiting the drug selection. Thus shortage of clarification in knowledge regarding these co-occurring medical issues adds to greater death rates. Based on very limited research many pathways were presented to make

clear the potential integrative associations in glucose and insulin metabolism and PCa but the impact of varying extracellular degrees of glucose on malignant cells is weakly stated till now. Although, the role of insulin is confirmed but how exactly it is affecting is still unclear. Whereas the growth-encouraging role of insulin via canonical mTOR/Akt/ ribosomal protein S6 kinase signaling in malignant cells has undoubtedly shown contribution in cancer development but the possible straight away metabolic role of insulin in malignant cells is still under the soil. Though, the associations between both of them are evident but more research is demanding to come out with some constant inferences and potential mechanisms could be drawn regarding an association between PCa and metabolic abnormalities. Future analyses should include the detection of more biomarkers for prostate neoplasm. Next generation sequencing technology has already identified some genetic variants in Caucasians but Asian India has not been explored yet.

Hence, the present study is aimed to decipher the known potential targets and identify the splice variants of the non-coding RNA players such as lncRNAs of modulation in tumor metabolism that cross talk between PCa and diabetes.

2.16.2. Aim:

In Silico Analysis for Genomics Understanding to Ascertain Impact of Diabetes risk in PCa

2.16.3. Objectives:

1. To screen variants in PCa patients affected with diabetes, thereby establishing a causal relationship between them using integrated systems genomic approaches
2. To identify and explore the role of lncRNAs as potential targets using lncRNA-protein interactions
3. To perform protein-protein interaction (PPI) network analyses and identify pathways associated with T2D and PCa *in silico* method

CHAPTER 3: METHODOLOGY

3.1. To screen variants in PCa patients affected with diabetes, thereby establishing a causal relationship between them using integrated systems genomic approaches

3.1.1. Integrated machine learning approaches on Publicly Available Dataset for identification of Plausible Candidate

3.1.1.1. NCBI dbSNP

The dbSNP stands for single nucleotide polymorphism database which is a publically available repository of gene variations data at NCBI (National Center for Biotechnology Information) (https://www.ncbi.nlm.nih.gov/snp/ last accessed on April 2, 2024). As suggested by its name, it includes all the changes at nucleotides as well as other small insertions or deletions and short tandem repeats (Bhagwat, 2010).

This helps to match the variations and handles more than 2 billion of variants of a human each with different submission IDs (Bradley Holmes et al., 2020). When submitted SNP matches the same location as the genome, they are then reported to Reference SNP (refSNP) and receive an ID number starting with "rs". We need to put query in search box of dbSNP in the form of text and then multiple filters (like organism, SNP type, and functional aspects) can be selected, which will be added on by Boolean operators "AND", "OR" or "NOT" (Bhagwat, 2010). It links all the variations of clinical mutations and polymorphisms with several other databases like sequence resources through E-PCR and BLAST, literature with their time of submissions or PubMed, GenBank, Entrez and Locus Link (Sherry et al., 2001).

3.1.1.2. ClinVar

National Institute of Health (NIH) has maintained publically available record of all the clinically validated variants (ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/ last accessed on April 12, 2024) (Harrison et al., 2016). More than 1300 different organizations which include several laboratories (for clinical tests and research), clinicians, LSDBs (locus-specific databases), panel experts, registries of patients and many others submit their identified variants to ClinVar (Landrum et al., 2020). Further, each submission has the record of conditions associated with it and disease for which it has been identified, its clinical significance, status of review and evidence of identification (Harrison et al., 2016). The

scope of ClinVar is restricted to only those variations that were thoroughly evaluated for their both functional and clinical significance but not just discovered (Landrum et al., 2020).

We went to the website of the ClinVar and a home page was displayed with a search box at the top for our queries. We provided the variants with rs IDs as a query and hit the search option. From the homepage we were directed to the page where both variation report and record report were displayed with filter options available on the left side for more refinement. We selected both pathogenic and likely pathogenic for clinical significance and then filtered for only RefSeq variants. Later, we download the searched variant results as tab delimited variant summary file (Harrison et al., 2016).

3.1.1.3. Dataset

Separate catalogs of clinically verified variants (ClinVar) were first prepared for PCa, diabetes, and obesity from the NCBI with searches using keywords, *"Prostate Cancer", "Diabetes", "Obesity", and* Boolean expressions, *viz.* AND, OR, NOT were used wherever needed. Also, we surveyed the literature and identified the associated clinical parameters of comorbidities leading to cancer progression, especially for the prostate malignancy was carefully chosen. The ClinVar datasets were then further categorized into unknown significance, likely benign, benign, likely pathogenic and pathogenic based on their clinical significance and similarly the clinical parameters were categorized based on threshold values. Later, the annotated data for all three diseases is transformed into binary/semi-binary scores, *viz*.-1, -0.5, 0, 0.5, and 1 (**Figure 3.1) and (Figure 3.2).** While these binary scores are given based on the risk of acquiring PCa increases with rising in the pathogenicity of a variant, the general stratification of low grade, intermediate but less risk, intermediate, moderate risk, and high risk corresponds with -1, -0.5, 0, 0.5, and 1 respectively were checked and tabulated. The training model was prepared from the converted binary/semi-binary datasets and then subjected to seven different machine learning algorithms that are linear regression, multilayer perceptron, simple linear regression, random forest, random tree, REP tree, and Zero R and further compared to predict the chance of developing PCa when there are any change variables from the control population (Kour et al., 2023).

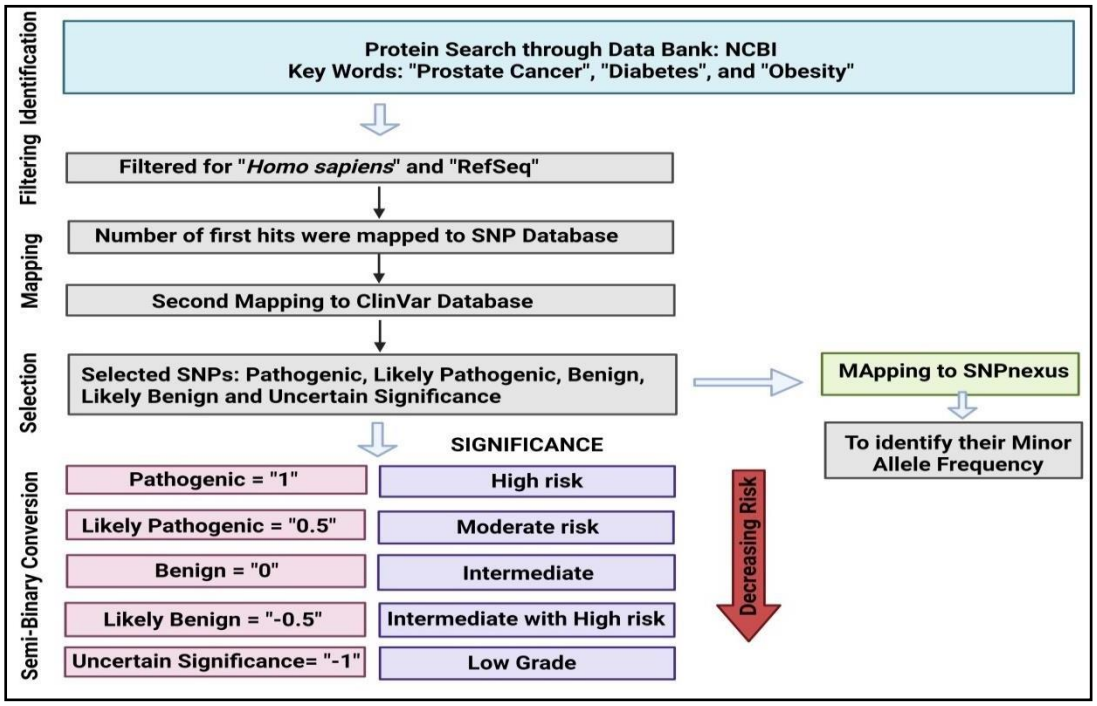


Figure 3.1: Flow chart representing identification, filtering, mapping (to SNP database, ClinVar, and SNPnexus), selecting, and finally, semi-binary data conversion of (1, 0.5, 0, -0.5, -1) with respective significance.

3.1.1.4. Clinical parameters of PCa:

3.1.1.4.1. Gleason Grading:

Originally, Gleason grading was based on anatomical patterns seen in hematoxylin and eosin (H and E) stained sections of prostate adenocarcinoma, instead of cellular characteristics. In that system, pattern 1 was regarded as a well-confined lump consisting of somewhat even, tightly packed, separate, well distinguished and moderate-sized glands whereas pattern 2 showed several variations in neoplastic glands size with increased stroma in them and irregularities in their lump's circumference. In pattern 3 some Polyporus form glands in gland structures called glomerulations while fuse glands were seen in pattern 4. A blemish outgrowth demarcated it as pattern 5 where solid cord growth and tumor cell infiltration were seen later (N. Chen & Zhou, 2016). Modern biopsies approaches demanded more advances in Gleason grading to interpret and score biopsies. Hence, on the new grading system, Gleason's score less than or equal to 6 is categorized under grade group 1, a score of 3+4=7 under group 2, a score of 4+3=7 under group 3, a score of 4+4=8 under group 4, and a score of 9 -10 under grade 5. This latest advanced grading system is incorporated as a new addition to the World Health Organization classification for prostate tumor. Scoring is the sum of highly frequent common and secondary patterns 3+5 (Bulten et al., 2020). As per current developments in

immunohistochemistry, grades 1, 2, and 3 are almost identified as similar, hence are not much considered. However, in grade five, an interface between grade 3 or 4 and 6 or 7 is marginal between low risk and high risk of cancer (Iczkowski & Lucia, 2011) **(Figure 3.2.A).**

3.1.1.4.2. Prostate-Specific Antigen (PSA):

PSA is used in PCa screening at earlier stages to reduce the overall mortality rate but specific mortality and improve treatment approaches. However, PSA screening has still not shown remarkable results in saving patient lives. Besides this, PSA screening is accompanied by several problems like over diagnosis (false positive or false negative) which can lead to prolonged side effects of treatments (Velonas et al., 2013) **(Figure 3.2.A)**.

3.1.1.4.3. Digital Rectal examination (DRE):

Digital rectal examination was largely used for earlier diagnosis of PCa before the advent of PSA. However, DRE can only detect several tumors because of its inefficacy in correlating location nodules with tumor locations in biopsy results. Therefore, these days DRE tests are less recommended in routine PCa screenings **(Figure 3.2.A)** (Okotie et al., 2007).

A. Clinical parameters of Prostate cancer

Clinical Significances of a Gene	Binary Scoring
Uncertain Significance	-1
Likely Benign	-0.5
Benign	0
Likely Pathogenic	0.5
Pathogenic	1

a. Clinical Significance

Threshold Values	Significance	Binary Scoring
0-2.5ng/ml	Safe	-1
2.6-4ng/ml	Safe but risky	-0.5
4-10ng/ml	Suspicious/Low risk	0
10-20ng/ml	Intermediate Risk	0.5
above 20ng/ml	High risk	1

b. PSA level

Grading Scores (0-10)	Staging	Binary Score
Group1 (0-6)	Low grade	-1
Group2 (3+4=7)	Intermediate with less risk	-0.5
Group3 (4+3=7)	Intermediate with more favourable	0
Group 4 (8)	Moderate risk	0.5
Group5 (9-10)	High risk	1

c. Gleason Grading groups

Clinical Significances of a Gene	Frequency Threshold values	Binary Score
Uncertain Significance	-	-
Likely Benign	-	-
Benign	-	-
Likely Pathogenic	>0.05	0
Pathogenic	<0.05	1

d. Minor allele frequency

DRE Physical Inferences	Binary Score
-	-
-	-
Only one side	0.5
Both sides	0
Spread beyond prostate	1

e. DRE

B. Clinical parameters of diabetes associated with Cancer/Prostate cancer

Clinical Significances of a Gene	Binary Scoring
Uncertain Significance	-1
Likely Benign	-0.5
Benign	0
Likely Pathogenic	0.5
Pathogenic	1

a. Clinical Significance

Clinical Significances of a Gene	Frequency Threshold values	Binary Score
Uncertain Significance	-	-
Likely Benign	-	-
Benign	-	-
Likely Pathogenic	>0.05	0
Pathogenic	<0.05	1

b. Minor allele frequency

Concentration Level	Significance	Binary Score
-	-	-
-	-	-
≥5.7, <6.5	Prediabetes (Likely Benign)	0
≥6.5%	Diabetes (Benign)	1

c. Glycated hemoglobin (HbA1c)

WBC Count Testing		
Range	Interpretations	Scoring
<4000 cells/microlitre	Lowest	0
4,500-11000 cells/microlitre	Normal	0.5
Above 11000 with every 1000cells/mm3	Diabeties risk inrease by 75%	1

d. WBC Count Testing

Blood Glucose Range	Interpretations	Scoring
<70mg/dl	Hypoglycaemia	0
84-150mg/dl	Normoglycemia	0.5
250-400mg/dl	Hyperglycaemia	1

e. Fasting Blood Glucose Testing

Range Kg/m2	Significance	Score
18.5-24.9 kg/m2	Healthy	0
≤ 25.0	Overweight	0.5
≥30kg/m2	Lifetime increased risk of diabetes	1

f. BMI

C. Clinical parameters of obesity associated with cancer/prostate cancer

Clinical Significances of a Gene	Binary Scoring
Uncertain Significance	-1
Likely Benign	-0.5
Benign	0
Likely Pathogenic	0.5
Pathogenic	1

a. Clinical Significance

Clinical Significances of a Gene	Frequency Threshold values	Binary Score
Uncertain Significance	-	-
Likely Benign	-	-
Benign	-	-
Likely Pathogenic	>0.05	0
Pathogenic	<0.05	1

b. Minor allele frequency

Clinical Significances of a Gene	Range Kg/m2	Significance	Score
Uncertain Significance	-	-	-
Likely Benign	-	-	-
Benign	18.5-24.9 kg/m2	Healthy	0
Likely Pathogenic	≤ 25.0	Overweight	0.5
Pathogenic	≥30kg/m2	Lifetime increased risk of diabetes	1

e. BMI

FAT type	Threshold Levels	Obesity Risk	PCa Risk	Scoring
Total Cholesterol	125 to 200mg/dL	-	-	-
Non-HDL	Less than 130mg/dL	High risk of obesity	Increased risk of cancer	0.5
LDL	Less than 100mg/dL	Obesity	Increased risk of cancer	1
HDL	40mg/dL or higher	Non-obese	Lower risk of cancer	0

d. HDL/LDL

Figure 3.2: Different clinical parameters of **A**. PCa (Clinical Significance, PSA Level, Gleason Grading, Minor Allele Frequency, DRE Value), **B**. Different clinical parameters of Diabetes (Clinical Significance, Minor Allele Frequency, Glycated Hemoglobin: HbA1c,

WBC Count Test, Fasting Blood Glucose, Body Mass Index); **C**. Different clinical parameters of Obesity (Clinical Significance, Minor Allele Frequency, Body Mass Index, LDL/HDL) for PCa, and their different ranges based on risk, and their conversion into semi-binary data (-1, -0.5, 0, 0.5, 1).

3.1.1.5. Clinical parameters of diabetes associated with cancer/PCa:

3.1.1.5.1. Glycated hemoglobin (HbA1c):

Glycated hemoglobin (HbA1c) testing is considered a gold standard for evaluating glycemic control in diabetic patients. It gives the average estimation of plasma glucose (Y. Hu et al., 2010) (Schnedl et al., 2001). A high level of HbA1c is interlinked with a chance of having hepatocellular carcinoma (HCC) among pre-existing diabetic patients. With every rise of 1% HbA1c level, the possibility of having HCC elevates by 26-50%. In insulin resistance (IR) in T2DM patients, because of prolonged use of antidiabetic therapies, exposure to free circulating insulin increases, and cellular mitosis gets stimulated by the insulin growth factor (IGF-1) intracellular pathway, a key mitogenic and antiapoptotic trigger in cancer development (Donadon et al., 2010) (Kour et al., 2023). A threshold value of 6.1% is the optimum sensitivity and specificity and 6.5% is the finest specificity to diagnose diabetes as indexed in American Diabetes Association (ADA) recommendations (P. R. Kumar et al., 2010) (Lippi & Targher, 2010)**.** Its limitations include its association with poor performance in pregnant females, old age, and the chance of overdoing in anemia and genetically predisposed ones (Kour et al., 2023). A study under UK Biobank has also suggested that high HbA1c is associated with several types of cancer with increased risk for stomach, liver, colon, bladder, esophagous, lungs, endometrium, pancreas, and kidneys and decreased risk for PCa suggesting that diabetes and glycemic control is crucial in limiting cancer risk **(Figure 3.2.B)** (Peila & Rohan, 2020).

3.1.1.5.2. WBC count test:

WBC count testing alone can predict diabetes even in non-glycemic men. It has been estimated that for every 1000 cell/mm^3 rise within the normal range, the chance of diabetes rises by 7.6% (Twig et al., 2013). In addition, chorionic inflammation increases the likelihood of diabetes even without the obesity in autoimmune ailed patients. So, WBC count is considered an independent risk factor for diabetes in young (Vozarova et al., n.d.). A high WBC count is associated with an increased risk of venous thromboembolism (VTE) (arterial thrombosis and pulmonary embolism) in cancer patients. These cancer patients who

developed VTE showed a short life span compared with those who didn't develop VTE (Kour et al., 2023) (Blix et al., 2013)**(Figure 3.2.B)**.

3.1.1.5.3. Fasting Blood Glucose:

Fasting or exposure to a nutrient-deprived (fast mimicking diet, FMD) environment of cancer cells brings alteration in growth factors and metabolites which could lower the tendency of cancer to adapt and survive. This can be a possible way of refining the cancer treatment approaches (Nencioni et al., 2018). Some epidemiological studies have presented that T2DM has an inconsistent effect on the risk of PCa at different points in time. It was suggested in some cases that, over time, diabetes has shown a protective impact on PCa development because of poor serum levels or less availability and activity of IGF-1 in the late stages of T2DM. Fasting blood sugar testing is more reliable than HbA1c (Ghazanfari et al., 2010) **(Figure 3.2.B)**.

3.1.1.5.4. Body Mass Index (BMI):

Overweight or obesity in adults has shown considerable chances of acquiring diabetes in a lifetime. However, with aging its impact on the risk of diabetes, life span and period of diabetes will weaken. Adults are affected and have a higher chance of mortality due to diabetes if the BMI level is above or equal to 30kg/m^2 (Ghazanfari et al., 2010). A number of genetic variants are found common in GWAS which confirms their associations **(Figure 3.2.B)** (L. Cheng et al., 2019).

3.1.1.6. Clinical parameters of obesity-associated with cancer/PCa:

3.1.1.6.1. BMI (Body Mass Index)

Accumulation of adipose tissue in excess amounts as a result of a high intake of calories as compared to the energy expenditure of the body is considered obesity (Khandekar et al., 2011). It is quite evident that along with the risk of T2DM and cardiovascular diseases, several cancer types' risk is also directly proportional to increasing body weight. Their interlinking can be explained based on altered endogenous hormone metabolisms like insulin, IGF and steroids which deviate from living processes; cell proliferation, differentiation, and apoptosis from the normal equilibrium. Hence, checking on weight gain could significantly help in lowering cancer risk. A BMI of 18.5-25kg/m2 is suggested to escape from this risk, even as, in some studies, it is found that there is a high risk of cancer even in the range of 20-25kg/m2. Therefore, it has been highly advised to maintain weight in lower fields (Ancellin &

Bessette, 2013). A systemic pro-inflammatory environment caused by abdominal adiposity might initiate diabetes and cancer as internal metabolic alterations in combination with several environmental factors trigger various other processes in the body that are required in the initiation of tumor development (Khandekar et al., 2011) (Hopkins et al., 2016). If there is a family history of PCa there are more likely chances of having the same with increasing BMI. So, it has been indicated that BMI is one of the clinical factors that could predict PCa while biopsies (Liang et al., 2014). This is supported by a study clarifying that a higher BMI results in more mortality and moderate to high short duration annual changes in BMI linked with less mortality rate in any cancer **(Figure 3.2.C)** (Kour et al., 2023) (Taghizadeh et al., 2015).

3.1.1.6.2. LDL/HDL:

Several types of cancers, including aggressive PCa, are known to be caused because of obesity. Cholesterol is a known precursor of androgens which plays a key role in PCa development. Cholesterol-related comorbidity called hypercholesterolemia in association with obesity is a promoter of both tumor proliferation and inflammation. Serum cholesterol is related to PSA and ends up in a high rate of PSA-based biopsies and diagnosis resulting in high cholesterol in men. It can be said that high total serum cholesterol or HDL (high-density lipoprotein) is a risk factor for having a more aggressive form of PCa **(Figure 3.2.C)** (Jamnagerwalla et al., 2018).

3.1.1.7. Machine learning Algorithms:

We have used Waikato Environment for Knowledge Analysis (Weka) for implementing the machine learning algorithms on training model prepared earlier. The algorithms used for the study included Linear regression, Multilayer perceptron, Random forest, Random tree and REPTree for the regression analysis (**Table 3.1)**.

Table 3.1: List of different regression algorithms used for regression analysis

Algorithm	Details
Linear regression	This algorithm is used to model a linear relationship between multiple independent variables and the target variable using a linear function. The main objective is to minimize the sum of squared errors between the original values and the predicted values of the target variable. This algorithm differs from simple linear regression in the way that the latter handles only one input feature while the

	former is capable of handling multiple input features (Kour et al., 2023).
Multilayer perceptron (MLP)	MLP is an artificial neural network which can be used both for classification and regression tasks. It is used to model the non-linear relationships between the input features and the output/target variable. A typical MLP network contains an input layer, an output layer and a number of intermediate hidden layers. With the Weka tool, one can define several configurations of the network, while specifying the activation function, number of hidden layers, and number of nodes per hidden layer learning rate parameters (Kour et al., 2023).
Random forest	Random forest algorithm is an ensemble of decision trees each trained on a random set of features and random set of samples. This sampling procedure tends to minimize the over fitting and induces diversity in the ensemble. The random forest can be configured using WEKA in terms of number of trees, number of input features at each split, and tree depth parameters. It can also be used to model non-linear relationships between the input features and the target variable both for classification and regression tasks (Kour et al., 2023).
Random tree	This algorithm is used for both classification and regression tasks. It considers a set of decision trees each constructed using a subset of input features. The predictions from the individual trees are aggregated to generate the final predictions. As for random forest, WEKA provides parameters to configure the results of the algorithm (Kour et al., 2023).
REPTree	REPTree (Reduced Error Pruning Tree) is a decision tree algorithm that works by partitioning the input data by considering the feature that produces the best split at each node of the tree. After building the tree, the algorithm applies reduced error pruning to discard the branches that do not improve the accuracy of the tree. This process involves removing each subtree of the tree while evaluating the performance of the pruned tree on a validation set. If the performance of the pruned tree is same or better than the original tree on the validation set, the subtree is discarded (Kour et al., 2023).

Note: The configuration of each algorithm in WEKA was customized based on the task, including parameters like learning rate, number of features and tree depth. Algorithms listed are applicable for both classification and regression tasks where applicable.

The annotated PCa, diabetes, and obesity data sets consist of three attributes: protein change, clinical significance (last reviewed), and semi binary value to clinical relevance **(Figure 3.3).** The first two attributes form the independent variables while the third denotes the defendant variable. We evaluated the following regression algorithms (with the default set of parameters) on each of the PCa, diabetes, and obesity datasets using a train test split of 70:30. Based on the RMSE values **(Table 3.2)** and the average, the standard deviation (PCa=0.001; DM=0.012; Obesity=0.0013) and then the normalized deviation value is calculated and plotted against each other by line graphs.

	A	B	C	D	E
1	Gene(s)	Protein change	Clinical significance (Last reviewed)	Condition(s)	Semi Binary value to clinical significance
2	EPHB2	1	1	1	1
3	MUTYH	1	1	1	1
4	MUTYH	1	1	1	1
5	DNAJC6	1	1	1	1
6	RNASEL	1	1	1	1
7	MSH2	1	1	1	1
8	MSH2	1	1	1	1
9	MSH2	1	1	1	1
10	MSH2	0	1	1	1
11	MSH2	1	1	1	1
12	MSH2	0	1	1	1
13	MSH2	1	1	1	1
14	MSH2	1	1	1	1
15	MSH2	1	1	1	1
16	MSH6	1	1	1	1
17	MSH6	1	1	1	1
18	MSH6	1	1	1	1
19	MSH6	1	1	1	1
20	MSH6	1	1	1	1
21	MSH6	0	1	1	1
22	MSH6	1	1	1	1
23	MSH6	0	1	1	1

Figure 3.3: Representation of Matrices we prepared by semi binary conversion of data

Table 3.2: List of different regression algorithms applied in Weka software applied on datasets of PCa, diabetes mellitus, and obesity and their Root mean square error (RMSE) value

Dataset used	**Algorithm**	**Root Mean Square Error (RMSE)**
PCa	Linear Regression	0.5419
	Multilayer Perceptron	0.5434
	Random Forest	0.5419
	Random Tree	0.5419
	Reptree	0.5419
Diabetes	Linear Regression	0.5975
	Multilayer Perceptron	0.5977
	Random Forest	0.6022
	Random Tree	0.6022
	Reptree	0.6022
Obesity	Linear Regression	0.2982
	Multilayer Perceptron	0.2985
	Random Forest	0.2981
	Random Tree	0.2981
	Reptree	0.2981

Note: The Root Mean Square Error (RMSE) values indicate the prediction error of each algorithm for different datasets (PCa, Diabetes and Obesity). Lower RMSE values represent better model performance. Algorithms were evaluated using WEKA with optimized settings.

3.1.1.8. Correlational study:

To categorize this, the risk variants were retrieved from ClinVar, published GWAS data (from the PRACTICAL Consortium and GWAS central), and our own Exome data (Kour et al., 2023), (Saxena et al., 2021) and Diabetes Type 2 Mellitus data **(Figure 3.4)**. A cross-check is

performed to identify common key players using Venn plots (https://bioinfogp.cnb.csic.es/tools/venny/).

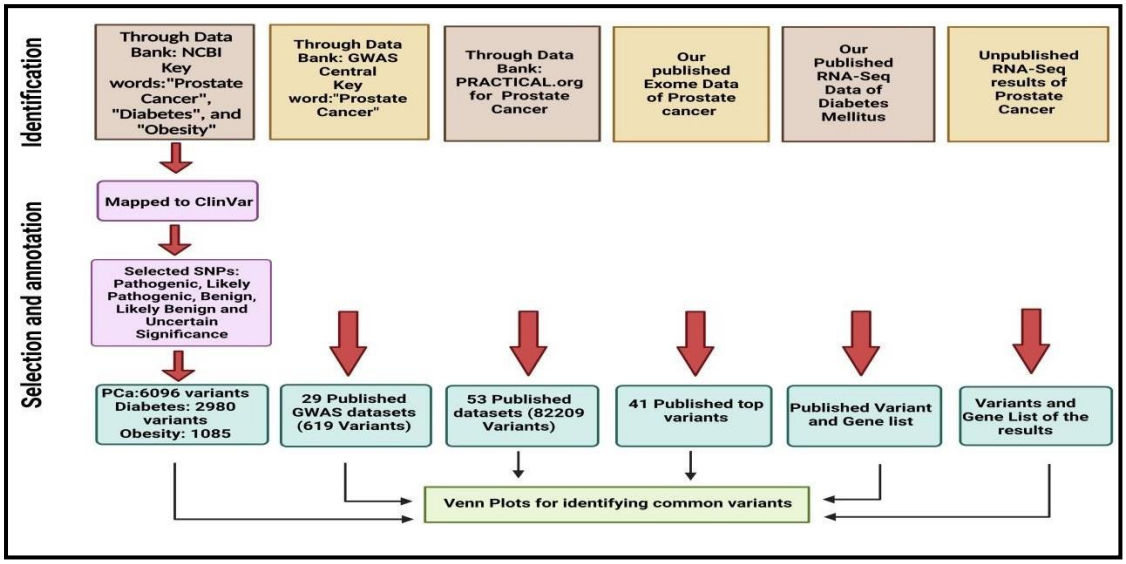


Figure 3.4: Flow chart for the correlational study of ClinVar variants (PCa, diabetes, and obesity) and published GWAS data from GWAS Central, PRACTICAL Consortium, Exome data, and RNA Seq data of both PCa and diabetes mellitus for identifying common variants

3.1.1.9. Differential Analysis: GEPIA 2 (Gene Expression Profiling Interactive Analysis):

GEPIA 2 facilitates the comprehensive analysis and complex data mining tasks of expression datasets from TCGA (The Cancer Genome Atlas) and GTEx (Genotype-Tissue Expression) (Lonsdale et al., 2013) (http://gepia2.cancer-pku.cn/#index) last accessed on April 12, 2024. Box plots were analysed for studying transcription profiles of different cancers in humans and normal tissues using the datasets of TCGA and GTEx in the GEPIA tool. It is one of the important publicly available and personalized tools for the functions like correlation, survival, profiling, plotting, analysis, dimension reductional or differential expression analysis, and detection of a similar gene (**Figure 3.5)** (Z. Tang et al., 2017) (Kour et al., 2023)**.**

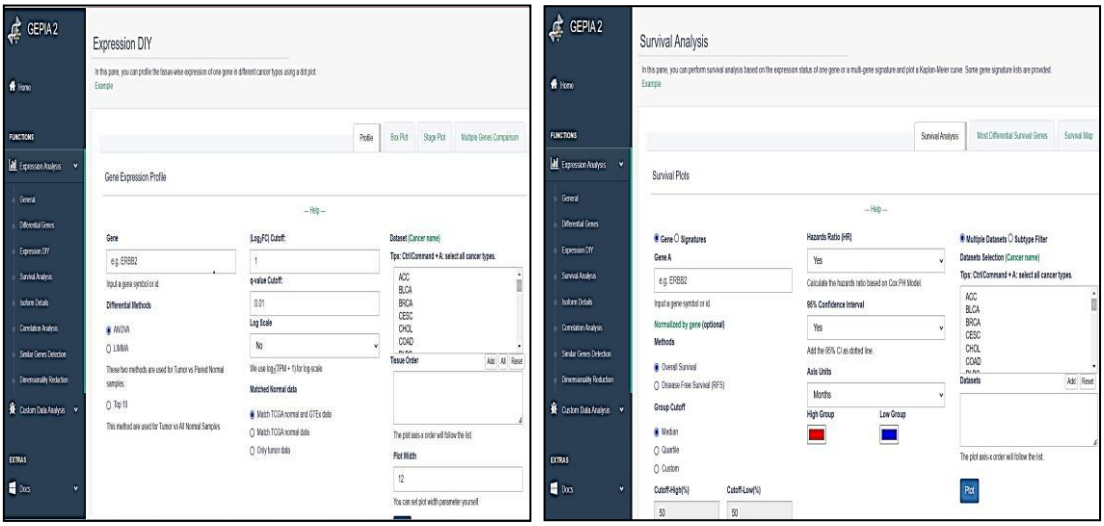


Figure 3.5: Interface of the GEPIA 2 tool used for expression and survival analysis. For expression analysis we have to first enter the key gene name, select the differential method, log2 fold change cutoff, q value cutoff, matching data TCGA data, and the type of prostate cancer dataset. For survival analysis plots, key gene name, method selection either overall or disease frees survival, group cutoff, hazard ratio, axis units and the type of prostate cancer dataset.

3.1.1.10. CBioportal*:*

We used Armenia et al., (2018) datasets wherein they have identified 97 significantly mutated genes (SMGs), 70 of which were not earlier involved in PCa and followed by several mutations that were seen in less than 3% of the cases and the other study used is of TCGA, Cell 2015 (https://www.cbioportal.org/datasets) last accessed on April 12, 2024.

3.1.1.11. Survival Analysis:

For the survival analysis we have used GEPIA2 wherein the log-rank below 0.05 ($p<0.05$) is referred to as significant. GEPIA2 uses RNA-Seq data from TCGA, and divides the patients into high expression groups (whose gene expression value is above the threshold) and low expression groups (whose gene expression value is below the threshold) and then applies Kaplan-Meier survival analysis to estimate survival function. Survival curves are formed with Kaplan-Meier (KM curve), showing the proportion of patients surviving over time for both high and low expression groups. Later, it is compared with the log-rank test (Mantel-Cox test) to find out the differences between the two groups. The test provides a P-value indicating whether the difference in survival of groups is statistically significant (C. Li et al., 2021).

3.1.2. To establish a causal relationship between PCa (PCa) and type-2-diabetes through WES

3.1.2.1. Clinical cases

Case 1:

A 70-year old man was diagnosed with prostate adenocarcinoma, with a PSA level of 55ng/mL in routine evaluations. The patient was complaining with lower back pain and frequent urination.

His physical examination with DRE revealed enlargement in the prostate gland. Histological evaluations of the biopsy specimen revealed a Gleason score grade of 6 (3+3). The laboratory data showed the patient was having diabetes mellitus as comorbidity.

Sample taken: FFPE block

Case 2:

An 80-year old man diagnosed with Prostate adenocarcinoma, with a PSA level of 200ng/mL in his routine laboratory evaluations. The patient was complaining with weakness, loss of appetite, back pain, and urinary incompetence.

His physical examination with DRE and MRI revealed enlargement in both sides of the prostate gland. Histological evaluations of biopsy specimens revealed a Gleason grade score of 8 (4+4). The other laboratory test and the clinical history showed the patient was having diabetes mellitus as comorbidity.

Sample taken: FFPE block

Case 3:

A 68-year old man diagnosed with Prostate adenocarcinoma, with a PSA level of 38ng/ml in his routine laboratory evaluation. The patient was complaining with body aches, increased frequency of micturition, fever and weakness.

His physical examination with DRE and MRI revealed enlargement in both sides of the prostate gland. Histological evaluations of biopsy specimens revealed a Gleason grade score of 7 (3+4). The other laboratory test and the clinical history showed the patient was having diabetes mellitus as comorbidity.

Sample taken: FFPE block

Case 4:

A 65-year old man diagnosed with Prostate adenocarcinoma, with a PSA level of 12ng/mL in his routine laboratory evaluations. The patient was complaining with loss of appetite, abdomen pain and urinary frequency.

His physical examination with DRE and MRI revealed enlargement in both side of the prostate gland. Histological evaluations of biopsy specimens revealed a Gleason grade score of 6 (3+3). The other laboratory test and the clinical history showed the patient was having diabetes mellitus as comorbidity.

Sample taken: FFPE block

Case 5:

A 60-year old man diagnosed with Prostate adenocarcinoma, with a PSA level of 60ng/mL in his routine laboratory evaluations. The patient was complaining of burning micturition, cough, back pain, insomnia and fever.

His physical examination with DRE and MRI revealed enlargement in the left side of the prostate gland. Histological evaluations of biopsy specimens revealed a Gleason grade score of 7 (3+4). The other laboratory test and the clinical history showed the patient was having diabetes mellitus as comorbidity.

Sample taken: FFPE block

The FFPE blocks of the cases were obtained from Rukmani Birla Hospital, RBH Jaipur (Rajasthan), India. After taking informed consent prescribed in Performa of each patient, clinical data were collected from the registry with the help of a medical oncologist. The ethical committee was approved for biomedical research which was received once in 2018 under registration number (RBH/IEC/18/004) and next in 2021 under the registration number (RBH/IEC/21/008) from RBH. The inclusion criteria were; male age should be above 52 having PCa with comorbid/immunomodulatory response/diabetes. The exclusion criteria were; healthy, benign prostate hyperplasia (BPH), and no detected tumors or malignant growth.

3.1.2.2. Next Generation Sequencing (NGS)-WES

NGS is a platform to perform sequencing of millions of small DNA fragments together in parallel. Later, we put together these fragments with the help of a bioinformatic approach which maps each read to the reference of the human genome. Every base undergoes sequencing a number of times to provide the higher depths, so that the unexpected variant can

be explored and the data is delivered with highest accuracy. Old Sanger sequencing has limitations of identifying only substitution and small deletions and insertions (Behjati & Tarpey, 2013). The basic principle of NGS is the same as Sanger validation. It works on the principle of chain termination with chain terminating and fluorescent labelling dideoxyneucleotides, separation of fragments on the basis of size and then finally electrophoresis (capillary gel) to analyse the fragments (T. Hu et al., 2021).

When it comes to discovering disease related genes that are resistant to linkage analysis, WES has shown to be a particularly helpful diagnostic technique in the study of genetic disorders. The coding region of the genome is obtained using WES, through which we can locate possible disease causing genetic mutations in exome (Xiaolin Zhu et al., 2015) **(Figure 3.6).**

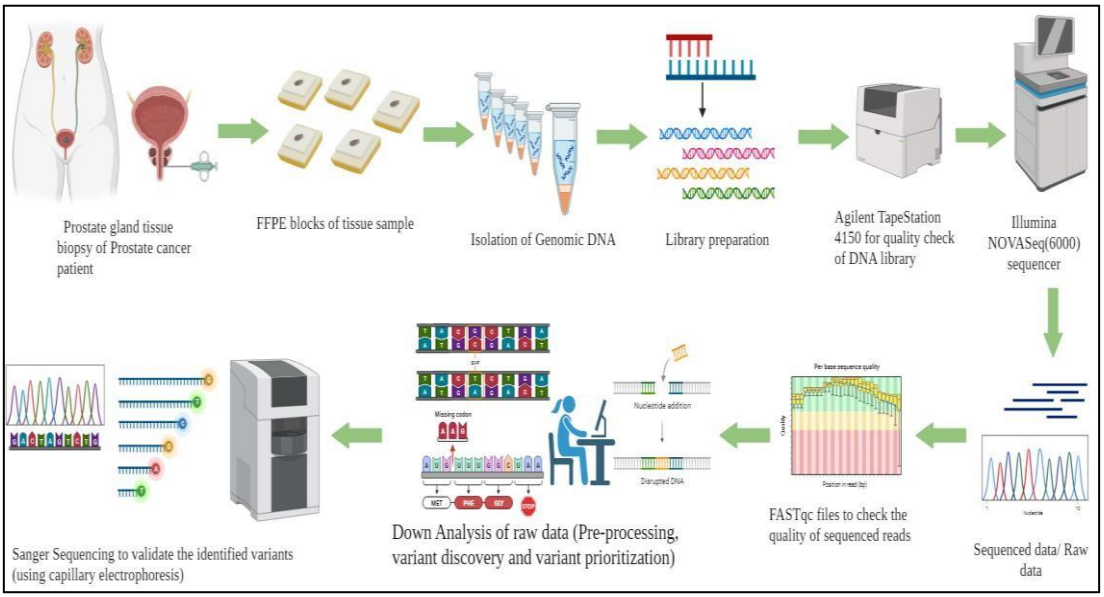


Figure 3.6: Diagrammatic illustration of WES approach (Figure created with BioRender https://www.biorender.com/)

3.1.2.2.1. DNA extraction, DNA quantification, and DNA QC

The quantity of Extracted DNA from FFPE blocks (**Figure 3.8)** was checked with the help of NanoDrop 1000 (Thermo Fisher scientific, Washington, DE, USA). DNA concentrations, D260/OD280 and OD260/OD230 ratios of solutions were calculated by NanoDrop. PCR was then performed which involved the initial denaturation at 94°C for one minute, then 35 cycles of 98°C for ten seconds and 65°C for one minute and finally followed by extension at 72°C for five minutes.

The quality of the quantified DNA was confirmed on the 1% agarose gel. In brief, 2 µls of the DNA is mixed with 2 µls of 6x Loading dye (Invitrogen) and subjected to electrophoresis at 120 volts for 30 mins.

3.1.2.2.2. Preparation of libraries for exome capture and evaluation of Molarity of electronic ladder and sample

Final libraries were quantified using Qubit 4.0 fluorometer (Thermofisher #Q33238) using a DNA HS assay kit (Thermofisher #Q32851) following the manufacturer's protocol. To identify the insert size of the library, we queried it on Tapestation 4150 (Agilent) utilizing high sensitivity D1000 screentape (Agilent # 5067- 5582) following manufacturers' protocol (**Figure 3.7)**.

Tapestation Electropherogram: In order to make sure of effective sequencing outcomes, it is essential to evaluate the quality and quantity of the NGS library manufacturing process. For the purpose of offering a reproducible QC approach to analyse samples while the workflow of library preparation is going on, Agilent created the Genomic DNA ScreenTape (D1000 ScreenTape used) and new D1000 ScreenTape assays. In less than two minutes per sample, the Agilent 4150 TapeStation device and Genomic ScreenTape automates the evaluation (automated electrophoresis) of initial genomic DNA with volume of sample as little as 1 µl and produces digital results (M. H. Liu et al., 2014).

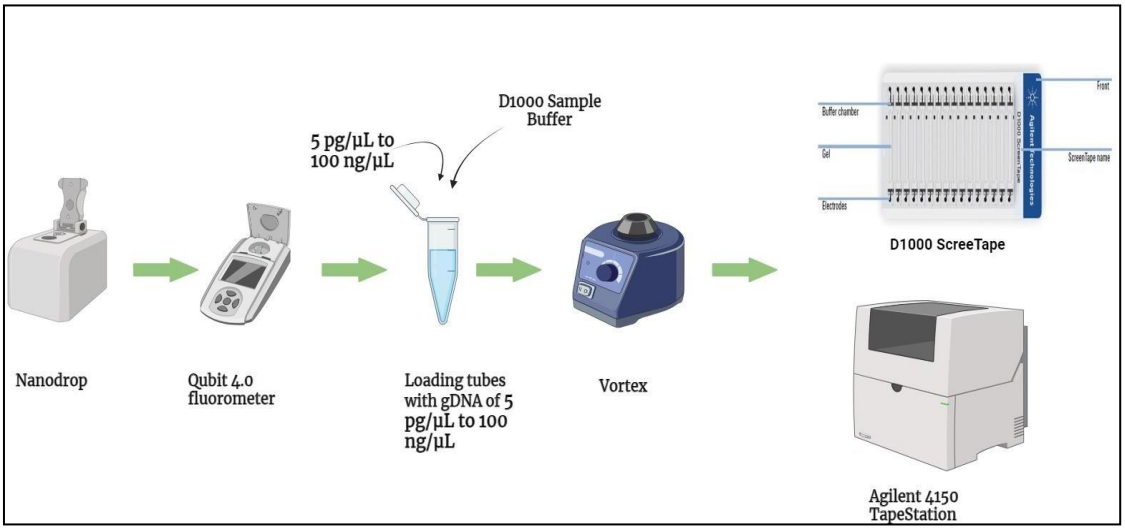


Figure 3.7: Workflow of library quality check using Agilent 4150 Tapestation (Figure created with BioRender https://www.biorender.com/)

Table 3.3: Summary of sequenced data including total number of reads both forward and backward and the size of total sequenced data

SAMPLE	Raw Tot Reads	Raw Total Bases	Read1 length	Read 2 length	Total GB	Trimmed Total Reads	Trimmed Total Base	Trimmed Read 1 length	Trimmed Read 2 length	Total Data
S1	159926532	2542831858	159	159	25.42832	143541822	20771198403	144	144	25.42832
S2	162418144	25824484896	159	159	25.82448	151617042	22467159525	148	148	25.82448
S3	182630124	29038189716	159	159	29.03819	156136314	21943308209	140	140	29.03819
S4	232925348	37035130332	159	159	37.03513	217367520	32073748997	147	147	37.03513
S5	129863620	20648315580	159	159	20.64832	113252290	16112317189	142	142	20.64832

Note: The table shows sequencing data for samples S1 to S5, including raw and trimmed reads, bases counts, read lengths, and total data in gigabase (GB). "Raw" refers to the unprocessed data, while "Trimmed" refers to data that has been processed to remove low-quality sections or adapters, improving the overall quality for downstream analysis.

3.1.2.2.3. Pro*cessing Sequenced data:*

All samples were sequenced using Illumina (NOVASEQ 6000), sequencing depth asked for 6GB/sample. The read length of both forward and backward reads was 159 bp. **Table 3.3** illustrates the total number of reads (both forward and backward), the trimming size of read and the total size of data generated after the sequencing per sample.

The WES pipeline is a combination of software *viz*, Fastqc, Bowtie2, Samtools, Bamtools, VarScan, Bcftools, Vcftools, and dbSNP and a reference human genome (hg38).

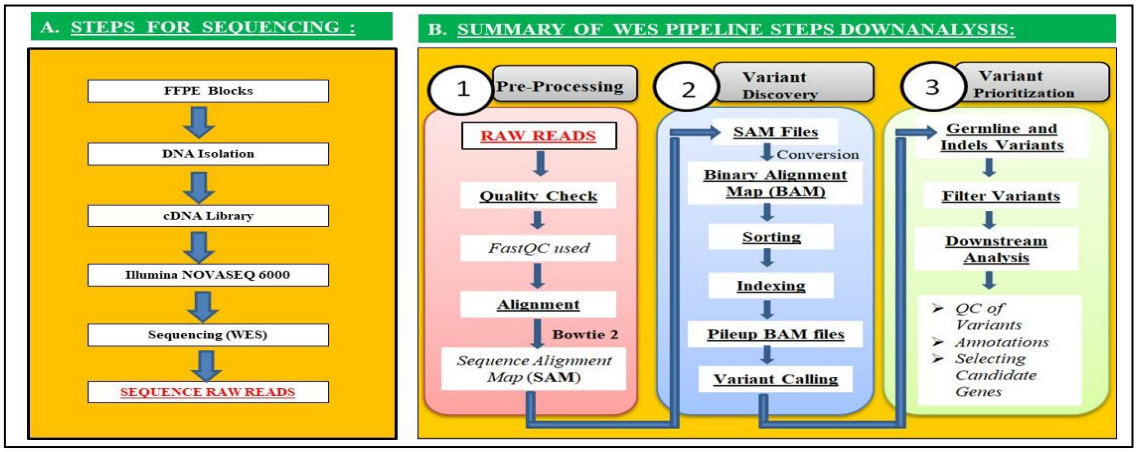


Figure 3.8: Work flow of WES benched marked pipeline. After obtaining the raw reads the sequenced data downstream analysis performed in 3 phases; Pre-processing, Variant discovery and Variant Prioritization

```
#Indexing already done using bowtie2, BWA and samtools: /home/ngs/Data/hg38
#All scripts and commands are to be run from Expipe
#fastqc already one for all samples. Pl check the folder
#bowtie2 -x /home/ngs/Data/hg38/hg38  -1 S1_1.fastq -2 S1_2.fastq -S S1.sam
#samtools import /home/ngs/Data/hg38/hg38.fna S1.sam S1.bam
#samtools sort S1.bam >S1.sorted.bam
samtools index S1.sorted.bam S1.sorted.bam.bai &
samtools merge S1.merged.bam S1.*
samtools mpileup S1.bam > S1.mpileup.bam
varscan mpileup2snp S1.mpileup.bam > S1.mpileup.snps &
varscan mpileup2indel S1.mpileup.bam > S1.mpileup.indels
varscan filter S1.mpileup.snps >S1.mpileup.snps.filter
varscan readcounts S1.mpileup.bam >S1.mpileup.readcounts
samtools mpileup -uf /home/ngs/Data/hg38/hg38.fna S1.sorted.bam | bcftools view ->S1.raw.bcf
#samtools calmd -Abr S1.mpileup.bam/Data/hg38/hg38.fa > S1.baq.bam
bcftools view S1.raw.bcf >S1.vcf
```

Figure 3.9: Set of commands used for WES pipelines, a benchmarked pipeline using bowtie2, BAM tools and SAM tools, Varscan, bcftools and vcftools as cited previously (Misawa et al., 2017)

3.1.2.2.4. **Data pre-processing**:

Pre-processing of raw data initiates with sequence quality check using FastQC html files of both forward and reverse read. Html files contain plots (box and graphs) representing mean values of quality scores for read length, depth, intended coverage, yield, base per read, and sequences and statistical inference is made. Bowtie2 first indexed the reference genome and then both forward and reverse reads were then aligned to indexed reference genome. Alignment gave the sequence alignment map (SAM) format files **(Figure 3.8) (Figure 3.9) (Figure 4.4)**.

a. Variant discovery: Bamtools converted the SAM files to binary mapped (BAM) format to save memory. We used sorting BAM to remove extra alignment or overlappings and samtools mpileup to get the coverage of the mapped reads on the reference genome across all the samples at single base pair resolution **(Figure 3.8) (Figure 3.9) (Figure 4.4)**. VarScan helped to call the desired variants from the pool of sequenced data that matched the required threshold values for base quality, allele frequency, read depth, and statistical significance (*P*-value of less than 0.005).

b. Variant prioritization: We identified the germline variants (SNPs), Indels from the pool and remove the false positive indels and SNPs to obtain highly assured SNVs and indels. Later variant calling files (VCF) were created to store the data of all the SNPs and indels with the help of bcftools and vcftools **(Figure 3.8 and 4.4)**.

3.1.2.3. Downstream analysis of VCF files

These files were then subjected to downstream processing which involves their annotation of the data through different tools and databases *viz*, SNPnexus, NCBI-dbSNP, Batch Entrez, and ClinVar (**Figure 3.10)**.

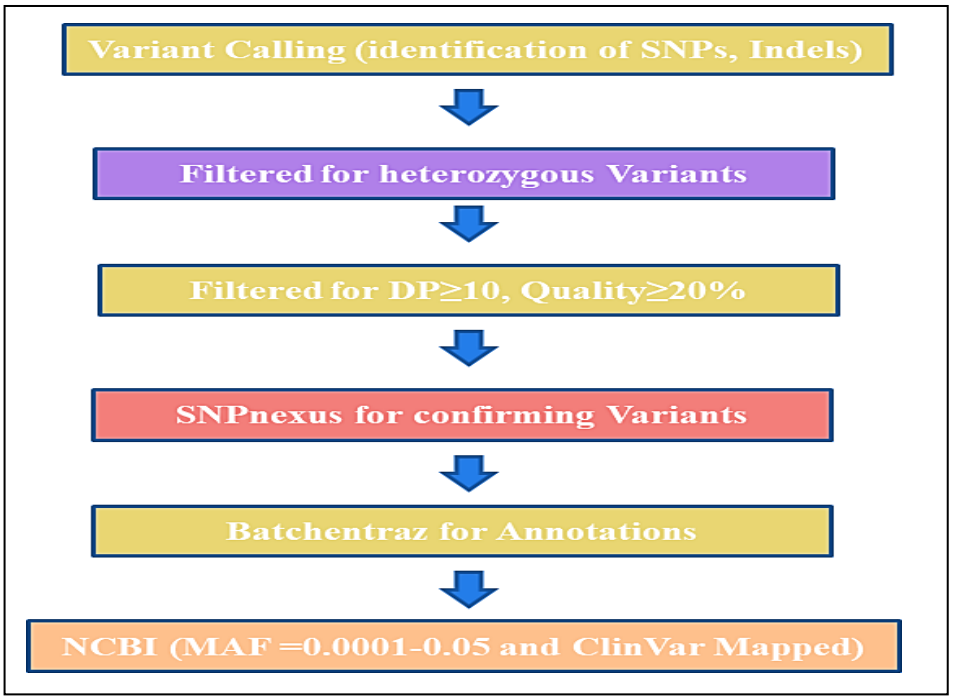


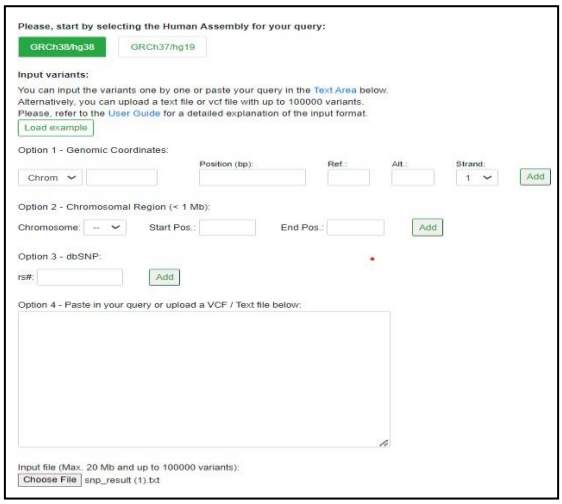
Figure 3.10: Workflow of vcf files analysis. The called variants including several SNPs, Indels were filtered to select only heterozygous variants, and then selected those variants which were having mapped read depth above or equal to 10 and with good quality of greater than or equal to 20%. The screened variants were queried using SNPnexus to confirm each

variant and for annotations we used Batch Entrez which retrieves a large number of SNP data in one mode. Annotated variants were mapped to NCBI and ClinVar to know the clinical significance and clinical conditions associated.

3.1.2.3.1. SNPnexus

Perhaps, it is the most robust tool accessible at present for evaluation of functional consequences of both already publicly available and novel variants using transcriptomic, proteomic, regulatory and the structural variant models. The annotation in SNPnexus (https://www.snp-nexus.org/v4/ last accessed on April 2, 2024) is made by using data from five more annotation platforms (RefSeq, AceView, VEGA, UCSC and Ensemble) (Chelala et al., 2009) (Dayem Ullah et al., 2012). Recently, the range of SNPnexus is extended to interpret sequenced data of cancer patients more efficiently by identifying the driver variants for cancer and those variations which can be targeted as biomarkers (Oscanoa et al., 2020).

After obtaining the filtered vcf files we uploaded that vcf file format as “My Query”, entered username, email detail and University name. Next, we selected all for the annotations comprising Gene or protein consequences, effect of non-synonymous coding SNP, population data, regulatory elements, conversation, non-coding scoring, structural variants, pathway analysis and biological/clinical interpretations. Then finally we hit the submit query (**Figure 3.11).**



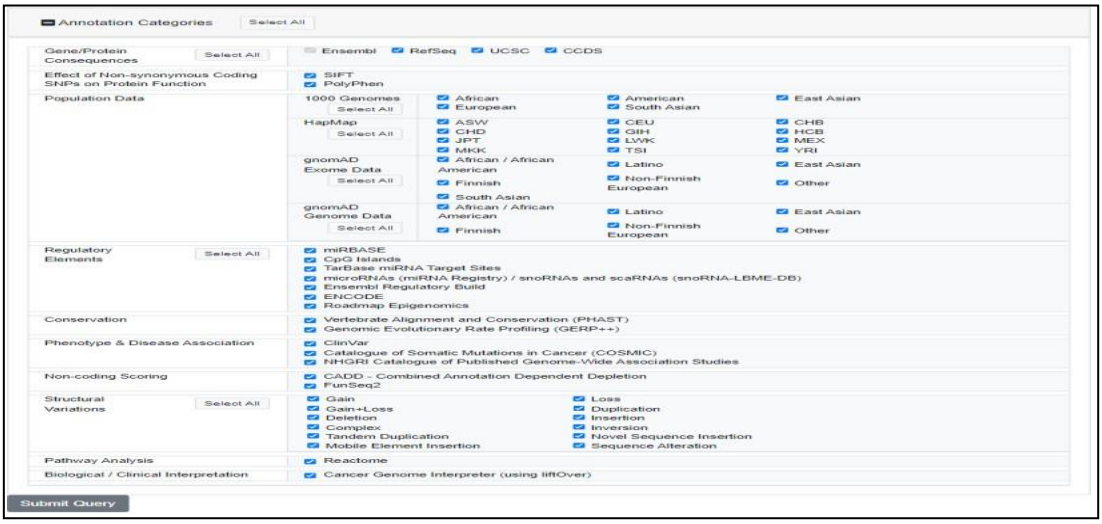


Figure 3.11: Three steps to run SNPnexus: First information, providing the query format (genomic coordinates/chromosome regions/dbSNP/vcf file) and selection of all the annotation categories and finally submitting the query.

3.1.2.3.2. Batch Entrez

The Entrez and Batch Entrez (https://www.ncbi.nlm.nih.gov/sites/batchentrez) programmes from NCBI are the two major ways to retrieve this data quickly (Crow et al., 2001). It is a data retrieval system of NCBI (Bhagwat, 2010) **(Figure 3.12)**.

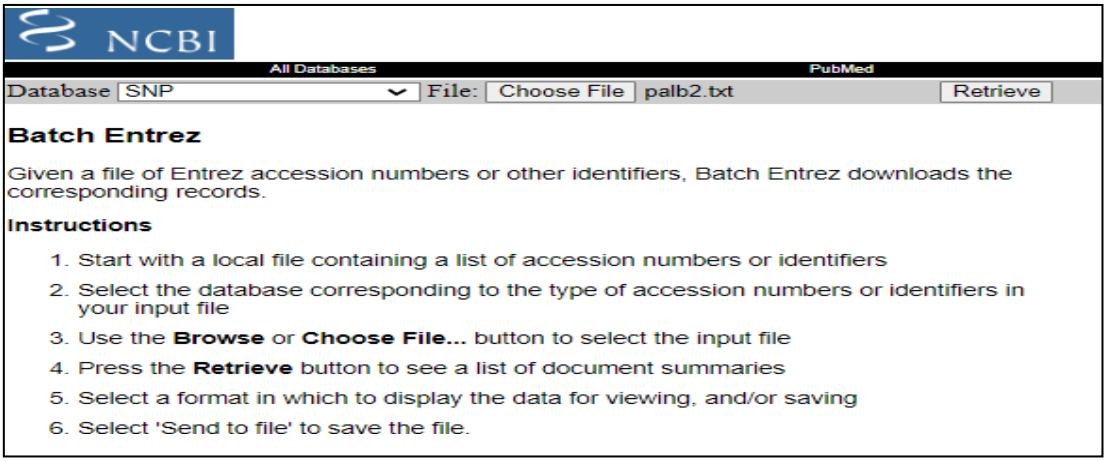


Figure 3.12: Image showing the interface of Batch Entrez on which we selected database on SNP to provide the data type and uploaded .text file for analysis.

3.1.3. Sanger Sequencing for Validation of WES results:

Sanger sequencing or dideoxy terminator sequencing is a gold standard approach for validating results of next generation sequencing. First the genes are selected for validation

and then an equal set of primers were designed and synthesized for both PCR and Sanger sequencing. Flanking regions were taken from dbSNP database (Beck et al., 2016). Amplicons for PCR were then formed, purified, and sequenced for validation (De Cario et al., 2020) (**Figure 3.13)**.

Steps for Sanger sequencing includes primer annealing, DNA amplification by deoxynucleotides (dNTPs), chain termination, generation of amplicons by cDNA amplification and its quantification by using Qubit® 4.0 Fluorometer (**Table 3.4)** and using master mix (**Table 3.5)** (De Cario et al., 2020) (Crossley et al., 2020). The PCR cycling conditions are summarized in **Table 3.6**. The details of primer sequence used and their lengths are given in **Table 3.7.** Later the sequencing template identification, isolation, purification amplicons by removing all the free dNTPs, unbound primers, salt, polymerase enzyme, or any other impurities and quantification of amplicons were done and the capillary gel electrophoresis suggests the homogeneity of the product. Next, the amplicon visualization using gel electrophoresis, quantification assessment by spectrophotometry and ultraviolet fluorescence tagging was done. Finally, sequencing analysis was received in the format of the chromatogram generated as either ".fa" and or ".ab1". Primers are removed once the chromatogram was assessed by the software leaving behind a target sequence of high quality (Crossley et al., 2020).

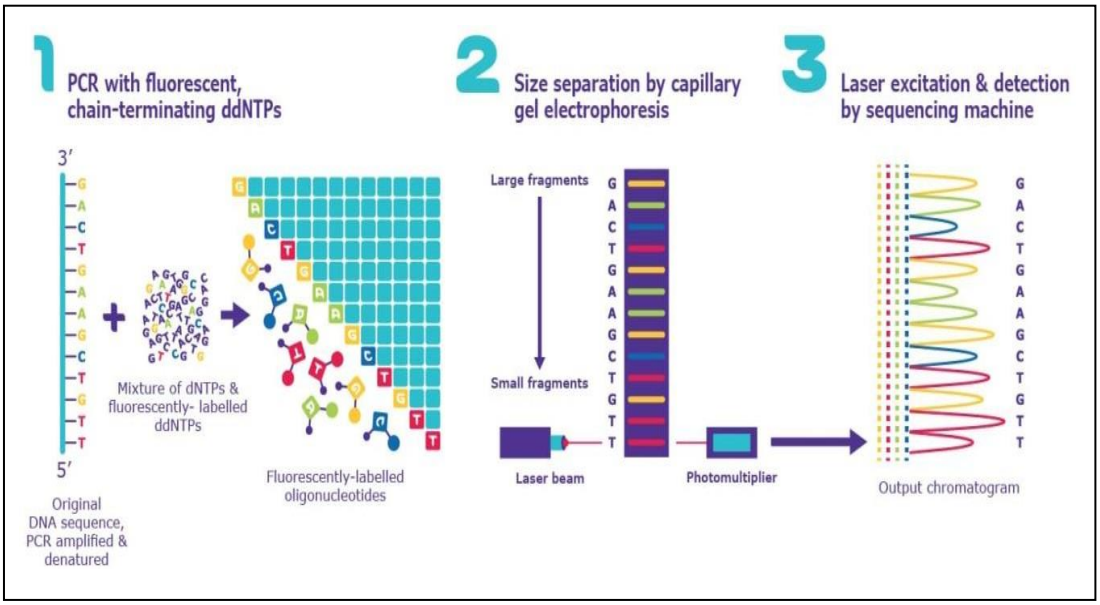


Figure 3.13: Steps involved in Sanger sequencing using capillary gel electrophoresis (https://www.sigmaaldrich.com/IN/en/technical-documents/protocol/genomics/sequencing/sanger-sequencing)

Table 3.4: Quantification by Qubit® 4.0 Fluorometer

S. No.	FFPE Block No.	DNA Con. (ng/µl)	Volume (µl)	Yield (ng)
1	H20/6591	41.6	50	2080
2	H20/8329	5.3	50	265
3	H20/8030	7.82	50	391
4	H20/7111	4.56	50	228
5	H20/7255	15.5	50	755

Note: This table presents the DNA concentration (ng/µl), and yield (ng) for various FFPE (formalin-fixed, paraffin-embedded) block samples. The yield is calculated based on the DNA concentration and volume, indicating the total amount of DNA extracted from each block.

Table 3.5: Master Mix for Amplification

S. No.	Reagent	Volume in µl
1	Premix Master mix	**12.5**
2	Betaine	**1.0**
3	FP Primer (10 pmol/ µl)	**1.0**
4	RP Primer (10 pmol/ µl)	**1.0**
5	DNA Sample (10-20ng)	**Variable**
6	DH2O	**Variable**
Final Volume		**25**

Note: This table outlines the components and corresponding volumes of reagents used in the PCR (polymerase chain reaction) setup. The final volume is adjusted to 25µl, with the DNA sample and DH2O volumes varying based on the specific requirements

Table 3.6: PCR cycling conditions

No. of Cycles	Name of Step	Temperature (°C)	Time (mm:ss)
1	Initial Denaturation	95	**05:00**
35	Denaturation	95	**00:30**
	Annealing	60	**00:45**
	Extension	72	**01:00**

1	Final Extension	72	**05:00**
1	**Hold**	**4**	∞

Note: This table summarizes the thermal cycling conditions for the PCR process, detailing the number of cycles, step names, temperatures, and durations. The initial denaturation occurs at 95ºC for 5 minutes, followed by 35 cycles of denaturation, annealing, and extension steps. The final extension is performed at 72ºC for 5 minutes, and the reaction is held at 4ºC

Table 3.7 Primer details designed for the Sanger validation

S. No	Gene Name	SNP	Primer	Primer sequence	Primer Length (bp)	Mutation (Hotspot region)	Product size
1	MYRF	rs370887875	Forward	TGGGTGGAGATTCAGAGGCG	20	**c.3207C>T**	**193bp**
			Reverse	GTGGGTGTGGAGACTCTGTG	20		
2	MYO15A	rs375290498	Forward	TTGCTTGAGTGTGGCCGCCTTG	22	**c.5925G>A**	**250bp**
			Reverse	TAGAGGGACAGTGCATGGGAC	21		
3	ITGB4	rs199620842	Forward	AGCGGTCAGTGTAGACATGCC	21	**c.4559-4C>G**	**204bp**
			Reverse	TCCTGCCAGCTCACTCTGAG	20		
4	ATM	rs587779865	Forward	AATTAGCCCTGCGTGCACTG	20	**c.7456C>T**	**200bp**
			Reverse	GGGTAGAATATTGGGCTGAG	20		
5	GJB2	rs111033186	Forward	TTTAAGGACATCGAGGAGATC	21	**c.457G>A**	**202bp**
			Reverse	CAAAGCAGTCCACAGTGTTG	20		
6	HNF1A	rs1169305	Forward	TCTTCACCTCAGACACTGAGG	21	**c.1720A>G**	**228bp**
			Reverse	GTGACGGACAG	20		

				CAACAGAAG			
7	MPO	rs35897051	Forward	GGCTTTGTTATA TCCTGGGAG	21	c.2031-2A>C	**193bp**
			Reverse	**TGATGCCTGT GTTGTCGCAG**	**20**		

Note: This table lists the details of specific genes and their associated single nucleotide polymorphisms (SNPs), including primer sequences, primer lengths, and mutations in hotspot regions, and product sizes for PCR amplification. Each entry provides information on the forward and reverse primers used for amplification, essential for studying genetic variants.

3.1.4. Study of Mutational Hotspots of PTEN:

By using SNPnexus we identified first all the single nucleotide variants (SNVs). Those variants were then annotated using ANNOVAR (https://annovar.openbioinformatics.org), gnomAD v21.1, 1000 Genomes Project, dbNSFP (https://sites.google.com/site/jpopgen/dbNSFP, and Exome Aggregation Consortium (ExAC). 0.1% Minor allele frequency (MAF). Then we used CADD scoring to score the variants to find the deleterious and potential pathogenic variants. This scoring is based on quantitative prioritizing of the variant data as per function, deleterious effect and disease caused. It uses the PHRED score, for instance, a score of ≥10 indicates 10% of the deleterious variants in the genome and a score of ≥20 indicates the top 1% of deleterious variants (Rentzsch et al., 2019). For evolutionary conservation screening we used Genomic Evolutionary Rate Profiling (GREP) and Phastcons. This indicated the functional importance and helped in predicting whether the variant is deleterious or not (Skopelitou et al., 2021). Then we performed the functional screening by SIFT (analysed the substitution mutation impact), PolyPhen (showed the impact of Amino acid change either loss of function or gain of function), ClinVar, and Mutation Taster scoring (Sim et al., 2012) (Flanagan et al., 2010). At last we screened deleterious variants and validated them using the ClinVar and Ensemble database (Skopelitou et al., 2021) **(Figure 3.14).**

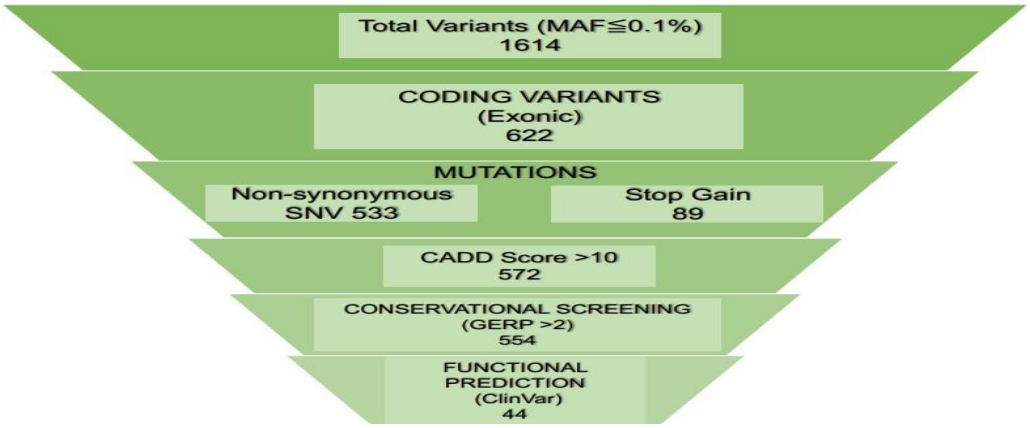


Figure 3.14: Overview of cancer variant prioritization based on coding variants, conservational screening, and functional prediction.

3.1.5. Categorizing the variants or candidate genes for PCa in Urolithiasis as Co-morbidity

We have systematically reviewed and searched for published literature using PubMed using the keywords, *viz.* "Urolithiasis" and "PCa" in NCBI. A gene list is then prepared from the published National Centre of Biotechnology Information (NCBI) dataset, comprising all the genes related to Urolithiasis primarily with mutations (both pathogenic and likely pathogenic Single nucleotide polymorphisms, SNP's) for every particular gene screened from the published datasets. To see the interactions among all the potential genetic factors, GeneMANIA and String databases were used and an interaction map was prepared. After which, determination of the cellular and subcellular location of all the potential genes was done by using target p2.0 and UniProt databases respectively. For the characterization of mutations representing the associations of urolithiasis and PCa, a correlational study is applied. For this purpose, we have used gnomAD (https://gnomad.broadinstitute.org/ last accessed on April 2, 2024) for verifying all the SNPs whether they are synonymous or nonsynonymous mutations (**Figure 3.15)** (X. Sun et al., 2021).

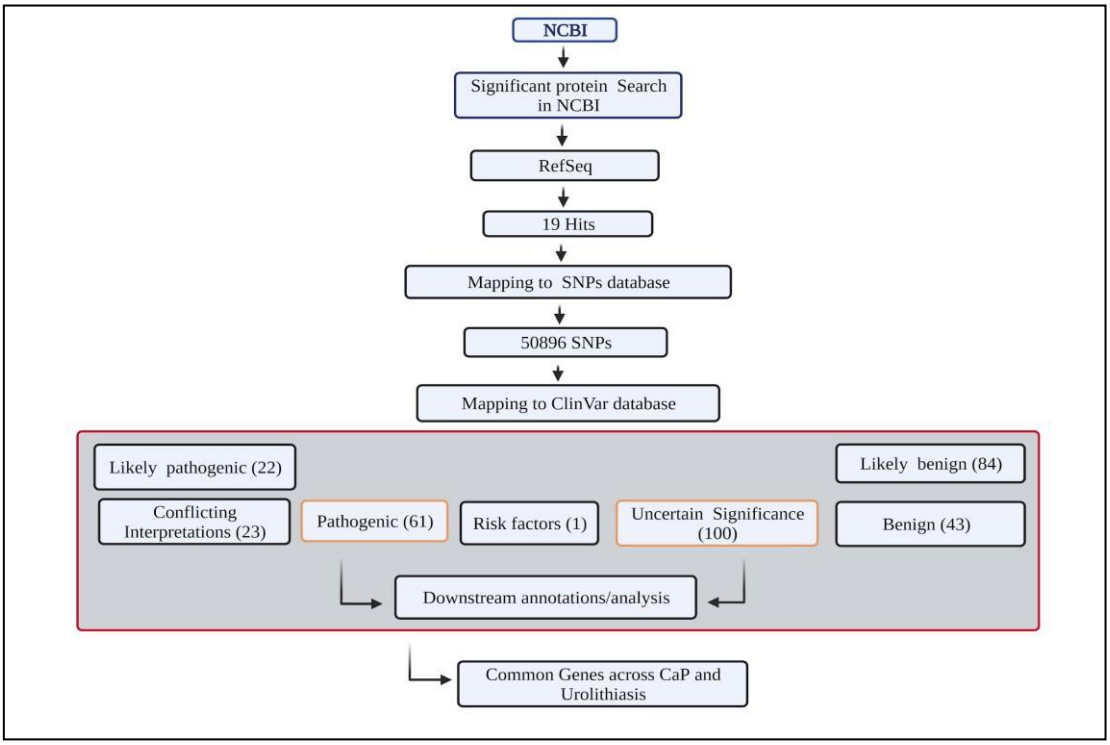


Figure 3.15: A flowchart for the steps of gnomAD to identify the variant types and their rarity of Urolithiasis related genes. We mapped RefSeq proteins (19) of urolithiasis to the SNP database which resulted in the identification of 50896 different SNPs, which were subjected to ClinVar/dbVar. Later from the given categorized list (pathogenic, likely pathogenic, conflicting predictions, uncertain significant, benign and likely benign) we filtered out those genes which are pathogenic (61) and are having uncertain significance (100). These SNPs can be verified on the basis of type of variation (missense or nonsense) or can also be analysed for their associations with other diseases (Figure created through BioRender https://www.biorender.com/).

3.2. To identify and explore the role of lncRNAs as potential targets using lncRNA-protein interactions

3.2.1. PCa Causal Proteins, LncRNAs and Androgen Receptors (AR):

FASTA files of all the 28 prostate specific proteins as follows: *ADA, ANG, BRCA1 , CTNS, HBB, GNPTAB, COL6A1, OTOF , TP53, CYP11B2, CYP1B1 , GJB6, RHAG, DNAAF1, BRCA2, NF1 MCM8, MCCC1, CAPN3, MYO15A, MRE11, KRIT1, HEXB, SCN9A, PRLR, OPA1, ATP6V0A2 and USH2A* **(Table 3.6)** were taken from NCBI (Khilwani et al., n.d.), (Unpublished Data).

The clinical samples data was screened to identify pathogenic related lncRNAs in which 11 lncRNAs were identified (Shukla *et al.*, 2023). The sequence data of 11 lncRNAs (*SCARNA10, LINC01973, LINC00940, NPBWR1, FLJ16779, ANKRD20A9P, LINC00298, SNHG19, LOC341056, TLX1NB, LINC00662:60*) were extracted from NONCODE (http://www.noncode.org/ last accessed on April 2, 2024), LNCipedia (https://lncipedia.org/) and RNAcentral (https://rnacentral.org/ last accessed on April 2, 2024) to get their HSAT id's and their FASTA files. Here, proteins were taken as receptors and LncRNAs as ligands. These fasta files were then given as input to the HDOCK server (Khilwani et al., n.d.), (Unpublished Data).

We had selected PCa associated 8 ARs (2Q71, 5V8Q, 4QL8, 2PNU, 5CJ6, 2AM9, 1E3G, 7KW7) and we gave their PDB id and chain IDs as input **(Table 3.8).** Both PDB and chain ID were taken from RSCB PDB protein data bank (https://www.rcsb.org/ last accessed on April 2, 2024). All the ARs were then considered as receptors and lncRNAs as ligands (Khilwani et al., n.d.), (Unpublished Data).

Table 3.8: List of PCa Proteins, Androgen Receptors and PDB Ids used in the molecular docking study

PCa Proteins		
PROTEIN	**PDB ID**	**REFERENCES**
ADA	3IAR	(Khilwani et al., n.d.) (Unpublished Data).
ANG	4AOH	(Khilwani et al., n.d.) (Unpublished Data).
BRCA1	6GVW	(Khilwani et al., n.d.) (Unpublished Data).
CTNS	7ZKW	(Khilwani et al., n.d.) (Unpublished Data).
HBB	6LCX	(Khilwani et al., n.d.) (Unpublished Data).
GNPTAB	2N6D	(Khilwani et al., n.d.) (Unpublished Data).
COL6A1	1KNT	(Khilwani et al., n.d.) (Unpublished Data).
OTOF	3L9B	(Khilwani et al., n.d.) (Unpublished Data).
TP53	6VTC	(Khilwani et al., n.d.) (Unpublished Data).
CYP11B2	7M8I	(Khilwani et al., n.d.) (Unpublished Data).
CYP1B1	6OYV	(Khilwani et al., n.d.) (Unpublished Data).
GJB6	5ER7	(Khilwani et al., n.d.) (Unpublished Data).
RHAG	8CSX	(Khilwani et al., n.d.) (Unpublished Data).
BRCA2	1MIU	(Khilwani et al., n.d.) (Unpublished Data).

NF1	1NF1	(Khilwani et al., n.d.) (Unpublished Data).
MCM8	6L0O	(Khilwani et al., n.d.) (Unpublished Data).
MCCC1	2EJM	(Khilwani et al., n.d.) (Unpublished Data).
CAPN3	6BGP	(Khilwani et al., n.d.) (Unpublished Data).
MYO15A	7UDU	(Khilwani et al., n.d.) (Unpublished Data).
MRE11	3T1I	(Khilwani et al., n.d.) (Unpublished Data).
KRIT1	5D68	(Khilwani et al., n.d.) (Unpublished Data).
HEXB	3LMY	(Khilwani et al., n.d.) (Unpublished Data).
SCN9A	7W9M	(Khilwani et al., n.d.) (Unpublished Data).
PRLR	3NPZ	(Khilwani et al., n.d.) (Unpublished Data).
OPA1	6JTG	(Khilwani et al., n.d.) (Unpublished Data).
ATP6V0A2	3RRK	(Khilwani et al., n.d.) (Unpublished Data).
Androgen Receptors		
Uroporphyrinogen Decarboxylase (URO-D)	2Q71	(Khilwani et al., n.d.) (Unpublished Data).
Androgen Receptor (AR)	4QL8	(Khilwani et al., n.d.) (Unpublished Data).
Selective Androgen Receptor Modulator (SARM)	5V8Q	(Khilwani et al., n.d.) (Unpublished Data).
Androgen receptor (AR)	2AM9	(Khilwani et al., n.d.) (Unpublished Data).
Androgen receptor (AR)	2PNU	(Khilwani et al., n.d.) (Unpublished Data).
Selective Androgen Receptor Modulator (SARM)	5CJ6	(Khilwani et al., n.d.) (Unpublished Data).
Androgen receptor (AR)	1E3G	(Khilwani et al., n.d.) (Unpublished Data).

Note: *This table lists various proteins and androgen receptors (AR) associated with prostate cancer (PCa) along with their corresponding PDB (Protein Data Bank) Ida nd references used in molecular docking studies*.

3.2.2. Molecular Docking studies:

Docking analysis was performed using HDOCK (http://hdock.phys.hust.edu.cn/), a docking tool. Numerous biological activities, including signal transmission, cell control, protein synthesis, DNA replication and repair, RNA transcription, etc., depend on interactions between nucleic acids and proteins. Thus, understanding their intricate structure will help researchers design treatment strategies or medications that specifically target these interactions. Because experimental approaches are expensive and technically challenging, molecular docking has become crucial in the identification of complex structures (Khilwani et al., n.d.), (Unpublished Data).

For protein-protein docking, the HDOCK server (http://hdock.phys.hust.edu.cn/ last accessed on April 12, 2024) combines homology search, template-based modeling, structure prediction, macromolecular docking, biological information incorporation, and task administration **(Figure 3.16).** The server automatically predicts the interaction between receptor and ligand molecules using input data for both molecules (amino acid sequences or Protein Data Bank structures). This was done using a hybrid method of template-based and template-free docking (Khilwani et al., n.d.), (Unpublished Data).

3.2.3. Work flow of molecular docking

3.2.3.1. *Input*:

Both protein sequences and structures were accepted as input data in the workflow's initial step. The HDOCK server was built to take inputs for both protein sequences and structures, which makes it easier for both inexperienced and regular users to operate. The server takes two types of inputs for structures and two types of inputs for sequences for every molecule, given: (1) A PDB-formatted pdb file, (ii) pdb file with chainID in PDB (e.g. 1CGI:E), (iii) Copy the protein sequence and paste it in the FASTA format, (iv) Uploading a FASTA-formatted protein sequence file. Each molecule just requires one kind of input and with automated modeling of DNA/RNA structures from sequences currently difficult; the service only accepts structure inputs for DNAs and RNAs at this (Khilwani et al., n.d.), (Unpublished Data).

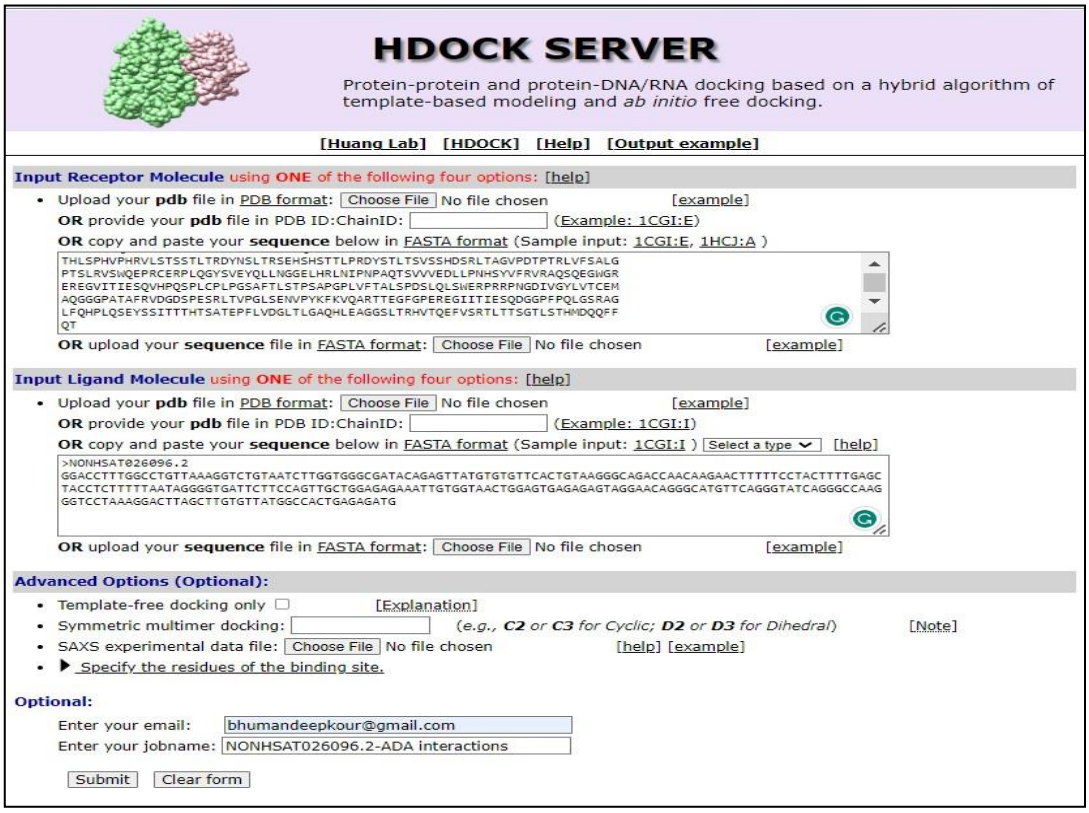


Figure 3.16: Showing interface of HDOCK server and the uploaded FASTA format sequence of the both Receptor (PCa protein) and ligand (LncRNA).

3.2.3.2. *Sequencing similarity*:

The workflow's second phase is the search for similarity in sequence. To determine the homologous sequences for receptor and ligand molecules, this similarity match was carried out against the PDB sequence database using the sequences from input or converted from structures. The HHSuite software was used for protein sequence searches since it is widely known because of its effectiveness in locating distant homologs. Since FASTA (version 3.6) is a powerful and user-friendly tool for both protein and DNA/RNA sequence search, it is utilized for DNA/RNA and thus two sets of homologous templates are produced as a result of this process (Khilwani et al., n.d.), (Unpublished Data).

3.2.3.3. Template selection:

After that, the process moves on to the third stage, this involves comparing two set of templates to check if they share any entries with similar PDB codes. A similar template will be chosen both for the receptor and the ligand if there are any such PDB codes. The best templates for the receptor protein and/or ligand protein will be chosen from two sets of homologous templates, respectively, assuming that there is no link between the two sets.

When there are many templates present, then the one with maximum sequence coverage, sequence similarity, and resolution is chosen. Models are constructed using MODELLER with the chosen templates, and ClustalW was used for sequence alignment (Khilwani et al., n.d.), (Unpublished Data).

3.2.3.4. *Result*:

HDOCK server adds docking tasks to the queue on providing input three task status including "QUEUED," "RUNNING," and "RESULTS," are yielded and finally the docking results may be found at http://hdock.phys.hust.edu.cn/date/jobid, where "jobid" is the specific job id displayed on the web page of status (Khilwani et al., n.d.), (Unpublished Data).

3.2.3.5. *Output*:

The docking output consists of three fundamental files: Receptor PDB file created by the server using the users' FASTA sequence or supplied by users, Ligand PDB file created by the server using the user-provided FASTA sequence or supplied by users and ligand binding modes reflected by their transformations in the HDOCK output. Additionally, the result page displays a docking summary of the top 10 models at the bottom and the template information for the receptor and ligand at the top. Based on the highest ligand receptor binding energy interacting model of each interaction was selected. Those models were then subjected for visualization (**Figure 3.17)** (Khilwani et al., n.d.), (Unpublished Data).

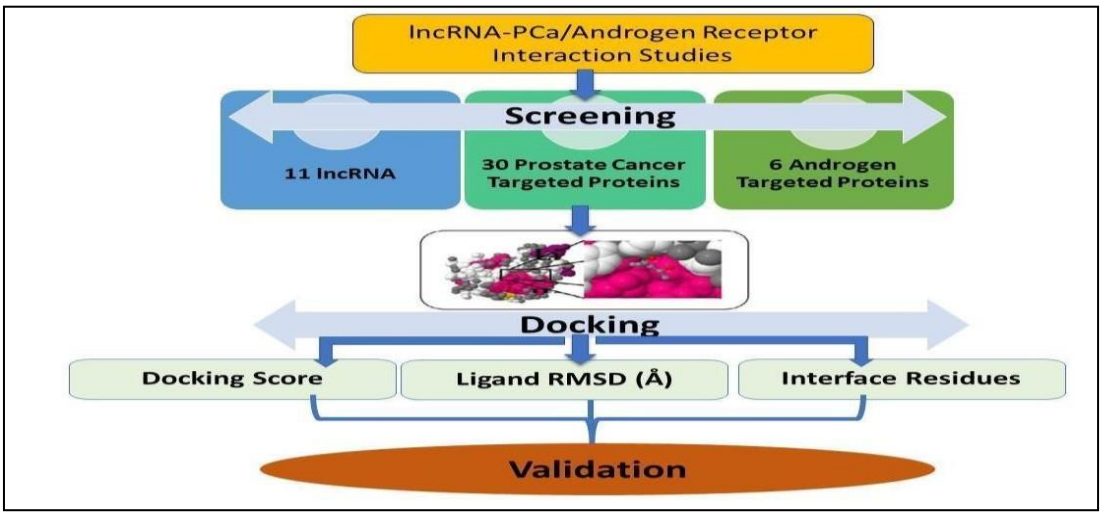


Figure 3.17: Systematic work flow of the 11 lncRNAs and the 36 targeted PCa and AR Protein.

3.2.3.6. *Visualization*:

The docked scores of lncRNA with proteins generated 180 complexes with 10 best poses for each lncRNA-protein resulting in 1800 models. Screening of the models was carried by least-bind energy complexes; 12 models of the complexes were considered to be identified of which, 5 complexes with PCa and 7 complexes with AR. The lowest energy known to have more stability was considered for 3D visualization with Pymol software. The parameters were assigned to ligand site hydrogen bonds with below 3Å. Bond distances were identified that indicate the high intensity and the possible orientation of the proteins resulting in the stable complex formation (Khilwani et al., n.d.), (Unpublished Data).

3.2.3.7. Coding potential:

To check if there is any coding potential attributing to the interfacial residues, we performed checks using intrinsic feature estimation by employing CPC2 (Kang *et al.,* 2017). The tool works on the premise that the open reading frame (ORF) length coverage is estimated along with Fickett and hexamer scores which serve as a prudent classifier for estimating non coding RNAs (ncRNAs). The Fickett score is based on frequencies of A, T, G, Cs learning upon the intrinsic divergence between ncRNAs (Khilwani et al., n.d.), (Unpublished Data).

3.2.4. Correlation of T2DM associated lncRNAs with PCa related lncRNAs identified through RNA-seq study:

We have also used a publically available database generated for type-2-diabetes associated lncRNAs to correlate with the PCa linked lncRNAs identified in our RNA-seq analysis. T2DB database was developed by using three GEO datasets of RNA-seq studies with accession number GSE164416, GSE75988, and GSE115601 in order to discover and annotate genes of T2D related lncRNAs (http://rebeccadistefano.shinyapps.io/T2DM/; last accessed on 17 May 2024). Both the GSE164416 and GSE75988 datasets belongs to expression profiles of lncRNAs identified in T2D and Impaired Glucose Tolerance (IGT) while GSE115601 dataset contain expression data of Diabetic Gastroparesis (Distefano et al., 2023). Therefore, we considered only GSE164416 and GSE75988 datasets and compared them with our RNA-seq study.

3.3. To perform protein-protein interaction (PPI) network analyses and identify pathways associated with T2D and PCa *in silico* method

3.3.1. Interactome network:

For visualizing the interaction network of the commonalities to find associations through their interacting partners, GeneMania is used wherein different types of gene-gene interactions by providing a seed list of our interest and which is then extended to incorporate other genes as interacting partners which are predicted to share the same function based on their overlapping connections in biological pathways (Kour et al., 2023).

3.3.2. Phenolyzer:

In order to prioritize genes related to human diseases or phenotypic data that a user provides in the text form as input, a computational tool was introduced called Phenolyzer (https://phenolyzer.wglab.org/ last accessed on April 12, 2024). It can relate input phenotype to other linked phenotypes, predict the genes which were earlier unknown for their association to any particular disease, integrate large information on previously known genes of a disease, number of factors to help in scoring and prioritizing every potential genes and visualize networks for investigating links between genes and genotype-phenotype relationships (H. Yang et al., 2015).

3.3.3. Gene/Pathway enrichment:

(PEA) Pathway enrichment analysis is based completely on computational biology and a common approach to discover gene functions and disease causing pathways (Chicco & Agapito, 2022) (H. Liu et al., 2022). It determines the biological functions which are over-represented in a gene list and functions are ranked by significance (Chicco & Agapito, 2022). It can also help to get molecular insights of gene sets produced by genomic studies (Reimand et al., 2019) (Ulgen et al., 2018). By using this strategy, it is possible to find the pathways of all biological processes which are more prevalent in a given gene set than would be predicted by chance. Enrichment study is basically performed in 3 steps: selection of genes from the genomics studies data, identification of statistically significant pathways and then result visualization and their interpretations. There are several free tools or software available for pathway enrichment: reactome, g:Profiler, Cytoscape, Gene Set Enrichment Analysis (GSEA), KEGG pathway, DAVID, gene ontology (GO) and reactome (Reimand et al., 2019) (Mubeen et al., 2022) (Hong et al., 2014). It is suggested to always use more than one PEA tool because there is a chance of relegation of results with databases related to specific PEA

software. Therefore, using more than one can probably give more information of associated pathways, some pathways can be confirmed, some can be complementary and some could be conflicting (Chicco & Agapito, 2022).

Steps:

a. Defining a gene list:

The raw data generated from large scale genomic studies are then normalized and scored with a sequence process of computational tools which leaves us with identification of the gene list of our interest (Reimand et al., 2019).

b. Enrichment studies:

After finding the pathways with help of any pathway database, they are filtered by the *P* value <0.05 and fold change value >2 (Reimand et al., 2019).

c. Pathway visualization and result interpretations:

Visualization is like the main pillar in bioinformatics; it not only represents the data but can also deliver us with new or alternative data (Chicco & Agapito, 2022). A gene of interest is possibly involved in more than one pathway and the available databases maintain the hierarchy of pathways by including both specific and non-specific pathways of shared genes. Interpretation is made easier by condensing multiple biological processes into a single matter of concern. To combat redundancy like this visualization is done by Enrichment map using Cytoscape (Reimand et al., 2019).

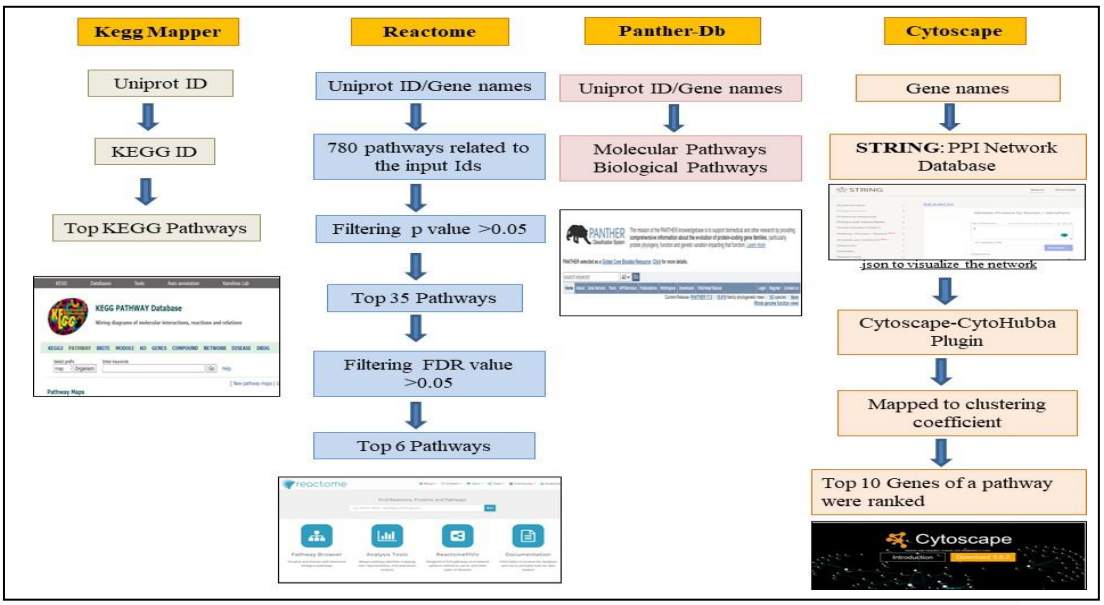


Figure 3.18: Work flow for Gene Enrichment Analysis using Kegg Mapper and Kegg Pathway database, Reactome, Panther-Db and Cytoscape software.

3.3.4. Reactome

Reactome is a database containing details of life processes like molecular signaling transmission, replication, transport, metabolism and several other cellular level processes (https://reactome.org/) last accessed on April 12, 2024. All these processes are arranged in a network of molecular changes in a single model database (Fabregat et al., 2018). Reactome serves as a resource of biological mechanisms as well as a method for identifying unanticipated functional links, for instance, expression level or registries of somatic mutations in tumors (Fabregat et al., 2018). It includes both acquired and inherited disease data. Recently the improved version also incorporated the drugs that could target the disease related signaling pathways (Gillespie et al., 2022). PhD scientists or peer reviewers before curating the reactome annotations manually (Rothfels et al., 2023).

We retrieved the Uniprot IDs from the NCBI Uniprot database and put the list of all the IDs as input into the Analysis tool portal of the Reactome database. We obtained 780 pathways involved with our input. Then we filtered all pathways on the basis of the *p* value and FDR value and the final list narrowed down to the top 6 significant pathways (**Figure 3.18).**

3.3.5. KEGG pathway

All living organisms have protein-protein interactions (PPI), which are one of the most crucial and common biological processes. These interactions (protein-protein) result in several regulatory and biological processes. PPIs are generally categorised into five groups based on various biological implications like; signal transmission, transfer of electrons, contraction of muscles and cell metabolism (Y. H. Zhang et al., 2021). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway is known as a diagram which schematically wires molecules of biological systems to describe the cell functions resulting from the molecular interaction in a network and could be used as guide for the functional reconstruction. Additionally, KEGG includes binary equations that depict molecular relationships and interactions. The KEGG pathway can also be used to compute and make comparisons of the pathways. The KEGG pathway separately displays metabolic pathways and regulatory pathways (https://www.genome.jp/kegg/pathway.html) last accessed on April 12, 2024 (Ogata et al., 1998).

First we took all the Uniprot ID of our gene set and uploaded in the KEGG Mapper portal to obtain the KEGG Ids and then we feed those Ids into KEGG pathway database and selected the *Homo sapiens* as organism to map all KEGG Ids and then we obtained the different pathways with their accession codes. We enlisted all the pathways in which our gene set was involved (**Figure 3.18)**.

3.3.6. Panther database:

PANTHER is a curated database of pathways (https://pantherdb.org/) (Mi et al., 2009). It stands for Protein Analysis through Evolutionary Relationships and is a broadly used publicly available database to classify proteins comprehensively in terms of evolutionary history and their functional roles (Ontology). It offers tools for processing biological data at a large scale. It uses UniProt reference proteomes data set which is organized in homologous gene families. Every family has a phylogenetic tree to depict the evolutionary relation among the genes of a family by accessing all the processes of divergence of gene or protein (speciation, gene duplication, horizontal and vertical transfer) of an organism (Mi et al., 2016). PANTHER GO-slim is the upgraded PANTHER to provide 4 times more gene ontology (GO) for functional classification (Mi et al., 2019).

We have uploaded the list of genes on PANTHER homepage, hit the run and retrieved the functional classification (biological and molecular) and its visualization as pie chart or bar chart or enrichment analysis (**Figure 3.18)**.

3.3.7. Cytoscape:

Cytoscape is free software for integrating molecular interaction networks, enabling network querying, visualization, and integration with phenotypes and biomolecule data. It connects to functional annotation databases, utilizing annotations for static data like protein function ontology and attributes for dynamic molecular interactions. Plug-ins extends its functionality. In Cytoscape's network graphs, molecules are nodes and intermolecular relations are edges. The visualization of nodes and edges as a 2D network is one of the most essential tools to understand molecular interactions (Shannon et al., 2003). Filtering is also supported by Cytoscape by a tool called "Boolean Meta-Filter" which helps in pathway analysis to select the nodes which show large differences (Kohl et al., 2011).

CytoHubba plugin in Cytoscape was integrated that ranked network nodes by features, enabled topological analysis and subnetworks retrieval. Nodes' topological attributes are computed and saved. Clustering coefficient and node color were determined based on degree (Assenov et al., 2008). PPI networks from STRING databases were visualized, and subnetworks formed. Network Analyzer Plugin calculated topological parameter like connectivity, node number, connecting edges, clustering coefficient, average clustering coefficient, centralization, connectivity degree, etc. (Gollapalli et al., 2021) (**Figure 3.18)**.

CHAPTER 4: RESULTS

4.1. To screen variants in PCa and diabetes using integrated systems genomic approaches

4.1.1. Results of Meta-analysis Study

Comparison of vivid datasets yield candidate genes common to Diabetes and PCa:

By comparing variants of all the individual comorbidities with PCa, we identified several common genes and variants along with some common protein change (Kour et al., 2023).

4.1.1.1. Comparison of ClinVar unveiled BLM as a key player in Diabetes, PCa and Obesity:

It showed 27 common genes between PCa and diabetes, 8 common genes between PCa and obesity, 69 common genes between diabetes and obesity, and 4 among all the three diseases. *DNAJC6, SDHA, WRN, RET, MRE11, NF1, GNRHR, HSPB1, ACAT1, SHH, BIN1, STAT3, ZIC2, FBP1, PARK7, RGS2, SYNE2, ALDH6A1, BBOF1, HNF1B, AKT1, CACNA1H, KDM6B, KDM3B, HP, TXNL4B, BMP4, IRS1, and NC*L are common in PCa and diabetes. *BRCA2, FANCI, BAP1, AR, and PMS2* are common in PCa and obesity. *AHDC1, TOP3A, LEP, BBS12, FOXP3, BBS7, GATA6, GCK, POLG, GHRL, BBS1, SCN1A, BTK, CYP19A1, ASTN2, PDE4D, SHANK3, WDPCP, RPL36A-HNRNPH2, MKKS, IFT74, MC4R, BBS9, SLC12A3, GHRLOS, DMXL2, PAX6, TTC8, BBS4, NUDC, CEP19, NR0B2, IL1RN, BBS5, GLI2, TGIF1, LPL, MIR4713HG, ALMS1, UCP3, APC, ARL6, DARS2, LZTR1, SDCCAG8, PHKG2, MECP2, PROK2, NPHP1, SMARCB1, ATP7A, BBS10, ZDHHC24, APOA5, MPO, AVPR2, LOC106694316, DYRK1B, TMEM67, BBS2, CLCN5, GJB2, CORIN, ABCD1, ABCB4, TRIM32, MAGEL2, INSR, FOXP1, DDX3X, and APPL1* are common in diabetes and obesity with variants. *BLM, TMEM67, RFX6, NR0B2,* and *NUDC* were found common among PCa, diabetes mellitus, and obesity (**Figure 4.1**) (Kour et al., 2023).

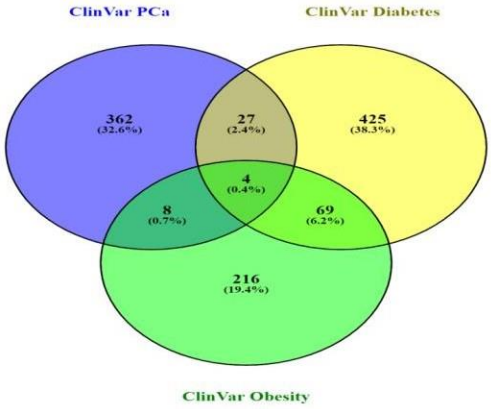


Figure 4.1: Venn plot for identifying commonalities between Clinvar datasets of PCa, diabetes, and obesity. A total of 27 common genes between PCa and diabetes, 8 common genes between PCa and obesity, 69 common genes between diabetes and obesity, and 4 among all the three diseases.

4.1.1.2. Overlapping variants in PCa ClinVar and GWAS central:

GWAS comparison yielded **rs721048** (*EHBP1*) and **rs138213197** (*HOXB13*), which were found to be in ClinVar of PCa and GWAS central. In addition, 5 variants of *BRCA2* from exome data and PCa ClinVar data are seen as common, i.e. **rs145988146, rs80358600, rs276174854, rs276174889, rs771203198,** and 1 variant of *BRCA1* **(rs28897696)** GWAS Central common to both of them (**Figure 4.2**) (Kour et al., 2023).

4.1.1.3. More variant alliances between PRACTICAL Consortium GWAS Central and Clinvar of PCa:

Whereas PRACTICAL consortium comparison *yielded* **rs721048** (*EHBP1*) and **rs138213197** (*HOXB13*) were found earlier common to ClinVar PCa and GWAS Central, 17 other variants are also found common in the PRACTICAL consortium and ClinVar. Among them, 8 belongs to *BRCA1* (**rs147297981, rs181430678, rs147509580, rs182524124, rs148500539, rs147856441, rs148068102, rs149141411,** 2 of *FANCM (***rs147021911** and **rs144567652**), and 1 of each *MAD1L1* (**rs121908982**), *MSH6* (**rs1800937**), *PALB2* (**rs45494092**), *PNKP* (**rs201872477**), *XPC* (**rs182616621**), *POLD1* (**rs149366027**), and *EXOC8* (**rs148264842**) (**Figure 4.2.B**). No commonality is observed between exome data and obesity (**Figure 4.2.A**). One common variant **rs61816761** associated with the *FLAG* gene is found common to obesity and diabetes (**Figure 4.2.C and D**) (Kour et al., 2023).

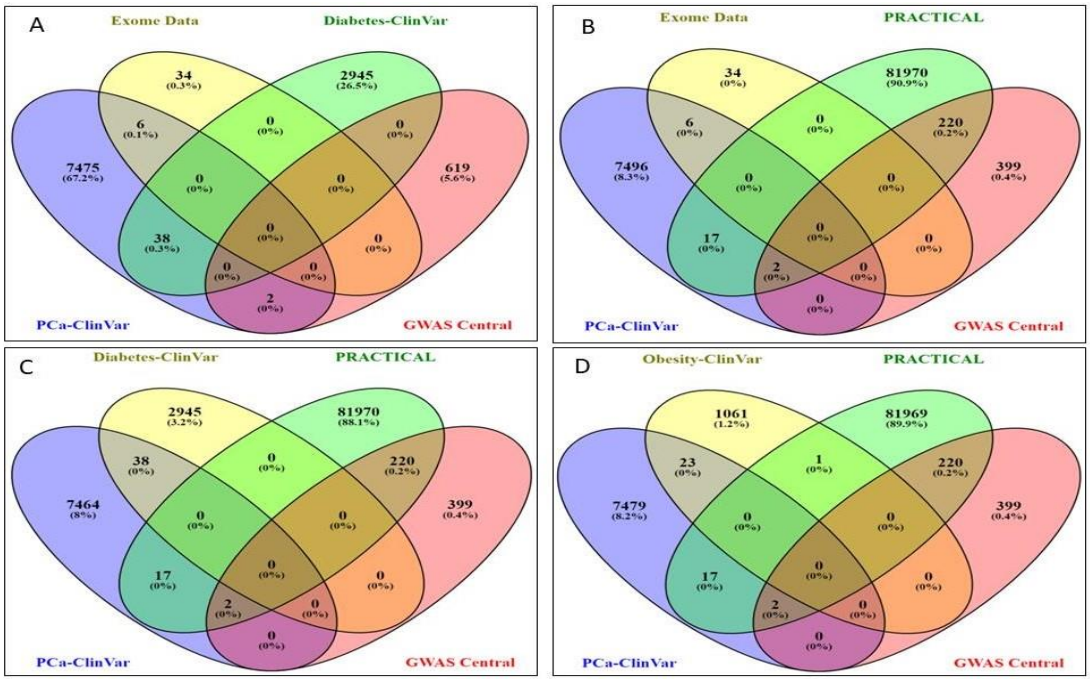


Figure 4.2: A, B: Venn Plot to identify common variants among ClinVar PCa, diabetes, GWAS central, PRACTICAL consortium for PCa, and PCa exome data. No commonality is observed between exome data and obesity. PRACTICAL consortium comparison ***yielded rs721048*** (*EHBP1*) and **rs138213197** (*HOXB13*) were found earlier common to ClinVar PCa and GWAS Central, 17 other variants are also found common in the PRACTICAL consortium and ClinVar; **C, D**: Venn Plot to identify common variants among ClinVar PCa, Diabetes, Obesity, GWAS central, PRACTICAL consortium for PCa and PCa exome data. One common variant **rs61816761** associated with the *FLAG* gene is found common to obesity and diabetes.

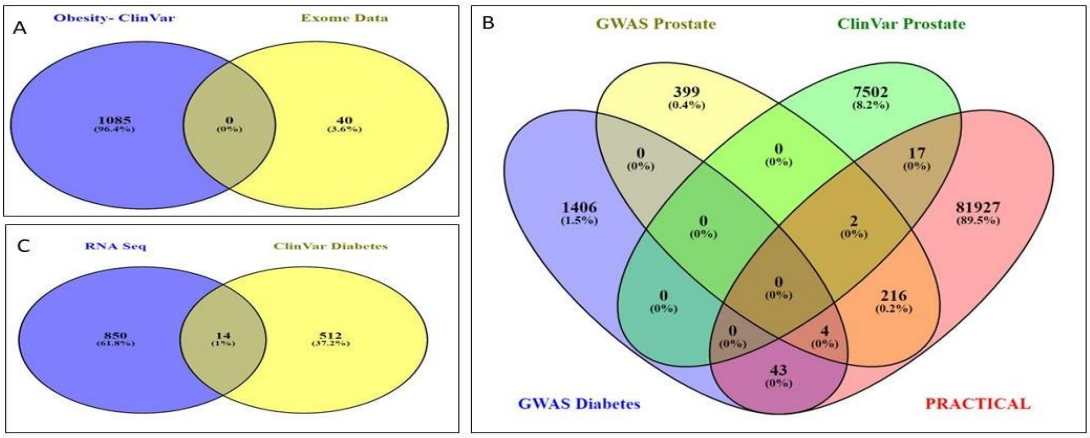


Figure 4.3: A: Venn plot showing no commonality between exome data and obesity ClinVar data; **B:** Venn plot for GWAS central data of Diabetes and PCa, PRACTICAL Consortium,

and ClinVar PCa; **C:** Venn plot showing some commonalities between RNAseq results of three PCa samples and ClinVar data of Diabetes

No commonality was seen between ClinVar Obesity and exome data (**Figure 4.3.A**), 2 commonalities were seen among ClinVar PCa data, GWAS data of PCa and PRACTICAL Consortium, 216 common variants between GWAS Central PCa and PRACTICAL Consortium and only 4 variants common among both the GWAS Central of PCa and Diabetes with PRACTICAL Consortium data (**Figure 4.3B**). In **Figure 4.3.C,** we observed 14 commonalities in RNA Seq and ClinVar Diabetes (Kour et al., 2023).

4.1.1.4. ***Comparison of RNA-Seq data revealed PP1MB and SFTPC as common:***

On comparing RNA-seq data of both PCa and DT2M with all the datasets ***PP1MB*** and ***SFTPC*** were common in both types of RNA seq and PCa ClinVar data.

We used Gene cards to study the pathways involved with the identified genes (**Table 4.1 and Table 4.2)** (Kour et al., 2023).

4.1.1.5. Regression algorithms on PCa, diabetes and obesity data sets suggest potential improvements for predictive models:

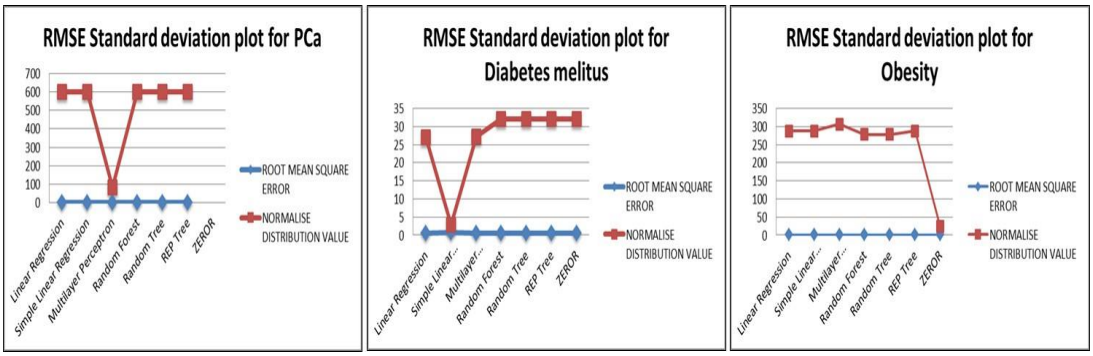


Figure 4.4: Linear plots for RMSE standard deviations of PCa, DM and Obesity

The annotated PCa, diabetes, and obesity data sets consist of three attributes: protein change, clinical significance (last reviewed), and semi binary value to clinical relevance. The first two attributes form the independent variables while the third denotes the dependent variable. We evaluated the following regression algorithms (with the default set of parameters) on each of the PCa, diabetes, and obesity datasets using a train test split of 70:30. Based on the RMSE values and the average, the standard deviation (PCa=0.001; DM=0.012; Obesity=0.0013) and then the normalized deviation value is calculated and plotted against each other by line graphs (**Figure 4.4**). The results can be further improved by including more independent variables

with respect to each dataset. The linear regression and tree-based algorithms have a lower RMSE than the multilayer perceptron algorithm for the PCa dataset. The linear regression has the lowest RMSE as compared to multilayer perceptron and tree-based algorithms for the diabetes dataset. The tree-based algorithms have the lowest RMSE as compared to linear regression and multilayer perceptron algorithms for the obesity dataset (Kour et al., 2023).

Table 4.1: Functions and pathways involved in the common genes among ClinVar PCa and ClinVar diabetes from Genecards (Kour et al., 2023)

GENE	FUNCTION	PATHWAYS INVOLVED
WRN	Magnesium and ATP-dependent DNA multifunctional helicase enzyme, 3'->5' exonuclease (ds DNA) at 5' overhangs. Plays key roles in joint DNA molecule dissociation which might end up giving homologous recombination products, formation of focal centers while replication, and ds break repair after gamma-irradiation. Also, it increases DNA polymerase obstruction at DNA lesion sites.	. Homology Directed Repair . Homologous DNA Pairing and Strand Exchange . Regulation of TP53 Activity . Resolution of D-Loop Structures . SUMOylation . Cell Cycle Checkpoints . Gene Expression . Metabolism of proteins . DNA Damage . Cell Cycle, Mitotic . Telomere C-strand (Lagging Strand) Synthesis . DNA damage NHEJ mechanisms of DSBs repair . Regulation of Telomerase
DNAJC6	Promotes uncoating of clathrin-coated vesicles by recruiting HSPA8 or HSC70 and clathrin-mediated endocytosis in neurons	. Clathrin derived vesicle budding . Vesicle-mediated transport . Clathrin-mediated endocytosis
RET	Plays role in neuronal navigation, cell proliferation, cell differentiation, and cell migration, when upon binding with glial cell-derived neurotrophic factor family ligands, PTK2/FAK1 phosphorylation and regulates the balance between both cell death and survival. Active without ligand and triggers apoptosis by a process that needs receptor intracellular caspase cleavage. Behave as a dependence receptor in the	. RET signaling . Developmental Biology . Cytokine Signaling in Immune system . Innate Immune System . Signaling by GPCR . Tyrosine Kinases / Adaptors . VEGF Pathway (Tocris) . G-protein signaling-RAS

	presence of the ligand GDNF in somatotrophs and promotes survival and downregulates growth hormone (GH) production, and if GDNF is absent triggers apoptosis	regulation pathway . Aryl Hydrocarbon Receptor . Dopaminergic Neurogenesis . Signaling events regulated by Ret tyrosine kinase . Sudden Infant Death Syndrome (SIDS) Susceptibility Pathways
NF1	NF1 is known to stimulate the Ras GTPase activity. It might be regulating the activity of Ras, has more affinity for Ras GAP, and lessens its particular activity.	. MAP Kinase Signaling . Ras Signaling . Endometrial Cancer . Development of VEGF signaling and activation . Oncogenic MAPK signaling . Integrated breast cancer Pathway . Prolactin Signaling
TMPO	Encodes many different LEM domains comprising isoforms of proteins that are involved in gene expressions, replication, cell cycle control, and chromatin organization. Alpha isoform encoded by it is mostly diffused in the nucleus and has lamin binding domain, whereas beta and gamma isoforms get located on nuclear membranes containing HDAC3 interaction domains.	. Cell Cycle, Mitosis (M Phase) . Nuclear envelope reassembly . Transport of the SLBP independent Mature mRNA . Mitotic Metaphase and Anaphase (depolymerisation of the nuclear lamina) . 5. Apoptosis and Autophagy

Table 4.2: Details of common genes retrieved using genecards (Kour et al., 2023)

GENE	FUNCTION (UniProtKB)
SFTPC	Elevates alveolar stability by reducing surface tension at the air-liquid interface in the peripheral air spaces.
PPM1B	It encodes an enzyme that has large specificity. This enzyme can dephosphorylate PRKAA1 and PRKAA2; CDK2 and CDK6 in vitro. Its dephosphorylation at 'Ser-17'in can inhibit TBK1-mediated antiviral signaling. Has an important role in terminating TNF-alpha-mediated NF-kappa-B activation by dephosphorylating and inactivating IKBKB/IKKB
PPP2CA	Important phosphatase for microtubule-associated proteins (MAPs), modulates the phosphorylase B kinase casein kinase 2 activity, MAP-2 kinase, and mitogen-stimulated S6 kinase; protects centromeric cohesion in oocytes especially during meiosis I; can

	dephosphorylate SV40 large T antigen as well as p53/Tp53; activation of RAF1 by dephosphorylating it at 'Ser-259'; dephosphorylation of WEE1 which prevents its ubiquitin-mediated proteolysis; increase levels of WEE1 protein; G2/M checkpoint promotion; dephosphorylation of MYC and its ubiquitin-mediated proteolysis; dephosphorylation of FOXO3 which promotes its stabilization
PPARG	It is a nuclear receptor and binds peroxisome proliferators like fatty acids and hypolipidemic drugs; modulates the transcription of its target genes like acyl-CoA oxidase; important regulator of glucose homeostasis and adipocyte differentiation; critical regulator of gut homeostasis through NF-kappa-B-mediated proinflammatory responses suppression; regulates the transcription of ARNTL/BMAL1 in the blood vessels which controls cardiovascular circadian rhythms
TMEM67	It is crucial for the structure and operation of cilia; potentially involved in the regulation of the composition of ciliary membranes. It facilitates the migration of the centrosome to the apical cell surface in the early stages of ciliogenesis. *TMEM67* also participates in the maintenance of optimal cilia length and the correct number by overseeing centrosome duplication. Further it helps in development of the cell branching morphology and plays important role in endoplasmic reticulum-associated degradation (ERAD) process of surfactant
FOXP1	As a transcriptional repressor, it collaborates with CTBP1 to influence lung epithelium differentiation, regulating secretory cell fate and inhibiting goblet cell lineage. It also plays a vital role in B-cell development, cardiac muscle cell proliferation, and spinal motor neuron organization. Additionally, it contributes to midbrain identify, T follicular helper cell regulation, and hair follicle stem maintenance, promoting B-cell expansion by inhibiting caspase-dependent apoptosis.

4.1.1.6. Mutational analysis

4.1.1.6.1. FOXP1: a highly mutated in Prostate adenocarcinoma

On putting different queries of individual genes, it summarizes the genomic alterations across the whole sample lot, given the details about the frequency of gene mRNA (RNA-seq FPKM) related to its mutations from the selected study. Also, the graphical representation of protein domains and specific regions of a particular mutation in a gene is provided. Graphs showed *FOXP1* with the highest 8 % of alteration (amplification, deep deletions, in frame and missense mutation with unknown significance, and truncated mutations in putative driver), *SFTPC* with an overall 5% of alteration which includes amplification and deep deletions (shown in blue), 1.8 % in *PPP2CA* (amplification, deletions, missense mutation with

unknown significance), *PPARG* with an overall 1.5 % of alteration (amplification, deletions, missense mutation with unknown significance), *TMEM67* with overall 6 % of alteration which includes amplification, deep deletions, missense and truncated mutation with unknown significance, 1.5 % of alteration (amplification, deletions, and missense shown in green) for *PPM1B* gene, *TMPO* with 0.1% of truncated mutation and *BLM* with 0.3% of mutations (missense with unknown significance, truncation as putative driver and deep deletion) (**Figure 4.5)** in a prostate adenocarcinoma study of Armenia *et al.,* (2018) (Kour et al., 2023).

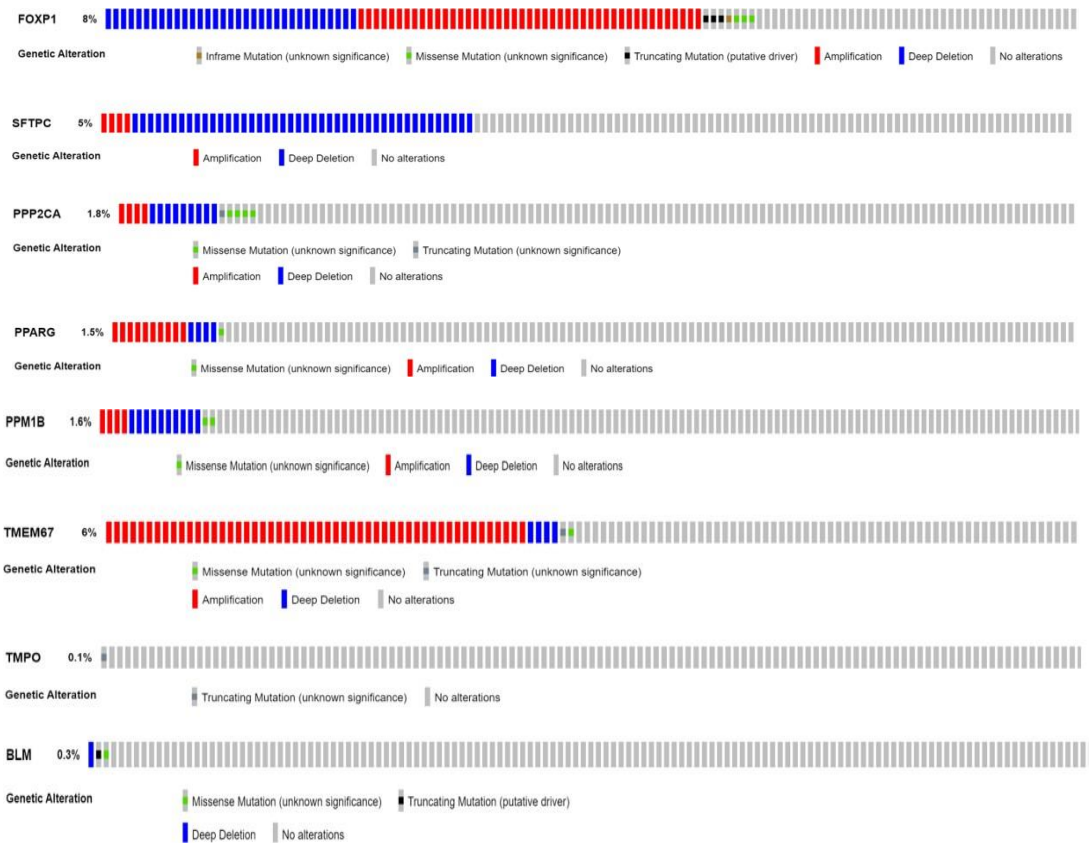


Figure 4.5: Overall percent and type of genetic mutations in *FOXP1* (8%), *SFTPC* (5%), *PPP2CA* (1.8%), *PPARG* (1.5%), *PPM1B* (1.6%), *TMEM67* (6%), *TMPO* (0.1%), and *BLM* (0.3%) related to prostate adenocarcinoma.

4.1.1.7. Gene expression patterns highlighted PPP2CA and FOXP1 as highly expressed genes:

Using normal expression profile graphs (transcripts per million mapped reads, a.k.a., TPM) and box plots with the help of GEPIA 2, we performed a comparative expression analyzed. Box Plots were divided based on quartiles, with every box depicting the median range of

expression of a particular gene in both normal and tumor samples separately. A horizontal bar in the middle of all boxes is the actual median of the expression and both medians of tumor and normal are different. Outside the box both below and above sets a deviation limit, beyond that is known for outlier regions (abundant expression) (**Figure 4.6a**). Outliers in normal sample expressions might be a chance of experimental error or error in replicates. The prostate adenocarcinoma (PRAD) dataset was used to compare 492 tumors with 152 normal sample expression data and later the multigene expression comparison was rendered based on Z scores (Kour et al., 2023).

In comparison, only tumor tissue expression and by matching TCGA normal and GTEx data, we found that the *PPP2CA* (5.5) and *FOXP1*(4.9) is highly expressing genes as compared to others, whereas TMEM67, PPARG are low expressing, and *TMPO*, *SFTC* and *PPM1B* are least expressing genes (**Figure 4.6b**)TMPO showed a very small change in expression profile from the normal but with a slightly higher deviation from the median expression in tumor (**Figure 4.6c**) and the most outlied expression is the seen for *PPARG* gene (**Figure 4.6a**) (Kour et al., 2023).

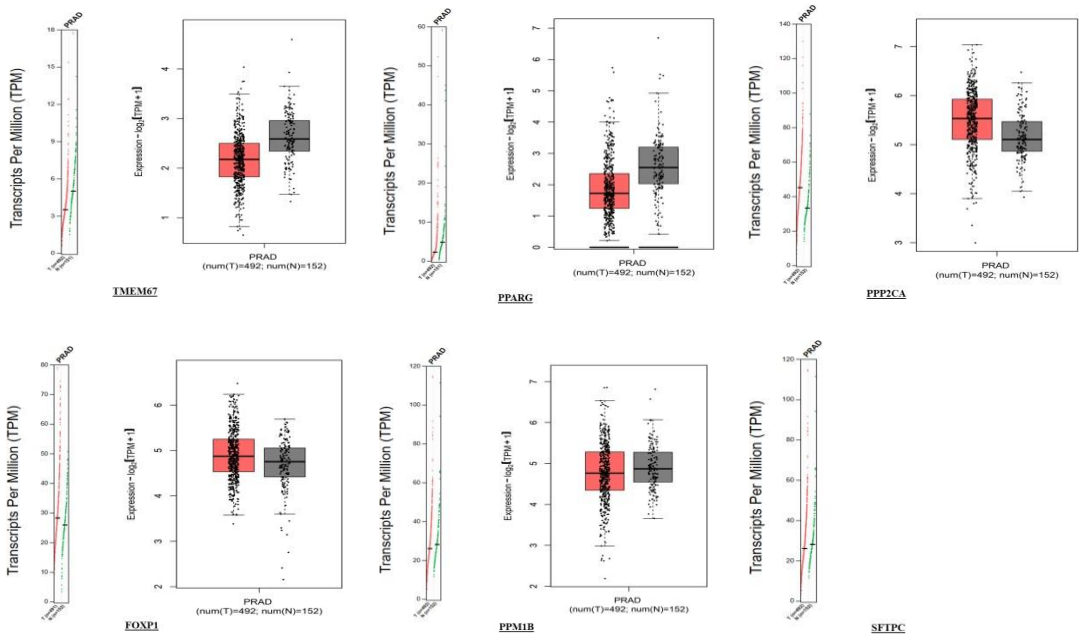


Figure 4.6.a: Box plots for expression of common genes compared to TCGA tumor-normal datasets of Prostate adenocarcinomas (PRAD). Every box depicts the median range of expression of a particular gene in both normal and tumor samples separately. A horizontal bar in the middle of all boxes is the actual median of the expression and both medians of

tumor and normal are different. Outside the box both below and above sets a deviation limit, beyond that is known for outlier regions (abundant expression)

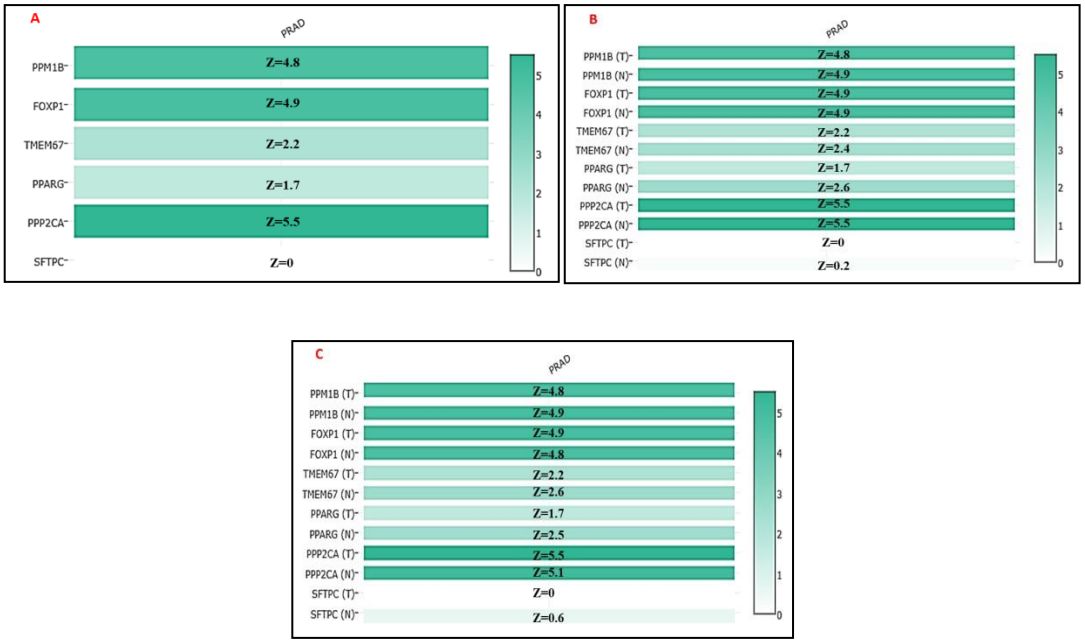


Figure 4.6.b: Box plots for expression of common genes in compared to TCGA tumor-normal datasets of Prostate adenocarcinomas (PRAD) and GTEx data (multiple genes based on Z scores) **A**. Only tumor tissue expression; **B**. Match TCGA normal data; **C**. Match TCGA normal and GTEx data, the plots showed *PPP2CA* (5.5) and *FOXP1* (4.9) is highly expressing genes as compared to others, whereas, the *TMPO, SFTC* and *PPM1B* are the least expressing genes.

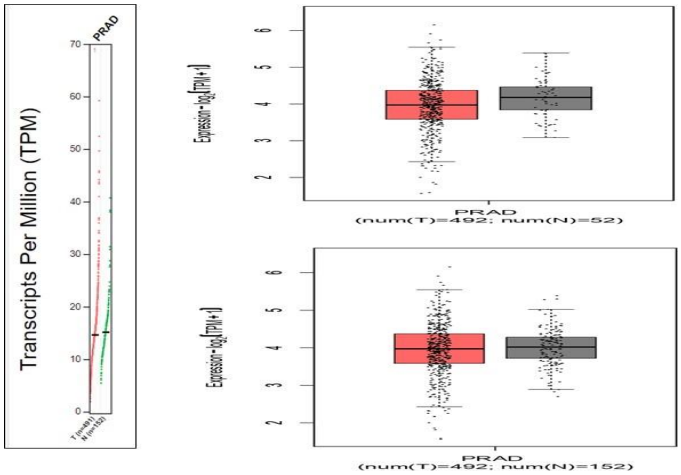


Figure 4.6.c: Expression in comparison to TCGA tumor-normal dataset of Prostate Adenocarcinomas (PRAD) dataset and GTEx data for *TMPO* gene

4.1.1.8. ***BLM showed significant impact on disease free survival in survival analysis:***

On analysing we found that the overall survival for *FOXP1* (log rank *p*=0.39), *TMPO* (log rank *p*=0.14), *TMEM67* (log rank *p*=0.24), *BLM* (log rank p=0.12), *PPP2CA* (log rank *p*=0.12), *PPARG* (log rank *p*=0.53) and *PP1MB* (log rank *p*=0.87).

Disease-free survival for *FOXP1* (log rank p=0.005), *TMPO* (log rank *p*=0.21), *TMEM67* (log rank *p*=0.55), ***BLM*** (log rank ***p*=0.00065**), *PPP2CA* (log rank *p*=0.5), *PPARG* (log rank *p*=0.13) and *PPM1B* (log rank *p*=0.88) (**Figure 4.7**). In a comparison of overall survival and disease-free survival in reference to the significance, only *BLM* is observed to have a significant p-value in disease-free survival. The insignificance with respect to other genes of *p*-value is may be because of the individual genetic variability, each study exhibits clinical heterogeneity. Therefore we need to see higher sample size studies (Kour et al., 2023).

OVERALL SURVIVAL **DISEASE FREE SURVIVAL**

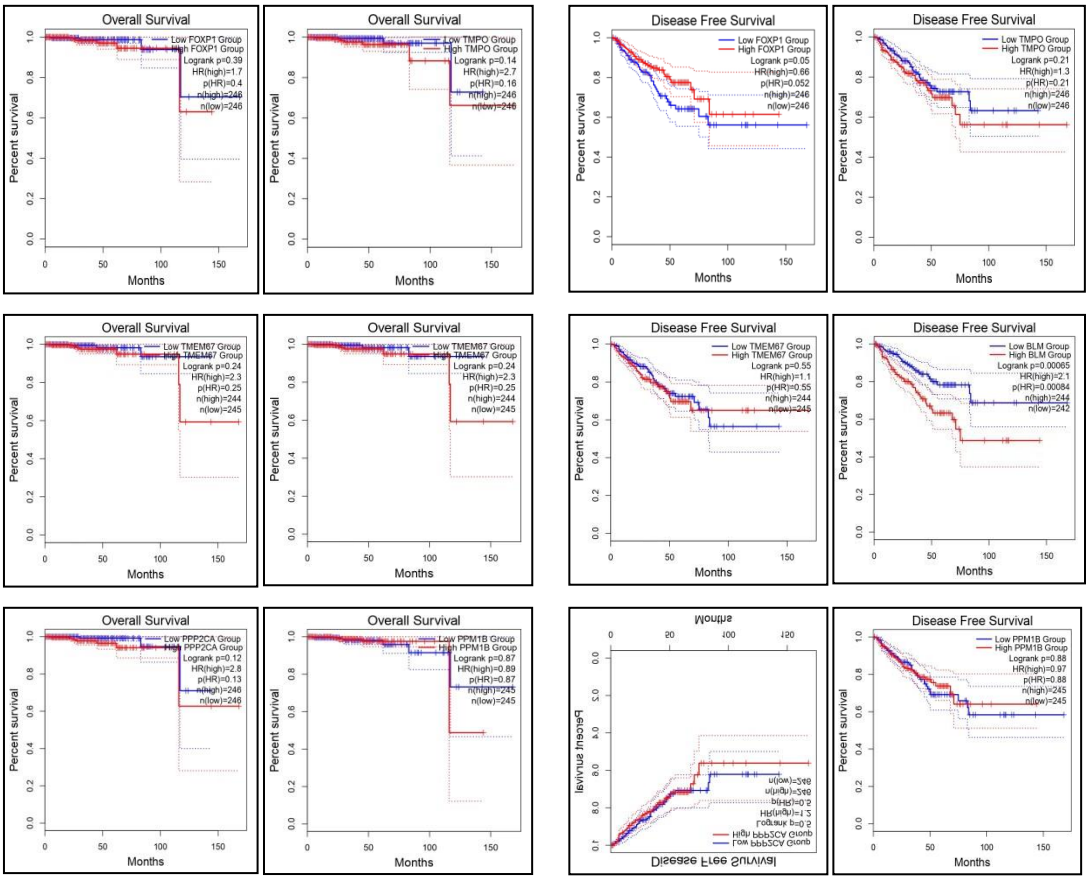


Figure 4.7: Comparison of overall survival (OS) and disease-free survival (DFS) bases on log *p* value among *FOXP1, TMPO, TMEM67, TMPO, BLM, PPP2CA*, *PPM1B*, and *PPARG using* GEPIA 2 tool with PRAD dataset for both TCGA normalized data and GTEx data.

4.1.2. Causal relationship between PCa (PCa) and type-2-diabetes (T2D) through WES

The DNA was isolated from FFPE blocks and then it was checked with the help of NanoDrop 1000 which showed good quality of isolated DNA on the basis of OD260/OD280 and OD260/OD230 ratios. Also, the quantified DNA using PCR passed the quality check when we runned it on 1% agarose gel.

4.1.2.1. Quality Check

4.1.2.1.1. DNA QC report:

DNA quantity check: Extracted DNA quantity was checked with the help of Nanodrop 1000, average of two individual data points is provided below in the table 5.1. By looking at both the A260/280 ratio (1.9) and A260/230 ratio (2.0-2.2) it suggested good quality as per the report given and passed the check **(Table 4.3)**. **Figure 4.8** shows the all 5 sample bands in the smear on running agarose gel electrophoresis.

Table 4.3: Table representing the DNA quality at both ratio A260/280 (1.9) and A260/230 (2.0-2.2) by using Nanodrop 1000

Sr. No.	Case ID	Case PPF Block Name	Nanodrop (ng/µl)	A260/280	A260/230	QC Remarks
1	20307401271	H20/6591	3088.7	1.98	2.29	Pass
2	20307401272	H20/8329	1866.5	1.96	2.25	Pass
3	20307401273	H20/8030	1564	1.97	2.24	Pass
4	20307401274	H20/7111	1784.2	1.98	2.24	Pass
5	20307401275	H20/7255	3070.6	1.98	2.29	Pass

Note: This table provides the quality control (QC) metrics for DNA samples extracted from FFPE blocks, including cases IDs, Nanodrop concentrations (ng/µl), and absorbance ratios (A260/280 and A260/230). The A260/280 ratio indicated protein contamination, while, the A260/230 ratio reflected the presence of organic compounds and impurities. All samples passed QC, demonstrating acceptable purity and concentration for downstream application.

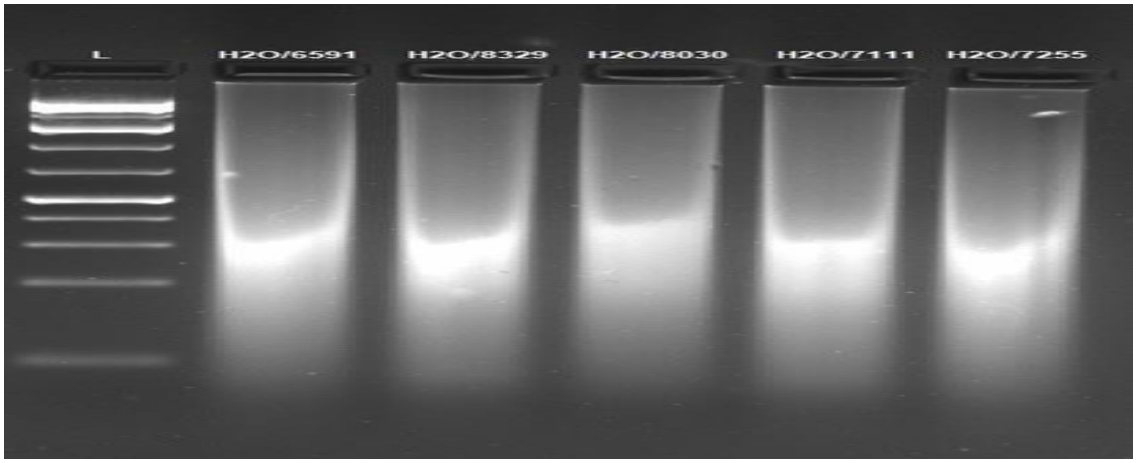


Figure 4.8: Image of agarose gel electrophoresis of all tumor samples and Ladder

4.1.2.1.2. cDNA QC report:

Acquired sizes of all libraries are reported in the following (**Table 4.4).**

Table 4.4: Table showing details of quality check of cDNA

Sr. No.	Case PPFE Block Name	ng/µl	Insert Size	Index	QC remarks
Sample 1	H20/6591	**18.5**	302	C10	**Pass**
Sample 2	H20/8329	**15.5**	368	D10	**Pass**
Sample 3	H20/8030	**25.0**	429	E10	**Pass**
Sample 4	H20/7111	**10.2**	426	F10	**Pass**
Sample 5	H20/7255	**20.0**	**296**	**G10**	**Pass**

Note: The table summarizes the quality control (QC) metrics for DNA samples extracted from FFPE blocks, including the concentration (ng/µl), insert size (bp), index designation, and QC remarks for each sample. The values indicate acceptable DNA concentration and insert sizes for subsequent analysis, with all samples passing the QC assessment.

4.1.2.2. Molarity of electronic ladder and sample:

Different calibration concentrations ranging from 78.4pg/µl (lowest) up to 1110pg/µl (highest) as per the different fragment lengths were used and to determine the modal fragment length of sequencing pool by running on D1000 HS screen tape of Agilent Tapestation 4150. Later, calculation of the concentrations and molarity of the sequencing pool was done. An electronic ladder was included in each run which contained one lower and one upper mark to align the ladder data against the samples so that size of a fragment can be determined. The size of a library was calculated within two regions and later normalized before sequencing. Each region showed a normalized peak value. The upper value determines the molar sample concentration. The electropherogram report shows the size of different fragments on x-axis mostly ranging from 150-200 base pairs in a smear **Figure 4.9-4.14** shows TapeStation Electropherogram report of sample representing the size of different fragments, sample intensity, peak values of lower mark, and the a major peak values of ladder and all 5 samples.

EL1: Electronic Ladder:

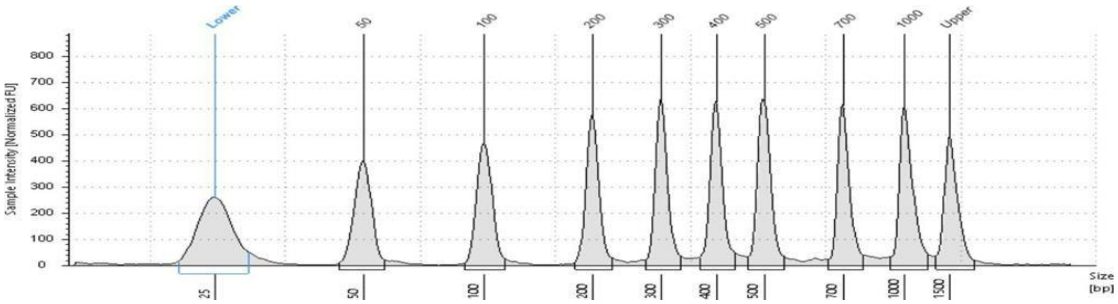


Figure 4.9: The TapeStation Electropherogram report of electronic ladder. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values as per the different size of fragments seen ranging from 50-1500 bps.

Size of fragments, calibrated concentrations, peak molarity, % age integrated area and their concerned observations are tabulated separately (sample table and peak table) for each sample and electronic ladder from **Table 4.5-Table 4.16.**

Table 4.5: Sample table for electronic ladder

Size (bp.)	Calibrate Conc. [pg/µl]	Assigned Conc. [pg/µl]	Peak Molarity [pmol/l]	%Integrated Area	Peak Comment	Observation
25	340	-	20900	-		Lower Marker
50	265	-	8160	11.28		
100	278	-	4270	11.82		
200	290	-	2230	12.32		
300	304	-	1560	12.95		
400	306	-	1180	13.00		
500	312	-	961	13.29		
700	286	-	629	12.19		
1000	309	-	476	13.15		
1500	250	250	256	-		Upper Marker

Note: This table represents the characterization of DNA fragments based on size (in bp) and their corresponding calibrated and assigned concentration (pg/µl), peak molarity (pmol/l), and integrated area percentages. The data include lower and upper marker observations, with the last row indicating and assigned concentration for the 1500 bp fragment. This information is crucial for assessing the quality and quantity of DNA samples.

Table 4.6: Peak Table for ladder

Well	Sample Description	Alert	Observation
EL1	Electronic Ladder		Ladder

Note: Summarizes the peak analysis for the electronic ladder EL1, which serves as reference for determining the sizes of DNA fragments. No alerts indicate no issues with the ladder's performance.

H20/6591

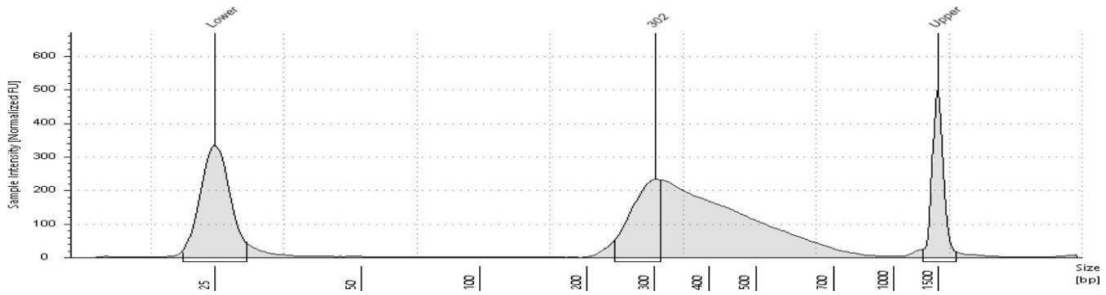


Figure 4.10: The TapeStation Electropherogram report of sample (gDNA) H20/6591. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values of lower mark, a major peak in the region of 275-305 bps

Table 4.7: Sample table for sample 1

Well	Sample Description	Alert	Observation
B2	NCGM:794-H20/6591		

Note: Table presents the sample information for sample 1 analyzed in the study. There are no alerts and observations noted, suggesting that the sample was processed without any issues or anomalies during the analysis.

Table 4.8: Peak table for sample 1

Size (bp.)	Calibrate Conc. [pg/µl]	Assigned Conc. [pg/µl]	Peak Molarity [pmol/l]	%Integrated Area	Peak Comment	Observation
25	457	-	28100	-		Lower Marker
302	301	-	1560	100.00		
1500	250	250	256	-		Upper Marker

Note: This table represents the characterization of DNA fragments of sample 1 based on size (in bp) and their corresponding calibrated and assigned concentration (pg/µl), peak molarity (pmol/l), and integrated area percentages, The data indicates lower mark at 25bp, a sample peak at 302 bp with full integrated area, and upper marker at 1500 bp with an assigned concentration. This information is crucial for assessing quality and quantity of DNA fragments in samples.

H20/8329

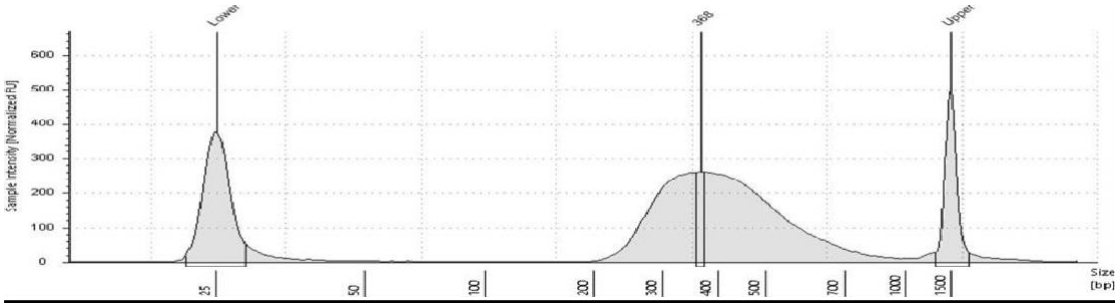


Figure 4.11: The TapeStation Electropherogram report of sample (gDNA) H20/8329. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values of lower mark, a major peak in the region of 360-380 bps

Table 4.9: Sample table for sample 2

Well	Sample Description	Alert	Observation
A1	NCGM:794-H20/8329		

Note: Table presents the sample information for sample 2 analyzed in the study. There are no alerts and observations noted, suggesting that the sample was processed without any issues or anomalies during the analysis.

Table 4.10: Peak Table for sample 2

Size (bp.)	Calibrate Conc. [pg/µl]	Assigned Conc. [pg/µl]	Peak Molarity [pmol/l]	%Integrated Area	Peak Comment	Observation
25	442	-	27200	-		Lower Marker
368	78.4	-	328	100.00		
1500	250	250	256	-		Upper Marker

Note: This table represents the characterization of DNA fragments of sample 1 based on size (in bp) and their corresponding calibrated and assigned concentration (pg/µl), peak molarity (pmol/l), and integrated area percentages, The data indicates lower mark at 25bp, a sample peak at 368 bp with full integrated area, and upper marker at 1500 bp with an assigned concentration. This information is crucial for assessing quality and quantity of DNA fragments in samples.

H20/8030

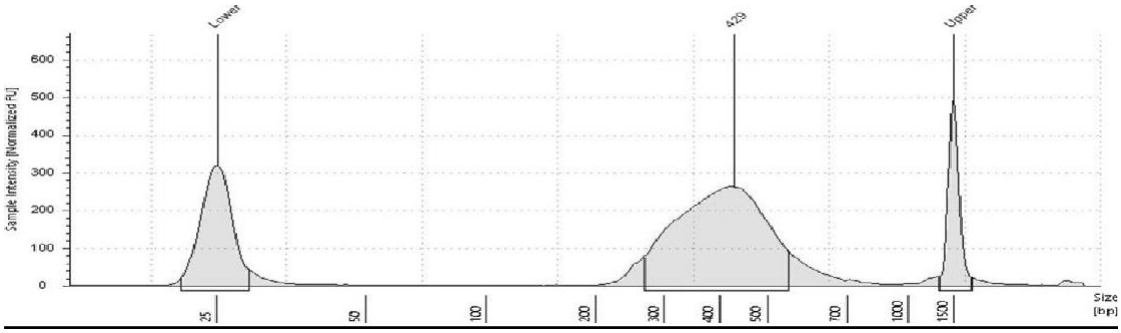


Figure 4.12 The TapeStation Electropherogram report of sample (gDNA) H20/8030. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values of lower mark, a major peak in the region of 275-530 bps

Table 4.11: Sample Table for Sample 3

Well	Sample Description	Alert	Observation
G2	NCGM:794-H20/8030		

Note: Table presents the sample information for sample 3 analyzed in the study. There are no alerts and observations noted, suggesting that the sample was processed without any issues or anomalies during the analysis.

Table 4.12: Peak Table for sample 3

Size (bp.)	Calibrate Conc. [pg/µl]	Assigned Conc. [pg/µl]	Peak Molarity [pmol/l]	%Integrated Area	Peak Comment	Observation
25	466	-	28700	-		Lower Marker
429	1110	-	3980	100.00		
1500	250	250	256	-		Upper Marker

Note: This table represents the characterization of DNA fragments of sample 1 based on size (in bp) and their corresponding calibrated and assigned concentration (pg/µl), peak molarity (pmol/l), and integrated area percentages, The data indicates lower mark at 25bp, a sample peak at 429 bp with full integrated area, and upper marker at 1500 bp with an assigned

concentration. This information is crucial for assessing quality and quantity of DNA fragments in samples.

H20/7111

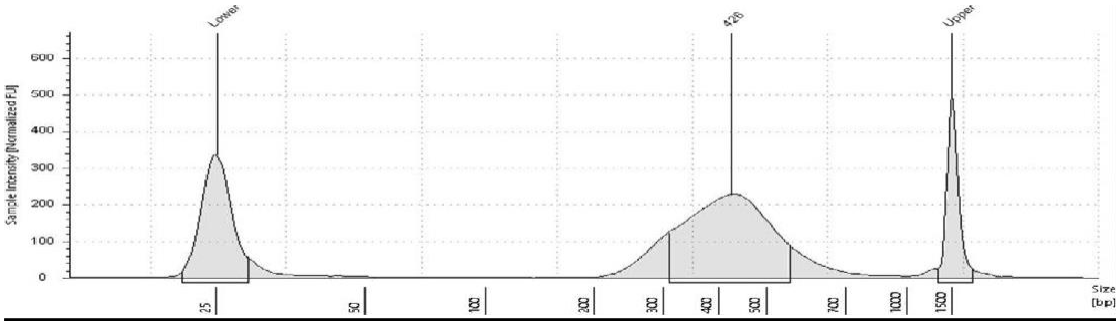


Figure 4.13: The TapeStation Electropherogram report of sample (gDNA) H20/7111. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values of lower mark, a major peak in the region of 305-540 bps

Table 4.13: Sample table for Sample 4

Well	Sample Description	Alert	Observation
F1	NCGM:794-H20/7111		

Note: Table presents the sample information for sample 4 analyzed in the study. There are no alerts and observations noted, suggesting that the sample was processed without any issues or anomalies during the analysis.

Table 4.14: Peak table for Sample 4

Size (bp.)	Calibrate Conc. [pg/µl]	Assigned Conc. [pg/µl]	Peak Molarity [pmol/l]	%Integrated Area	Peak Comment	Observation
25	444	-	27300	-		Lower Marker
426	804	-	2900	100.00		
1500	250	250	256	-		Upper Marker

Note: This table represents the characterization of DNA fragments of sample 1 based on size (in bp) and their corresponding calibrated and assigned concentration (pg/µl), peak molarity (pmol/l), and integrated area percentages, The data indicates lower mark at 25 bp, a sample peak at 426 bp with full integrated area, and upper marker at 1500 bp with an assigned

concentration. This information is crucial for assessing quality and quantity of DNA fragments in samples.

H20/7255

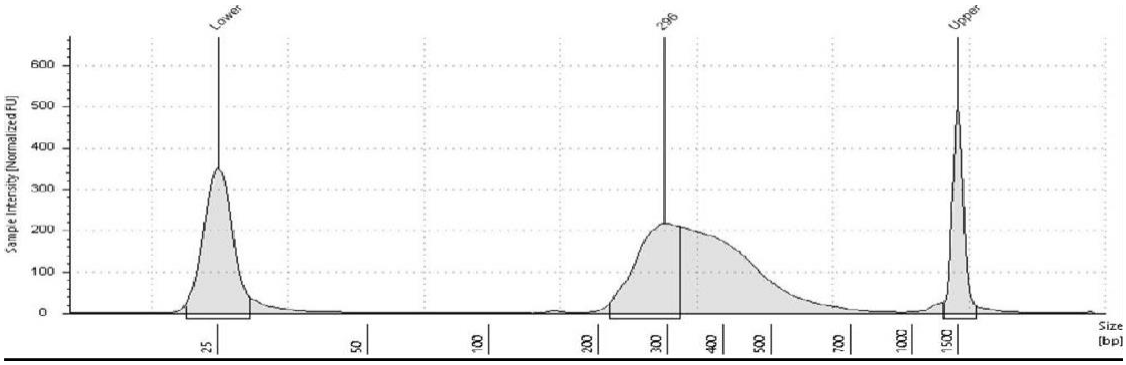


Figure 4.14: The TapeStation Electropherogram report of sample (gDNA) H20/7255. It shows the size of different fragments on x-axis, sample intensity at y-axis and peak values of lower mark, a major peak in the region of 210-320bps

Table 4.15: Sample table for sample 5

Well	Sample Description	Alert	Observation
F2	NCGM:794-H20/7255		

Note: Table presents the sample information for sample 5 analyzed in the study. There are no alerts and observations noted, suggesting that the sample was processed without any issues or anomalies during the analysis.

Table 4.16: Peak table for sample 5

Size (bp.)	Calibrate Conc. [pg/µl]	Assigned Conc. [pg/µl]	Peak Molarity [pmol/l]	%Integrated Area	Peak Comment	Observation
25	458	-	28200	-		Lower Marker
296	410	-	2130	100.00		
1500	250	250	256	-		Upper Marker

Note: This table represents the characterization of DNA fragments of sample 1 based on size (in bp) and their corresponding calibrated and assigned concentration (pg/µl), peak molarity

(pmol/l), and integrated area percentages, The data indicates lower mark at 25 bp, a sample peak at 296 bp with full integrated area, and upper marker at 1500 bp with an assigned concentration. This information is crucial for assessing quality and quantity of DNA fragments in samples.

4.1.2.3. Sample QC report

The NGS data we obtained by WES was checked for quality by running FastQC and we received the .html formatted files of quality check reports for every individual read (forward and backward separately). We transfer the html files from the server to the local system. The very first table of the report provided "**basic statistics"** for the particular sample like name of file, read length, number of the sequences to keep us on track. The plot of **"per base sequence quality"** showed quality scores on the y-axis of each read at each position on the x-axis. The color segregation of the plot differentiates the high, medium and low scores of the quality. The blue line showed the average quality score of a base and the red line represented the median. While "**per sequence quality score"** provided average quality score on x-axis against the number of sequences on y-axis. No bumps in low quality scores are a sign of good quality of data. The **"Per base sequence content"** has a given percentage of individual nucleotides (A, C, T. G) at every single position across all the reads. **Per sequence GC** content plot determined the GC distribution in all the sequences. **Per base N content** plot showed the percentage of bases at each position or with no base call (N). The **"Sequence length distribution"** plot presented the distribution of fragments of different sizes in an analysed file. "**Sequence duplication level**" identified the number of duplicated sequences in the library resulting from over cycles of PCR amplification than expected or too little initial material for the reaction. "**Overrepresented sequences**" showed a minimum of at least 20 bp sequences that could be present in above 0.1 percent of the total sequences. Once the quality of the read is determined we had to check from which gene or mRNA the reads belong by mapping. This also determined the contamination of any vector, "**adaptor content"** and any low quality bases at the reads end **(Supplementary Figures 1-10).**

4.1.2.4. Results of down-analysis on sequenced data:

After our benched marked pipeline for pre-processing (read alignment, variant calling, and variant prioritization) we obtained the variants in the form of a vcf file. Vcf files were then subjected to various *in silico* tools (like SNPnexus, gnomAD, ClinVar and dbSNP of NCBI data bank) to characterize the variants and to annotate. We first check how many variants are commonly mutated in the sample **(Figure 4.15).**

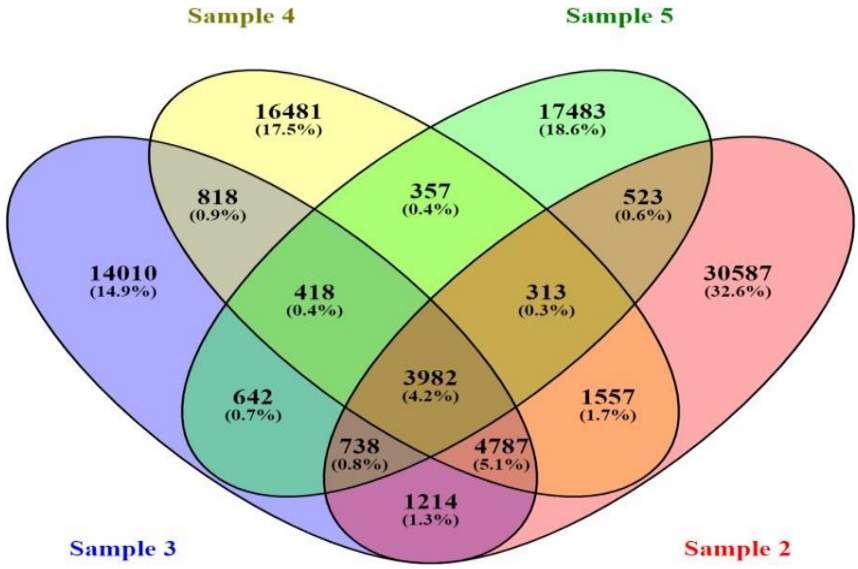


Figure 4.15: Venn plot showing the common number of variants among the sequenced samples of the 5 cases (Sample 2, Sample 3, Sample 4, and Sample 5). A total of 3982 variants were found common in all four samples. From Sample 1 of case 1 didn't identify pathogenic variants.

4.1.2.5. MYO15A and BRCA2 mutations appeared common with the previous pilot study:

We have identified a pathogenic, a germline, missense and noncoding transcript variant, **rs375290498** (A=0.00005/7) of *MYO15A* which is common among our cohort. Besides this, two germline variants of the *BRCA2* gene were identified common with the previous cohort study but with different mutations; benign rs148341992 (0.00129), pathogenic rs780919805 (T=0.000004/1), and a single base pair pathogenic duplication rs80359668. In the previous cohort, we found rs276174854, rs276174889, rs80358600, rs80359171, rs771203198, and rs145988146 (**Table 4.17**).

Table 4.17: Significant common variants and genes with previous cohort

Gene	Reference ID	MAF	Clinical Significance	Functional Consequence	Allele change	Description
MYO15A	rs375290498	A=0.00005/7	Pathogenic, Conflicting: Interpretations of Pathogenicit	Non_coding_transcript_variant_missense_variant	G> T, A	SNV

			y, Uncertain Significance			
BRCA2	rs276174854	-0.000004/1	Pathogenic	Coding_sequence_Variant,frameshift_varaint	**TATG**>-	4bp deletion
BRCA2	rs276174889	G=0/0	Pathogenic	Stop_gained,missen_variant,coding_sequence_varaint	C>**G**,T	SNV
BRCA2	rs80358600	T=0.000626/3	Pathogenic	Stop_gained,missen_variant,coding_sequence_varaint	C>A,G,**T**	SNV
BRCA2	rs80359171	A=0.000014/2	Uncertain Significance, Pathogenic, Likely-Benign	Missense_varaint,coding_sequence_variant	G>**A**,C,T	SNV
BRCA2	rs771203198	C=0.000004/1	Pathogenic	Stop_gained,missen_variant,coding_sequence_varaint	G>A,**C**,T	SNV
BRCA2	rs145988146	T=0.000012/3	Pathogenic	Coding_sequence_varaint,synonymous_varisnt,stop_gained	C>A,G,**T**	SNV
BRCA2	rs780919805	T=0.000004/1	Pathogenic	Stop_gained,synonymous_varian t,coding_sequence_varaint	C>A,G,**T**	SNV
BRCA2	rs80359668	-	Pathogenic	Coding_sequence_Variant,frameshift_varaint	**->T**	1bp dup
BRCA1	rs148341992	G=0.001324	Benign	Inton_Variant	A>**G**	1bp dup

Note: The table lists significant common variants and associated genes identified in a previous cohort study (Last Accessed on April 2, 2024). Each entry includes the gene name,

reference SNP (rs) ID, minor allele frequency (MAF), clinical significance, functional consequence, allele changes, and a brief description. The data indicates various pathogenic and uncertain significance variants in the BRCA1 and BRCA2 genes, highlighting their relevance in clinical genetics and implication for disease risk assessment.

<u>*MYO15A*'s plausible role on hearing loss and PCa:</u>

MYO15A is known for second most causing non-syndromic hearing loss (HL) (Vanniya. S et al., 2022)**.** Familial inheritance of mutated *MYO15A* is associated with high severity (Y. Chen et al., 2018). In a study of HL in 61 consanguineous Egyptian families' linkage analysis wherein the majority of variants were identified in *MYO15A* and among which one family had mutated the *WFS1* gene causing Wolfram syndrome (WS) and the variant was inherited in an autosomal recessive manner. Wolfram syndrome is characterized by childhood-onset diabetes mellitus, diabetes insipidus optic atrophy, and sensorineural HL SNHL. However, the affected siblings of the same family didn't express any other complication except WS (Budde et al., 2020). Another study was conducted on four South Indian families which are mating assortatively, in which one individual was having a history of 10 years of diabetes (Vanniya. S et al., 2022)**.**

The movement and migration of cells depend on the broad and varied family of molecular motors known as myosins (Makowska et al., 2015). A study of W. Zhang *et al.*, in the year 2022 found *MYO15A* as a novel contributor in the evolution of a hormone sensitive PCa (HSPC) to CRPC (W. Zhang et al., 2022). Mutation in *MYO15A* is one of the frequent somatic mutations of PCa and commonly related to advanced stage PCa (Zarzour et al., 2020) (Mamidi et al., 2019).

<u>Regulatory potential of *BRCA1* and *BRCA2* in PCa:</u>

While the link between *BRCA1* and 2 and PCa remains enigmatic, the genetic alterations identified so far that the greatest PCa risk particularly made over the age of 65, are mutations in the *BRCA2* gene, which is associated with an 8.6 fold increased risk (Castro & Eeles, 2012). *BRCA1* holds the moderate risk of 1.35-fol for PCa indicating very less association with PCa (Oh et al., 2019). Aside from the DNA damage response, repair, chromatin modelling, and transcriptional control, *BRCA1* has recently been found as an AR co-regulator. It has also been proposed that it may affect another key route in PCa by regulating *IGF-1R* in an AR-dependent manner. *BRCA2* may potentially operate as a tumor suppressor in epithelial prostate tissue, leading to premalignant prostatic lesions. *BRCA2* may reduce cancer chances

of spreading (metastasis) by inhibiting PI3-kinase/Akt and activating *MAP/ERK*, thereby reducing cancer cell migration and invasion (Castro & Eeles, 2012).

Significant genomic mutations were explored from our cases:

A total of 37 more significant pathogenic variants were identified in our cohort (**Table 4.18)** and among them were 31 Nonsynonymous mutations 19 missense mutations in the coding region (rs5110, rs6339, rs6336, rs1169305, rs41265017, rs45517213, rs61753021, rs104894493, rs111033186, rs120074126, rs121913601, rs145525174, rs201518227, rs387907228, rs764790770, rs770066171, rs771578775, rs987916591, rs778295360); 5 stop gain missense mutations (rs9514067, rs61749438, rs200098356, rs786205776, rs9536062); 3 stop gain missense mutation in the downstream transcript (rs9536062, rs587779865, rs267606953); one intronic downstream transcript mutation (rs1161370391); one, stop gain missense mutation in the upstream transcript (rs781580050); one missense downstream transcript variant (rs12021720) and one noncoding transcript missense variant (rs377715702). Furthermore, 4 synonymous mutations were seen in which one was stop gained variation in the coding region (rs74315369), 1 stop gained in the downstream transcript (rs764389018), and a stop gain intron variant (rs151212477). Two intron splice donor variants (rs724159829, rs758623165), one intron variant (rs765990217), one frameshift in 5' UTR (rs886041816), and stop gain mutation of non-coding transcript in the downstream transcript (rs1394131270) were also seen. Later we checked the associations of these genes with the PCa and diabetes by looking into the reported literature (**Table 4.19).**

Table 4.18: Other 37 significant pathogenic variants from our cohort

Gene	**Reference ID**	**MAF**	**Clinical significance**	**Functional Consequence**	**Allele Change**	**Description**
APOA4	rs5110	A=0.052285/13128	Pathogenic	coding_sequence_variant;missense_variant	C>A,G	SNV
NTRK1	rs6339	T=0.02436/122	Pathogenic; Benign	missense_variant;coding_sequence_variant	G>T	SNV, Haplotype
NTRK1	rs6336	T=0.036964/5180	Pathogenic; Benign	missense_variant;coding_sequence_variant	C> A,G,T	SNV, Haplotype

HNF1A	rs1169305	A=0.013899/1950	Pathogenic; Likely Pathogenic, Benign	coding_sequence_variant; missense_variant	A>C,G,T	SNV
ERCC5; BIVM-ERCC5	rs9514067	G=0.002342/12	Likely-Pathogenic; not-provided	stop_gained;coding_sequence_variant; missense_variant	G>A,C,T	SNV
ATM; C11orf65	rs587779865	T=0.000012/3	Pathogenic; Pathogenic-Likely-Pathogenic	genic_downstream_transcript_variant;intron_variant;coding_sequence_variant;stop_gained	C> A,G,T	SNV
SEMA4A	rs41265017	A=0.020924/105	Pathogenic; Benign	missense_variant;coding_sequence_variant	C> A,G,T	SNV
TSC2	rs45517213		Pathogenic	stop_gained;coding_sequence_variant	G>A	SNV
ABCA4	rs61749438	A=0.000004/1	Pathogenic; not-provided	stop_gained;missense_variant;coding_sequence_variant	C>A,T	SNV
ABCA4	rs61753021	T=0.000036/5	Likely-Pathogenic; not-provided	missense_variant;coding_sequence_variant	C>A,T	SNV
SDHB	rs74315369	T=0./0 (ALFA)	Pathogenic; uncertain-significance; Benign	synonymous_variant;coding_sequence_variant;missense_variant;stop_gained	G>A,C,T	SNV
NR2E3	rs104894493	A=0.000285/40	Pathogenic; Uncertain-Significance	coding_sequence_variant;missense_variant	G>A	SNV
GJB2	rs111033186	T=0.002474/347	Likely-Pathogenic; Benign	missense_variant;coding_sequence_variant	C>A,T	SNV

SMPD1	rs120074126	T=0.000007/1	Pathogenic	coding_sequence_variant; non_coding_transcript_variant;missense_variant	C>T	SNV
MPZ	rs121913601	T=0.000781/4	Pathogenic; Likely-Pathogenic; Uncertain-Significance	missense_variant; coding_sequence_variant	G>A,C	SNV
ABCA4	rs145525174	T=0.002461/345	Likely-Pathogenic; Likely Benign	coding_sequence_variant;missense_variant	C>T	SNV
PKP2	rs151212477	A=0.000007/1 T=0./0	Pathogenic; Likely Benign	synonymous_variant;intron_variant;stop_gained;coding_sequence_variant	G>A,T	SNV
CACNA2D4	rs200098356	A=0.000468/2	Likely-Pathogenic; Uncertain Significance	stop_gained;missense_variant;coding_sequence_variant	G>A,C	SNV
TOR1AIP1	rs201518227	T=0.000057/8	Likely Pathogenic,	missense_variant;coding_sequence_variant	C>T	SNV
FGD4	rs281865063	A=0./0	Pathogenic; Uncertain Significance	non_coding_transcript_variant;genic_downstream_transcript_variant;missense_variant;coding_sequence_variant	G>A	SNV
RNPC3	rs370930012	A=0.000029/4	Pathogenic	missense_variant;coding_sequence_variant	C>A	SNV
POLR3B	rs371453512	T=0.000007/1	Likely Pathogenic	missense_variant;coding_sequence_variant	C>T	SNV
TMX2;TMX2-CTNND1	rs377715702	T=0.000043/6	Likely Pathogenic	coding_sequence_variant;missense_variant;non_coding_transcript_variant;intron_variant;sto	C>T	SNV

				p_gained		
ABCC9	rs387907228	A=0./0 (ALFA)	Pathogenic; Likely Pathogenic	coding_sequence_variant;missense_variant	G>A	SNV
ATM;C11orf65	rs587779865	T=0.000008/2	Pathogenic; Pathogenic-Likely-Pathogenic	genic_downstream_transcript_variant;intron_variant;coding_sequence_variant;stop_gained	C> A,G,T	SNV
PDHX	rs724159829		Pathogenic	intron_variant;splice_donor_variant	G>A	SNV
UBR1	rs758623165	G=0.000007/1	Likely Pathogenic	splice_donor_variant	C> A,G,T	SNV
ATM	rs764389018	T=0.000007/1	Pathogenic; likely Pathogenic	coding_sequence_variant;genic_downstream_transcript_variant;stop_gained;non_coding_transcript_variant;missense_variant;synonymous_variant	C> A,G,T	SNV
CLN5;FBXL3	rs764790770	C=0./0	Uncertain Significance, Likely Benign; Pathogenic,	500B_downstream_variant;coding_sequence_variant;stop_gained;missense_variant;downstream_transcript_variant	G>A,C	SNV
TSC2	rs765990217	T=0.000014/2	Likely Pathogenic; Benign	intron_variant	C> A,G,T	SNV
PGM1	rs770066171	T=0.000012/3	Pathogenic	coding_sequence_variant;stop_gained	C>T	SNV
COQ8A	rs771578775	T=0.000057/8	Pathogenic	coding_sequence_variant;stop_gained	C>T	SNV
AGL	rs781580050	T=0.000007/1	Pathogenic-Likely-Pathogenic	stop_gained;coding_sequence_variant;genic_upstream_transcript_variant	C>T	

CACNA1C	rs786205776		Likely Pathogenic; Uncertain-Significance	coding_sequence_variant;missense_variant;stop_gained	G>A,C,T	
PTEN	rs886041816		Pathogenic	5_prime_UTR_variant;coding_sequence_variant;frameshift_variant	->CT	2bp duplication (DELINS)
PDE2A	rs987916591		Pathogenic	coding_sequence_variant;missense_variant	G>A	SNV
DOCK7	rs1161370391		Likely Pathogenic	genic_downstream_transcript_variant;intron_variant	T>-	DEL
SZT2	rs1394131270	T=0.000004/1	Pathogenic	coding_sequence_variant;non_coding_transcript_variant;genic_downstream_transcript_variant;stop_gained	C>T	SNV
NDUFV1	rs778295360	T=0.000014/2	Likely Pathogenic	stop_gained;coding_sequence_variant;missense_variant	G>A,T	SNV
THSD1	rs9536062	C=0.054029/271	Pathogenic	genic_downstream_transcript_variant;coding_sequence_variant;stop_gained;intron_variant	G>A,C	SNV
PLCE1; NOC3L	rs267606953	T=0./0 (ALFA) T=0.000004/1 (GnomAD_exomes)	Likely Pathogenic	coding_sequence_variant;missense_variant	C>T	SNV

Note: The table lists 37 significant pathogenic variants identified in our cohort. Each entry include the gene name, reference SNP (rs) ID, minor allele frequency (MAF), clinical significance, functional consequence, and allele changes (Last Accessed on April 2, 2024).

Table 4.19: Tabular summary of the general associations of identified mutated gene with PCa and diabetes

Gene	Association with PCa	Association with Diabetes	Disease conditions linked to Gene
APOA4	• The specific function of apolipoprotein, APOA4 has not been completely investigated. However it is suggested that the APOA4 plays a role in metabolism and transporting lipids (including cholesterol and triglycerides) particularly in the lymphatic and circulatory system. • Found upregulated in mCRPC. (Kakkat et al., 2023). • Cardiovascular diseases (CVDs) are commonly observed in PCa patients and a vice versa relation is also seen, that the pre-existing CVD patients are more likely to have PCa. It could be possible that there might be a molecular connection between CVD and PCa (Kakkat et al., 2023). • Physiological role of Thrombosis and aggregation of platelets is also seen (Buscher, n.d.) .	• In the type-2-diabetes patients APOA4 can be a novel biomarker in plasma for identification of renal problems (Peters et al., 2017). • It can decrease the glucose • Level is found lower in prediabetes • Has anti-inflammatory and antioxidant properties suggesting its potency to prevent development of type-2-diabetes (von Toerne et al., 2016).	Cardiovascular diseases and renal complications.
ITGB4	• Particularly in the context of tumour aggressiveness, metastasis (bone), and therapeutic response (resistance to chemotherapies), ITGB4 appears to be important in PCa. It is more abundantly expressed in advanced PCa and correlated with increased motility and invasiveness (Wilkinson et al., 2020). • ITGB4 was also suggested in a		

	study to interact with Prostate-Specific membrane Antigen (PSMA, a unique glycoprotein and highly expressed in PCa cells) and regulates the angiogenesis through activating NF-κB (Y. Gao et al., 2021).		
ABCC9	• Downregulated in PCa (Demidenko et al., 2015).	Found mutated in those myocardial infarction patients who were having history of diabetes mellitus, hypercholesterolemia, and hypertensions (Minoretti et al., 2006).	Coronary vasomotor dysfunction, ischemic disease and myocardial infarction
SEMA4A	• Initiates the migration of endothelial cells (Nkyimbeng-Takwi & Chapoval, 2011). • Plays role in EMT and promotes invasion and metastasis (X. Liu et al., 2023).	Different semaphorins are involved in different diabetes related complications like diabetic retinopathy, neuropathy, and osteoporosis and wound healing. SEMA4A is found to play role in regulation of inflammatory colitis (Lu & Zhu, 2020).	Inflammatory colitis
ATM	• Loss of ATM is found in 10% of the advanced PCa, cells more sensitive to the PARP inhibitors and can bring genetic instability as the ATR will get inhibited and its sensitivity also increases (Neeb et al., 2021).	• By phosphorylating p53 regulates the insulin resistance and deficiency of ATM can develop DM • Through alteration in metabolism of glucose can cause cardiac problems (Espach et al., 2015), and a role in metformin response (glycemic) (Yee et al., 2012).	Cardiovascular disease (Ischemic conditions, fibrosis and hypertrophy), insulin resistance and diabetes
ERCC5	• ERCC5 is found in Asian cohort studies. Identified polymorphisms didn't suggest the risk associations; however, it is believed that its deficiency can modulate the susceptibility for PCa (Y. Liu et al., 2018).	No clear associations	Xeroderma pigmentosum, bladder cancer, breast cancer, CS and TTD
HNF1A	• HNF4G together with HNF1A initiates a new enhancer	• HNF1A is a most commonly mutated gene among the key	GD, T2D and MODY

	transcription circuit in gastrointestinal cancer of mCRPCs which offers more AR independent growth and resistance and rapid development of CRPC (S. Shukla et al., 2017). • HNF1A regulates the glucuronosyltransferase (UGT) which can modulate androgen activity suggesting AR and HNF1A regulated CRPC development (Yun et al., 2017).	genes related to maturity onset diabetes in the young (MODY) (Valkovicova et al., 2019). • It can trigger the risk of GD and T2D (Ding et al., 2022).	
NTRK1	• NTRK1 is reported in fusion with IRF2BP2 gene translocation in PCa (Yeh et al., 2019). • It is also found playing a role in tumour progression together with NGF by involving into a loop of autocrine signaling in prostate cells (Pierotti & Greco, 2006). • It is related to a poor prognosis and downregulated in PCa patients (Bagherabadi et al., 2022).	It is identified in serving as a potential biomarker of diabetes (type 3 diabetes) caused by Alzheimer disease (Pandiyan et al., 2021).	type 3 diabetes
TSC2	In PCa cells isoform of TSC2 (TSC2A) functions as an active target of androgen receptors and is down regulated. Its down regulation stimulates cell proliferation (Munkley et al., 2014).	TSC2 regulates mTORC1 by inhibiting it which decreases insulin resistance. Further for the action of metformin both TSC2 and RAPTOR is required to inhibit mTORC1suggesting TSC2 importance in metformin pathway and diabetes management (van Nostrand et al., 2020).	T2D, tuber sclerosis and diabetic neuropathy
ABCA4	No association found	No association found	Stargardt disease
SDHB	• It forms one of the catalytic domains of succinate dehydrogenase (SDH) in a matrix of mitochondria. SDHB	It is down-regulated in prediabetes and β cells of T2DM which reduces the activity of SDH enzymes and leads to development of type-2-	Mitochondrial dysfunction and T2DM

	is strongly linked to cancer development. Its inhibition causes succinate accumulation as result of respiration and TCA cycle impairment. This succinate accumulation inside PCa is linked to PTEN deletion (Sant'anna-Silva et al., 2021).	diabetes (S. Lee et al., 2022).	
NR2E3	No association found	Diabetes can negatively affect the clock system of the retina and the output of the circadian cycle. Diabetes is found to alter clock but not at the gene expression level of NR2E3 (Vancura et al., 2021).	Retinopathies
GJB2	• GJB2 is downregulated in PCa and helps in transport and cell-cell signaling in CRPC cells resistant to docetaxel (Marín-Aguilera et al., 2012). • It is one of the frequently mutated genes in PCa. It codes a gap junction protein, and it has been assumed that it might promote in metastasis of PCa (Tang *et al.,* 2022)	It is observed that the GJB2 is mutated in those who have hearing loss and maternally inherited diabetes (Frei et al., 2005).	Inherited diabetes and Hearing loss
SMPD1	Down regulated in PCa	Found upregulated and associated with diabetic related and alcoholic cardiomyopathy (R. Liu et al., 2023)	Alcoholic cardiomyopathy and cardiac dysfunction
PKP2	It is one of the most commonly distributed desmosomal plaque proteins found in both basal and luminal cells present in pseudostratified epithelium of the prostate gland. However, its expression never changes with increasing tumor grade which suggests its indirect role in PCa (Breuninger et al., 2010).	It is interrelated with diabetes and arrhythmic cardiomyopathy (Mathiyalagan et al., 2014).	Diabetic retinopathy and arrhythmic cardiomyopathy
PDHX	No association	Plays role in glucose metabolism. MiRNA-26a inhibits PDHX to stop citric acid cycle pyruvate to convert	

		into acetyl coenzyme A (B. Chen et al., 2014).	
CACNA2D4	It is rarely identified because of epigenetic mutation and found related to PCa specific mortality (PCSM). Its promoter gets methylated and downregulated and suggested to be a potential biomarker (Pidsley et al., 2022).	It plays a role in diabetes induced atherosclerosis (L. Wang et al., 2019).	Cardiovascular diseases and atherosclerosis
TOR1AIP	No association, not reported till now in PCa	Mutated in distal myopathy (Finsterer & Stöllberger, 2016).	Distal myopathy and Alzheimer disease
FGD4	It is upregulated in PCa cells and expressed as the advancement of the PCa. FGD4 basically helps in PCa cell migration and expression of mesenchymal cell markers (Bossan et al., 2018).	• Found mutated in Amyotrophic lateral sclerosis of the Chinese population and associated with Charcot-Marie Tooth disease with diabetes (Wei et al., 2019). • Also, reported in mutated esophageal squamous cell carcinoma with a history of diabetes (Ying Yang et al., 2021).	Amyotrophic lateral sclerosis
RNPC3	It is a minor component of spliceosomes and its expressions counterparts the tumor progression. Its expression is lower in hormone sensitive cells of PCa and benign tumors, intermediately expressed in CRPC. However, hyper-expressed in metastatic neuroendocrine PCa its expression (Augspach et al., 2021).	It encodes spliceosomes and is found linked to growth hormone deficiency in children who had diabetes insipidus (Murray & Clayton, 2015).	Growth hormone deficiency
POLR3B	No clear association	No clear association	
TMX2	No association	TMX2 regulates redox signaling where endoplasmic reticulum (ER) and mitochondria contacts called Mitochondrial-associated endoplasmic reticulum, MAMs (site for influencing energy production). In the liver these MAMs increase in	Neurodevelopmental disorders

		obesity and insulin resistance. Over formation of MAMs leads Ca2+ load in mitochondria and mitochondrial dysfunction. Mutation in TMX2-like transmembrane ER redox genes causes neurodevelopmental disorders (N. Sharma et al., 2020).	
CTNND1	No association	CTNND1 relates to diabetes by influencing the migration of neutrophil cells in diabetic skin wounds (Kang et al., 2021).	Diabetic skin wounds
CLN5/FBXL3	A glycoprotein which is downregulated in PCa (Shah et al., 2015)	• CLN5 is downregulated in diabetic foot and down regulated in DM (W. Zhao et al., 2020). • Inhibition of CLN5 can cause suppression of Wnt-pathway leading to *in vitro* endothelial cell formations and retinopathy (J. Chen et al., 2011).	DM, diabetic foot and retinopathies
PLCE1	PLCE1 mutations are associated with high grade PCa (Edwards *et al.*, 2013). Found commonly mutated in stomach cancer (H. Sun et al., 2015)	It is a podocyte cell marker of glomeruli whose expression decreases in diabetic neuropathy and diabetic kidney disease (Eadon et al., 2022).	Diabetic neuropathy and diabetic kidney disease
CACNA1C	Its lower expression could be a potential biomarker for PCa (Phan et al., 2017).	CACNA1C is a calcium voltage-gated channel subunit and its polymorphisms linked to hypertension and coronary artery disease (Beitelshees et al., 2009). It is also associated with diabetic cataract by calcium channels	Coronary artery disease and diabetic cataract
PTEN	Most frequently mutated in PCa and brings altering the PI3K signaling (Wise et al., 2017) (Jamaspishvili et al., 2018).	It is found responsible for DN and in the PI3K pathway it regulates Akt and Fak (Yan et al., 2019).	Diabetic neuropathy
DOCK7	DOCK7 gene fusion with OLR1 gene was found in promoting PCa metastasis and its recurrence (Y. P. Yu et al., 2023).	DM patients who have dyslipidemia as comorbidity with risk of developing CVDs have DOCK7 mutations in Chinese population (Kong et al., 2015).	Dyslipidemia

SZT2	SZT2 is one of the protein complexes which plays a role in mTORC1 signaling (Yin et al., 2021)	This gene is downregulated in the plasma of CVD and also linked to epilepsy and development of the human brain. Loss of SZT2 causes the increase in mTORC1, important for diabetes and cardiac systems (Lygirou et al., 2018).	CVD
NDUFV1	NDUFV1 is supposed to be one of the 11 protein panels which can differentiate patients with low grade PCa from high grade (Kawahara et al., 2019).	No Specific associations	
THSD1	No association	No association	

Rare insights of *MPZ*, *UBR1*, *PGM1*, *COQ8A*, *AGL* and *PDE2A* anomalies associated with diabetes and PCa:

MPZ protein is produced by Schwan cells and is essential for retaining the firmness and integrity of myelin sheaths in peripheral nerves (Haddad et al., 2022). In diabetic patients, phospholipids, cholesterol, and fatty acid content of myelin gets altered, affecting the fluidity of the membrane and leading to a condition called diabetic neuropathy. Myelin protein zero coding gene *MPZ* is seen suppressed in diabetic patients (Cermenati et al., 2012). Type 1 diabetes-associated myelin abnormalities lower the signal conduction velocity in nerves of the peripheral nervous system (Cermenati et al., 2012). In a study on non-obese T2DM mice (MKR) it has been found that metformin can restore the *MPZ* level by regulating the ROS produced in hyperglycaemic conditions suggesting a correlation of Metformin, ROS, and alteration in *MPZ* protein (Haddad et al., 2022). Therefore, it has been marked as the biomarker of diabetic peripheral neuropathy (Haddad et al., 2022).

The gene *UBR1* codes ubiquitin ligase protein and plays a role in stabilizing a protein called adipose triglyceride lipase (*ATGL*) which works as a rate-limiting lipase, crucial for the breakdown of lipids. In a study, UBR1 and 2 were found in reducing the level of *ATGL* by ubiquitination and degradation and promoting lipid storage (Bingham, 2023). This regulation of ATGL is important for maintaining a cellular level of lipids in obese patients (Bingham, 2023). However, in PCa cell lines like PC3, it has been observed to destabilize and inhibit a pro-apoptotic truncated bone marrow kinase (BMXΔN) of bone marrow kinase (BMX) by ubiquitination. BMX in a study was found to get cleaved into its truncated form which makes

the PCa cells more sensitive towards apoptosis as a response to several apoptosis stimuli (Eldeeb & Fahlman, 2016).

Phoshoglucomutase (*PGM1*) is a glycogen metabolism enzyme and it catalyzes the transformation of glucose-1-phosphate to glucose-6-phosphate in a reversible manner. Among all the isoforms, *PGM1* is majorly (90-95%) is responsible for its activity (Gloria-Bottini F et al., 2019). *PGM1* has been shown to have an important function in regulating glucose metabolism and cancer development in different types of cancers. *PGM1* has been related to cell proliferation and metabolism (Bo Cao et al., 2021). The maternal-effect is more significant on T2DM susceptibility in offspring. This effect has been observed in association with polymorphism in phoshpoglucomutase when transmitted from a type 2 diabetic mother (Akbarzadeh et al., 2022).

An unusual kinase-like protein called *COQ8A* facilitates the production of coenzyme Q, a crucial antioxidant and cofactor for cells. *COQ8A's* mode of action remains unclear, in part due to the lack of small molecule tools to probe its function. The absence of small molecule instruments to investigate the function of *COQ8A* contributes to the uncertainty surrounding its method of action (N. H. Murray et al., 2022).

COQ8A is an important lipid soluble electron transporter, however in diabetic cardiomyopathies it has been observed that its level decreased therefore leading to impaired production of mitochondrial ATP (Gomes et al., 2022).

Genes involved in amylo-alpha-1, 6-glucosidase activity, 4-alpha-glucanotransferase activity, were downregulated in DPH (Diabetes with parental history) with respect to obese people (Das & Rao, 2007). As a part of CD44 pathogenesis in cancer, those cells which have lost the *AGL* gene, CD44 interacts with hyaluronic acid (HA) which increases the hyaluronic acid synthase 2 (HAS2). Eventually producing more HA and CD44-HA interactions involved in inflammatory responses (Sottnik & Theodorescu, 2016).

PDE2A is one of the 100 isoforms of superfamily phosphodiesterases (PDEs) is known to be involved in degradation of cAMP and cGMP, expressed in different tissues (brain, liver, heart, lung etc.). Different isoforms are expressed within different sites of cells and consequently are regulating differentially cAMPs local level at specific locations, determination of which *PKA* to be targeted, and its downstream signaling. One of its three variants (PDE2A1, PDE2A2 and PDE2A3), PDE2A2, localized in mitochondria controls the

production of ATP by regulating cAMP generation from the plasma membrane. Impaired cAMP signaling is linked with several diseases (Monterisi et al., 2017).

4.1.2.6. Correlation of sequencing results using meta-analysis:

4.1.2.6.1. Comparison with ClinVar data of PCa:

Among cohort 5 significant variants are seen as common with ClinVar Datasets of PCa; rs148341992, rs199620842, and rs186753161 which were benign and 2 were pathogenic rs780919805 and rs80359668 (**Table 4.20**) (**Figure 4.16**).

Table 4.20: ClinVar and common of 5 samples

Gene	Reference ID	MAF	Clinical Significance	Variant type	Mutation type	Allele change
BRCA 2	rs148341992	0.00129 (G)	Benign	**SNV**	**Intron variant**	A>G
ITGB4	rs199620842	0.00619 (G)	Benign	SNV	Downstream transcript variant	C>G
CTPS1	rs186753161	0.000114/16 (A)	Benign	SNV	Downstream transcript variant	G>C,T
BRCA 2	rs780919805	0.000004/1 (T)	**Pathogenic**	**1bp dup**	Synonymous stop gain variant	
BRCA 2	rs80359668		**Pathogenic**		**Frame shift**	

Note: The table lists significant variants identified common in our cohort and ClinVar data of PCa. Each entry include the gene name, reference SNP (rs) ID, minor allele frequency (MAF), clinical significance, functional consequence, allele changes. The mutations includes 3with benign and 2 with pathogenic significance (Last Accessed on April2, 2024).

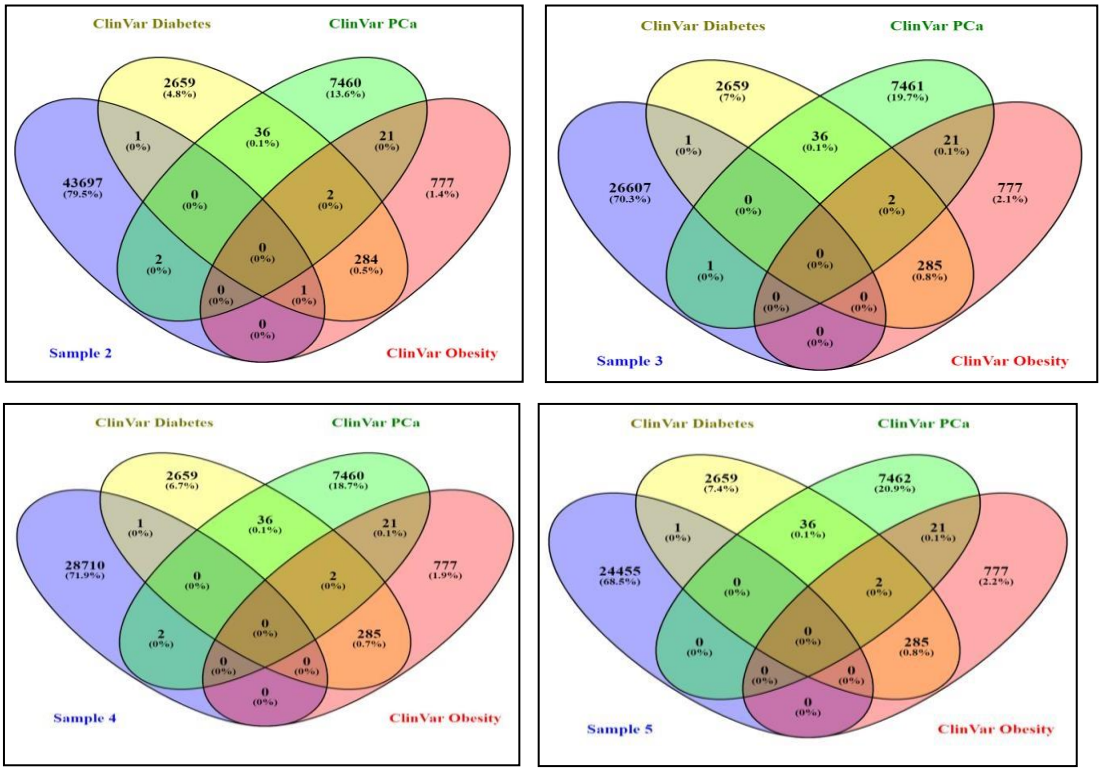


Figure 4.16: Venn plots of comparison of each case sample variant with the ClinVar data of PCa, diabetes and obesity. Sample 2, 3, 4 and 5 each showed 1 commonality with ClinVar diabetes. Sample 2 showed 2, Sample 3 showed 1, Sample 4 showed 2 common variants with PCa ClinVar. In case of ClinVar obesity, Sample 2 has 1 commonality

4.1.2.6.2. Comparison of CAPCI with PRACTICAL consortium yielded significant variants

In total 866 variants were found similar in our cohort and PRACTICAL consortium data sets.

The prostate cancer association group to investigate cancer associated alternations in the genome, (PRACTICAL) consortium was established by several scientific collaborations in 2008 with the goal to identify genes increasing the risk of inherited PCa. The consortium aimed to combine data of different studies to provide a reliable assessment of the risk associated with these genes, and to validate new findings. The PRACTICAL involves 134 different research groups from Europe, North America, Asia, Australia, South America, and Africa.

Among which very few were clinically verified and benign. Those variants were rs506504 (*CHEK1*), rs117773969 (*TUBGCP4*), rs6416927 (*BRCA1*), rs1800058 (*ATM*), rs1793959 (*COL2A1*), rs1800574 (HNF1A), and rs61750984 of PARP1 which falls under the significant *p-value* (less than or equals to 0.05) with MAF 0.00419 (A) (**Figure 4.17) (Table 4.21).**

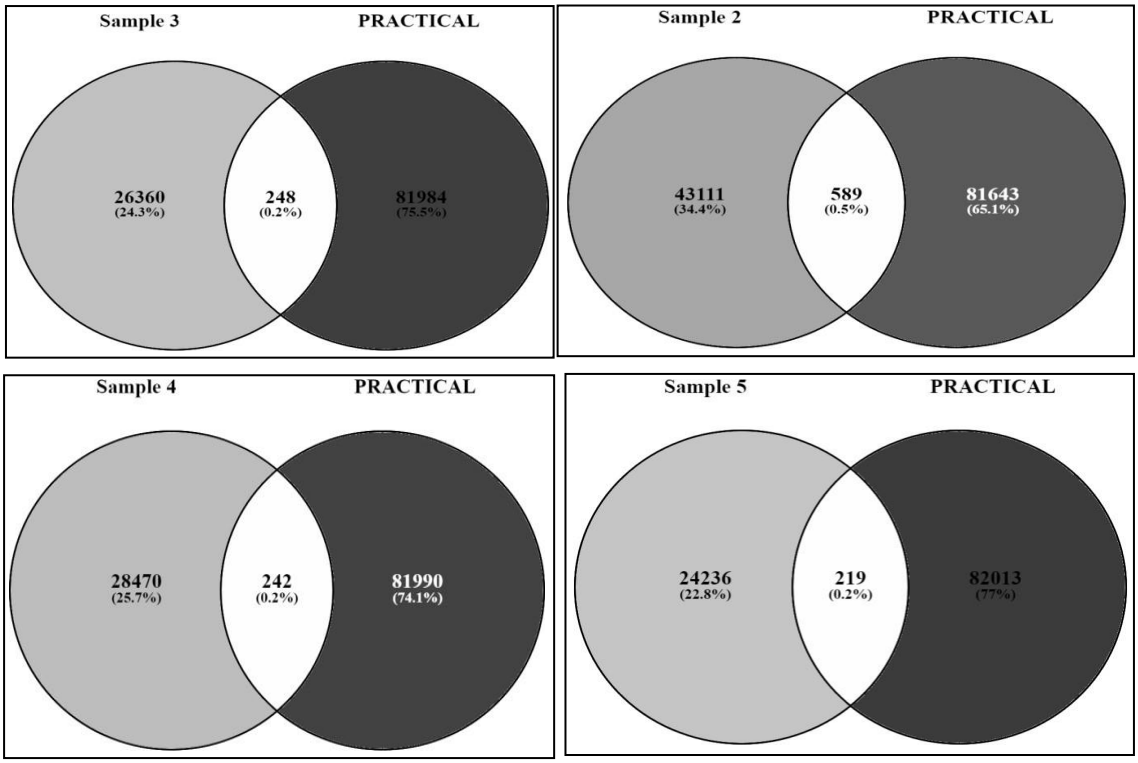


Figure 4.17: Venn plots for cohort and PRACTICAL consortium. A total of 866 variants were common with PRACTICAL consortium data from our case samples.

Table 4.21: Variants common with PRACTICAL Consortium

Gene	**Reference ID**	**MAF**	**Clinical significance**	**Functional Consequence**	**Allele change**
CHEK1	rs506504	A=0.014054/70	Benign	non coding transcript missense variant	A>C,G,T
TUBGCP4	s117773969	A=0.00203/10	Likely benign	intron variant	G>A
BRCA1	rs6416927	G=0.003629/509	Benign	intron variant	G>A,C,T
ATM	rs1800058	T=0.006715/34	Benign	downstream transcript variant, missense variant	C>G,T

HNF1A	rs1800574	T=0.020649/2896	Benign	missense variant	C> A,G,T
COL2A1	rs1793959	T=0.03654/183	Benign	intron variant	T>A.C.G
PARP1	rs61750984	0.00419 (A)	Benign	synonymous missense variant	G>A,T

Note: This table presents variants identified in common with PRACTICAL Consortium, detailing gene name, reference SNP (rs) ID, minor allele frequency (MAF), clinical significance, functional consequence, and allele change. The majority of variants listed, including those in CHEK1, BRCA1, and ATM are classified as benign or likely benign, highlighting the genetic diversity and low pathogenic impact of these variants in the studies population (Last Accessed on April2, 2024).

4.1.2.6.3. Comparison with GWAS Central PCa:

8 variants of unknown significance were seen as similar to GWAS central data from our cohort rs2230552 (G= 0.290935/1457), rs12421354 (T=0.152955/766, *ROBO4*), rs4924675 (A=0.123306/17273, *CAPN3*), rs7920517 (A=0.425319/2130), rs1243647 (A=0.247005/1237, *RNASE9*), rs4245739 (C=0.214058/1072, *MDM4*), rs2901964 (G=0.335736/46989, *CELA2A*;LOC105376767), rs636291 (G=0.455272/2280, *PEX14*) and a clinically verified, with significant risk factor MAF 0.459445, *MSMB*) was also seen similar (**Figure 4.18) (Table 4.22).**

Table 4.22: Variants common with GWAS central data of PCa

Gene	**Reference ID**	**MAF**	**Variant type**	**Allele change**
MSMB	rs10993994	A= 0.459445/64348	Upstream transcript variant	A>G
CCT6B	rs2230552	G=0.290935/1457	Missense upstream transcript variant	A>C,G
ROBO4	rs12421354	T=0.152955/766	Upstream transcript variant	C>A,T
CAPN3	rs4924675	A=0.123306/17273	Upstream transcript variant	G>A
Gene not	rs7920517	A=0.425319/2130		C>G,T

specified				
RNASE9	rs1243647	A=0.247005/1237	Missense variant	A>C,G
MDM4	rs4245739	C=0.214058/1072	Downstream transcript variant	C> A,G,T
CELA2A; LOC105376 767	rs2901964	G=0.335736/46989	Intron variant	C>G,T
PEX14	rs636291	G=0.455272/2280	Upstream transcript variant	G>A

Note: The table summarizes variants associated with PCa as reported in the GWAS central database. Each entry includes gene name, reference SNP (rs) ID, minor allele frequency (MAF), clinical significance, variant type, and allele change. The high MAF values in genes underscore their potential relevance in the genetic predisposition to PCa (Last Accessed on April2, 2024).

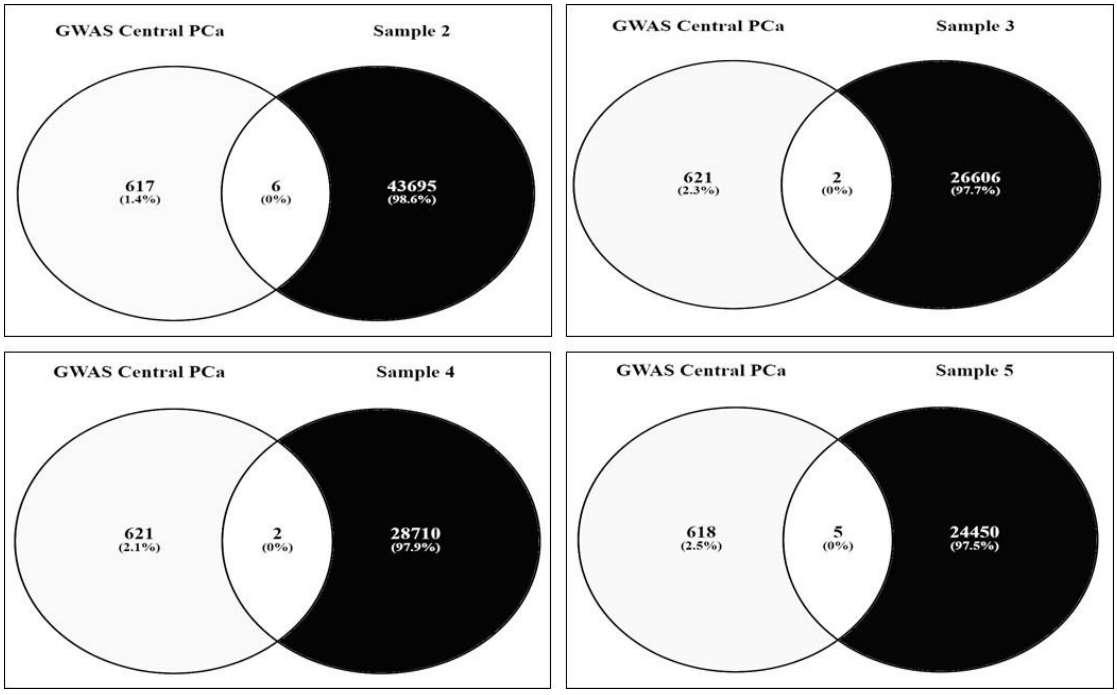


Figure 4.18: Commonalities with GWAS data sets of PCa. 6 variants in sample 2, 2 variants in sample 3, 2 variants in sample 4 and 5 variants were found common with GWAS Central data of PCa

4.1.2.6.4. A suspicious mutation of *SLC16A13* discovered on comparison with GWAS central Diabetes Datasets:

36 variants were identified commonly with GWAS central Diabetes *viz*. rs3729842, rs689, rs2476601, rs757110, rs5215, rs846111, rs2298632, rs1805096, rs5219, rs4148646, rs5213, rs657317, rs2275620, rs3740878, rs11037909, rs2271586, rs10509201, rs10741243, rs3184504, rs883079, rs2259816, rs1060105, rs1800574, rs11597086, rs1504907, rs1775368, rs12029454, rs3101336, rs2934381, rs937254, rs312457, rs1058018, rs781852, rs781831, rs10278 and rs2032844. Rs2476601 of *PTPN22* (A=0.068394/9585), rs657317 of LOC105369526 (A=0.0454/525), rs1800574 of *HNF1A* (T=0.020649/2896), rs312457 of *SLC16A13* (G=0.033428/797) (**Figure 4.19**) (**Table 4.23**).

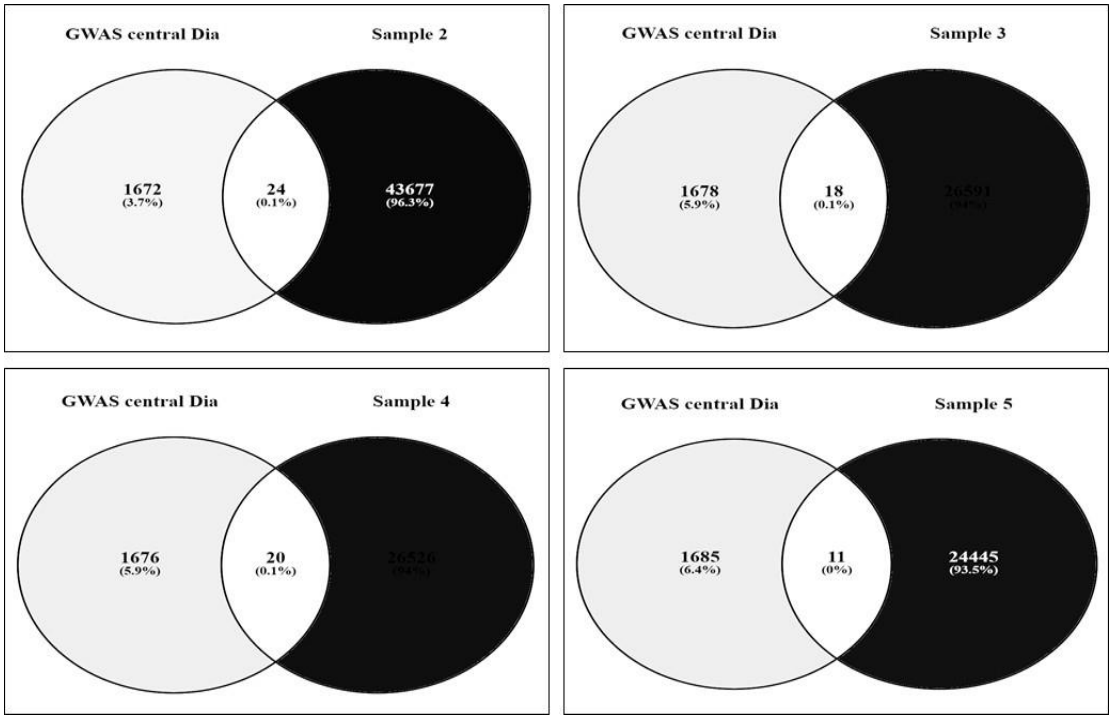


Figure 4.19: Commonalities with GWAS data sets of diabetes. 24 variants in sample 2, 18 variants in sample 3, 20 variants in sample 4 and 11 variants in sample 5 were found common with GWAS Central data of Diabetes

Table 4.23: Variants common with GWAS central Diabetes

Gene	Reference ID	Clinical Significance	MAF
TNNT2	rs3729842	Benign	A=0.158499/22201

INS	rs689	Benign	A=0.32427/9040
PTNP22	rs2476601	Benign	A=0.068394/9585
ABCC8	rs757110	Benign	C=0.280648/39303
KCNJ11	rs5215	Benign	C=0.286488/40137
RNF207	rs846111	-	C=0.194283/27198
TCEA3	rs2298632	-	T=0.401959/56312
LEPR	rs1805096	Benign	A=0.431398/60363
KCNJ11	rs5219	Likely-benign, drug-response	T=0.279516/39187
ABCC8	rs4148646	-	C=0.284537/39883
KCNJ11	rs5213	BENIGN	C=0.26258/1315
LOC105369526	rs657317	-	A=0.0454/525
CYP2C8	rs2275620	-	A=0.422025/105275
EXT2	rs3740878	BENIGN	C=0.238554/33409
EXT2	rs11037909	BENIGN	C=0.238554/33409
ART5	rs2271586	-	T=0.299521/1500
Gene Not Specified	rs10509201	-	C=0.488506/68428
TCERG1L	rs10741243	-	C=0.188898/946
SH2B3	rs3184504	-	G=0./0 T=0.332057/46498
EXT2	rs883079	BENIGN	C=0.238554/33409
HNF1A	rs2259816	-	T=0.358826/1797
SBNO1	rs1060105	-	T=0.161355/22602
HNF1A	rs1800574	**BENIGN**	**T=0.020649/2896**
CHUK	rs11597086	-	C=0.293409/41109
PLEKHG7	rs1504907	-	G=0.147856/27309
LOC105378797	rs1775368	-	A=0.290027/40558
NOS1AP	rs12029454	-	A=0.20641/28904

LOC105378797	rs3101336	-	T=0.383145/53641
NOTCH2	rs2934381	-	A=0.171048/23959
GCOM1 (Varview), MYZAP (Varview)	rs937254	-	A=0.467277/65401
SLC16A13	rs312457	-	**G=0.033428/797**
UBE2Z	rs1058018	-	C=0.409629/102273
ZZEF1	rs781852	-	A=0.49508/69330
ZZEF1	rs781831	-	C=0.452122/113516
CALCOCO2	rs10278	-	G=0.324753/45429
TTLL6	rs2032844	-	A=0.265422/37115

Note: This table lists genetic variants common with GWAS central data of diabetes. Each entry includes gene name, reference SNP (rs) ID, minor allele frequency (MAF), clinical significance, and allele change. Majority if these variants are classified as benign suggesting they may not contribute significantly to PCa. Variants exhibiting high MAF indicated a more common occurrence in the population. The clinical significance of some variants remains unspecified, highlighting the need for further research (Last Accessed on April2, 2024).

4.1.2.6.5. Comparison with ClinVar Diabetes:

On comparing our cohort with ClinVar diabetes, we got a synonymous rs370887875 and a nonsynonymous missense rs1800450 as common. Two mutations, 1 intron variant (rs148341992) and downstream transcript intron variant (rs199620842) from our cohort were seen present in both ClinVar datasets of PCa and diabetes. Further, a splice acceptor variant (rs3589051) was identified common in ClinVar datasets of obesity also (**Table 4.24**) **(Figure 4.16).**

Table 4.24: Common Variants common with ClinVar diabetes

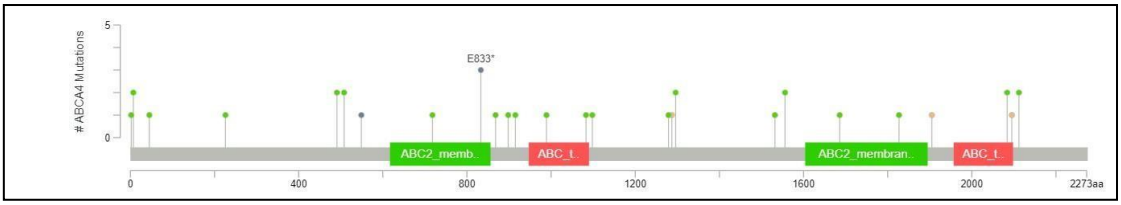
Gene	Reference ID	MAF	Clinical Significance
MYRF	rs370887875	0.00719 (T)	Benign
MBL2	rs1800450	0.12201 (T))	Conflicting interpretations of pathogenicity, benign, uncertain significance
BRCA2	rs148341992	0.0129(G)	Benign
ITGB4	rs199620842	0.00619 (G)	Benign

MPO	rs35897051	0.00080(G)	Likely pathogenic and pathogenic

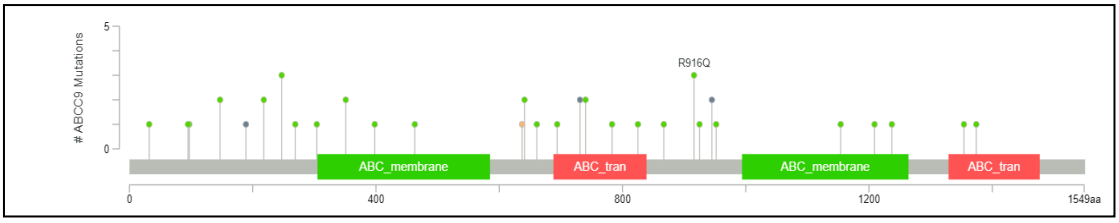
Note: The table summarizes variant common with ClinVar diabetes data. Each entry includes gene name, reference SNP (rs) ID, minor allele frequency (MAF), clinical significance, and allele change (Last Accessed on April2, 2024).

4.1.2.6.7. Mutation in our PCa Cohort:

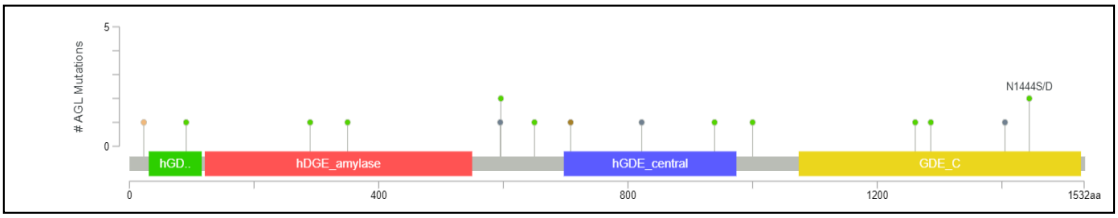
Using the cBioportal we mapped the mutated genes to the PCa datasets and found the different mutation regions on each gene related to the PCa. Each protein change and the region is represented by upright lollipop marking is on the results (**Figure 4.20 a-as**)



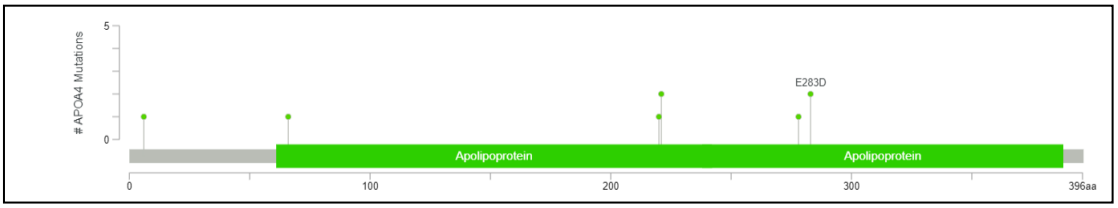
(a) ABCA4



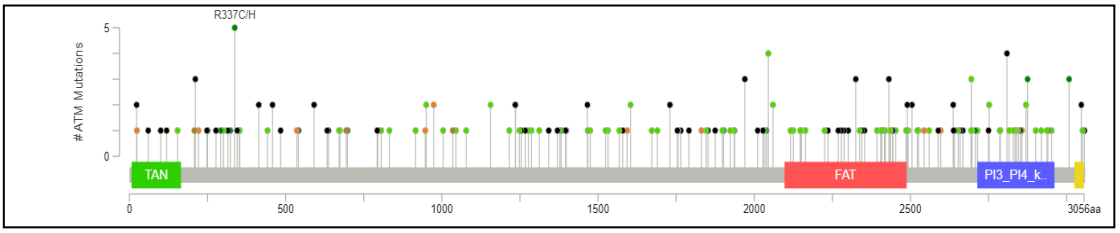
(b) ABCC9



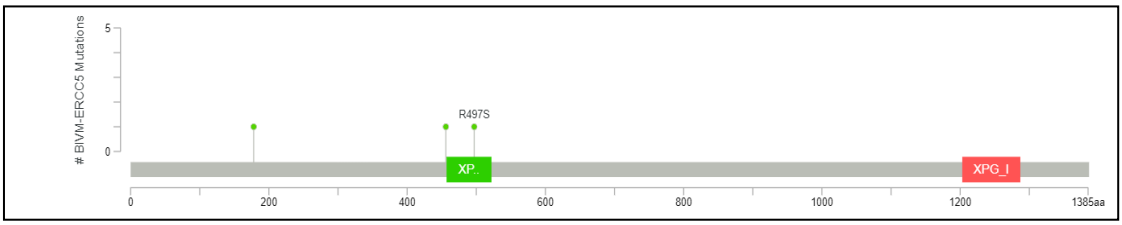
(c) AGL



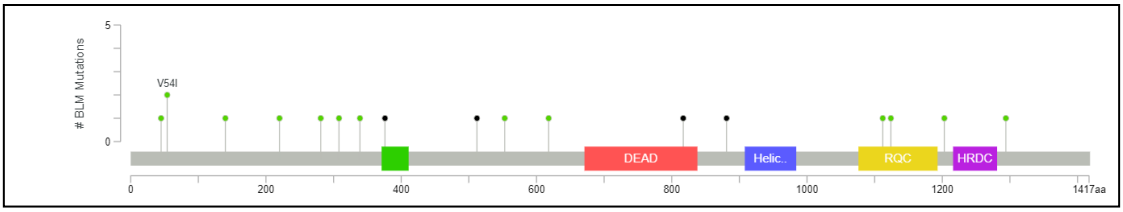
(d) APOA4



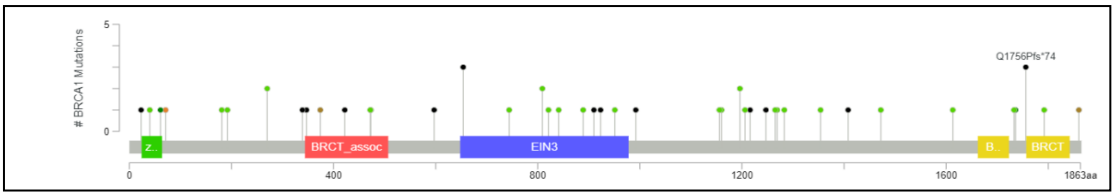
(e) ATM



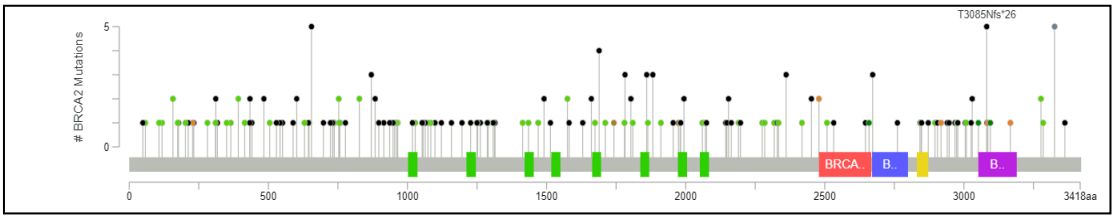
(f) BIVM-ERCC5



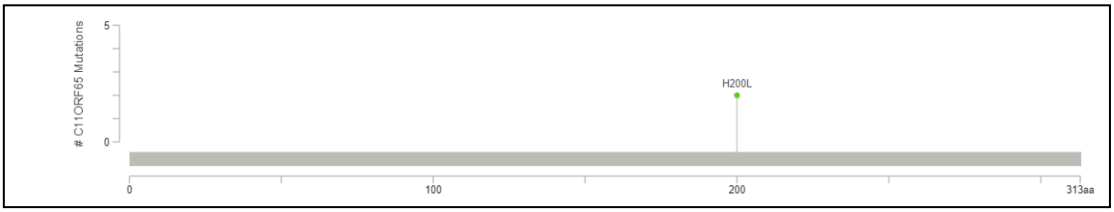
(g) BLM



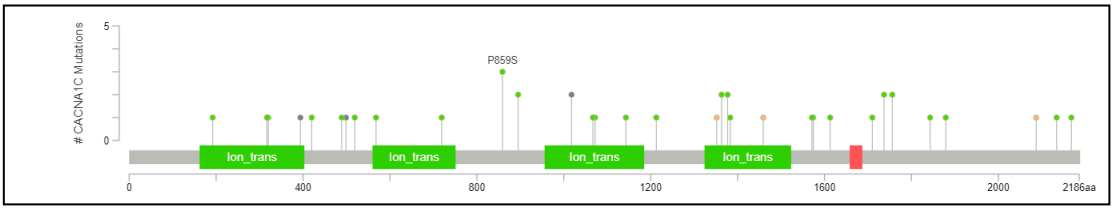
(h) BRCA1



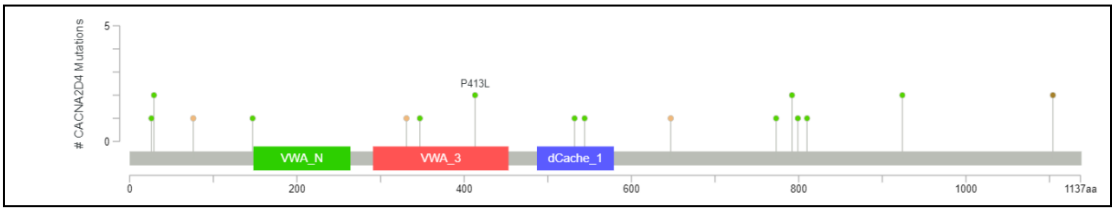
(i) BRCA2



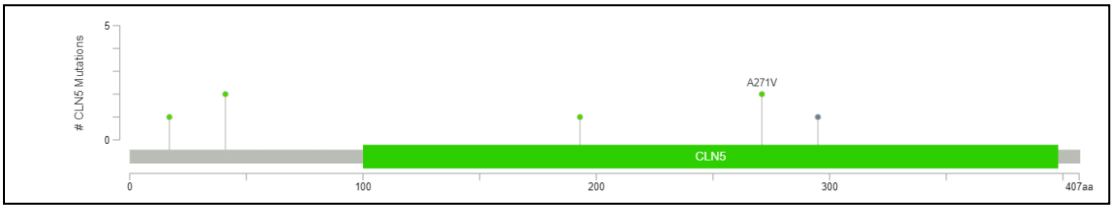
(j) C11ORF65



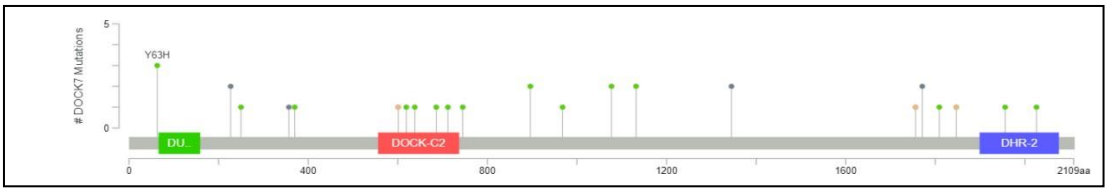
(k) CACNA1C



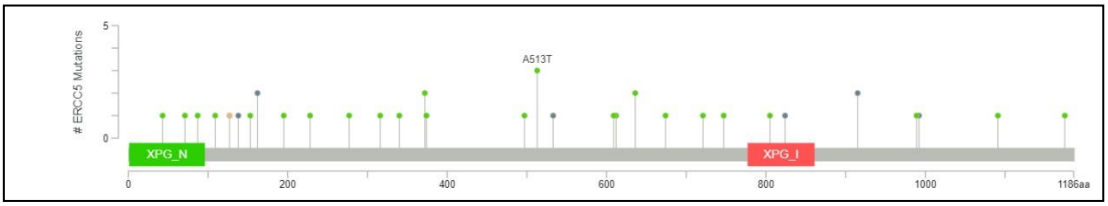
(l) CACNA2D4



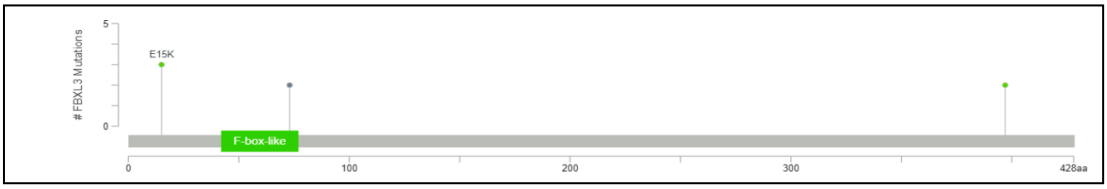
(m) CLN5



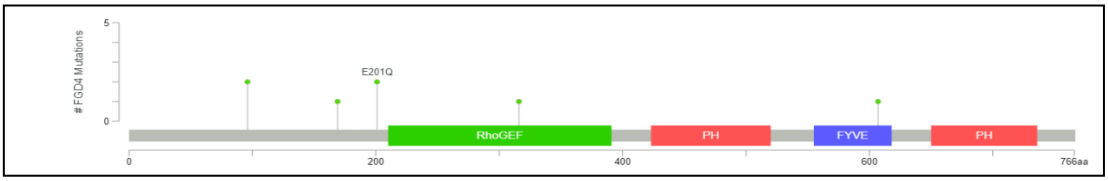
(n) DOCK7



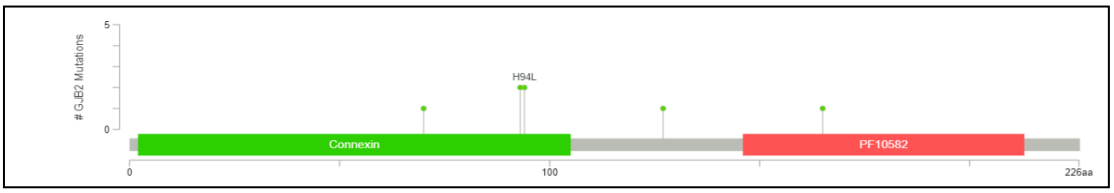
(o) ERCC5



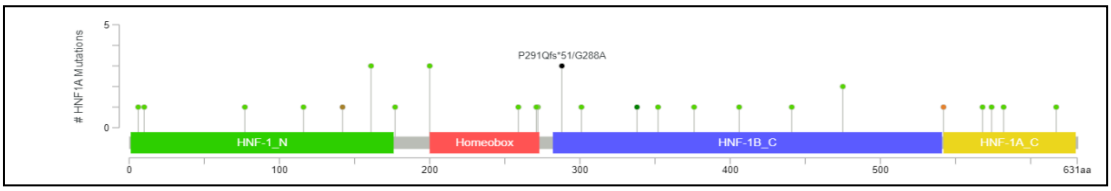
(p) FBXL3



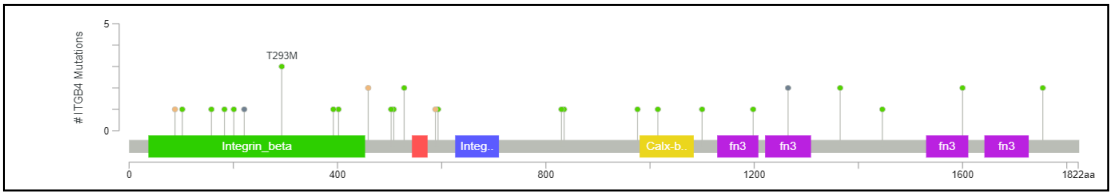
(q) FGD4



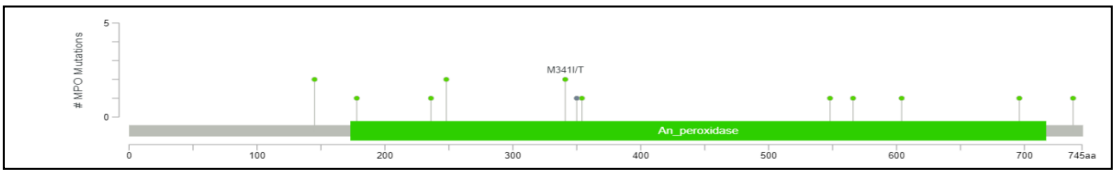
(r) GJB2



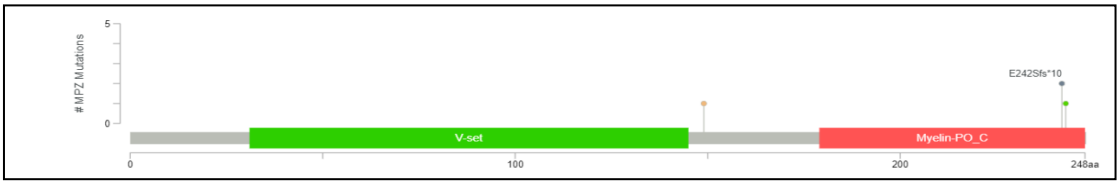
(s) HNF1A



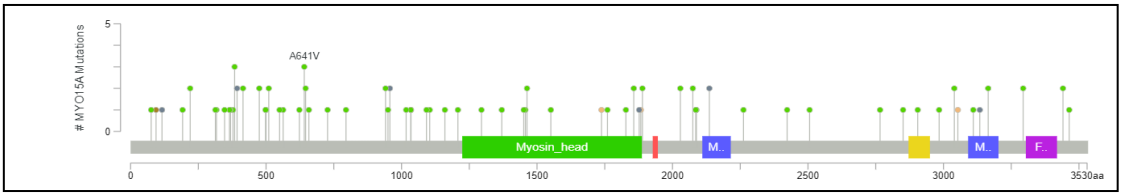
(t) ITGB4`



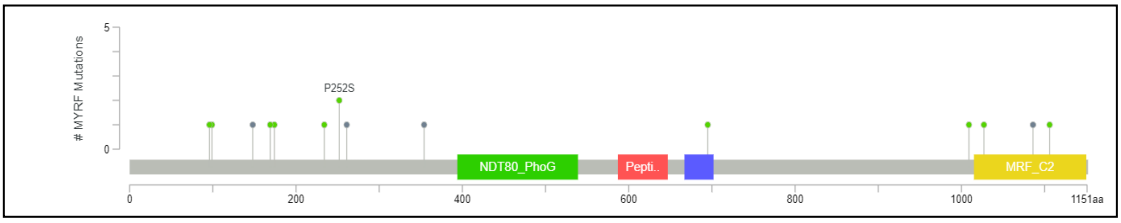
(u) MPO



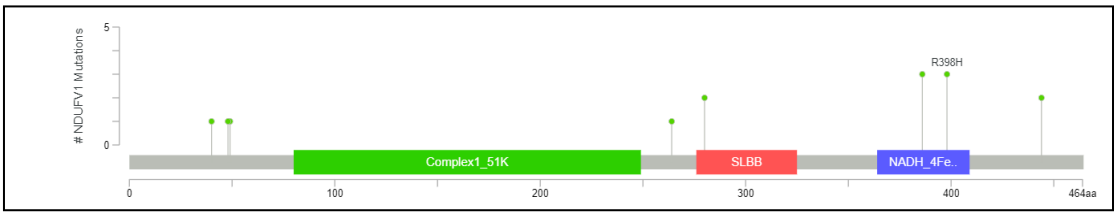
(v) MPZ



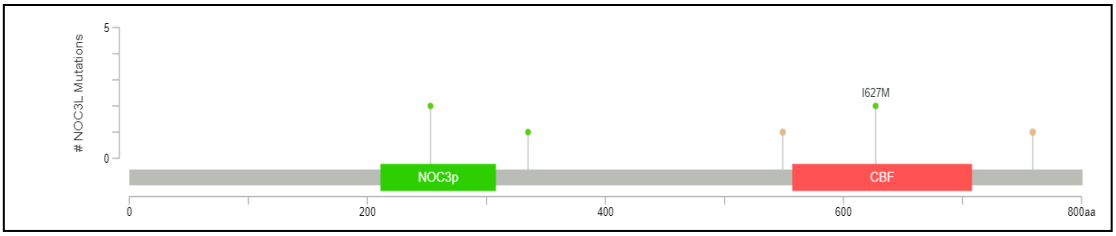
(w) MYO15A



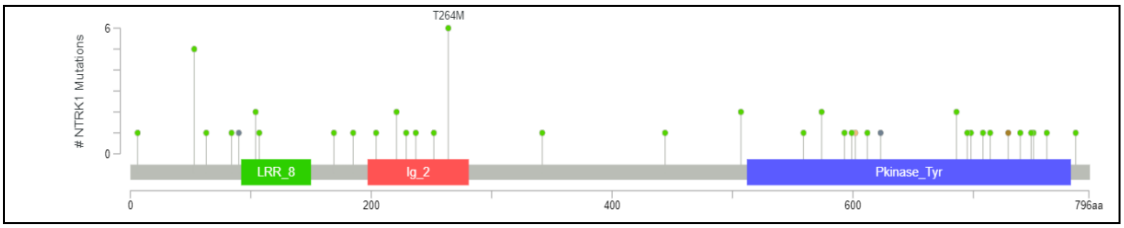
(x) MYRF



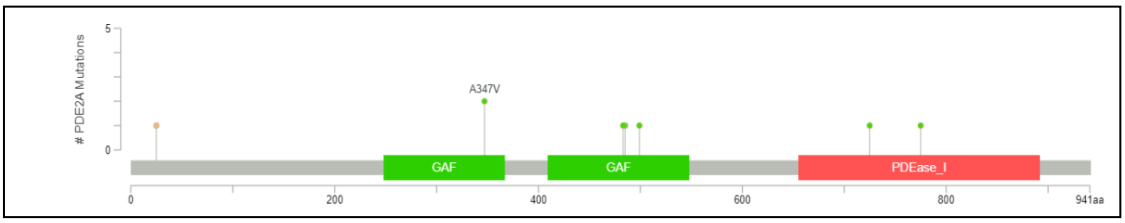
(y) NDUFV1



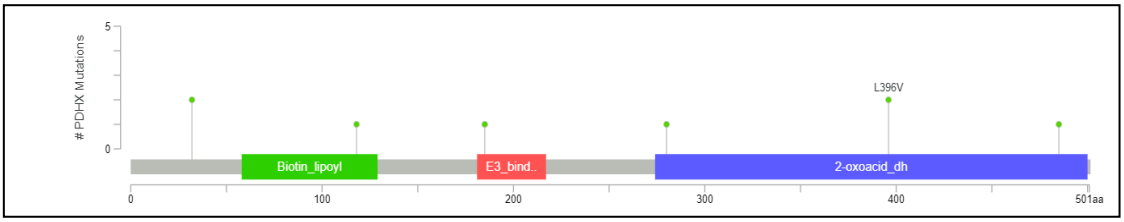
(z) NOC3L



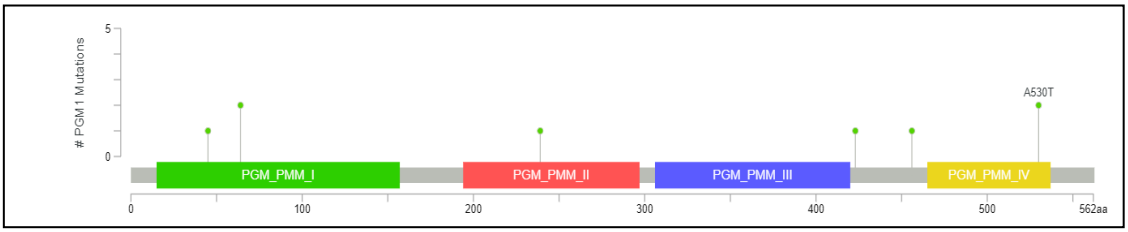
(aa) NTRK1



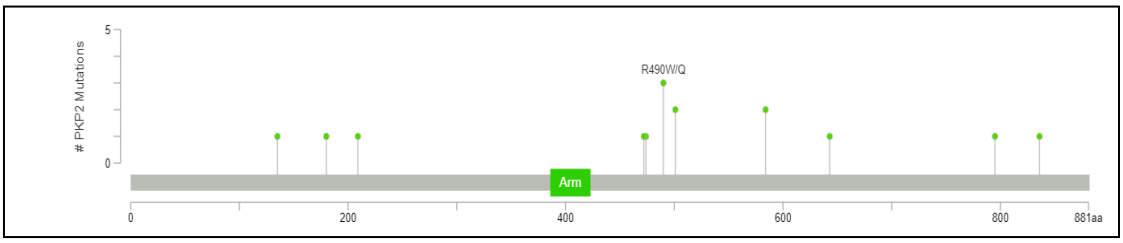
(ab) *PDE2A*



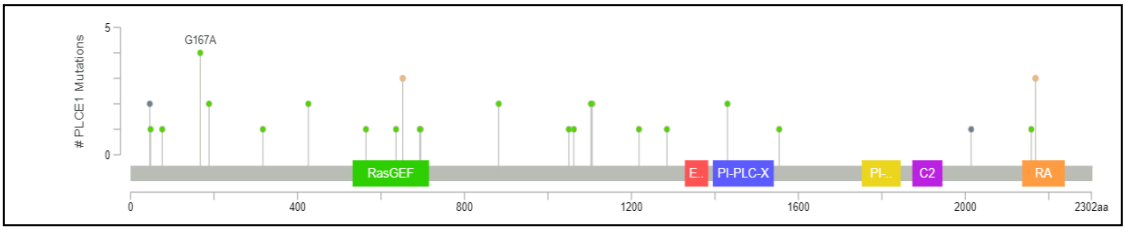
(ac) *PDHX*



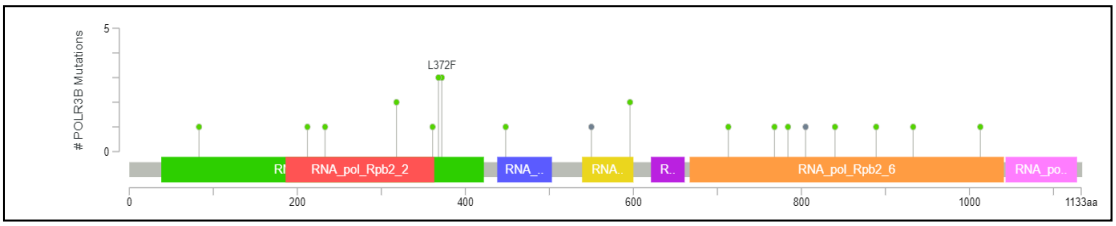
(ad) *PGM1*



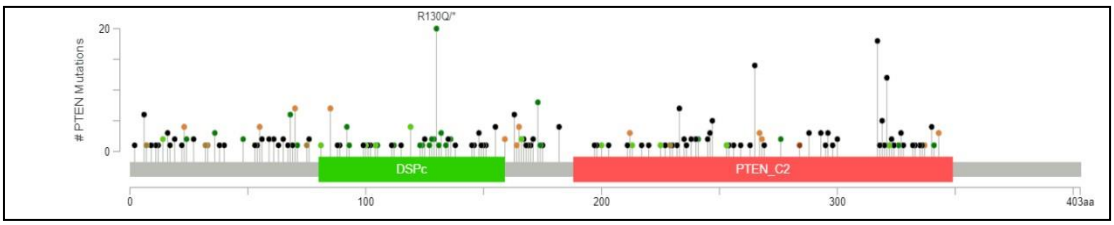
(ae) *PKP2*



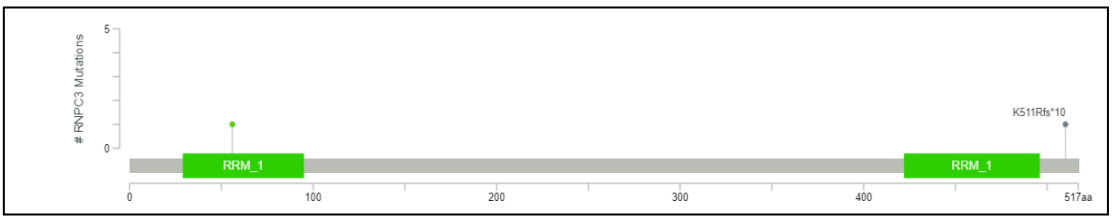
(af) *PLCE1*



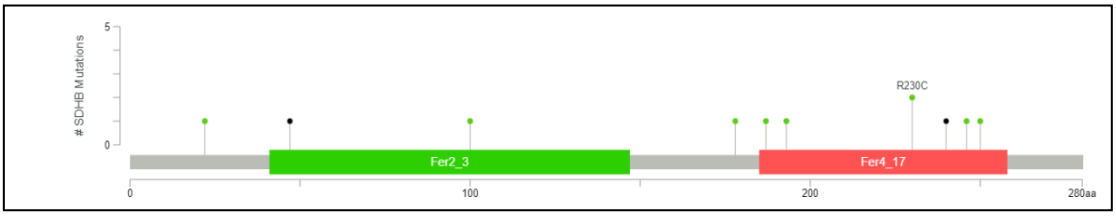
(ag) *POLR3B*



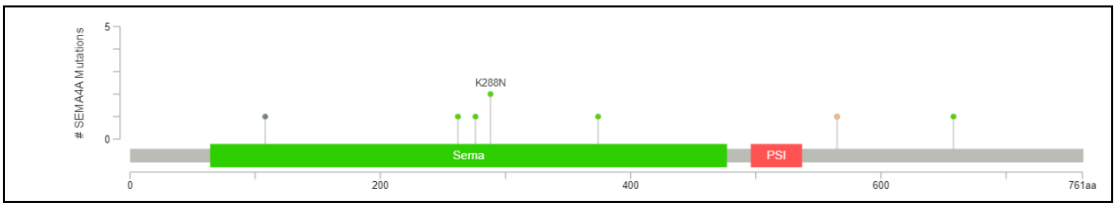
(ah) *PTEN*



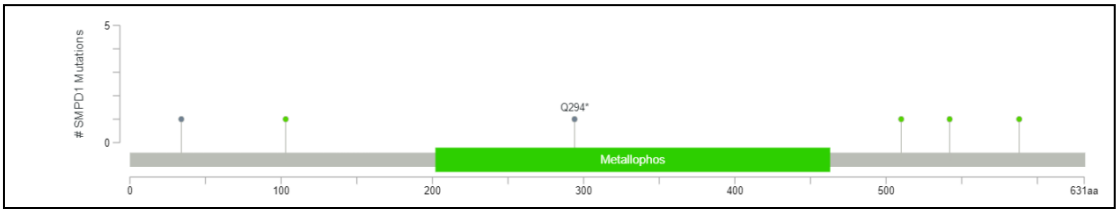
(ai) *RNPC3*



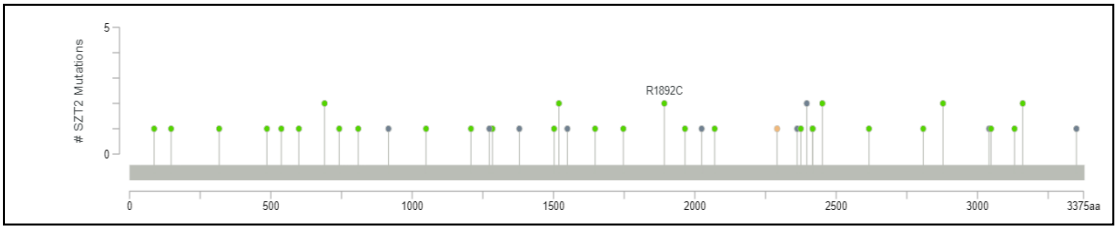
(aj) *SDHB*



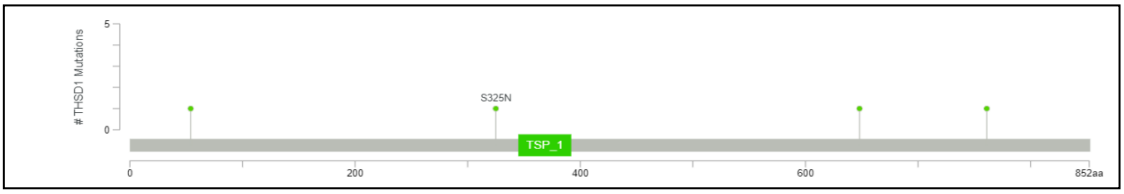
(ak) *SEMA4A*



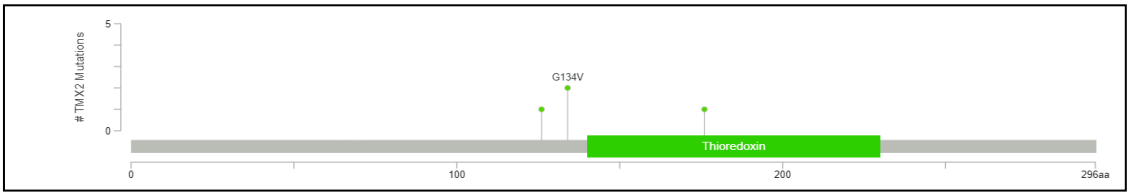
(al) *SMPD1*



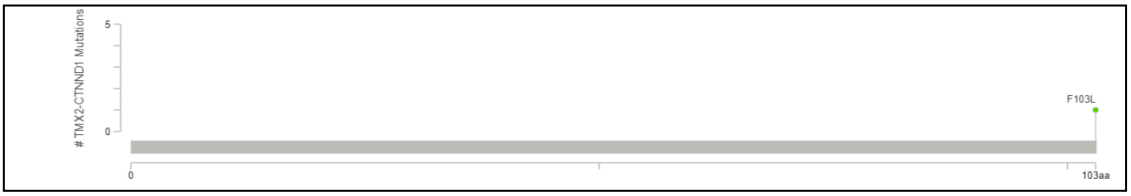
(am) *SZT2*



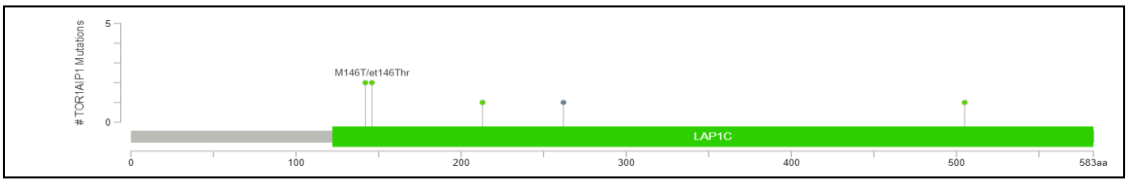
(an) *THSD1*



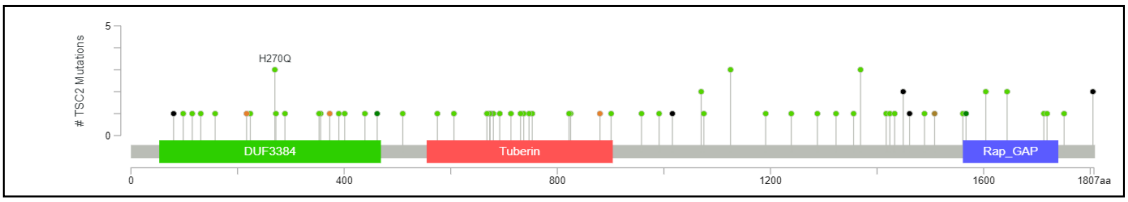
(ao) *TMX2*



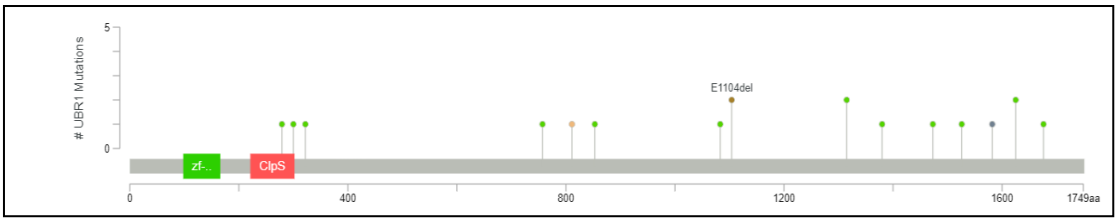
(ap) *TMX2-CTNND1*



(aq) *TOR1AIP1*



(ar) *TSC2*



(as) *UBR1*

Figure 4.20: All the results images **(a-as)** of cBioportal mutational analysis done mapping the identified genes to the PCa datasets. Upright lollipops indicate the position of the different protein changes caused by mutations related to PCa in that particular region. *PTEN*, *BRCA2*, *ATM*, *MYO15A*, *CACNA1C*, *ABCC9*, *TSC2*, *SZT2* and *PLCE1* are high in protein changes.

4.1.2.3. Results of Sanger validation:

As we considered heterozygous mutations, a few variants identified from our cohort study were validated using Sanger chemistry, *viz*; *MYRF*, *MYO15A*, *ATM*, *ITGB4*, *GJB2*, *MPO*, and *HNF1A* **(Table 4.25)**. Validated mutation was identified as heterozygous and there chromatographs helping in visualization of allele change are obtained and visualized using SnapGENE Viewer software (**Figure 4.21 a-i)**. While one mutation was found to be homozygous, it wasn't considered it assuming they are not pathogenic.

Table 4.25: Analysis and identification of Genotype

Gene name	Reference Id	Mutation region	Genotype observed	Remarks
MYRF	rs370887875	c.3207C>T	CT	Heterozygous
MYO15A	rs375290498	c.5925G>A	GA	Heterozygous
ITGB4	rs199620842	c.4559-4C>G	CG	Heterozygous
ATM	rs587779865	c.7456C>T	CT	Heterozygous
GJB2	rs111033186	c.457G>A	GA	Heterozygous
HNF1A	rs1169305	c.1720A>G	GG*	Homozygous
MPO	rs35897051	c.2031-2A>C	AC	Heterozygous
MYRF	rs370887875	c.3207C>T	CT	Heterozygous
MYRF	rs370887875	c.3207C>T	CT	Heterozygous
GJB2	rs111033186	c.457G>A	GA	Heterozygous
HNF1A	rs1169305	c.1720A>G	GG^{+}	Homozygous

Note: The table lists observed genetic variants in Sanger sequencing, including gene name, reference IDs, mutation regions, genotypes and remarks on allele inheritance (Heterozygous and Homozygous) (Last Accessed on April2, 2024)..

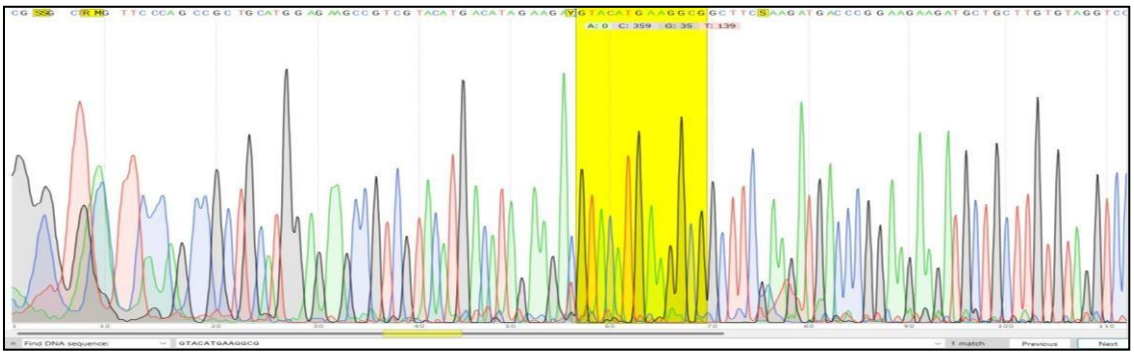


Figure 4.21 a: Chromatograph of Sanger validated gene ***GJB2***

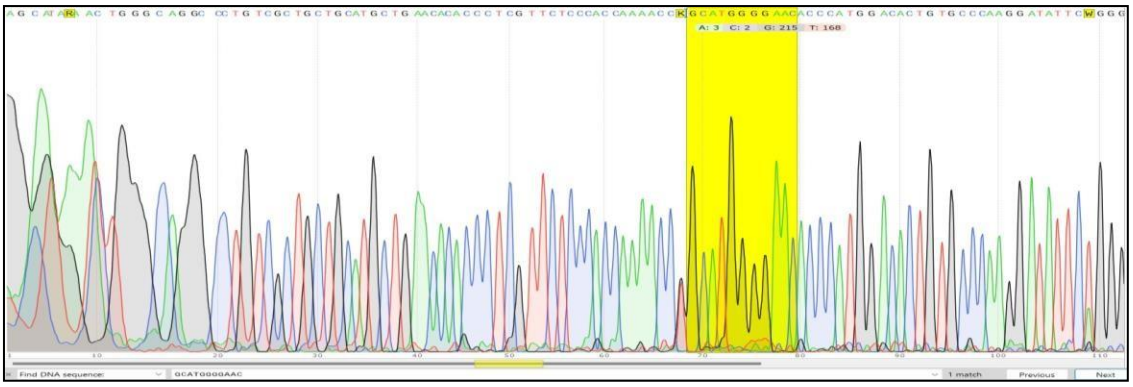


Figure 4.21 b: Chromatograph of Sanger validated gene ***MPO***

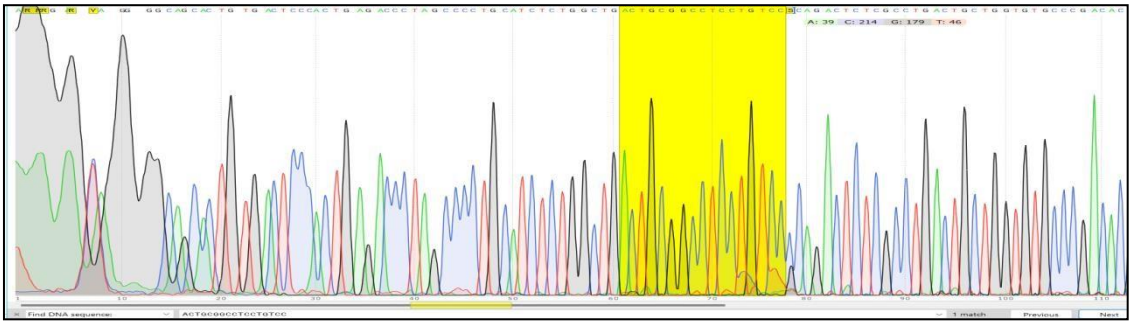


Figure 4.21 c: Chromatograph of Sanger validated gene ***ITGB4***

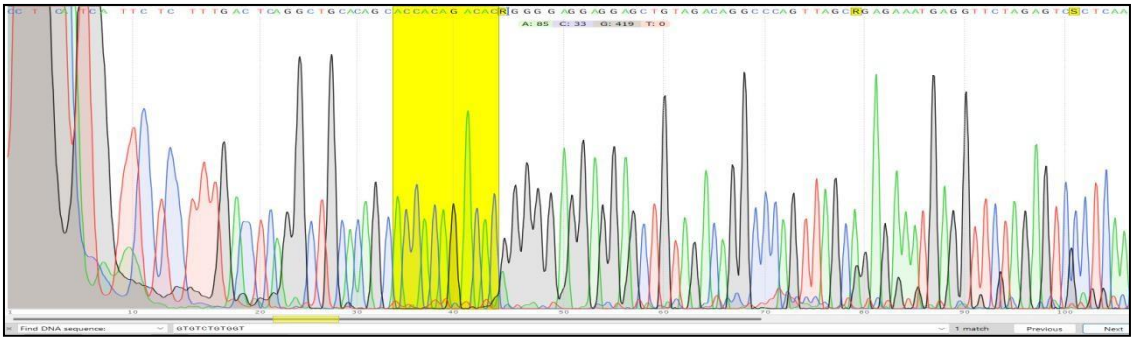


Figure 4.21 d: Chromatograph of Sanger validated gene ***MYRF***

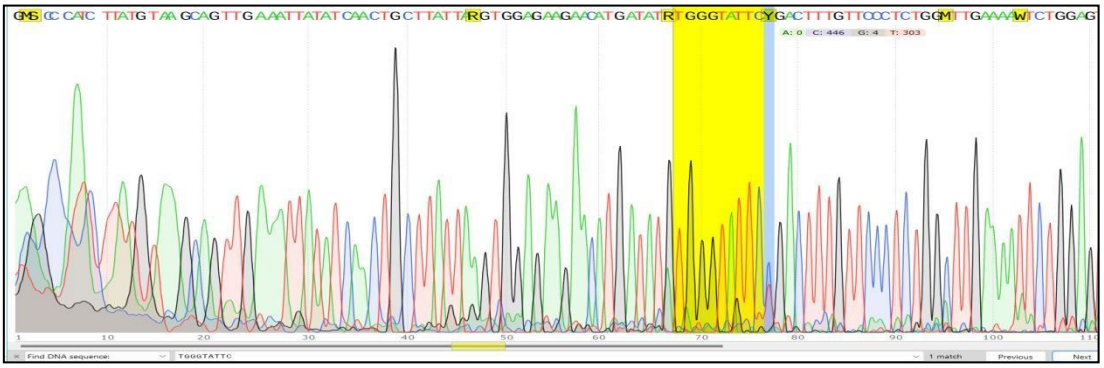


Figure 4.21 e: Chromatograph of Sanger validated gene ***ATM***

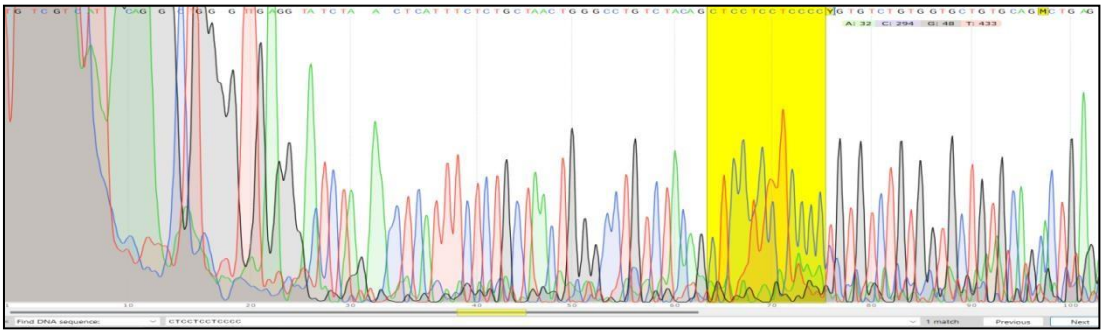


Figure 4.21 f: Chromatograph of Sanger validated gene ***MYRF***

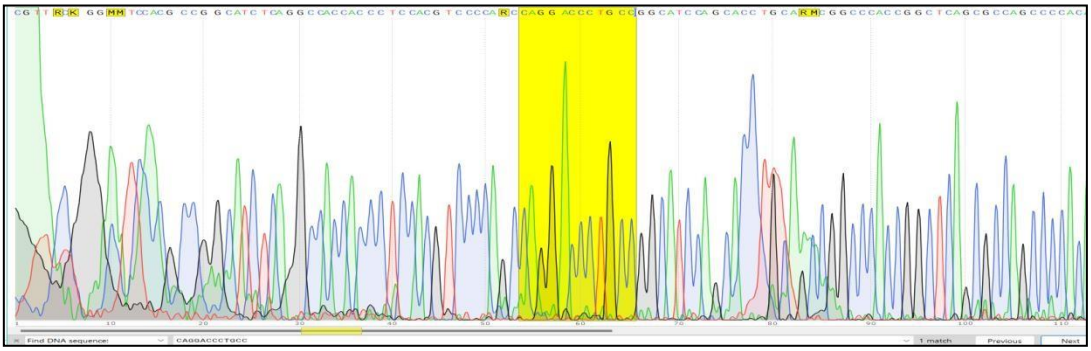


Figure 4.21 g: Chromatograph of Sanger validated gene ***HNF1A***

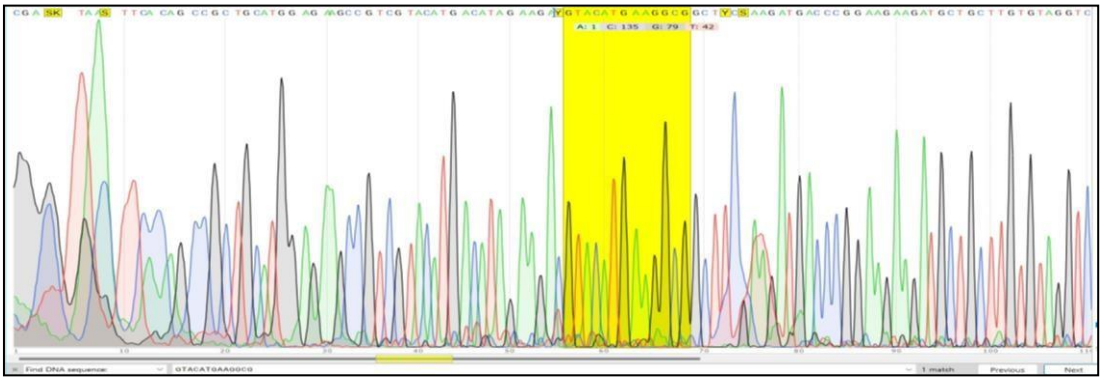


Figure 4.21 h: Chromatograph of Sanger validated gene ***GJB2***

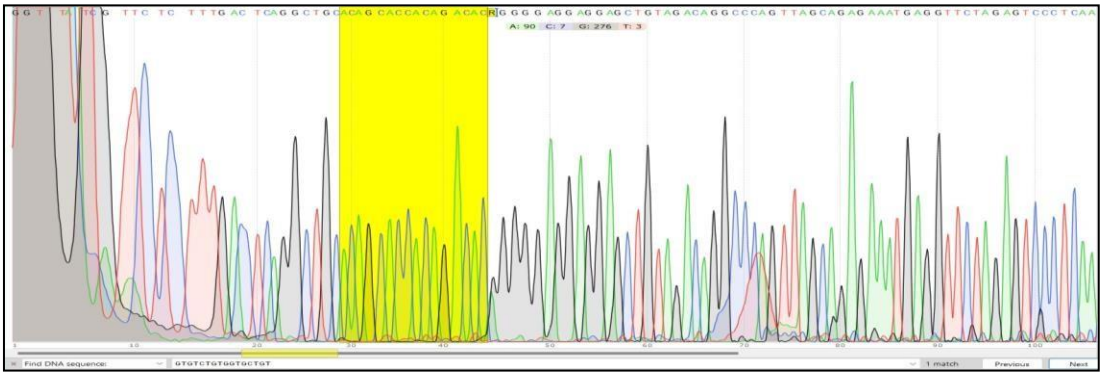


Figure 4.21 i: Chromatograph of Sanger validated gene ***MYRF***

4.2. To identify and explore the role of lncRNAs as potential targets using lncRNA-protein interactions

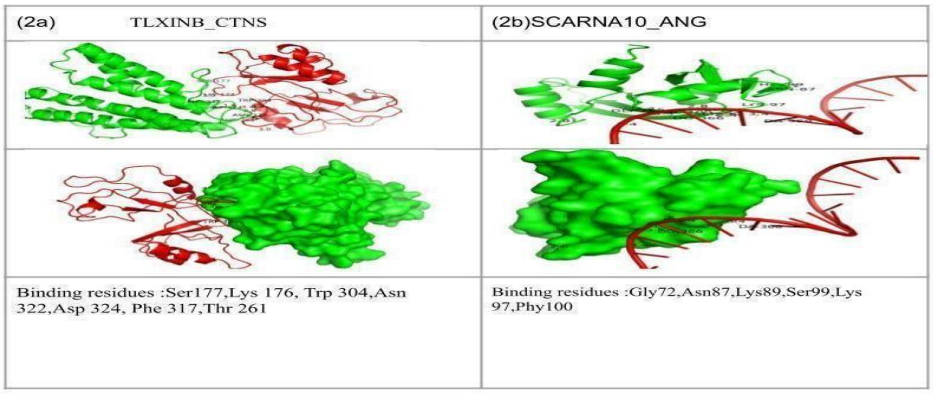
4.2.1. Molecular Docking and Visualization results:

We have considered the HDOCK web server for the molecular docking studies to identify the interactions of receptor-ligands. Among the 11 lncRNAs of them 6 lncRNAs *LINC01973, FLJ16779, LINC00298, SNHG19, LOC341056, and LINC00662:60* due to the restrictions

with sequence with a limit of 5000 residues were unable to perform the interactions and no alternate tools are available to consider. We performed docking on 5 lncRNAs (*SCARNA10, LINC00940, NPBWR1, ANKRD20A9P and TLX1NB*) with 28 PCa and AR targeted proteins (Refer materials: *PCa Causal Proteins, LncRNAs and Androgen Receptors (AR))*. Five lncRNAs (*SCARNA10, LINC00940, NPBWR1, ANKRD20A9P and TLX1NB*) with 27 PCa proteins except USH2A resulted best confirmers each generated 10 models of which CTNS (PDB ID: 5CTG); ANG (PDB ID: 4AOH); CYP1B1 (PDB ID: 3PM0) observed to show more stable complex formation (Khilwani et al., n.d.), (Unpublished data).

4.2.1.1. TLXINB and SCARNA10 LncRNA showed highest PCa related genes interactions

The docked complexes of TLXINB-CTNS (PDB ID: 5CTG) with binding energy-322.82 kcal/mol and interacting residues was identified as Ser177,Lys 176, Trp 304,Asn 322,Asp 324, Phe 317,Thr 261 as shown in figure (**Figure 4.22**a); SCARNA10-ANG(PDB ID: 4AOH) observed to have the least binding energy -259.38 kcal/mol and interacting residues was identified as Gly72, Asn87, Lys89, Ser99, Lys 9, Phy100 as shown in **Figure 4.22**b; in complex NPBWR1-ANG(PDB ID:4AOH) -230.33 kcal/mol and interacting residues was identified as Lys41, Gly72, Ser99, Arg146 as shown in **Figure 4.22**c; in complex ANKRD20A9P-ANG (PDB ID: 4AOH)-220.15 kcal/mol and interacting residues was identified as Asn 73,Ser 99, Lys 97 as shown in **Figure 4.22** d; While the LINC00940-CYP1B1 (PDB ID:3PM0) -201.96 kcal/mol and interacting residues were identified as Gly72,Ser99,Ile143 as shown in **Figure 4.22**e. Results are summarized in **Table 4.22** (Khilwani et al., n.d.)**,** (Unpublished data).



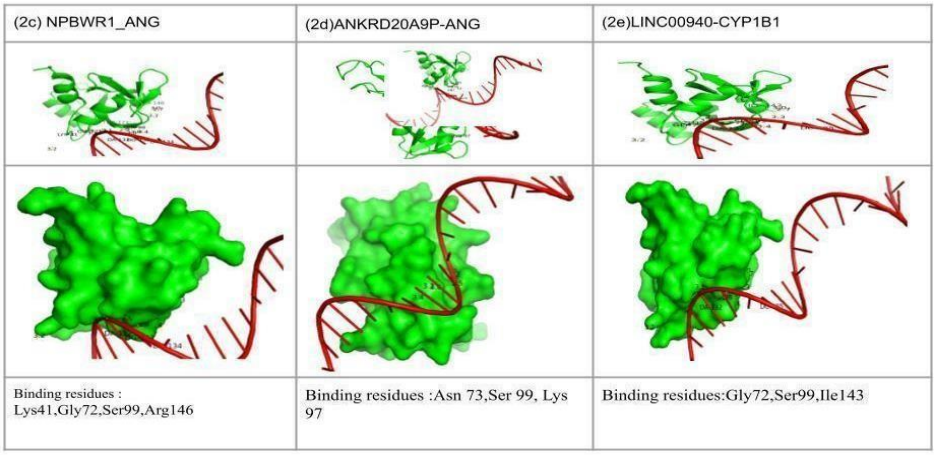


Figure 4.22: Representation of 3D visualizations of the complexes using Pymol for the potential 5-lncRNAs (TXLNIB, SCARNA10, NPBWR1, ANKRD20A9P, and LINC00940) and their interacting residues with 3-PCa proteins (CTNS, ANG, and CYP1B1) **(a-e).**

Table 4.26: The potential candidate 5-lncRNAs' interactions with 3-PCa proteins (PDB id) and binding energies, RMSD and interacting residues were summarized

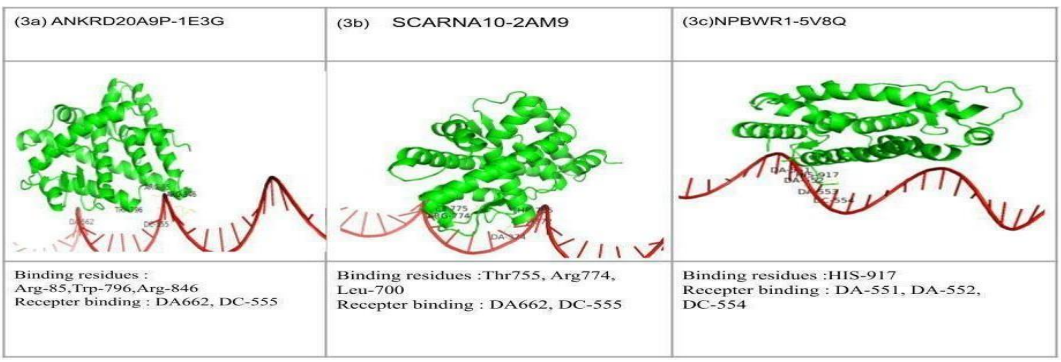
S. No.	LncRNA	PCa Protein	PDB ID	Docking Score	Ligand RMSD (Å)	Interface Residues
1	**TLXINB**	**CTNS**	**7ZKW**	**-322.82**	**57.44**	**Ser177, Lys 176, Trp 304, Asn 322, Asp 324, Phe 317, Thr 261**
2	**SCARNA 10**	**ANG**	**4AOH**	**-259.38**	**482.89**	**Gly72, Asn87, Lys89, Ser99, Lys 97, Phy100**
3	NPBWR 1	ANG	4AOH	-230.33	740.84	Lys41, Gly72, Ser99, Arg146
4	ANKRD 20A9P	ANG	4AOH	-220.15	803.24	Asn 73, Ser 99, Lys 97
5	LINC009 40	CYP1B1	6OYB	-201.96	3095.39	Gly72, Ser99, Ile143

Note: This table summarizes the interactions between 5 lncRNAs and 3 PCa proteins, detailing their PBD IDs, docking scores, ligand root mean square deviation RMSD) and their interface residues involved in the interactions. Negative docking scores indicate a favourable

binding affinity between the lncRNAs and the proteins, while RMSD values reflect the stability of the ligand-protein complex.

4.2.1.2. ANKRD20A9P, SCARNA10 and NPBWR1 *lncRNAs* showed more AR interactions affinity

The docked complexes of 5 lncRNA (*SCARNA10, LINC00940, NPBWR1, ANKRD20A9P, TLX1NB)* with 8 AR targeted proteins (PDB Id: 2Q71, 5V8Q, 4QL8, 2PNU, 5CJ6, 2AM9, 1E3G, 7KW7) excluding 1E3G and 4QL8 AR proteins against TLXINB and all the docking jobs regarding AR 7KW7 with all 5 LncRNAs. We identified least binding energy in 1E3G-ANKRD20A9P, 2AM9-SCARNA10, 5V8Q-NPBWR1 with -212--208 Kcal/mol binding energy as shown in Table 3. 1E3G-ANKRD20A9P has binding energy of -212.74 kcal/mol, with ligand RMSD of 742.99Å and residues are Arg-85,Trp-796,Arg-846 as shown in **Figure 4.23a**; 2AM9-SCARNA10 with binding energy of -208.93 kcal/mol, ligand RMSD as 536.64Å and the interaction residues are Thr755, Arg774, Leu-700 as shown in **Figure 4.23b**; While 5V8Q-NPBWR1 with -208.04 kcal/mol, ligand RMSD 677.49Å of and interacting residues was identified as His-917as shown in **Figure 4.23c**; 2PNU-SCARNA10 with binding energy as -201.39 kcal/mol, RMSD 613.81Å and interacting residues Thr-850, Ser-853,Tyr-857 as shown in **Figure 4.23d**; 5CJ6-SCARNA10 complex has binding energy of -197.78 kcal/mol, RMSD of 631.85Å and binding residues are Trp-796,Leu-797,Pro-868, Thr-918 as shown in **Figure 4.23e**; and 2Q71-SCARNA10 complex showed binding energy of -193.91 kcal/mol, RMSD of 634.84Å and its binding residues are Gln-919,Gln-792, Tyr-857 as shown in **Figure 4.23f** were visualized the 3D complexes with hydrogen interactions bellow 3 Å bond distance. While in 4QL8-ANKRD20A9P we have identified the highest binding energy of -185.99kcal/mol with 762.89+ Å as ligand RMSD value (**Table 4.27)** (Khilwani et al., n.d.), (Unpublished data).



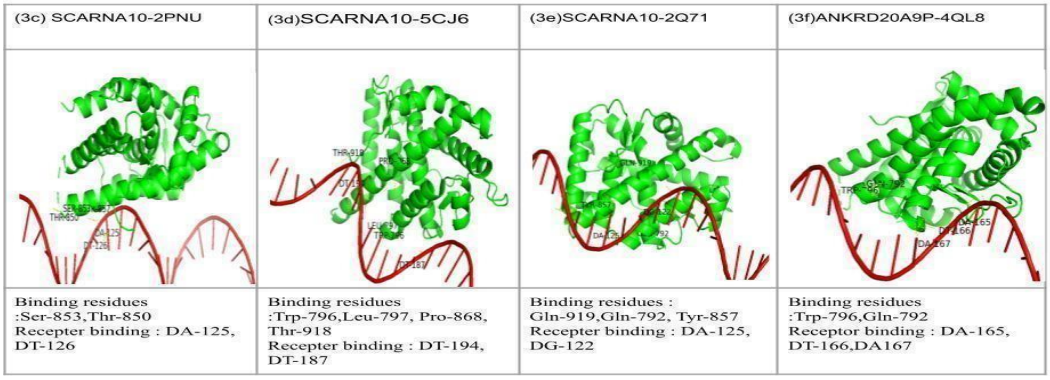


Figure 4.23: Representation of the 3D visualization using Pymol for the complexes of potential 5-lncRNAs (TXLNIB, SCARNA10, NPBWR1, ANKRD20A9P, and LINC00940) and interacting residues with 7 targeted AR proteins (1E3G, 2AM9, 5V8Q, 2PNU, 5CJ6, 2Q71, 4QL8) **(a-f).**

Table 4.27: The potential candidate 5-lncRNAs interactions with 7-AR proteins (PDB id) and binding energies, RMSD and interacting residues were summarized

S. No.	LncRNA	Androgen Receptor Protein	Docking Score	Ligand RMSD (Å)	Interface Residues
1	**ANKRD20A9P**	**1E3G**	**-212.74**	**742.99**	**Arg-85,Trp-796,Arg-846**
2	**SCARNA10**	**2AM9**	**-208.93**	**536.64**	**Thr755, Arg774, Leu-700**
3	**NPBWR1**	**5V8Q**	**-208.04**	**677.49**	**His-917**
4	SCARNA10	2PNU	-201.39	613.81	Thr-850, Ser-853,Tyr-857
5	SCARNA10	5CJ6	-197.78	631.85	Trp-796,Leu-797,Pro-868, Thr-918
6	SCARNA10	2Q71	-193.91	634.84	Gln-919,Gln-792, Tyr-857
7	ANKRD20A9P	4QL8	-185.99	762.89	Trp-796,Gln-792

Note: This table summarizes the interactions between 7 lncRNAs and 7 ARs, detailing their PBD IDs, docking scores, ligand root mean square deviation RMSD) and their interface residues involved in the interactions. Negative docking scores indicate a favourable binding

affinity between the lncRNAs and the ARs, while RMSD values reflect the stability of the ligand-protein complex.

Our study clearly highlighted that the SCARNA10 is a potential lncRNA which is showing lowest binding affinity energy with both PCa protein (ANG) and AR (2AM9). The binding score is -259.38 (SCARNA10-ANG) and -259.38 (SCARNA10-2AM9) with interacting residues which are suggesting their potency (**Figure 4.24)** (Khilwani et al., n.d.)**, (**Unpublished data).

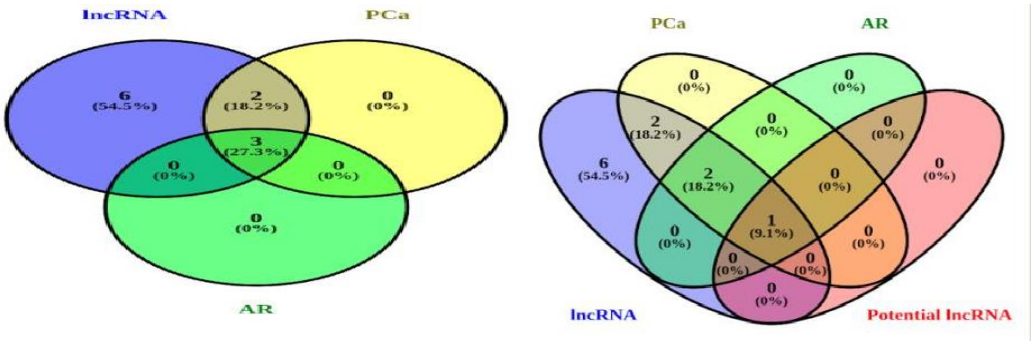


Figure 4.24: Statistical overview of 11 lncRNAs and 3 lncRNAs that are common, of which **SCARNA10** as potential biomarker using Venny 2.1.

4.2.2. Common LncRNAs identified through correlation of PCa related lncRNAs from RNA-seq study and publically available datasets of T2D lncRNAs:

On comparing identified 11 lncRNAs (*SCARNA10, LINC01973, LINC00940, NPBWR1, FLJ16779, ANKRD20A9P, LINC00298, SNHG19, LOC341056, TLX1NB, LINC00662:60)* discovered in RNA-seq study, we found 9 lncRNAs (*SCARNA10, LINC01973, LINC00940, NPBWR1, FLJ16779, ANKRD20A9P, LINC00298, SNHG19, and TLX1NB)* were common with datasets available in T2DB database (**Table 4.28).**

Table 4.28: List of common PCa related LncRNAs with T2D related lncRNAs datasets

Ensemble Gene ID	Gene Symbol	Biotype	Function	Reference
ENSG00000235665	*LINC00298*	lncRNA	Long Intergenic non coding RNA with unknown function	(Kenkpen et al., 2023) (Prokopenko et al., 2021)
ENSG00000288	*NPBWR1*	protein coding	Regulatory lncRNA	(Stein et al.,

11				2023)
ENSG00000236311	*TLX1NB*	lncRNA	Neighbouring lncRNA of TLX1 gene and regulates in chromatin binding, modification and transcription	(Verboom et al., 2018)
ENSG00000235049	*LINC00940*	lncRNA	Long Intergenic non coding RNA, with unknown function, and forms triple helices in promoter regions with AL157886.1 lncRNA	(KUNKLER et al., 2022)
ENSG00000239002	*SCARNA10*	snoRNA	Small Nucleolar RNA, serum lncRNA, regulates the transcription and linked to PRC2 complex	(K. Zhang et al., 2019)
ENSG00000206192	*ANKRD20A9P*	unprocessed_pseudogene	Chromatin remodeling	(Tindale, 2018)
ENSG00000260260	*SNHG19*	lncRNA	Regulation at both transcriptional and post-translational level	(G. Y. Zhao et al., 2021)
ENSG00000204283	*LINC01973*	lncRNA	Long Intergenic non coding RNA and with unknown function	https://www.genecards.org/cgi-bin/carddisp.pl?gene=LINC01973
ENSG00000275620	*FLJ16779*	lncRNA	An Uncharacterized lncRNA	https://www.genecards.org/cgi-bin/carddisp.pl?gene=FLJ16779&keywords=FLJ16779

Note: This table lists the lncRNAs associated with PCa and their relevance in T2DM. It includes the Ensemble Gene ID, gene symbol, biotype, functional annotations, and references for each lncRNA. Notably, some lncRNAs exhibit unknown functions, while others are linked to chromatin remodeling and regulatory processes.

4.3. To perform protein-protein interaction (PPI) network analyses and identify pathways associated with T2D and PCa *in silico* method

4.3.1. *Interactome networks using GeneMania* for Metadata analysis:

4.3.1.1. BLM, TMPO and FOXP1 as common interacting partners:

On applying GeneMANIA to the common genes of PCa and diabetes mellitus data of ClinVar, the interactome network showed a common interacting partner among them, which is ***TMPO*** **(Figure 4.25 A.).** The interactome network of *PP1MB* and *SFTPC* showed some commonalities in their interacting partners with other datasets. *SFTPC*'s interacting partner *TMEM67* is common with PCa ClinVar data and *FOXP1* with diabetes ClinVar data. *PPM1B*'s interacting partner PPP2CA is common with PCa ClinVar data and *PPARG* is common with diabetes ClinVar. *TMEM67* is also common among the ClinVar data sets and *SFTPC*, whereas, ***FOXP1*** is found as the common interacting partner of *PP1MB*, *SFTPC*, and *TMEM67* (**Figure 4.25**). *FOXP1* showed genetic interactions with both *SFTPC* and *PPM1B* (**Figure 4.25**). However, the rest of the interactome broadly showed physical interactions (red) with their functions and pathways involved enlisted from Genecards. (www.genecards.org last accessed on May 25, 2022). On the other hand, what was seen as interesting is the BLM gene which is found as the only common gene between PCa, and diabetes, interacting with both the earlier identified *FOXP1* (physical interactions shown in red color) and *TMPO* (co-expression shown in purple color) (**Figure 4.25 C-D**) (Kour et al., 2023).

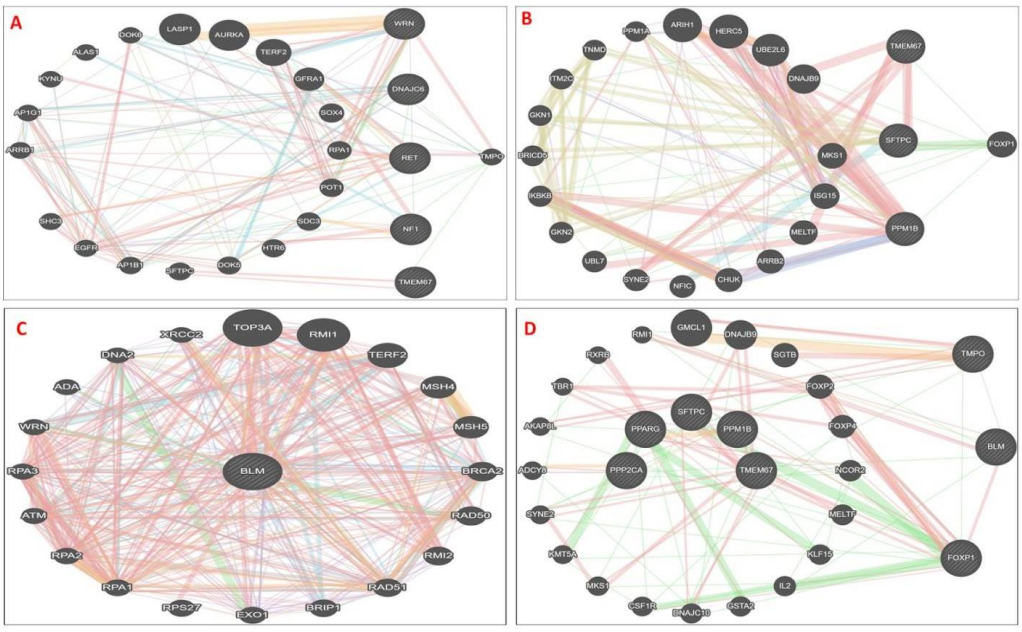


Figure 4.25: A. Interactome Network of commonly associated genes (*WRN, DNAJC6, RET, NF1AND TMEM67)* of PCa and diabetes from ClinVar data, *TMPO* is shown as a common interacting partner; **B**. Interactome network of *TMEM67*, *SFTPC*, and *PPMIB* showing mainly genetic interactions (green) with one common *FOXP1*; **C**. Interactome network of *BLM* gene found it as the only one common gene between PCa, diabetes, and obesity that is interacting with both the earlier identified *FOXP1*(physical interactions shown in red color) and *TMPO* (co-expression shown in purple color) which can be seen in **D**.

4.3.2. *Interactome networks using GeneMania revealed proteins with* co-expression of 49.23% associated with common pathways *:*

Apart from the commonalities of meta-analysis, we also studied the protein-protein interaction studies for our cohort mutated gene set also by using GeneMania and prepared the PPI network map. Majority of genes showed co-expression of 49.23%, common pathways 23.71%, genetic interactions 9.10%, co-localization 8.32%, physical interaction 7.91% and for predictive it was 1.7% (**Figure 4.26)** (Kour et al., 2023).

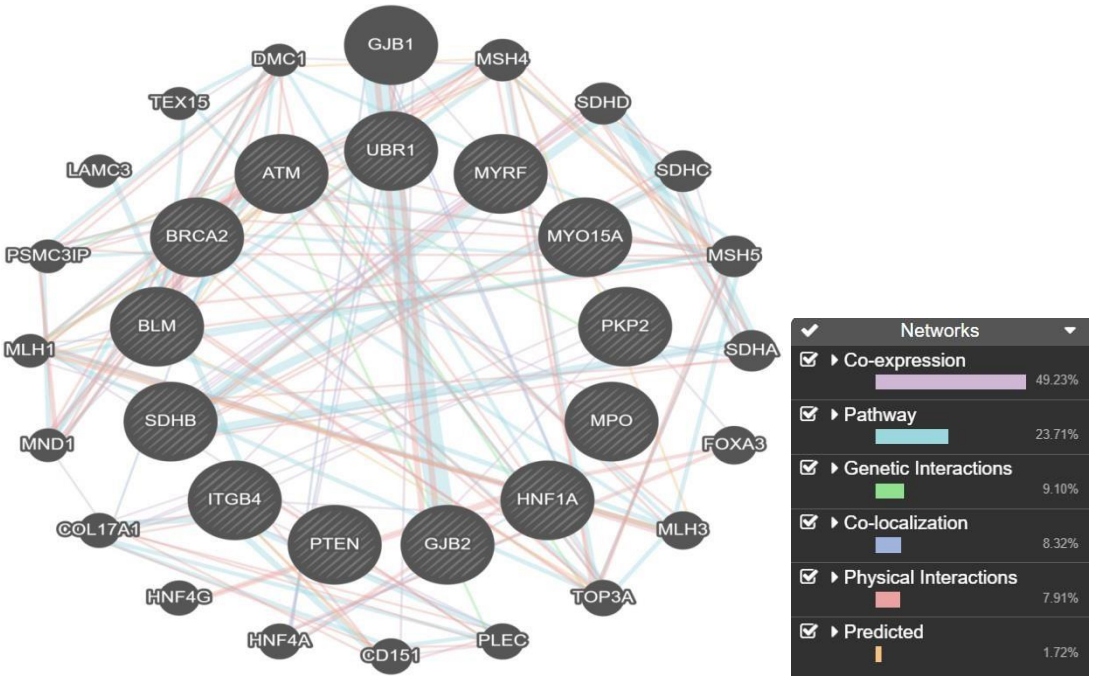


Figure 4.26: PPI map of selected Sanger validated gene set. Different color of edges indicates different kind of interactions among the proteins *viz.* Co-expression of 49.23% (purple), pathway of 23.71% (sky blue), genetic interactions of 9.10% (green), co-localization of 8.32% (blue), physical interactions of 7.91% (red) and predicted genes of 1.72% with orange color. PPI map indicates majority of genes are co-expressed.

4.3.3. Pathway Enrichment of Gene sets from our PCa Cohort:

4.3.3.1. Kegg Mapper and Kegg Pathway

The KEGG, a database of signaling pathways is used and we obtained that most of the genes are enriched in signaling pathways related to glucose metabolism and insulin resistance (**Table 4.29)**.

Table 4.29: List of pathways in which our input gene set playing role obtained by KEGG Mapper and pathway

PATHWAYS	**PATHWAYS**	**PATHWAYS**
hsa01100 Metabolic pathways (9)	hsa04261 Adrenergic signaling in cardiomyocytes (2)	hsa04142 Lysosome (2)
hsa04151 PI3K-Akt signaling pathway (5)	hsa04925 Aldosterone synthesis and secretion (2)	hsa05016 Huntington disease (2)
hsa05165 Human papillomavirus infection (4)	hsa05166 Human T-cell leukemia virus 1 infection (2)	hsa04015 Rap1 signaling pathway (2)
hsa05412 Arrhythmogenic right ventricular cardiomyopathy (4)	hsa00190 Oxidative phosphorylation (2)	hsa04260 Cardiac muscle contraction (2)
hsa05410 Hypertrophic cardiomyopathy (3)	hsa04932 Non-alcoholic fatty liver disease (2)	hsa04068 FoxO signaling pathway (2)
hsa05415 Diabetic cardiomyopathy (3)	hsa00500 Starch and sucrose metabolism (2)	hsa05230 Central carbon metabolism in cancer (2)
hsa05208 Chemical carcinogenesis - reactive oxygen species (3)	hsa04071 Sphingolipid signaling pathway (2)	hsa05202 Transcriptional misregulation in cancer (2)
hsa05200 Pathways in cancer (3)	hsa04210 Apoptosis (2)	hsa05012 Parkinson disease (2)
hsa05022 Pathways of neurodegeneration - multiple diseases (3)	hsa04510 Focal adhesion (2)	hsa01524 Platinum drug resistance (2)
hsa05224 Breast cancer (3)	hsa04024 cAMP signaling pathway (2)	hsa05131 Shigellosis (2)

hsa04714 Thermogenesis (3)	hsa00230 Purine metabolism (2)	hsa04919 Thyroid hormone signaling pathway (2)
hsa04010 MAPK signaling pathway (3)	hsa03460 Fanconi anemia pathway (2)	hsa04022 cGMP-PKG signaling pathway (2)
hsa05020 Prion disease (3)	hsa00562 Inositol phosphate metabolism (2)	hsa04950 Maturity onset diabetes of the young (1)
hsa05010 Alzheimer disease (3)	hsa04070 Phosphatidylinositol signaling system (2)	hsa04810 Regulation of actin cytoskeleton (1)
hsa04020 Calcium signaling pathway (3)	hsa04014 Ras signaling pathway (2)	hsa05216 Thyroid cancer (1)
hsa04115 p53 signaling pathway (3)	hsa04723 Retrograde endocannabinoid signaling (2)	hsa04270 Vascular smooth muscle contraction (1)
hsa03440 Homologous recombination (3)	hsa04150 mTOR signaling pathway (2)	hsa04750 Inflammatory mediator regulation of TRP channels (1)
hsa05414 Dilated cardiomyopathy (3)	hsa04921 Oxytocin signaling pathway (2)	hsa05163 Human cytomegalovirus infection (1)
hsa04218 Cellular senescence (3)	hsa05014 Amyotrophic lateral sclerosis (2)	hsa05231 Choline metabolism in cancer (1)
hsa05206 MicroRNAs in cancer (3)	hsa04140 Autophagy - animal (2)	hsa04724 Glutamatergic synapse (1)
hsa04072 Phospholipase D signaling pathway (1)	hsa01250 Biosynthesis of nucleotide sugars (1)	hsa04713 Circadian entrainment (1)
hsa05170 Human immunodeficiency virus 1 infection (1)	hsa05417 Lipid and atherosclerosis (1)	hsa04211 Longevity regulating pathway (1)
hsa05031 Amphetamine addiction (1)	hsa04720 Long-term potentiation (1)	hsa04360 Axon guidance (1)
hsa05212 Pancreatic cancer (1)	hsa04924 Renin secretion (1)	hsa05205 Proteoglycans in cancer (1)
hsa05168 Herpes simplex virus 1 infection (1)	hsa04910 Insulin signaling pathway (1)	hsa04979 Cholesterol metabolism (1)
hsa05207 Chemical carcinogenesis - receptor	hsa00010 Glycolysis / Gluconeogenesis (1)	hsa04110 Cell cycle (1)

activation (1)		
hsa04911 Insulin secretion (1)	hsa00030 Pentose phosphate pathway (1)	hsa04512 ECM-receptor interaction (1)
hsa04927 Cortisol synthesis and secretion (1)	hsa04725 Cholinergic synapse (1)	hsa05218 Melanoma (1)
hsa00020 Citrate cycle (TCA cycle) (1)	hsa03420 Nucleotide excision repair (1)	hsa04670 Leukocyte transendothelial migration (1)
hsa04722 Neurotrophin signaling pathway (1)	hsa04912 GnRH signaling pathway (1)	hsa00520 Amino sugar and nucleotide sugar metabolism (1)
hsa00600 Sphingolipid metabolism (1)	hsa04931 Insulin resistance (1)	hsa05213 Endometrial cancer (1)
hsa05032 Morphine addiction (1)	hsa04977 Vitamin digestion and absorption (1)	hsa04740 Olfactory transduction (1)
hsa04726 Serotonergic synapse (1)	hsa05214 Glioma (1)	hsa05225 Hepatocellular carcinoma (1)
hsa00052 Galactose metabolism (1)	hsa04935 Growth hormone synthesis, secretion and action (1)	hsa04727 GABAergic synapse (1)
hsa01521 EGFR tyrosine kinase inhibitor resistance (1)	hsa05235 PD-L1 expression and PD-1 checkpoint pathway in cancer (1)	hsa04933 AGE-RAGE signaling pathway in diabetic complications (1)
hsa03020 RNA polymerase (1)	hsa04217 Necroptosis (1)	hsa04814 Motor proteins (1)
hsa04975 Fat digestion and absorption (1)	hsa04520 Adherens junction (1)	hsa02010 ABC transporters (1)
hsa05215 PCa (1)	hsa04064 NF-kappa B signaling pathway (1)	hsa04728 Dopaminergic synapse (1)
hsa04742 Taste transduction (1)	hsa01200 Carbon metabolism (1)	hsa04934 Cushing syndrome (1)
hsa04152 AMPK signaling pathway (1)	hsa04514 Cell adhesion molecules (1)	hsa04930 Type II diabetes mellitus (1)
hsa04120 Ubiquitin mediated proteolysis (1)	hsa04623 Cytosolic DNA-sensing pathway (1)	hsa05222 Small cell lung cancer (1)

hsa05222 Small cell lung cancer (1)	hsa04929 GnRH secretion (1)	hsa04929 GnRH secretion (1)
hsa04072 Phospholipase D signaling pathway (1)	hsa01250 Biosynthesis of nucleotide sugars (1)	hsa04713 Circadian entrainment (1)
hsa05170 Human immunodeficiency virus 1 infection (1)	hsa05417 Lipid and atherosclerosis (1)	hsa04211 Longevity regulating pathway (1)
hsa05031 Amphetamine addiction (1)	hsa04720 Long-term potentiation (1)	hsa04360 Axon guidance (1)
hsa05212 Pancreatic cancer (1)	hsa04924 Renin secretion (1)	hsa05205 Proteoglycans in cancer (1)
hsa05168 Herpes simplex virus 1 infection (1)	hsa04910 Insulin signaling pathway (1)	hsa04979 Cholesterol metabolism (1)
hsa05207 Chemical carcinogenesis - receptor activation (1)	hsa00010 Glycolysis / Gluconeogenesis (1)	hsa04110 Cell cycle (1)
hsa04911 Insulin secretion (1)	hsa00030 Pentose phosphate pathway (1)	hsa04512 ECM-receptor interaction (1)
hsa04927 Cortisol synthesis and secretion (1)	hsa04725 Cholinergic synapse (1)	hsa05218 Melanoma (1)
hsa00020 Citrate cycle (TCA cycle) (1)	hsa03420 Nucleotide excision repair (1)	hsa04670 Leukocyte transendothelial migration (1)
hsa04722 Neurotrophin signaling pathway (1)	hsa04912 GnRH signaling pathway (1)	hsa00520 Amino sugar and nucleotide sugar metabolism
hsa00600 Sphingolipid metabolism (1)	hsa04931 Insulin resistance (1)	hsa05213 Endometrial cancer (1)
hsa05032 Morphine addiction (1)	hsa04977 Vitamin digestion and absorption (1)	hsa04740 Olfactory transduction (1)
hsa04726 Serotonergic synapse (1)	hsa05214 Glioma (1)	hsa05225 Hepatocellular carcinoma (1)
hsa00052 Galactose metabolism (1)	hsa04935 Growth hormone synthesis, secretion and action (1)	hsa04727 GABAergic synapse (1)
hsa01521 EGFR tyrosine	hsa05235 PD-L1 expression	hsa04933 AGE-RAGE

kinase inhibitor resistance (1)	and PD-1 checkpoint pathway in cancer (1)	signaling pathway in diabetic complications (1)
hsa03020 RNA polymerase (1)	hsa04217 Necroptosis (1)	hsa04814 Motor proteins (1)
hsa04975 Fat digestion and absorption (1)	hsa04520 Adherens junction (1)	hsa02010 ABC transporters (1)
hsa05215 PCa (1)	hsa04064 NF-kappa B signaling pathway (1)	hsa04728 Dopaminergic synapse (1)
hsa04742 Taste transduction (1)	hsa01200 Carbon metabolism (1)	hsa04934 Cushing syndrome (1)
hsa04152 AMPK signaling pathway (1)	hsa04514 Cell adhesion molecules (1)	hsa04930 Type II diabetes mellitus (1)
hsa04120 Ubiquitin mediated proteolysis (1)	hsa04623 Cytosolic DNA-sensing pathway (1)	

Note: This table summarizes the pathways in which the input gene set is implicated, as identified through KEGG Mapper analysis. Each pathway is listed alongside the number of genes associated with it from input, demonstrating the diverse biological processes and signaling pathways influenced by the gene set. These include those pathways related to metabolic pathways, cancer, diabetes, and various signaling mechanisms

On giving the input of "PCa" the KEGG pathway searched and displayed a complex network of multiple pathways related to PCa already explored and published in many studies (**Figure 4.27)**. Looking at the cross talk of pathways, we obtained 5 major pathways involved in PCa. These pathways also regulate glucose metabolism and insulin sensitivity. The identified pathways are Mutation-inactivated *PTEN* to *PI3K* signaling pathway, Loss *of NKX3-1* to *PI3K* signaling pathway, Amplified AR to androgen receptor signaling pathway, Mutation-activated AR to androgen receptor signaling pathway and Loss of *CDKN1B* to p27-cell cycle G1/S **(Table 4.30)**

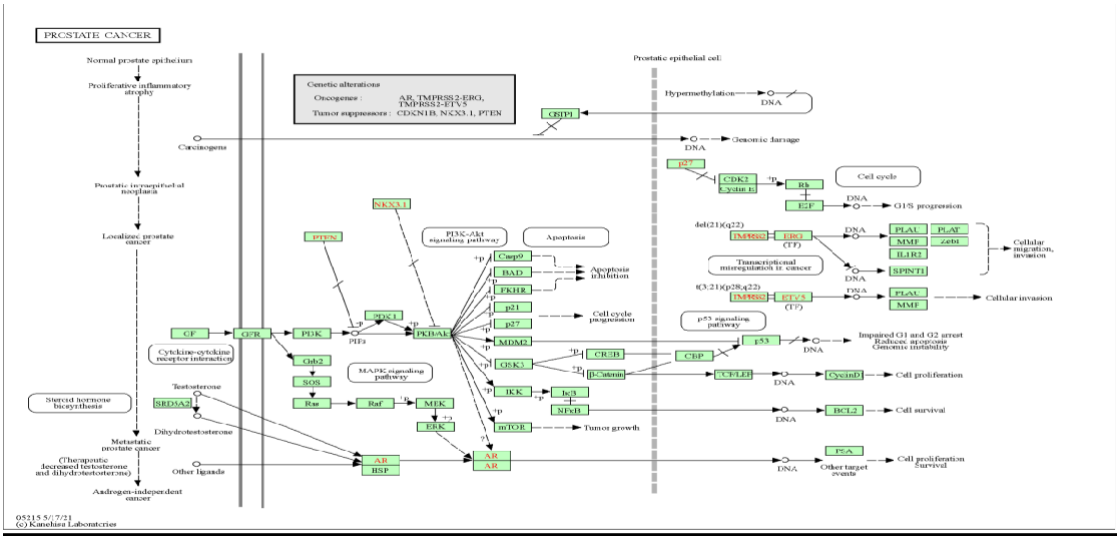


Figure 4.27: PCa pathway retrieved from KEGG pathway representing how normal prostate epithelium changes to PCa involving oncogenes, tumor suppressors and major signaling pathways playing roles like PI3K-Akt signaling, MAPK signaling, p53 signaling, transcriptional misregulation, hypermethylation, and Cytokine-cytokine receptor interaction to bring steroid hormone biosynthesis, genome damage, apoptosis inhibitions, tumor growth, cell survival, cell proliferation, cell invasions and impaired G1 and G2 arrest

Table 4.30: List of PCa related pathways from KEGG pathway database

Pathway id	Pathway involved in PCa
N00052	Mutation-inactivated PTEN to PI3K signaling pathway
N00082	Loss of NKX3-1 to PI3K signaling pathway
N00084	Amplified AR to androgen receptor signaling pathway
N00085	Mutation-activated AR to androgen receptor signaling pathway
N00093	Loss of CDKN1B to p27-cell cycle G1/S

4.3.3.2. Reactome identified top six significant pathways with focus on DNA repair genes *Tp53, PALB2, BRCA1, BRCA2* and *ATM :*

Pathway enrichment analysis using the reactome tool gave a list of associated pathways but we selected the top 6 pathways by filtering on the basis of FDR and *p* value below 0.05 **(Table 4.31).** The top six pathways significantly involved are *viz.*, regulation of transcription of DNA repair genes *Tp53*, defective homologous recombination repair (HRR) due to *PALB2*

loss of function, Defective HDR through Homologous Recombination Repair (HRR) due to *PALB2* loss of BRCA2/RAD51/RAD51C binding function, Defective HDR through Homologous Recombination Repair (HRR) due to PALB2 loss of BRCA1 binding function, Defective HDR through Homologous Recombination Repair (HRR) due to PALB2 loss of *BRCA1* binding function, Impaired *BRCA2* binding to PALB2 and Defective homologous recombination repair (HRR) due to *BRCA1* loss of function. *ATM*, *BRCA2* and *BRCA1* are playing a major role in the identified significant pathways (*©2019 TZEH KEONG FOO ALL RIGHTS RESERVED*, 2019).

Rarely reported *PALB2* mutations are found for hereditary PCa. *PALB2* works as a bridge for BRCA1 and *BRCA2* to form the BRCA complex in Homologous Recombination for DNA repair. 83% of its mutations are previously reported in Pancreatic and Breast Cancer and till now 0.4% mutations identified related to PCa and none of them are clinically verified yet for PCa but for familial pancreatic and breast cancer (DEL DOTTORATO Profssa Laura Stronati, 2023).

Table 4.31: List of pathways related to the gene set using the reactome database

Pathway identifier	Pathway name	Entities p Value	Entities FDR	Submitted entities found	Mapped entities
R-HSA-6796648	TP53 Regulates Transcription of DNA Repair Genes	0.0000611	0.021135	ATM;BRCA1	P38398;ENST00000278616;ENST00000357654;Q13315;ENSG00000149311;ENSG00000012048
R-HSA-9701193	Defective homologous recombination repair (HRR) due to PALB2 loss of function	0.00011	0.021135	ATM;BRCA1;BRCA2	
R-HSA-9704646	Defective HDR through Homologous Recombination Repair (HRR) due to PALB2 loss of BRCA2/RAD51/RAD51C binding function	0.00011	0.021135	ATM;BRCA1;BRCA2	
R-HSA-9704331	Defective HDR through Homologous Recombination Repair (HRR) due to PALB2	0.00011	0.021135	ATM;BRCA1;BRCA2	

	loss of BRCA1 binding function				
R-HSA-9709603	Impaired BRCA2 binding to PALB2	0.000198	0.027667	ATM;BRCA1;BRCA2	P38398;Q13315;P51587
R-HSA-9701192	Defective homologous recombination repair (HRR) due to BRCA1 loss of function	0.000216	0.027667	ATM;BRCA1;BRCA2	P38398;Q13315;P51587
R-HSA-9709570	Impaired BRCA2 binding to RAD51	0.000916	0.100792	ATM;BRCA1;BRCA2	P38398;Q13315;P51587
R-HSA-187042	TRKA activation by NGF	0.001361	0.130698	NTRK1	P04629-1;P04629
R-HSA-9701190	Defective homologous recombination repair (HRR) due to BRCA2 loss of function	0.002704	0.229832	ATM;BRCA1;BRCA2	P38398;Q13315;P51587
R-HSA-5628897	TP53 Regulates Metabolic Genes	0.005676	0.392895	PTEN;TSC2	ENST00000371953;P60484;P49815;ENSG00000171862
R-HSA-8982491	Glycogen metabolism	0.01688	0.50246	AGL;PGM1	
R-HSA-5693579	Homologous DNA Pairing and Strand Exchange	0.01753	0.50246	ATM;BRCA1;BRCA2	P38398;Q13315;P51587
R-HSA-1296025	ATP sensitive Potassium channels	0.01985	0.50246	ABCC9	
R-HSA-187687	Signalling to ERKs	0.02636	0.50246	NTRK1	
R-HSA-187706	Signalling to p38 via RIT and RIN	0.02638	0.50246	NTRK1	
R-HSA-8963889	Assembly of active LPL and LIPC lipase complexes	0.029	0.50246	APOA4	
R-HSA-9036866	Expression and Processing of Neurotrophins	0.03609	0.50246		
R-HSA-167060	NGF processing	0.03609	0.50246		
R-HSA-170984	ARMS-mediated activation	0.0393	0.50246	NTRK1	
R-HSA-4570571	Defective RFT1 causes CDG-1n	0.0393	0.50246		

R-HSA-190827	Transport of connexins along the secretory pathway	0.0393	0.50246	GJB2	
R-HSA-9675135	Diseases of DNA repair	0.04722	0.50246	ATM;BRCA1;BRCA2	P38398;Q13315;P51587
R-HSA-165181	Inhibition of TSC complex formation by PKB	0.04889	0.50246	TSC2	
R-HSA-187015	Activation of TRKA receptors	0.0066	0.3929	NTRK1	P04629-1;P04629
R-HSA-70221	Glycogen breakdown (glycogenolysis)	0.0066	0.3929	AGL;PGM1	
R-HSA-9709275	Impaired BRCA2 translocation to the nucleus	0.00666	0.3929	BRCA2	
R-HSA-1855204	Synthesis of IP3 and IP4 in the cytosol	0.0077	0.42339	PTEN;PLCE1	
R-HSA-5674404	PTEN Loss of Function in Cancer	0.00997	0.44876	PTEN	
R-HSA-5609974	Defective PGM1 causes PGM1-CDG	0.00997	0.44876	PGM1	
R-HSA-5678420	Defective ABCC9 causes CMD10, ATFB12 and Cantu syndrome	0.00997	0.44876	ABCC9	
R-HSA-9675136	Diseases of DNA Double-Strand Break Repair	0.01288	0.50246	ATM;BRCA1;BRCA2	P38398;Q13315;P51587
R-HSA-169893	Prolonged ERK activation events	0.01331	0.50246	NTRK1	
R-HSA-8943723	Regulation of PTEN mRNA translation	0.0138	0.50246	PTEN	ENST00000371953;P60484
R-HSA-5693616	Presynaptic phase of homologous DNA pairing and strand exchange	0.0151	0.50246	ATM;BRCA1;BRCA2	P38398;Q13315;P51587

Note: The table lists the pathways related to the gene set identified using the Reactome database. Each pathway is described by its identifier and name, along with statistical parameters, including the p-value and false discovery rate (FDR), which indicate the significance of the associations. The table also provides information about the entities involved in each pathway, including those submitted and mapped, highlighting the relevance of specific genes in biological processes such as DNA repair, metabolism, and signaling pathways.

4.3.3.3. Panther DB Analysis of pathways indicating molecular and biological processes

Panther DB tool is used to annotate genes and analyse their overall pathways, biological and molecular functions. The overall pathway analysis showed *ATM, PTEN, PLCE1, CACNA1C* are *TSC2* mostly involved in diabetes associated pathways. **Figure 4.28 and Table 4.32** showed the pie chart of overall pathways according to the proportion represented in different colours, **Table 4.33 Figure 4.29** gave the pie distribution of biological processes and **Table 4.34** and **Figure 4.30** molecular processes linked to our gene set.

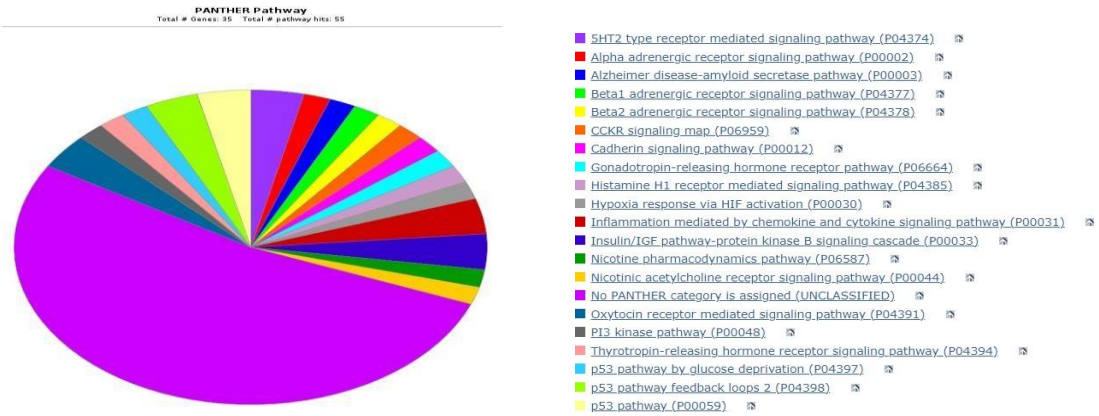


Figure 4.28: Pie chart distribution of the genes and the related pathways by different colours. Major portion represented that the genes are not well explored for the pathway associations (purple portion).

Table 4.32: List of overall pathways and genes playing role in

PATHWAYS	**GENES**
PI3K Kinase pathway,	*PTEN*
Inflammation mediated by chemokine and cytokine signaling pathway	*PTEN AND PLCE1*
Hypoxia response via HIF activation	*PTEN*
Insulin/IGF pathway-protein kinase B signaling cascade	*PTEN, TSC2*
p53 feedback loop	*PTEN and ATM*
p53 feedback loop	*PTEN and ATM*
Thyrotropin releasing hormone	*PLCE1*
Histamine H1 signaling	*PLCE1*
Oxytocin receptor mediated signaling pathway	*PLCE1, CACNA1C*

p53 pathway by glucose deprivation	*TSC2*
Gonadotropin releasing pathway	*CACNA1C*
Nicotine acetylcholine receptor signaling pathway	*CACNA1C*
Cadherin	*CTNND1*
Nicotine pharmacodynamics pathway	*CACNA1C*

Note: The table outlines the key pathways and their associated genes identified in the Panther DB analysis. Pathways include PI3K kinase inflammation signaling and hypoxia response, with significant roles played by genes like PTEN, PLCE1, and CACNA1C.

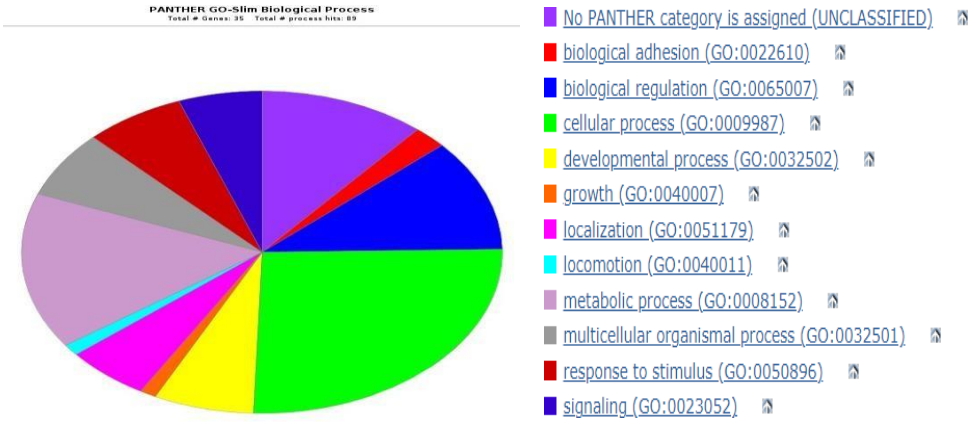


Figure 4.29: Pie chart distribution of the genes and the biological processes represented by different colours. Majorly gene are involved in cellular processes (green portion) and very least are playing role in locomotion (sky)

Table 4.33: List of Biological processes associated with genes set

Biological processes	**GENES**
Biological adhesions	*PKP2, CTNND1*
Biological Regulation	*APOA4, PLCE1, NR2E3, SEMA4A, DOCK7, HNF1A, CACNA1C, NTRK1, PDE2A, ATM*
Cellular process	*ATM, AGL, PTEN, APOA4, TOR1AIP1, PLCE1, UBR1, PKP2, FGD4, NR2E3, SEMA4A, ERCC5, DOCK7, HNF1A, MPZ, COQ8A, CACNA1C, RNPC3, CTNND1, NTRK1, PDE2A, CLN5, SDHB*
Development	*PKP2, NR2E3, SEMA4A, DOCK7, MPZ, NTRK1*
Growth	*SEMA4A*
Localization	*APOA4, PKP2, SEMA4A, CACNA1C, ABCA4*

Locomotion	*SEMA4A*
Metabolic	*APOA4, PGM1, PTEN, AGL, UBR1, NR2E3, ERCC5, HNF1A, COQ8A, RNPC3, NTRK1, PDE2A, SDHB, ATM*
Multicellular organismal process	*APOA4, PKP2, SEMA4A, DOCK7A, MPZ, NTRK1*
Response to stimulus	*PLCE1, SEMA4A, PDE2A, ATM, NTRK1, ERCC5*
Signaling	*PLCE1, SEMA4A, NTRK1, PDE2A, ATM*

Note: This table presents a summary of biological processes associated with the specified gene set, illustrating the diverse roles these genes play in cellular functions. Each process lists the corresponding gene involved, highlighting their contributions to biological adhesions regulations, cellular activity, development, growth, localization, locomotion, metabolic pathways, multicellular organismal processes, responses to stimuli and signaling mechanisms

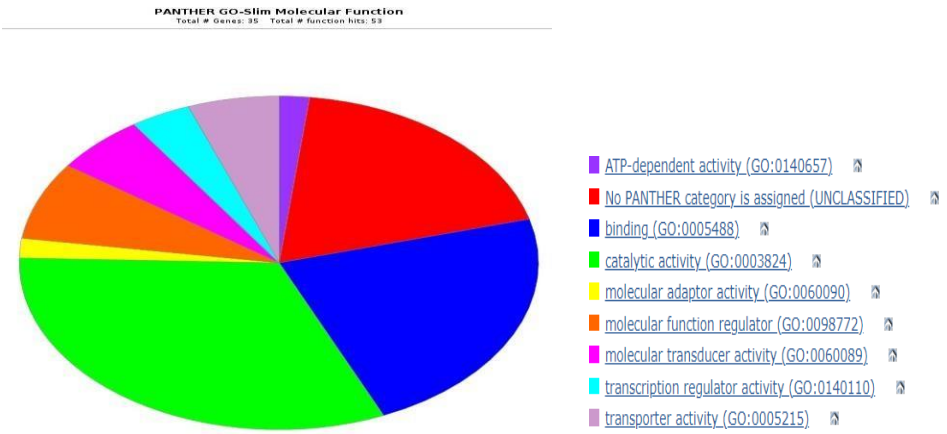


Figure 4.30: Pie chart distribution of the genes and the molecular processes represented by different colours. Majorly gene are involved in catalytic activities (green portion) followed by binding activities in blue portion and very less are playing role in molecular adaptor activities (yellow)

Table 4.34: List of Molecular processes associated with genes set

Molecular functions	**GENES**
ATP dependent	*ABCA4*
Binding	*APOA4, PKP2, NR2E3, SEMA4A, ERCC5, DOCK7, HNF1A, RNPC3, CTNND1, NTRK1, PDE2A, TSC2*
Catalytic	*POLR3B, TMX2, APOA4, PGM1, PTEN, AGL,*

	PLCE1, UBR1, RCC5 DOCK7, COQ8A, NTRK1, PDE2A, ABCA4, CLN5, ATM, TSC2
Molecular Adaptor	*PKP2*
Molecular Function Regulator	*APOA4, SEMA4A, DOCK7, TSC2*
Molecular Transducer	*NR2E3, SEMA4A, NTRK1*
Transcription Regulator	*NR2E3, HNF1A*
Transporter	*CACNA2D4, CACNA1C, ABCA4*

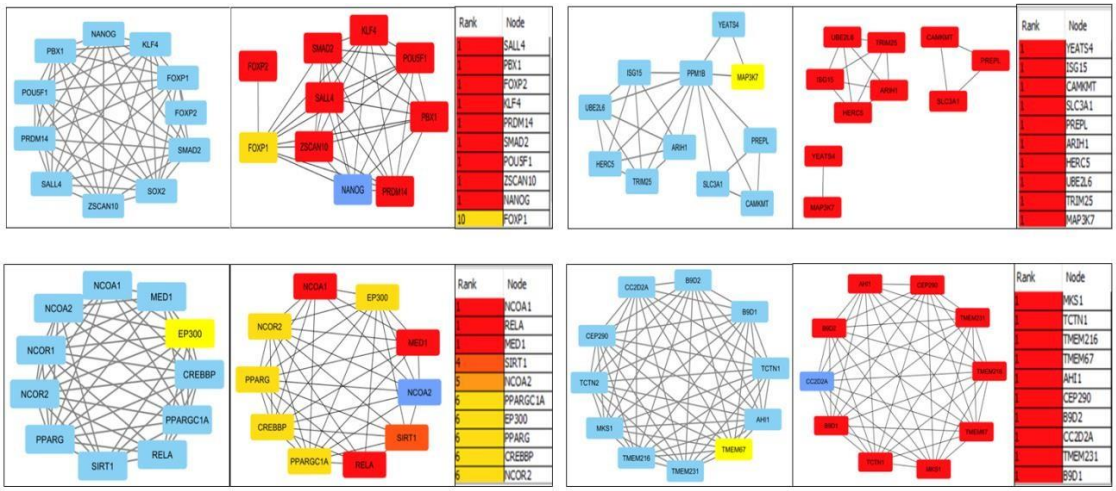
Note: This table outlines the molecular functions associated with the gene set, categorizing them into various processes such as ATP-dependent functions, binding activities, catalytic roles, molecular adaptation, regulation, transduction, transcription regulation, and transport mechanism. Each function lists the gene involved, emphasizing their contribution to essential molecular processes within the biological system.

4.3.3.4. Cytoscape-CytoHubba

4.3.3.4.1. Top ranking genes showed vivid clustering coefficients for ascertaining survival plots

By applying the clustering coefficients algorithm of plug-in cytoHubba, we screened the top 10 hub genes for each gene identified in the Meta-analysis.

The PPI visualization of all the major genes using Cytoscape-cytoHubba revealed a significant number of nodes and edges of the topmost stable and highest scored genes in their respective networks based on the degree clustering coefficient (**Figure 4.31)** (Kour et al., 2023).



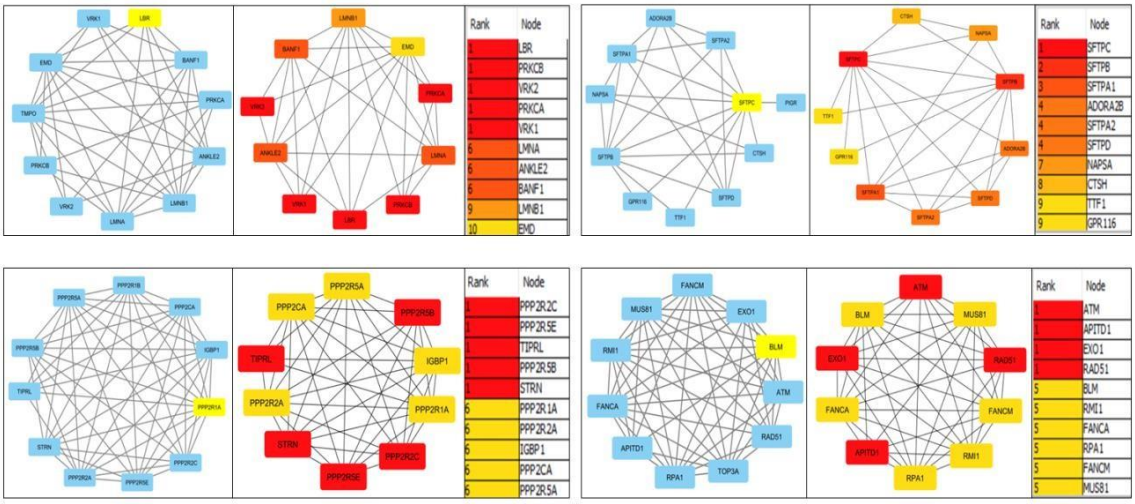


Figure 4.31: Clustering coefficient networks showing top 10 genes for *FOXP1*, *PPM1B*, *PPARG*, *TMEM67*, *TMPO*, *SFTPC*, *PPP2CA* and *BLM* genes identified in Meta-Analysis

Taking together the findings of meta-analysis studies and the genes discovered mutated in our current study we reformed the clustering coefficients cytoHubba plugged in analysis using Cytoscape to co relate the results, identify common pathway related genes and see the crosstalk (**Figure 4.32**).

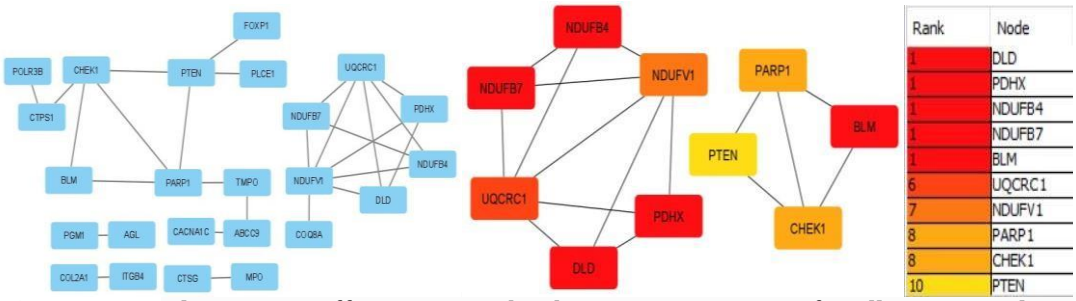


Figure 4.32: Clustering coefficient networks showing top 10 genes for all 42 mutated genes and *FOXP1*, *TMPO* and *BLM* identified from the meta-analysis study. *DLD*, *PDHX*, *NDUFB4*, *NDUFB7*, *BLM*, *UQCRC1*, *NDUFV1*, *PARP1*, *CHEK1* and *PTEN* are the top 10 genes that are functioning together. The network analysis of significant genes using the Cytoscape-cytoHubba plugin ranked the genes by the clustering coefficient to show the hierarchy in high confidence interactions in PPI networks. The network represents the degree of clustering coefficient through a color scale ranging from red to yellow based on which we can validate the experimental protein targets in the network and prioritize potential candidate genes. Clustering PPI networks can be helpful for identifying groups of interacting proteins that take part in the same biological process or cooperate together to carry out a particular biological function. These candidate genes with highest clustering coefficient are noted in red,

high to moderate clustering coefficients are in orange and those with low are shown in yellow. Some of the important6 genes were identified through all. The greater the colour the more they are connected.

PTEN: A Modulator of PI3K/AKT pathway (negative) regulating glucose homeostasis, insulin sensitivity and tumor development (PCa)

The pathway enrichment studies show that *PI3K/Akt* is majorly linked to mutated genes (*PTEN* particular) of PCa and also well-established relations in glucose metabolism and insulin regulation pathways. We observed that the link between the PCa and diabetes can be established through *PTEN*, a modulator of *PI3K/Akt* signaling.

First, decrease in *PTEN* expression was seen associated with the worsening of DN. However, *PTEN* also regulates the ETM and TIF where *PTEN* transcription is further regulated by ***PPARɤ***. PPARɤ shows dual regulation i.e., regulates *PTEN* and through *PTEN* regulates EMT and TIF; which can further regulate *AKT* and *FAK* phosphorylation level and *FAK* transcription. *PTEN* has two distinct phosphatase activities, protein phosphatase (regulates cell cycle, proliferation, migration and apoptosis) and lipid phosphatase (regulates cell growth, survival, migration and proliferation). Deletion in *PTEN* in chromosome 10 is a tumor suppressor gene. It regulates the *PI3K/Akt* (pathway which is presented frequently by our enrichment analysis) signaling however in DN it is altered. Peroxisome proliferator-activated receptors (*PPARs*) are nuclear hormone receptors (α, β, and ɤ). Tumor studies have revealed that the *PPARɤ* attaches to a *PTEN* upstream promoter *PPRE* (peroxisome proliferator responsive element) to regulate its expression. As a transcription factor, *PPARɤ* also exhibits both tissue and disease specificity. In exposure to high glucose (25 mmol/L of glucose and mannitol), Yan et al. in 2019 observed that the *PPARɤ* and *PTEN* expression decreased at both mRNA and protein level. High glucose initiates the cell proliferation, production of extracellular matrix and trans-differentiation. This suggests in diabetic patients, variation in expression of *PPARɤ* level regulates the *PTEN* expression. For suppression of invasion and migration in metastasis *PTEN* inhibits *FAK* expression. *PTEN* also exhibits the dual activity of dephosphorylating *FAK* and *AKT* and *FAK* transcription regulation (Yan et al., 2019) **Figure 4.33**.

In cancers *PTEN* is often mutated and dysregulated in *PIP3* metabolism in diseases like diabetes. Generally, *PTEN* is called an *AKT* pathway antagonist by turning *PIP3* to *PIP2* which reverses the PI3K effect. Through reversing the PI3K effect it can modulate the lipogenesis, glycolysis and other cellular processes. Hence the regulatory balance between

PI3K and *PTEN* is very important for cell survival and its proliferation by regulating *AKT* signaling (via *GSK3, FOXO, p53* and *mTORC1*) (S. L. Smith et al., 2021) (**Figure 4.34)**.

In T2DM, some polymorphism in the 5'-UTR region of *PTEN* was found associated in a Japanese population study. *PTEN* plays a role in both insulin resistance and sensitivity. On stimulation of Insulin receptor by a ligand, the activated β-subunit of IR stimulates the autophosphorylation of insulin receptor substrates (IRS) which acts as scaffolds to recruit the signaling complexes by phosphotyrosine binding and passes signal to *PI3K/AKT* to produce PIP3. Suppression of *PTEN* also increases the *GLUT4* which are insulin induced glucose transporters. Further, its suppression increases the pro and anti-inflammatory conditions in obesity associated with IR. Even hypo methylations in the promoter region of *PTEN* were seen in mild T2DM suggesting its potential biomarker for T2DM diagnosis (Khokhar et al., 2020).

In response to radiation treatment, the early growth regulated transcription factor *EGR1* and *IGF2* up-regulates *PTEN*. AN anti-diabetic drug rosiglitazone acts as a selective ligand and activate *PPARɤ* which then binds the two regions of PTEN promoter (PRE1 and PRE2) to promote the *PTEN* upregulation in both normal and cancerous cells. This upregulation goes along with the fall in *AKT* phosphorylation and cell proliferation suggesting the antagonist activity of PPARɤ could be better target in treating the cancers. Another up-regulator of *PTEN* is p53 (Tamguney & Stokoe, 2007).

A cytokine of IR and inflammation called resistin, also up-regulates the *PTEN* expression. In a study on Japanese cohort germline variation was identified in the 5'UTR region of promoter and linked with DT2M by increasing *PTEN* expression and decreasing the *PI3K* signaling on exposure to insulin stimuli (Tamguney & Stokoe, 2007). Further it has been observed that in the presence of higher levels of PTEN, glucose and translocation of *GLUT4* uptake stops. This suggested the importance of PTEN in maintaining the glucose homeostasis (Butler et al., 2002).

Glucose homeostasis is also maintained by skeletal muscles and adipose tissues. *PTEN* variant is linked to peripheral insulin resistance. It was observed that the glucose tolerance and insulin sensitivity are improved when one allele of *PTEN* (+/-) is altered or *PTEN* is depleted in *in vivo* study of lover tissue. This implies that the *PTEN* is a negative regulator of insulin stimulated glucose absorption and insulin sensitivity. Major sites for glucose metabolism and insulin sensitivity are skeletal tissues. *PTEN* expression in skeletal muscle is reduced in acute diabetes and rises in insulin resistance or chronic diabetes. Deficiency of

PTEN causes fatty liver by hepatocytes however increases the glucose uptake and insulin sensitivity increases in the skeletal muscles. Studies have found that the knocked out PTEN in muscles controls the degradation of proteins in diabetics by inhibiting ubiquitin-proteasome and caspase-3 activation (Shan et al., 2019).

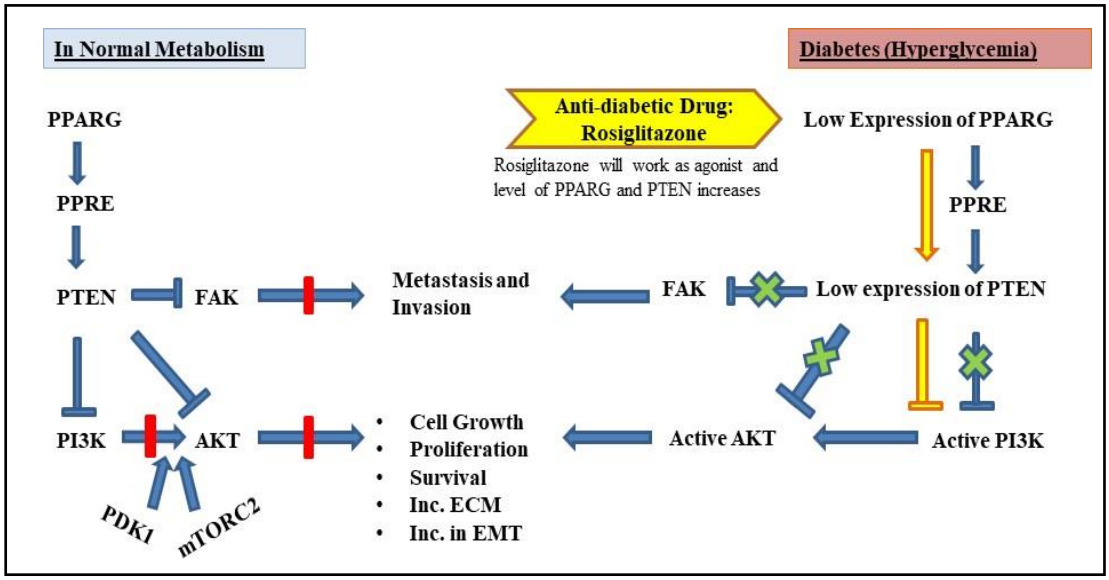


Figure 4.33: Pictorial representation illustrating the difference between the normal metabolism and hyperglycaemic (diabetes) situation in *PTEN* pathway. In normal state the *PPARG* and *PPRE* both in down signaling pathway stimulates the *PTEN* which then blocks *PI3K* from activating Akt and can also directly in inhibit Akt to stop the cellular processes like cell growth, cell proliferation, cell survival, increase in extracellular matrix and EMT. Further *PTEN* can also inhibit *FAK* to avoid metastasis and invasions. However, in the hyperglycemia when the *PPARG* expression is lower and hence inactive *PTEN* fails to block the *FAK*, *Akt* and *PI3K*. When antidiabetic drug like rosiglitazone is used it works as agonist and raise the level of *PPARG* and active *PTEN* to regain the function suggesting molecular interconnections of diabetes and *PTEN*.

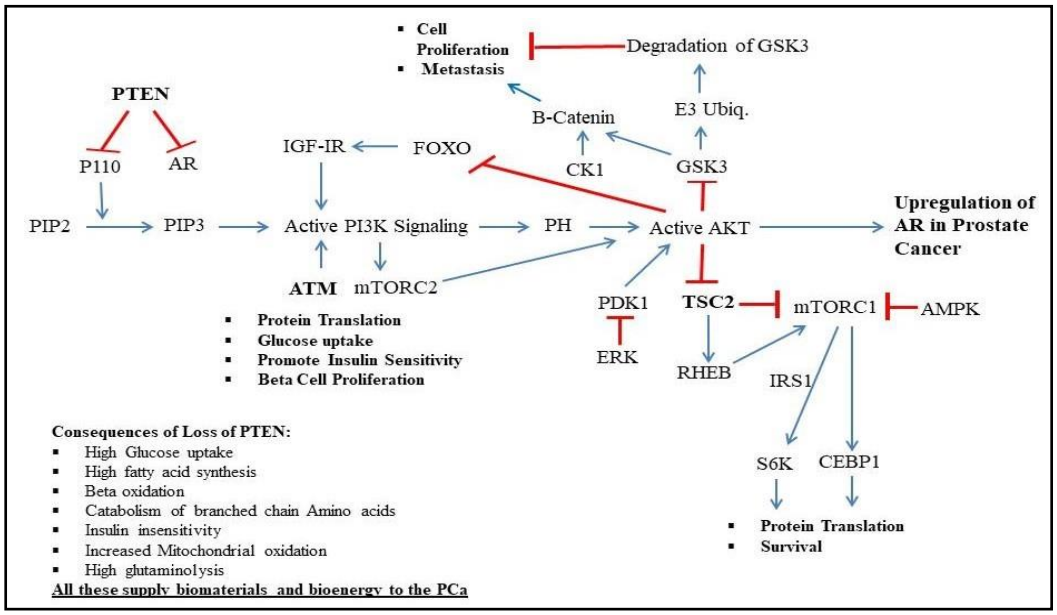


Figure 4.34: Figure representing the signaling pathway involving few mutated genes from our study playing a role in cancer development and diabetes. Blue arrows depicting the activating effect and red arrows are for halt in the signaling.

4.4. Results of urolithiasis:

GeneMANIA analysis revealed 20 other associated genes which are; *HRH1*, *A2M*, *FGA*, *NPL*, *ZBTB16*, *VKORC1L1*, *IL37*, *ORAI3*, *ORAI2*, *PROZ*, *PON2*, *HPRT1*,*PON3*, *IL1A*, *CAT*, *LCTL*, *IL18BP*, *KLB*, *APOA1* and *IL1RL2*. Based on several network studies till date, the identified urolithiasis genes are categorized on their functional grounds such as phosphate or calcium regulation *(VDR*, *KL*, Calcitonin receptor etc.), calcium sensing receptor related to calcium regulation (*CLDN14*, *ORAI1*), stone matrix (*OPN*), stone formation inhibitors (*SLC13A2*, *F2*), uric acid stone related (*CARD8*), for stones having atazanavir (*UGT1A1*) and also the anti-inflammatory and antioxidant stress (*IL-1*, *PON1*) related ones **(Figure 4.35).** Many genome wide association studies have been performed to identify the other candidate genes and their SNPs causing nephrolithiasis (*SLC34A1*, *CLDN14*, *AQP1*, *DGKH* and *ALPL*). We also checked this from the gene database of NCBI by which 22 urolithiasis responsible genes were identified and enlisted and also their pathogenic and likely pathogenic SNPs were identified, *viz-VDR*, *IL6*, *IL1B*, *SPP1*, *PON1*, *SOD2*, *F2*, *IL1RN*, *IL18*, *CASR*, *PLAU*, *KL*,*VKORC1*, *ORAI1*, *CP*, *HSPG2*, *GGCX*, *CALCR*, *SLC26A1*, *APRT*, *ZNF365* and *HOGA1* (Sun et al., 2021).

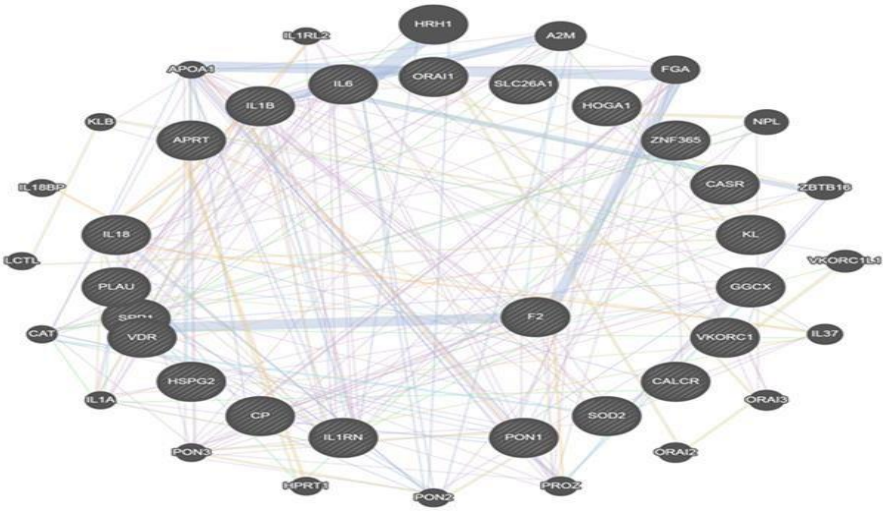


Figure 4.35: A schematic representation of an interaction map of all the 22 urolithiasis and other 20 associated genes using GeneMania, forming a network of urolithiasis causal genes with other genes. The purple edges show that those genes are co-expressed for example *IL-18* and *VDR; CALCR* and *ZNF365*. The blue edges show that they are localized to the same organelle (blue edges connected by nodes, viz. *HRH1* and I*L-6; F2-SPP1*).

Disease-Gene Interactions

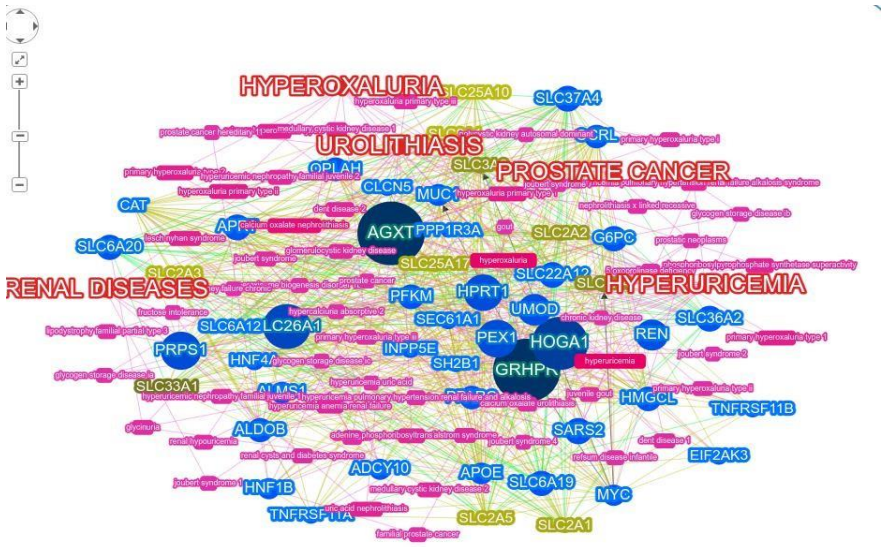


Figure 4.36: Disease gene based network analysis using Phenolyzer. All the edges indicate the propensity of interactions. By turning off gene annotation, interactions of all specific diseases and input terms; by turning off disease annotation, network of all the genes; Protein-protein interaction network of genes from the same gene family; and all genes of same family can be analysed.

While analysing the gene-disease (gene-phenotype) interaction networks (**Figure 4.36)**, 5 keywords (disease names) were given, viz. urolithiasis, Hyperuricemia, hyperoxaluria, PCa and renal diseases against the searches. In the resultant image, we found an approximate 2000 genes, among which ***CREBP*** was found to be interacting most (Sun et al., 2021).

CHAPTER 5: DISCUSSIONS

PCa was ranked fifth greatest reason of mortality in men caused by cancer by 2020 (Sung et al., 2021). It is projected to be diagnosed in one in seven American men and one in every 25 men worldwide (Barsouk et al., 2020). According to India's most recent Population Based Cancer Registries (PBCR) 2020 data, there have been 41,532 cases of PCa overall, with a crude rate of 5.7 among males (Sung et al., 2021). Comorbidities are very poorly considered in PCa due to which the patient faces a dilemma for definite treatment, if he should go for aggressive treatment or not. As a consequence, patients with comorbidity have to go through post treatment traumas. As reviewed, the men who are having single comorbidities like obesity, alcoholism, smoking and mobility problems are receiving aggressive treatments as compared to those who are without comorbidities. The reason for such discrepancy could be because of false assumptions by both doctors and patients of long survival period with single comorbidity as they are not going to impact much on men and hence neglecting proper estimation of mortality while decision making of treatment. However, men who have peripheral vascular problems, diabetes with organ damage and chronic obstructive pulmonary disease (COPD) and having chances of 10 years of life expectancy over the age of 75 years, than aggressive treatment (surgery or radiation therapy) may not be a right option for them (Chamie et al., 2012). Moreover, the obese PCa patient would face adverse during or post treatment effects and would more likely lead to CRPC. Besides this, obesity can also invite other comorbidities like diabetes and cardiovascular diseases (Wilson et al., 2022). Hormonal imbalances associated with obesity can also promote PCa progression instead of becoming a risk aspect directly (Mistry et al., 2007). Other comorbidities include hypertension and urolithiasis. To investigate the association of co-existing medical conditions, particularly, diabetes mellitus with PCa, we performed two small prospective studies one with meta-analysis study of already published datasets and other with WES of clinical samples of 5 case studies.

Meta-analysis:

We first studied already published datasets (ClinVar data of PCa, diabetes and obesity; GWAS Central of both PCa and diabetes; PRACTICAL Consortium datasets; Published WES dataset and RNA-Seq dataset from our lab) to check whether candidate genes could be considered for precision scale modeling, and therefore we brought this to prediction scale and analytics. The analysis revealed several common genetic factors shared among PCa, diabetes mellitus and obesity. Specifically, *BLM*, *TMEM67*, *RFX6* and *NUDC* found common genes

among these conditions when comparing their data from ClinVar. A single variant, rs61816761, associated with the *FLAG* gene, was found common to both obesity and diabetes with GWAS Central data of PCa. Our RNA-seq data showed *PP1MB* and *SFTPC* as common in PCa and diabetes. By using GeneMANIA to commonalities we obtained a network with common interacting partners between *TMPO* and *FOXP1*. While *FOXP1* was found to be a common interaction partner of *PP1MB, SFTPC* and *TMEM67,* an intriguing finding was that the BLM gene was the only common gene between PCa, diabetes mellitus and obesity, interacting with both *FOXP1* and *TMPO*. The strength of our study lies in the comprehensive analysis of genetic factors in PCa, diabetes mellitus and obesity, as well as its effective integration of diverse datasets, identification of common genetic variants, and the application of advanced analytical techniques. The current study employed a range of statistical and ML algorithms to analyse the data that identified common genetic factors and assessed predictive models. The lower RMSE values for specific algorithms in each dataset demonstrated the feasibility of using these methods for future research and clinical application.

NGS-WES Study:

Our study using WES has revealed 44 genes harbouring putative somatic and significant mutations. These genes includes *MYO15A, BRCA1, BRCA2, HNF1A, ERCC5, ATM, SEMA4A, TSC2, ABAC4, SDHB, NR2E3, GJB2, SMPD1, MPZ, PKP2, CACNA2D4, TOR1AIP1, FGD4, POLR3B, TMX2, ABCC9, PDHX, UBR1, CLN5, PGM1, COQ8A, AGL, CACNA1C, PTEN, PDE2A, DOCK7, SZT2, NDUFV1, TSHD1, PLCE1, CTPS1, ITGB4, CHEK1, COL2A1, PARP1, PTNP22, SLC16A13, MYRF* and *MPO.*

Association of genes with PCa

MYO15A has been identified in our previous study (Gupta et al., 2020) and with the same mutation '*rs375290498'*. *MYO15A* is known for second most causing non-syndromic hearing loss (HL) (Vanniya. S et al., 2022)**.** Familial inheritance of mutated *MYO15A* is associated with high severity (Chen et al., 2018). In a study of HL in 61 consanguineous Egyptian families' linkage analysis wherein the majority of variants were identified in *MYO15A* and among which one family had mutated the *WFS1* gene causing Wolfram syndrome (WS) and the variant was inherited in an autosomal recessive manner. Wolfram syndrome is characterized by childhood-onset diabetes mellitus, diabetes insipidus optic atrophy, and sensorineural HL SNHL. However, the affected siblings of the same family didn't express any other complication except WS (Budde et al., 2020). Another study was conducted on four

South Indian families which were mating assortatively, in which one individual was having a history of 10 years of diabetes (Vanniya. S et al., 2022).

The movement and migration of cells depend on the broad and varied family of molecular motors known as myosins (Makowska et al., 2015). A study of W. Zhang et al., in the year 2022 found *MYO15A* as a novel contributor in the evolution of a hormone sensitive PCa (HSPC) to CRPC (Zhang et al., 2022). Mutation in *MYO15A* is one of the frequent somatic mutations of PCa and commonly related to advanced stage PCa (Zarzour et al., 2020) (Mamidi et al., 2019).

BRCA1 and *BRCA2* were also mutated in previous cohort studies. While the link between *BRCA1* and *BRCA2* and PCa remains enigmatic, the genetic alterations identified so far that the greatest PCa risk particularly made over the age of 65, are mutations in the *BRCA2* gene, which is associated with an 8.6 fold increased risk (Castro & Eeles, 2012). *BRCA1* holds the moderate risk of 1.35-fol for PCa (Oh et al., 2019). Aside from the DNA damage response, repair, chromatin modelling, and transcriptional control, *BRCA1* has recently been found as an AR co-regulator. It may affect another key route in PCa by regulating IGF-1R in an AR-dependent manner. *BRCA2* may potentially operate as a tumor suppressor in epithelial prostate tissue, leading to premalignant prostatic lesions (Castro & Eeles, 2012). 10% of PCa are harbouring the *ATM* mutations in *PARP* inhibitor sensitive cells and bring more genetic instability of PCa cells (Neeb et al., 2021). *TSC2* stimulates cell proliferation in PCa (Munkley et al., 2014).

Particularly in the context of tumour aggressiveness, metastasis (bone), and therapeutic response (resistance to chemotherapies), *ITGB4* appears to be important in PCa. It is more abundantly expressed in advanced PCa and correlated with increased motility and invasiveness (Wilkinson et al., 2020). *ITGB4* also interacts with a glycoprotein, Prostate-Specific membrane Antigen (PSMA), and regulates the angiogenesis (Gao et al., 2021). A lipid *APOA4* transporter found which transports cholesterol and triglyceride gets upregulated in mCRPC. Being a lipid transporter it suggested it can connect PCa with CVDs (Kakkat et al., 2023). *SEMA4A* stimulates migration in endothelial cells and EMT to promote metastasis (Nkyimbeng-Takwi & Chapoval, 2011) (Liu et al., 2023). *GJB2* encodes junction proteins. It is a frequently mutated gene in PCa playing a role in transport, cell-cell signaling in castration resistant PCa and metastasis (Tang et al., 2022) (Marín-Aguilera et al., 2012). *PTEN* is a tumor suppressor gene; loss of function mutation causes alteration in PI3K signaling and is considered as the most frequent driving mutation found in PCa. Its mutation is mostly observed in metastasis PCa (Wise et al., 2017) (Jamaspishvili et al., 2018). *RNPC3* is a minor

component of spliceosomes and its expressions counterparts the tumor progression. Its expression is lower in hormone sensitive cells of PCa and benign tumors, intermediately expressed in CRPC. However, it is hyper-expressed in metastatic neuroendocrine PCa, its expression (Augspach et al., 2021). *HNF1A* regulates the glucuronosyltransferase (UGT) which can modulate androgen activity suggesting AR and *HNF1A* regulated CRPC development (Yun et al., 2017). *FGD4* is upregulated in PCa cells and expressed as the advancement of the PCa. *FGD4* basically helps in PCa cell migration and expression of mesenchymal cell markers (Bossan et al., 2018)**.** *PKP2*, a desmosomal protein, is present in both basal and luminal cells of prostate gland stratified epithelium (Breuninger et al., 2010). An indirect relation has been suggested but no clear association has been made till now. *PLCE1* mutations are associated with high grade PCa (Edwards et al., 2013). Found commonly mutated in stomach cancer (Sun et al., 2015). *DOCK7* gene fusion with OLR1 gene was found in promoting PCa metastasis and its recurrence (Y. P. Yu et al., 2023).

Mutation in *CACNA2D4* is rarely identified because of epigenetic mutation and found related to PCa specific mortality (PCSM). Its promoter gets methylated and downregulated and suggested to be a potential biomarker (Pidsley et al., 2022). *ERCC5* is linked to increasing PCa susceptibility and till now it has been found mutated in Asian populations only (Y. Liu et al., 2018). *NTRK1* helps in autocrine signaling of prostate cells, is downregulated and causes poor prognosis of PCa (Pierotti & Greco, 2006) (Bagherabadi et al., 2022). *SZT2* is one of the protein complexes which play a role in mTORC1 signaling (Yin et al., 2021).

Apart from this, few genes which are rarely mutated in PCa are (*UBR1*, *PDE2A*, *MPZ*, *COQ8A*, and *AGL*). In PCa cell lines like PC3, it has been observed to destabilize and inhibit a pro-apoptotic truncated bone marrow kinase (BMXΔN) of bone marrow kinase (BMX) by ubiquitination. BMX in a study was found to get cleaved into its truncated form which makes the PCa cells more sensitive towards apoptosis as a response to several apoptosis stimuli (Eldeeb & Fahlman, 2016). *MPZ* protein is produced by Schwan cells and is essential for retaining the firmness and integrity of myelin sheaths in peripheral nerves (Haddad et al., 2022). *COQ8A* is an unusual kinase-like protein called COQ8A facilitates the production of coenzyme Q, a crucial antioxidant and cofactor for cells (N. H. Murray et al., 2022). *AGL* involved in amylo alpha-1, 6-glucosidase activity, 4-alpha-glucanotransferase activity (Das & Rao, 2007). Further, few genes were never reported in association with PCa and even have no role related to prostate but found mutated and changes in their expression level. These genes includes; *ABCC9* (upregulated) (Demidenko et al., 2015). *SMPD1* (downregulated), *CLN5*, a glycoprotein and downregulated (Shah et al., 2015) and *CACNA1C* expresses lower and could

be a potential biomarker for PCa (Phan et al., 2017). However, *ABCA4, NR2E3, PDHX, TOR1AIP, POLR3B, TMX2, CTNND1, NDUFV1, PGM1* are those few these which are found mutated in our study but has no study clearly providing evidence of their role in PCa development or progression.

Association of genes with Diabetes

Our study identified some mutated genes regulating glucose level. *PGM1* has an important function in regulating glucose metabolism and cancer development in different types of cancers through cell proliferation and metabolism (Cao et al., 2021). The maternal-effect is more significant on T2DM susceptibility in offspring. This effect has been observed in association with polymorphism in phoshoglucomutase when transmitted from a Type 2 diabetic mother (Akbarzadeh et al., 2022). *APOA4* also has the potential to lower the glucose level and has an anti-inflammatory property which suggests its ability to prevent type-2-diabetes but its level decreases in prediabetes patients (von Toerne et al., 2016). *PDE2A* is known to involve degradation of cAMP and cGMP, expressed in different tissues (brain, liver, heart, lung etc.) Different isoforms are expressed within different sites of cells and consequently are regulating differentially cAMPs local level at specific locations. One of its three variants (PDE2A1, PDE2A2 and PDE2A3), PDE2A2, localized in mitochondria controls the production of ATP by regulating cAMP generation from the plasma membrane. Impaired cAMP signaling is linked with several clinical conditions (Monterisi et al., 2017).

Besides glucose regulation, some are involved in insulin regulation. By phosphorylating p53, *ATM* regulates the insulin resistance and deficiency of *ATM* can develop DM and has a role in metformin response (glycemic) (Yee et al., 2012). *HNF1A* is a most commonly mutated gene among the key genes related to maturity onset diabetes in the young (MODY) (Valkovicova et al., 2019). *TSC2* regulates *mTORC1* by inhibiting it which decreases the insulin resistance. Further for the action of metformin both *TSC2* and *RAPTOR* is required to inhibit mTORC1 suggesting *TSC2* importance in metformin pathway and diabetes management (van Nostrand et al., 2020). *TMX2* also regulates Mitochondrial-associated endoplasmic reticulum, MAMs (site for influencing energy production). In the liver, these MAMs increase in obesity, insulin resistance and due to Ca2+ load in mitochondria lead to mitochondrial dysfunction (N. Sharma et al., 2020). Type 1 diabetes-associated myelin abnormalities can lower the signal conduction velocity in nerves of the peripheral nervous system (Cermenati et al., 2012). In a non-obese T2DM mice (MKR) it has been found that metformin can restore the *MPZ* level by regulating the ROS produced in hyperglycaemic conditions suggesting a correlation of

Metformin, ROS, and alteration in *MPZ* protein (Haddad et al., 2022). *NTRK1* is identified in serving as a potential biomarker of diabetes (type 3 diabetes) caused by Alzheimer disease (Pandiyan et al., 2021). *SHDB* is downregulated in prediabetes and β cells of T2DM which reduces the activity of SDH enzyme and leads to development of type 2-diabetes (S. Lee et al., 2022). *CLN5* is also downregulated in diabetic foot and down regulated in DM (W. Zhao et al., 2020).

Interestingly, we have also noticed some mutated genes from our cohort associates diabetes associated with PCa through the cardiovascular abnormalities indicating, cardiac complications as an important comorbidity to consider. *ABBC9* is found mutated in those myocardial infarction patients who were having history of diabetes mellitus, hypercholesterolemia, and hypertension (Minoretti et al., 2006). *SMPD1* shows upregulation and is associated with diabetic related and alcoholic cardiomyopathy (R. Liu et al., 2023). *PDHX* plays a role in glucose metabolism. MiRNA-26a inhibits *PDHX* to stop citric acid cycle pyruvate to convert into acetyl coenzyme A (B. Chen et al., 2014). *FGD4* is found to be mutated in patients who were having history of diabetes and having Amyotrophic lateral sclerosis (Wei et al., 2019). *CACNA1C* is a calcium voltage-gated channel subunit and its polymorphisms linked to hypertension and coronary artery disease (Beitelshees et al., 2009). It is also associated with diabetic cataract by calcium channels. DM patients who have dyslipidemia as comorbidity with risk of developing CVDs have *DOCK7* mutations in the Chinese population (Kong et al., 2015). *SZT2* gene is downregulated in the plasma of CVD and also linked to epilepsy and development of the human brain. Loss of *SZT2* causes the increase in mTORC1, important for diabetes and cardiac systems (Lygirou et al., 2018). *COQ8A* is an important lipid soluble electron transporter, however in diabetic cardiomyopathies it has been observed that its level decreased therefore leading to impaired production of mitochondrial ATP (Gomes et al., 2022). The gene *UBR1* stabilizes a protein called adipose triglyceride lipase (ATGL) which works as a rate-limiting lipase, crucial for the breakdown of lipids. Changes in the ATGL bring change in the cellular lipid level of obese patients (Bingham, 2023).

There were some genes which showed the co-existence of diabetic comorbidities like Neuropathies and Renal diseases a PCa. Tubular epithelial mesenchymal transition (ETM) and tubuloiterstitial fibrosis (TIF) in the renal system are the two important pathological characteristics of diabetic neuropathy (DN). Decrease in *PTEN* expression was seen associated with the worsening of DN. However, PTEN also regulates the ETM and TIF where *PTEN* transcription is further regulated by PPARɤ. Thence, it is clear PPARɤ shows dual

regulation i.e., *PTEN* and through *PTEN* regulates EMT and TIF; which can further regulate *AKT* and *FAK* phosphorylation level and *FAK* transcription (Yan et al., 2019). *MPZ* protein is produced by Schwan cells and is essential for retaining the firmness and integrity of myelin sheaths in peripheral nerves (Haddad et al., 2022). In diabetic patients phospholipids, cholesterol, and fatty acid content of myelin gets altered, affecting the fluidity of the membrane and leading to a condition called diabetic neuropathy. Myelin protein zero coding gene *MPZ* is suppressed in diabetic patients (Cermenati et al., 2012). Therefore, it has been marked as the biomarker of diabetic peripheral neuropathy (Haddad et al., 2022). *PLCE1* is a podocyte cell marker of glomeruli whose expression decreases in diabetic neuropathy and diabetic kidney disease (Eadon et al., 2022). *SEMA4A* regulates inflammatory colitis. Different semaphorins are involved in different diabetes related complications like diabetic retinopathy, neuropathy, and osteoporosis and wound healing (Lu & Zhu, 2020) and in the type-2-diabetes patients *APOA4* can be novel biomarker in plasma for identification of renal problems (Peters et al., 2017).

We found two genes, *MYO15A* and *GJB2* linked with hearing loss. *MYO15A* association with diabetes is not clear but it is found mutated in those patients who have a history of diabetes with NON-syndromic hearing loss and Wolfram syndrome (Vanniya. S et al., 2022) (Budde et al., 2020). *GJB2* is also mutated in those who have hearing loss and maternally inherited diabetes (Frei et al., 2005).

CLN5, *PKP2* and *NR2E3* are interrelated with diabetes and generally cause diabetic retinopathy and are mutated in our cohort. Diabetes can negatively affect the clock system of the retina and the output of the circadian cycle. Diabetes is found to alter clock but not at the gene expression level of NR2E3 (Vancura et al., 2021). Other includes *FGD4* linked to esophageal squamous cell carcinoma (Y. Gao et al., 2021), *CACNA2D* which plays role in diabetes induced atherosclerosis (Wang et al., 2019), *TOR1AIP* causing distal myopathy (Finsterer & Stöllberger, 2016), *SEMA4A* which regulates of inflammatory colitis (Lu 2020), *RNPC3* encoding spliceosomes and found linked to growth hormone deficiency in children who had diabetes insipidus (Murray and Clayton 2015), *CTNND1* relates to diabetes by influencing the migration of neutrophil cells in diabetic skin wounds (Kang et al., 2021) and *AGL* through hyaluronic acid synthase 2 (HAS2) and CD44-HA regulates inflammatory responses (Sottnik & Theodorescu, 2016). However few genes *ITGB4*, *ERCC5*, *ABCA4*, *POLR3B*, *NDUFV1*, *THSD1*, *BRCA1* and *BRCA2* are not yet reported in clear association with diabetes and in future they must be explored to understand the mechanisms underlying which associate these genes with diabetes in PCa.

Disease conditions associated with altered alleles

In genetic profiling of our cases, we found heterozygous mutations and among them we validated some of them using Sanger sequencing as per the second objective requirement. A missense allele change c.5925G>A of rs375290498 in *MYO15A* is reported earlier and found in our cohort for non-syndromic hearing loss where c.5925G>A mutation truncates the MYO15A protein and causes functional impairments of myosin-XV (Fattahi et al., 2012). Another variant associated with hearing loss is c.457C>T of rs111033186 in *GJB2* (Fattahi et al., 2012). SNV c.4559-4C>G allele change of rs199620842 in *ITGB4* is not reported with any condition however it is supposed to be associated with junctional epidermolysis bullosa with pyloric atresia (https://www.ncbi.nlm.nih.gov/medgen/C5676875/ last accessed on November 5, 2023). A substitution allele change in 3' splice region c.2031-2A>C (rs35897051) of intron 11 in *MPO* is associated with inherited myeloperoxidase deficiency (MPOD) which is a very common defect in neutrophils (Marchetti et al., 2004). The nonsense variant c.7456C>T allele change of rs587779865 in *ATM* gene is reported in hereditary predisposed breast cancer, ovarian cancer and pancreatic cancer (Ohmoto et al., 2018). However, a variant position C>T of rs758623165 in UBR1 as per the ClinVar is reported in single Nigerian woman having inborn genetic diseases like underdeveloped nasal alae, exocrine pancreatic insufficiency, hearing impairment, and imperforate anus (https://www.ncbi.nlm.nih.gov/clinvar/variation/985540/?oq=rs758623165&m=NM_174916.3(UBR1):c.3848%201G%3ET last accessed on November 5, 2023). The variant region c.1720A>G of rs1169305 in *HNF1A* predisposes to MODY and cardiac problems in diabetic patients is validated in our cohort as homozygous with genotype GG (Koko et al., 2018). The ClinVar reported it with other clinical implications also *viz*. type-2-diabetes, hepatic adenocarcinomas, non-papillary renal cell carcinoma, diabetes mellitus type 1, reduced delayed hypersensitivity, and breast cancer.

Some of them were not provided with any record conditions suggesting it be as novel mutations identified in PCa from our cohort *viz.* c.3207C>T allele change of rs370887875 in *MYRF*, c.79G>A allele change of rs74315369 in *SDHB*, c.1489G>A allele of rs151212477 in *PKP2*, and c.392_39dup (ACT> ACTCT) of rs886041816 in *PTEN*. Since then, LncRNAs have been implicated as diagnostic and prognostic biomarkers in different cancers. Some well-known lncRNAs are identified and are being used as biomarkers in other cancers. In fact, some are also suggested in PCa but are not in use. Also, AR signaling pathway involves various mediators regulated by lncRNA by various mechanisms. Important role of lncRNAs has been identified in development of PCa, promotion of castration-resistant PCa (CRPC),

cell proliferation, invasion, and metastatic spread along with modulation of AR-mediated signaling (Gu et al., 2019). PCa antigen 3 (PCA3) modulates PCa cell survival via modulating AR signaling and is now used in PCa diagnosis (Ferreira et al., 2012). SChLAP1 (second chromosome locus associated with prostate-1) was identified as a highly prognostic lncRNA that differentially expressed in aggressive and indolent form of PCa (Malik & Feng, 2016). Considering the dynamic role of lncRNAs as novel prognostic, diagnostic and predictive markers in PCa, lncRNAs may also serve as therapeutic targets aiding in prevention, development and treatment of CRPC and metastasis of the disease (Ramnarine et al., 2019). Functional analysis of lncRNAs could be done by deciphering lncRNA-protein interaction as the function of most lncRNAs is dependent on interaction with protein-coding genes. Hence, there felt a need to discover a potential lncRNA biomarker for PCa diagnosis. Similarly in our parallel study which was going in our lab we identified 11 lncRNAs (SCARNA10, LINC01973, LINC00940, NPBWR1, FLJ16779, ANKRD20A9P, LINC00298, SNHG19, LOC341056, TLX1NB, LINC00662:60*)* using RNA sequencing (Shukla et al., 2023). Taking these lncRNAs we performed the molecular docking studies with already identified PCa causing genes using WES (*ADA, ANG, BRCA1 , CTNS, HBB, GNPTAB, COL6A1, OTOF , TP53, CYP11B2, CYP1B1 , GJB6, RHAG, DNAAF1, BRCA2, NF1 MCM8, MCCC1, CAPN3, MYO15A, MRE11, KRIT1, HEXB, SCN9A, PRLR, OPA1, ATP6V0A2 and USH2A*) (Gupta et al., 2020). Other than this, we also screened 7 ARs (2Q71, 5V8Q, 4QL8, 2PNU, 5CJ6, 2AM9, 1E3G, 7KW7) related to the PCa. Androgen signaling plays a vital role in PCa development and in treatment strategies (Jacob et al., 2021) even as AR splice variants, amplification/overexpression of androgens, AR-Ligand binding etc. stimulates development of CRPC (Efstathiou et al., 2020). Both PCa related proteins and ARs were taken as receptors and lncRNA as ligands. We used HDOCK server for this study and then visualized the binding interfacial sites and the binding energies were noted. Recent studies have revealed that lncRNAs can regulate androgen signaling through various mechanisms (S. Kumar et al., 2021). It has been reported that lncRNAs can transactivate AR by binding to its enhancer region. LncRNAs are significant in prostate tumorigenesis as gene transcription regulatory sequences. PCGEM1, HOX transcript antisense RNA (HOTAIR), PCa gene 3 (PCA3) are some examples of lncRNAs that function as oncogenic and/or tumor suppressor in PCa through AR signaling pathway (Zhang et al., 2016). In present study, docking analysis of prostate specific proteins and AR with lncRNAs was performed to explore binding potential towards identifying interacting residues and putative RNA-binding motifs.

From the molecular docking and visualization studies, ***SCARNA10*** (lncRNA) was identified as common in both the PCa and AR targeted proteins and our attempt to check the role of SCARNA10 (lncRNA) with PCa and AR targeted proteins were successful. SCARNA10 is a small cajal body specific RNA10, an lncRNA, observed majorly in liver fibrosis reported patients in tissues and serum. While it is known to inhibit targeted gene binding of a promoter to polycomb repressive complex 2 (PRC2) suppressing the TGF signaling, we aimed to understand the mechanism of SCARNA10, when silenced, reducing the levels of *TGF, TGF R1, KLF6* and *Smad2* and *Smad3* (Khilwani et al., n.d.), (Unpublished data). It was also reported that SCARNA10 has a major role in chromosomal mutation at 12p13.31 identified in PCa metastasis studies. In a few studies, SCARNA10 was reported to be up-regulated in breast and lung cancer and we have reported the role of *SCARNA10* in PCa considering LNCaP cell line (Khilwani et al., n.d.), (Unpublished data). Also, nine lncRNAs, *LINC01973, LINC00940, NPBWR1, FLJ16779, ANKRD20A9P, LINC00298, SNHG19, TLX1NB,* including *SCARNA10,* were found to be shared by publically available datasets in the T2DB database in a correlation analysis after comparing the 11 lncRNAs that were identified in the RNA-seq study. This suggests these lncRNA may be commonly differentially expressed in both the disease conditions and would provide a basis for further investigations to elucidate their functional significance, regulatory interactions in PCa progressions and identifying therapeutic targets.

Pathway enrichment study performed on the identified gene set using different online tools like KEGG pathway, Cytoscape, database Panther and reactome to discover gene functions and disease causing pathways. According to the result of the KEGG mapper, we identified that a majority of pathways are related to glucose metabolism and insulin resistance. The overall pathway of PCa available on KEGG showed 5 major pathways involved *viz*. mutation-inactivated *PTEN* to *PI3K* signaling pathway, loss of *NKX3-1* to *PI3K* signaling pathway, Amplified AR to androgen receptor signaling pathway, mutation-activated AR to androgen receptor signaling pathway and loss of *CDKN1B* to p27-cell cycle G1/S. Reactome study revealed top 6 pathways related to *ATM, BRCA1, BRCA2* and *PALB2* related homologous recombination DNA repair pathways with FDR value and *p* Value less than 0.05. *PALB2* is a very rarely mutated gene in PCa and *PALB2* works as a bridge for *BRCA1* and *BRCA2* to form BRCA complex in Homologous Recombination for DNA repair (*©2019 TZEH KEONG FOO ALL RIGHTS RESERVED*, 2019) (DEL DOTTORATO Profssa Laura Stronati, 2023). We identified its pathogenic variants in our cohort giving confirmations for the rare mutations with respect to PCa. Where with the help of Panther Db we could annotate the genes and the

type of pathways they are involved in, either biological or molecular. The results showed that overall the *PTEN, PLCE1* are *TSC2* mostly involved in diabetes associated pathways. Using the clustering coefficients plug-in cytoHubba, we screened the top 10 hub genes for our gene set. Cytoscape-cytoHubba plugin ranked the genes by the clustering coefficient to show the hierarchy in high confidence interactions in PPI networks. We identified *DLD, PDHX, NDUFB4, NDUFB7, BLM, UQCRC1, NDUFV1, PARP1, CHEK1* and *PTEN* as the top 10 candidate genes. While *BLM* was found to be among the top 10 candidate genes in our PPI network analyses, it was also shown to play a role in our meta-analyses. This could show a potential role of BLM in serving as inhibitor suppressing growth and metastasis of PCa (Ma et al., 2022). All of them except NDUFV1 and PDHX have no clear links identified, rest all 9 genes are PCa related and diabetes.

Association with Urolithiasis: a complication of metastasis in PCa

In a separate study to understand the PCa association with other comorbidities like urolithiasis, since it is well documented that bone is the most preferred site for metastasis in PCa. The fact that PCa typically exclusively spreads to skeletal areas, changes the structure of the bones, and raises extracellular levels of calcium, suggests that the bones provide PCa cells a suitable place to grow and localise (Sun et al., 2021). PCa is mostly localized in bone turnover regions, which are defined as areas with elevated rates of both bone creation and bone loss. The extracellular calcium content increases from 2.5 up to 7.5mmol/l. As a result of this, on-going turnover process, which releases a large amount of inorganic bioactive compounds. Its function as a potential mediator of PCa skeletal metastases is made obvious here. To check this, we considered a NCBI publicly available dataset and from the bioinformatics analyses, we found that 22 genes were linked to urolithiasis. Further, we utilised gnomAD to characterise them and determine if the mutations were synonymous or non-synonymous. After comparing the results with genes relevant to PCa, we carried out protein-protein interaction (PPI) investigations to evaluate the interactions and create PPI maps. Numerous genes were found to be shared in the research, and their subtle associations with immunomodulatory response are well established. Instead of urolithiasis, we discovered a correlation between vitamins and calcium based on our annotations. According to Dong (2006), *ORAI* 1 is an endogenous store-operated Ca2+ entry (SOCE) of human PCa cells that has demonstrated pro-apoptotic property rescue in PCa cells. PCa susceptibility is increased by sunlight tanning, demonstrating *VDR* participation and an inverse relationship between exposure and PCa risk, whereby an increase in exposure leads to a reduction in PCa risk. It is evident that vitamin D has protective effects if levels are not limited (Sun et al., 2021).

Angiogenesis and cell migration are two mechanisms linked to *VKORC1* that can promote invasion and metastasis, and the C subunit is implicated in raising PCa vulnerability. A transmembrane receptor called CASR is involved in homeostasis, PCa development, and PCa patients' increased risk of metastasis. This elevated calcium induces PI3K/Akt/mTOR signalling, which in turn drives the *CASR* and metastasis. Variants in the CASR gene are often associated and overexpressed in PCa cells (Sun et al., 2021). Differential roles of *IL6* include controlling VEGF expression, immunological responses, and cell proliferation and differentiation. It can even promote cell survival or prevent cell division. It is well established that *IL6* and its receptors are overexpressed in PCa, whether benign or malignant, and that *IL6* is a key positive growth factor for a large number of prostate cells (Sun et al., 2021). In prostate cell lines, a cytokine called *IL18* has been shown to evade immune surveillance. Prostate gland chronic inflammation is a common observation. By causing inflammation linked to DNA damage, angiogenesis, proliferation, invasion, and metastasis, inflammatory genes are known to increase the risk of PCa (Sun et al., 2021). *IL1RN* is an anti-inflammatory cytokine that suppresses the proinflammatory cytokines *IL1α* and *IL1β*. It provides evidence in favour of the theory that inflammation contributes significantly to the development of PCa and causes advanced PCa (both homozygous malignancies and non-cancerous conditions). Klotho transmembrane proteins in renal tubes are encoded by the anti-aging gene *KL*. Through many pathways, such as Wnt signalling, tumour growth factor (TGF β1), and insulin/insulin growth factor (IGF), it regulates the survival of malignant cells. It is muted or downregulated in PCa. According to (Sun et al., 2021), the expression of *PLAU* declines in PCa cells based on the relative expression of certain genes in PCa. The presumed role of serum lipids (HDL-cholesterol) and HDL cholesterol linked antioxidant enzymes (arylesterase and paraoxonase 1) in PCa promotion is also being investigated, with *PON1* being identified as an increasing factor. The findings unmistakably showed that *PON1* levels may raise the risk of PCa. Human paraoxonase 1 is associated with HDL and is produced in the liver. The number of neuroendocrine (NE) cells rises with the advancement of PCa and assumes a regulatory role in several processes; in a similar vein, the expression of mitochondrial *SOD2* increases with advanced PCa. For processes of differentiation and trans-differentiation, NE is crucial. Drug resistance and metastasis castration resistant PCa (mCPRC) have been shown to have much higher levels of *SPP1* expression, both mRNA and protein. For the remaining common genes *F2, IL1B, CALCR, SLC26A1, APRT, ZNF365* and *HOGA1* have not been reported yet with particular PCa links (Sun et al., 2021). The genotype phenotype study using Phenolyzer revealed *CREBP* is the most common interacting partner.

Polygenicity effect:

This work also attempted to check the role of polygenicity in cancers with respect to diabetes. Polygenicity refers to the combined effect of many genes, each contributing to a small proportion of the trait development (Wendt et al., 2020). Since the PCa is well known as a genetically heterogeneous disease, however, our study explains that the PCa is influenced by a combination of genetic factors and environmental factors interplaying at the background through comorbidities obesity, type-2-diabetes, cardiovascular problems and urolithiasis. Multiple genes are involved in various aspects of pathways underlying for instance in glucose metabolism, insulin sensitivity or energy storage and the presence of multiple risk alleles across the different genes must be contributing to the polygenic nature of the disease. Such variants are increasing the susceptibility of individuals even if they are contributing to the small effect. In addition, the summarized results also highlight the both types of heterogeneity, locus and allelic. For instance multiple genes like *ATM, PTEN, HNF1A, BRCA2, MYO15A* and *MPZ* are having allele change at specific location causing PCa shows the allelic heterogeneity while more than one mutation as in case of same gen *BRCA2* in our cohort suggesting locus heterogeneity.

Continued development in precision medicine demands thorough characterization of genetic heterogeneity in investigations for complicated diseases to understand the heterogeneous pattern of association.

CHAPTER 6: CONCLUSIONS AND FUTURE PERSPECTIVES

What we contemplated to achieve from the current study was (1) to comprehend the known potential targets and find variants associated with PCa and diabetes, (2) identify potential targets could later aid in establishing biomarkers other than PSA that may help in early detection and enhancing frequency rate records in the future and (3) provide indefatigable evidence for the correlation of phenotypic features with different variants.

With the initial meta-analysis and extensive *in silico* application of tools, we obtained major common interacting genes from a meta-analysis study of published datasets of PCa, diabetes and obesity *viz, TMPO, BLM,* and *FOXP1*. These 3 genes emerged as the potential target to establish the association in the meta-analysis. The meta-analysis study was aimed to check whether candidate genes could be considered for precision scale modeling, and therefore we brought this to prediction scale and analytics. Furthermore, a probabilistic machine learning model was also achieved to identify key candidates between diabetes, obesity, and PCa. This, we believe, would herald precision scale modeling for easy prognosis. The ML heuristics has set a precedent to bring a transformation in cancer with data-driven pipelines for understanding possible causal relationships.

Applying WES as another objective has led us to sequence the FFPE blocks of our clinical cases and control cases and we obtained 42 mutated genes with heterozygous putative somatic and germline variants *viz, MYO15A, BRCA1, BRCA2, HNF1A, ERCC5, ATM, SEMA4A, TSC2, ABAC4, SDHB, NR2E3, GJB2, SMPD1, MPZ, PKP2, CACNA2D4, TOR1AIP1, FGD4, POLR3B, TMX2, ABCC9, PDHX, UBR1, CLN5, PGM1, COQ8A, AGL, CACNA1C, PTEN, PDE2A, DOCK7, SZT2, NDUFV1, TSHD1, PLCE1, CTPS1, ITGB4, CHEK1, COL2A1, PARP1, PTNP22, SLC16A13, MYRF* and *MPO.* We used several *in silico* tools to characterize the variant and validate through Sanger sequencing, which has made us distinguish those which are in association with both diabetes and PCa and which are rare and novel in PCa association.

Together with the analysis of clinical cases, the results of WES and meta-analysis, we did the comparison study, we obtained *BRCA2* (rs148341992, rs780919805, rs80359668), *ITGB4* (rs199620842), and *CTPS1* (rs186753161) from ClinVar dataset of PCa, *CHEK1* (rs506504), *TUBGCP4* (rs117773969), *BRCA1* (rs6416927), *ATM* (rs1800058), *HNF1A* (rs1800574), *COL2A1* (rs1793959), *PARP1* (rs61750984) from PRACTICAL consortium, *SLC16A13*

(rs312457) from GWAS central diabetes and *MYRF* (rs370887875), *BRCA2* (rs148341992), *ITGB4* (rs199620842) and *MPO* (rs35897051) from ClinVar diabetes.

Moreover, our attempts on how candidates associated with AR signaling interacts with lncRNA networks in PCa biology were fulfilled by performing docking studies between selected lncRNAs from our RNA-seq data with AR targeted proteins as well as PCa causal proteins in Objective 3. Interestingly, we identified SCARNA10 as a common lncRNA between both PCa and AR. Characterizing the functional aspect of this lncRNA would give us some insights into PCa progression. We firmly hope that this lncRNA would serve as prognostic signatures for PCa detection in Indian phenotype. Further a comparison with database of lncRNA associated T2DM identified *LINC01973, LINC00940, NPBWR1, FLJ16779, ANKRD20A9P, LINC00298, SNHG19, TLX1NB,* including *SCARNA10* which needs future investigations understand the common regulatory mechanism of the T2DM and PCa to establish more associations.

Besides this we performed pathway enrichment analysis on results obtained from both clinical sample analysis and meta-analysis in objective four using KEGG pathway, Panther DB, and Reactome, we noticed most of the pathways associated with the mutated genes were related to glycolytic pathways and insulin metabolism and AR receptor pathway. By applying the clustering coefficients algorithm of plug-in cytoHubba, we screened the top 10 hub genes for gene identified in the Meta-analysis and the WES again we obtained a ranking of *DLD, PDHX, NDUFB4, NDUFB7, BLM, UQCRC1, NDUFV1, PARP1, CHEK1* and *PTEN* are the top 10 genes that are functioning together are well related to both PCa and Diabetes except *UQCRC, PDHX and PAPR1*. The network analysis of significant genes using the Cytoscape-cytoHubba plugin ranked the genes by the clustering coefficient to show the hierarchy in high confidence interactions in PPI networks. Clustering PPI networks can be helpful for identifying groups of interacting proteins that take part in the same biological process or cooperate together to carry out a particular biological function.

On a granular level, both investigations, whether meta-analysis or prospective sample analysis, have enabled us to gain insights into three distinct phenotypes: diabetes, PCa, and obesity. However, there is room for improvement in ML-based integration, and time will tell. WES research on clinical samples supports the existence of a link between PCa and comorbidities such as diabetes and cardiovascular disease. Aside from heterogeneity, polygenicity may now be postulated in PCa for the first time. The study also emphasises the

need of taking comorbidities into account during both diagnosis and therapy. Further investigation is warranted to understand more for the mechanism and early prediction of PCa.

Putting all together we have achieved what we expected. WES, meta-analysis and molecular docking studies identified the potential targets and potential biomarkers. Enrichment analysis confirmed and explains the possible interplay and co-existence of the PCa and diabetes and with the effect polygenicity.

The NGS holds promise for transforming and benefiting precision medicine. We aim to extend the molecular docking studies by taking the present study genes associated with both diabetes and PCa. Hopefully in the coming years, we would validate the AR-lncRNA-PCa protein results experimentally using cell line studies. Although we faced challenges to collect the samples and thereby end up performing a small preliminary study with 5 clinical samples, this thesis, in our humble opinion, has set an early inroads for understanding comorbidity associated with the risk of PCa and therefore, performing more studies with a large cohort could be on the anvil through our CAPCI efforts (biocues.org/capci last accessed on April 12, 2024). Since many complex ailments like diabetes and cancer are well known heterogeneous diseases which makes present study a novel resource for understanding undiscovered genetic heterogeneities, in figuring out missing inheritance, accuracy in risk prediction and biomarker identification for better treatment approaches. The future looks bright with such a study having done large scale meta-analysis to design better risk prediction models and treatment decisions using machine learning algorithms which could benefit from considering the genetic heterogeneity and polygenicity observed in our current study. Through the CAPCI consortium, we are providing such an opportunity to extend, improve and come with more precision for screening candidate variants associated with PCa diagnosis in India.

Overall, the study successfully identified potential genetic targets and biomarkers that link PCa with diabetes and obesity. The combination of meta-analysis, WES and machine learning, our research has highlighted key variants and pathways involved in these conditions. The findings support development of precision medicine approaches and set the foundation for future studies to validate these associations, with a focus on improving early detection, risk prediction and treatment strategies for PCa in populations like India.

BIBLIOGRAPHY

©2019 TZEH KEONG FOO ALL RIGHTS RESERVED. (n.d.).

Aird, J., Baird, A. M., Lim, M. C. J., McDermott, R., Finn, S. P., & Gray, S. G. (2018). Carcinogenesis in prostate cancer: The role of long non-coding RNAs. In *Non-coding RNA Research* (Vol. 3, Issue 1). https://doi.org/10.1016/j.ncrna.2018.01.001

Akbarzadeh, M., Riahi, P., Ramezankhani, A., Dehkordi, S. R., Roudbar, M. A., Zarkesh, M., Guity, K., Khalili, D., Zahedi, A. S., Azizi, F., & Daneshpour, M. S. (2022). Parental Transmission Plays the Major Role in High Aggregation of Type-2-diabetes in Iranian Families: Tehran Lipid and Glucose Study. *Canadian Journal of Diabetes*, *46*(1). https://doi.org/10.1016/j.jcjd.2021.05.009

Ancellin, R., & Bessette, D. (2013). Overweight, obesity and cancer risks. *Oncologie*, *15*(3–4). https://doi.org/10.1007/s10269-013-2261-2

Antunes, H. P., Teixo, R., Carvalho, J. A., Eliseu, M., Marques, I., Mamede, A., Neves, R., Oliveira, R., Tavares-da-Silva, E., Parada, B., Abrantes, A. M., Figueiredo, A., & Botelho, M. F. (2018). Diabetes mellitus and prostate cancer metabolism: Is there a relationship? *Archivio Italiano Di Urologia e Andrologia*, *90*(3), 184–190. https://doi.org/10.4081/aiua.2018.3.184

Assenov, Y., Ramírez, F., Schelhorn, S. E. S. E., Lengauer, T., & Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics*, *24*(2). https://doi.org/10.1093/bioinformatics/btm554

Augspach, A., Drake, K. D., Roma, L., Qian, E., Lee, S. R., Clarke, D., Kumar, S., Jaquet, M., Gallon, J., Bolis, M., Triscott, J., Galván, J. A., Chen, Y., Thalmann, G., Julio, M. K., Theurillat, J.-P. P., Wuchty, S., Gerstein, M., Piscuoglio, S., … Rubin, M. A. (2021). Minor intron splicing efficiency increases with the development of lethal prostate cancer. *BioRxiv*.

Baden, M., Lu, L., Drummond, F. J., Gavin, A., & Sharp, L. (2020). Pain, fatigue and depression symptom cluster in survivors of prostate cancer. *Supportive Care in Cancer*, *28*(10). https://doi.org/10.1007/s00520-019-05268-0

Bader, D. A., & McGuire, S. E. (2020). Tumour metabolism and its unique properties in prostate adenocarcinoma. In *Nature Reviews Urology* (Vol. 17, Issue 4). https://doi.org/10.1038/s41585-020-0288-x

Bagherabadi, A., Hooshmand, A., Shekari, N., Singh, P., Zolghadri, S., Stanek, A., & Dohare, R. (2022). Correlation of NTRK1 Downregulation with Low Levels of Tumor-

Infiltrating Immune Cells and Poor Prognosis of Prostate Cancer Revealed by Gene Network Analysis. *Genes*, *13*(5). https://doi.org/10.3390/genes13050840

Barry, M. J. (2001). Prostate-Specific–Antigen Testing for Early Diagnosis of Prostate Cancer. *New England Journal of Medicine*, *344*(18). https://doi.org/10.1056/nejm200105033441806

Barsouk, A., Padala, S. A., Vakiti, A., Mohammed, A., Saginala, K., Thandra, K. C., Rawla, P., & Barsouk, A. (2020). Epidemiology, Staging and Management of Prostate Cancer. In *Medical sciences (Basel, Switzerland)* (Vol. 8, Issue 3). https://doi.org/10.3390/medsci8030028

Baxi, P., Eaton, D. B., Puri, V., & Schoen, M. W. (2024). Survival and comorbid diabetes mellitus in metastatic hormone-sensitive prostate cancer. *Journal of Clinical Oncology*, *42*(4_suppl). https://doi.org/10.1200/jco.2024.42.4_suppl.74

Beck, T. F., Mullikin, J. C., & Biesecker, L. G. (2016). Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants on behalf of the NISC Comparative Sequencing Program. *Clinical Chemistry*, *62*(4).

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition*, *98*(6). https://doi.org/10.1136/archdischild-2013-304340

Beier, A. M. K., Puhr, M., Stope, M. B., Thomas, C., & Erb, H. H. H. (2023). Metabolic changes during prostate cancer development and progression. In *Journal of Cancer Research and Clinical Oncology* (Vol. 149, Issue 5). https://doi.org/10.1007/s00432-022-04371-w

Beitelshees, A. L., Navare, H., Wang, D., Gong, Y., Wessel, J., Moss, J. I., Langaee, T. Y., Cooper-DeHoff, R. M., Sadee, W., Pepine, C. J., Schork, N. J., & Johnson, J. A. (2009). CACNA1C gene polymorphisms, cardiovascular disease outcomes, and treatment response. *Circulation: Cardiovascular Genetics*, *2*(4). https://doi.org/10.1161/CIRCGENETICS.109.857839

Bejrananda, T., & Pliensiri, P. (2023). Prediction of biochemical recurrence after laparoscopic radical prostatectomy. *BMC Urology*, *23*(1). https://doi.org/10.1186/s12894-023-01350-2

Berenguer, C. V., Pereira, F., Câmara, J. S., & Pereira, J. A. M. (2023). Underlying Features of Prostate Cancer—Statistics, Risk Factors, and Emerging Methods for Its Diagnosis. In *Current Oncology* (Vol. 30, Issue 2). https://doi.org/10.3390/curroncol30020178

Bestmann, B., Loetters, C., Diemer, T., Weidner, W., Küchler, T., & Rohde, V. (2007). Prostate-specific symptoms of prostate cancer in a German general population. *Prostate*

Cancer and Prostatic Diseases, *10*(1). https://doi.org/10.1038/sj.pcan.4500921

Beylerli, O., Gareev, I., Sufianov, A., Ilyasova, T., & Guang, Y. (2022). Long noncoding RNAs as promising biomarkers in cancer: long non-coding RNAs and cancer. In *Non-coding RNA Research* (Vol. 7, Issue 2). https://doi.org/10.1016/j.ncrna.2022.02.004

Bhagwat, M. (2010). Searching NCBI's dbSNP database. *Current Protocols in Bioinformatics*, *SUPP.32*. https://doi.org/10.1002/0471250953.bi0119s32

Bingham, M. (2023). In This Issue of Diabetes Care . *Diabetes Care*, *46*(2). https://doi.org/10.2337/dc23-ti02

Birnbaum, M. J., & Shaw, R. J. (2011). Genomics: Drugs, diabetes and cancer. In *Nature* (Vol. 470, Issue 7334). https://doi.org/10.1038/470338a

Blix, K., Jensvoll, H., Brækkan, S. K., & Hansen, J. B. (2013). White Blood Cell Count Measured Prior to Cancer Development Is Associated with Future Risk of Venous Thromboembolism - The Tromsø Study. *PLoS ONE*, *8*(9). https://doi.org/10.1371/journal.pone.0073447

Bossan, A., Ottman, R., Andl, T., Hasan, M. F., Mahajan, N., Coppola, D., & Chakrabarti, R. (2018). Expression of FGD4 positively correlates with the aggressive phenotype of prostate cancer. *BMC Cancer*, *18*(1). https://doi.org/10.1186/s12885-018-5096-9

Bradley Holmes, J., Moyer, E., Phan, L., Maglott, D., & Kattman, B. (2020). SPDI: Data model for variants and applications at NCBI. *Bioinformatics*, *36*(6). https://doi.org/10.1093/bioinformatics/btz856

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Breuninger, S., Reidenbach, S., Sauer, C. G., Ströbel, P., Pfitzenmaier, J., Trojan, L., & Hofmann, I. (2010). Desmosomal plakophilins in the prostate and prostatic adenocarcinomas: Implications for diagnosis and tumor progression. *American Journal of Pathology*, *176*(5). https://doi.org/10.2353/ajpath.2010.090737

Bridges, M. C., Daulagala, A. C., & Kourtidis, A. (2021). LNCcation: lncRNA localization and function. In *Journal of Cell Biology* (Vol. 220, Issue 2). https://doi.org/10.1083/JCB.202009045

Brinkmann, A. O., Faber, P. W., van Rooij, H. C. J., Kuiper, G. G. J. M., Ris, C., Klaassen, P., van der Korput, J. A. G. M., Voorhorst, M. M., van Laar, J. H., Mulder, E., & Trapman, J. (1989). The human androgen receptor: Domain structure, genomic organization and regulation of expression. *Journal of Steroid Biochemistry*, *34*(1–6).

https://doi.org/10.1016/0022-4731(89)90098-8

Budde, B. S., Aly, M. A., Mohamed, M. R., Breß, A., Altmüller, J., Motameny, S., Kawalia, A., Thiele, H., Konrad, K., Becker, C., Toliat, M. R., Nürnberg, G., Sayed, E. A. F., Mohamed, E. S., Pfister, M., & Nürnberg, P. (2020). Comprehensive molecular analysis of 61 Egyptian families with hereditary nonsyndromic hearing loss. *Clinical Genetics*, *98*(1). https://doi.org/10.1111/cge.13754

Bulten, W., Pinckaers, H., van Boven, H., Vink, R., de Bel, T., van Ginneken, B., van der Laak, J., Hulsbergen-van de Kaa, C., & Litjens, G. (2020). Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *The Lancet Oncology*. https://doi.org/10.1016/S1470-2045(19)30739-9

Buscher, E. G. (n.d.). *ANALYSIS OF THE PROTEOMIC PROFILE AND COMPOSITION IN ANALYSIS OF THE PROTEOMIC PROFILE AND COMPOSITION IN PROSTATE CANCER PATIENTS PROSTATE CANCER PATIENTS*. https://scholarworks.uvm.edu/hcoltheses/445

Butler, M., McKay, R. A., Popoff, I. J., Gaarde, W. A., Witchell, D., Murray, S. F., Dean, N. M., Bhanot, S., & Monia, B. P. (2002). Specific inhibition of PTEN expression reverses hyperglycemia in diabetic mice. *Diabetes*, *51*(4). https://doi.org/10.2337/diabetes.51.4.1028

Cao, Bo, Deng, H., Cui, H., Zhao, R., Li, H., Wei, B., & Chen, L. (2021). Knockdown of PGM1 enhances anticancer effects of orlistat in gastric cancer under glucose deprivation. *Cancer Cell International*, *21*(1). https://doi.org/10.1186/s12935-021-02193-3

Cao, Bochen, Bray, F., Ilbawi, A., & Soerjomataram, I. (2018). Effect on longevity of one-third reduction in premature mortality from non-communicable diseases by 2030: a global analysis of the Sustainable Development Goal health target. *The Lancet Global Health*. https://doi.org/10.1016/S2214-109X(18)30411-X

Castro, E., & Eeles, R. (2012). The role of BRCA1 and BRCA2 in prostate cancer. In *Asian Journal of Andrology* (Vol. 14, Issue 3). https://doi.org/10.1038/aja.2011.150

Cermenati, G., Abbiati, F., Cermenati, S., Brioschi, E., Volonterio, A., Cavaletti, G., Saez, E., De Fabiani, E., Crestani, M., Garcia-Segura, L. M., Melcangi, R. C., Caruso, D., & Mitro, N. (2012). Diabetes-induced myelin abnormalities are associated with an altered lipid pattern: Protective effects of LXR activation. *Journal of Lipid Research*, *53*(2). https://doi.org/10.1194/jlr.M021188

Chamie, K., Daskivich, T. J., Kwan, L., Labo, J., Dash, A., Greenfield, S., & Litwin, M. S. (2012). Comorbidities, treatment and ensuing survival in men with prostate cancer. In

Journal of General Internal Medicine (Vol. 27, Issue 5). https://doi.org/10.1007/s11606-011-1869-4

Chelala, C., Khan, A., & Lemoine, N. R. (2009). SNPnexus: A web database for functional annotation of newly discovered and public domain single nucleotide polymorphisms. *Bioinformatics*, *25*(5). https://doi.org/10.1093/bioinformatics/btn653

Chen, B., Liu, Y., Jin, X., Lu, W., Liu, J., Xia, Z., Yuan, Q., Zhao, X., Xu, N., & Liang, S. (2014). MicroRNA-26a regulates glucose metabolism by direct targeting PDHX in colorectal cancer cells. *BMC Cancer*, *14*(1). https://doi.org/10.1186/1471-2407-14-443

Chen, J., Stahl, A., Krah, N. M., Seaward, M. R., Dennison, R. J., Sapieha, P., Hua, J., Hatton, C. J., Juan, A. M., Aderman, C. M., Willett, K. L., Guerin, K. I., Mammoto, A., Campbell, M., & Smith, L. E. H. (2011). Wnt signaling mediates pathological vascular growth in proliferative retinopathy. *Circulation*, *124*(17). https://doi.org/10.1161/CIRCULATIONAHA.111.040337

Chen, N., & Zhou, Q. (2016). The evolving gleason grading system. In *Chinese Journal of Cancer Research* (Vol. 28, Issue 1). https://doi.org/10.3978/j.issn.1000-9604.2016.02.04

Chen, Y., Lu, Y., Kuyaxi, P., Cheng, J., Zhao, J., Zhao, Q., Musha, P., Zhang, H., & Yuan, H. (2018). Identification of pathogenic genes of nonsyndromic hearing loss in Uyghur families using massively parallel DNA sequencing technique. *Disease Markers*, *2018*. https://doi.org/10.1155/2018/5298057

Cheng, L., Zhuang, H., Ju, H., Yang, S., Han, J., Tan, R., & Hu, Y. (2019). Exposing the causal effect of body mass index on the risk of type-2-diabetes mellitus: A mandelian randominzation study. *Frontiers in Genetics*, *10*(FEB). https://doi.org/10.3389/fgene.2019.00094

Cheng, X., Huang, Y., Yang, P., & Bu, L. (2020). miR-383 ameliorates high glucose-induced β-cells apoptosis and hyperglycemia in high-fat induced diabetic mice. *Life Sciences*, *263*. https://doi.org/10.1016/j.lfs.2020.118571

Chicco, D., & Agapito, G. (2022). Nine quick tips for pathway enrichment analysis. *PLoS Computational Biology*, *18*(8). https://doi.org/10.1371/journal.pcbi.1010348

Coakley, F. V., & Hricak, H. (2000). Radiologic anatomy of the prostate gland: A clinical approach. *Radiologic Clinics of North America*, *38*(1). https://doi.org/10.1016/S0033-8389(05)70147-0

Cohen, P., Graves, H. C. B., Peehl, D. M., Kamarei, M., Giudice, L. C., & Rosenfeld, R. G. (1992). Prostate-specific antigen (PSA) is an insulin-like growth factor binding protein-3 protease found in seminal plasma. *Journal of Clinical Endocrinology and Metabolism*, *75*(4). https://doi.org/10.1210/jcem.75.4.1383255

Crossley, B. M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M. L., Clement, T., & Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and molecular assay monitoring. *Journal of Veterinary Diagnostic Investigation*, *32*(6). https://doi.org/10.1177/1040638720905833

Crow, J. A., Staggs, R. A., Raghavan, S., Johnson, J. E., Silverstein, K. A. T., Mayer, M. A., Kunau, T. M., & Retzel, E. F. (2001). *Design and Implementation of a Simple Relational Database for GenBank-Derived Data*. www.gcg.commproductsseqstore.html

Culp, M. B. B., Soerjomataram, I., Efstathiou, J. A., Bray, F., & Jemal, A. (2020). Recent Global Patterns in Prostate Cancer Incidence and Mortality Rates. In *European Urology* (Vol. 77, Issue 1). https://doi.org/10.1016/j.eururo.2019.08.005

Cutruzzolà, F., Giardina, G., Marani, M., Macone, A., Paiardini, A., Rinaldo, S., & Paone, A. (2017). Glucose metabolism in the progression of prostate cancer. In *Frontiers in Physiology* (Vol. 8, Issue FEB). Frontiers Media S.A. https://doi.org/10.3389/fphys.2017.00097

Das, U. N., & Rao, A. A. (2007). Gene expression profile in obesity and type-2-diabetes mellitus. *Lipids in Health and Disease*, *6*. https://doi.org/10.1186/1476-511X-6-35

Davey, R. A., & Grossmann, M. (2016). Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *The Clinical Biochemist. Reviews*, *37*(1).

Dayem Ullah, A. Z., Lemoine, N. R., & Chelala, C. (2012). SNPnexus: A web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Research*, *40*(W1). https://doi.org/10.1093/nar/gks364

De Cario, R., Kura, A., Suraci, S., Magi, A., Volta, A., Marcucci, R., Gori, A. M., Pepe, G., Giusti, B., & Sticchi, E. (2020). Sanger Validation of High-Throughput Sequencing in Genetic Diagnosis: Still the Best Practice? *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.592588

DeCensi, A., Puntoni, M., Goodwin, P., Cazzaniga, M., Gennari, A., Bonanni, B., & Gandini, S. (2010). Metformin and cancer risk in diabetic patients: A systematic review and meta-analysis. *Cancer Prevention Research*, *3*(11). https://doi.org/10.1158/1940-6207.CAPR-10-0157

Deepthi, B., Sowjanya, K., Lidiya, B., Bhargavi, R., & Babu, P. (2017). A modern review of diabetes mellitus: An annihilatory metabolic disorder. *Journal of In Silico & In Vitro Pharmacology*, *3*(1).

DEL DOTTORATO Profssa Laura Stronati, C. (n.d.). *DOTTORATO DI RICERCA IN BIOLOGIA UMANA E GENETICA MEDICA XXXV CICLO DNA repair genes in*

cancer predisposition: detection of germline pathogenic variants by multigene panel testing DOTTORANDO DOCENTE GUIDA Dott.ssa Soha Sadeghi Prof. Antonio Pizzuti CO-DOCENTE GUIDA Dott.ssa Maria Piane.

Demidenko, R., Razanauskas, D., Daniunaite, K., Lazutka, J. R., Jankevicius, F., & Jarmalaite, S. (2015). Frequent down-regulation of ABC transporter genes in prostate cancer. *BMC Cancer*, *15*(1). https://doi.org/10.1186/s12885-015-1689-8

Deng, Y., Hu, H., Xiao, L., Cai, T., Gao, W., Zhu, H., Wang, S., & Liu, J. (2022). Identification of EMT-Related lncRNAs as Potential Prognostic Biomarkers and Therapeutic Targets for Pancreatic Adenocarcinoma. *Journal of Oncology*, *2022*. https://doi.org/10.1155/2022/8259951

Descotes, J. L. (2019). Diagnosis of prostate cancer. In *Asian Journal of Urology* (Vol. 6, Issue 2, pp. 129–136). Editorial Office of Asian Journal of Urology. https://doi.org/10.1016/j.ajur.2018.11.007

Di Sebastiano, K. M., Pinthus, J. H., Duivenvoorden, W. C. M., & Mourtzakis, M. (2018). Glucose impairments and insulin resistance in prostate cancer: the role of obesity, nutrition and exercise. *Obesity Reviews*, *19*(7), 1008–1016. https://doi.org/10.1111/obr.12674

Dieter, C., Lemos, N. E., Corrêa, N. R. de F., Assmann, T. S., & Crispim, D. (2021). The Impact of lncRNAs in Diabetes Mellitus: A Systematic Review and In Silico Analyses. In *Frontiers in Endocrinology* (Vol. 12). https://doi.org/10.3389/fendo.2021.602597

Ding, L., Barbetti, F., & Liu, M. (2022). Editorial: Monogenic diabetes: from genetics and cell biology to clinical practice. In *Frontiers in Endocrinology* (Vol. 13). https://doi.org/10.3389/fendo.2022.1022611

Distefano, R., Ilieva, M., Madsen, J. H., Ishii, H., Aikawa, M., Rennie, S., & Uchida, S. (2023). T2DB: A Web Database for Long Non-Coding RNA Genes in Type II Diabetes. *Non-Coding RNA*, *9*(3). https://doi.org/10.3390/ncrna9030030

Donadon, V., Balbi, M., Valent, F., & Avogaro, A. (2010). Glycated hemoglobin and antidiabetic strategies as risk factors for hepatocellular carcinoma. *World Journal of Gastroenterology*, *16*(24), 3025–3032. https://doi.org/10.3748/wjg.v16.i24.3025

Dong, Y., Liang, G., Yuan, B., Yang, C., Gao, R., & Zhou, X. (2014). MALAT1 promotes the proliferation and metastasis of osteosarcoma cells by activating the PI3K/Akt pathway. *Tumor Biology*, *36*(3). https://doi.org/10.1007/s13277-014-2631-4

Dovey, Z., Horowitz, A., & Waingankar, N. (2023). The influence of lifestyle changes (diet, exercise and stress reduction) on prostate cancer tumour biology and patient outcomes: A systematic review. In *BJUI Compass* (Vol. 4, Issue 4).

https://doi.org/10.1002/bco2.237

Eadon, M. T., Lampe, S., Baig, M. M., Collins, K. S., Melo Ferreira, R., Mang, H., Cheng, Y. H., Barwinska, D., El-Achkar, T. M., Schwantes-An, T. H., Winfree, S., Temm, C. J., Ferkowicz, M. J., Dunn, K. W., Kelly, K. J., Sutton, T. A., Moe, S. M., Moorthi, R. N., Phillips, C. L., & Dagher, P. C. (2022). Clinical, histopathologic and molecular features of idiopathic and diabetic nodular mesangial sclerosis in humans. *Nephrology Dialysis Transplantation*, *37*(1). https://doi.org/10.1093/ndt/gfaa331

Edwards, T. L., Giri, A., Motley, S., Duong, W., & Fowke, J. H. (2013). Pleiotropy between genetic markers of obesity and risk of prostate cancer. *Cancer Epidemiology Biomarkers and Prevention*, *22*(9). https://doi.org/10.1158/1055-9965.EPI-13-0123

Efstathiou, E., Titus, M., Wen, S., Troncoso, P., Hoang, A., Corn, P., Prokhorova, I., Araujo, J., Dmuchowski, C., Melhem-Bertrandt, A., Patil, S., & Logothetis, C. J. (2020). Enzalutamide in Combination with Abiraterone Acetate in Bone Metastatic Castration-resistant Prostate Cancer Patients. *European Urology Oncology*, *3*(1). https://doi.org/10.1016/j.euo.2019.01.008

Eidelman, E., Twum-Ampofo, J., Ansari, J., & Siddiqui, M. M. (2017). The metabolic phenotype of prostate cancer. In *Frontiers in Oncology* (Vol. 7, Issue JUN). Frontiers Media S.A. https://doi.org/10.3389/fonc.2017.00131

Eke, I., Bylicky, M. A., Sandfort, V., Chopra, S., Martello, S., Graves, E. E., Coleman, C. N., & Aryankalayil, M. J. (2021). The lncRNAs LINC00261 and LINC00665 are upregulated in long-term prostate cancer adaptation after radiotherapy. *Molecular Therapy Nucleic Acids*, *24*. https://doi.org/10.1016/j.omtn.2021.02.024

Eldeeb, M. A., & Fahlman, R. P. (2016). Phosphorylation Impacts N-end rule degradation of the proteolytically activated form of BMX kinase. *Journal of Biological Chemistry*, *291*(43). https://doi.org/10.1074/jbc.M116.737387

Espach, Y., Lochner, A., Strijdom, H., & Huisamen, B. (2015). ATM Protein Kinase Signaling, Type-2-diabetes and Cardiovascular Disease. *Cardiovascular Drugs and Therapy*, *29*(1). https://doi.org/10.1007/s10557-015-6571-z

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C. D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., … D'Eustachio, P. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Research*, *46*(D1). https://doi.org/10.1093/nar/gkx1132

Fattahi, Z., Shearer, A. E., Babanejad, M., Bazazzadegan, N., Almadani, S. N., Nikzat, N., Jalalvand, K., Arzhangi, S., Esteghamat, F., Abtahi, R., Azadeh, B., Smith, R. J. H.,

Kahrizi, K., & Najmabadi, H. (2012). Screening for MYO15A gene mutations in autosomal recessive nonsyndromic, GJB2 negative Iranian deaf population. *American Journal of Medical Genetics, Part A*, *158 A*(8). https://doi.org/10.1002/ajmg.a.34411

Ferrè, F., Colantoni, A., & Helmer-Citterich, M. (2016). Revealing protein-lncRNA interaction. *Briefings in Bioinformatics*, *17*(1). https://doi.org/10.1093/bib/bbv031

Ferreira, L. B., Palumbo, A., de Mello, K. D., Sternberg, C., Caetano, M. S., de Oliveira, F. L., Neves, A. F., Nasciutti, L. E., Goulart, L. R., & Gimba, E. R. P. (2012). PCA3 noncoding RNA is involved in the control of prostate-cancer cell survival and modulates androgen receptor signaling. *BMC Cancer*, *12*. https://doi.org/10.1186/1471-2407-12-507

Fervaha, G., Izard, J. P., Tripp, D. A., Rajan, S., Leong, D. P., & Siemens, D. R. (2019). Depression and prostate cancer: A focused review for the clinician. In *Urologic Oncology: Seminars and Original Investigations* (Vol. 37, Issue 4). https://doi.org/10.1016/j.urolonc.2018.12.020

Finsterer, J., & Stöllberger, C. (2016). Heart disease in disorders of muscle, neuromuscular transmission, and the nerves. In *Korean Circulation Journal* (Vol. 46, Issue 2). https://doi.org/10.4070/kcj.2016.46.2.117

Flanagan, S. E., Patch, A. M., & Ellard, S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genetic Testing and Molecular Biomarkers*, *14*(4). https://doi.org/10.1089/gtmb.2010.0036

Fleming, S. T., McDavid, K., Pearce, K., & Pavlov, D. (2006). Comorbidities and the risk of late-stage prostate cancer. *TheScientificWorldJournal*, *6*. https://doi.org/10.1100/tsw.2006.383

Freedland, S. J., & Aronson, W. J. (2004). Examining the relationship between obesity and prostate cancer. *Reviews in Urology*, *6*(2).

Frei, K., Ramsebner, R., Lucas, T., Hamader, G., Szuhai, K., Weipoltshammer, K., Baumgartner, W. D., Wachtler, F. J., & Kirschhofer, K. (2005). GJB2 mutations in hearing impairment: Identification of a broad clinical spectrum for improved genetic counseling. *Laryngoscope*, *115*(3). https://doi.org/10.1097/01.mlg.0000157855.47143.71

Gao, W., Bohl, C. E., & Dalton, J. T. (2005). Chemistry and structural biology of androgen receptor. In *Chemical Reviews* (Vol. 105, Issue 9). https://doi.org/10.1021/cr020456u

Gao, Y., Zheng, H., Li, L., Feng, M., Chen, X., Hao, B., Lv, Z., Zhou, X., & Cao, Y. (2021). Prostate-Specific Membrane Antigen (PSMA) Promotes Angiogenesis of Glioblastoma Through Interacting With ITGB4 and Regulating NF-κB Signaling Pathway. *Frontiers*

in Cell and Developmental Biology, *9*. https://doi.org/10.3389/fcell.2021.598377

Garg, H., Seth, A., & Kumar, R. (2022). Raised prostate-specific antigen alone may not be a true predictor in high-risk prostate cancer: A retrospective cohort analysis. *Indian Journal of Urology*, *38*(1). https://doi.org/10.4103/iju.iju_368_21

Ghazanfari, Z., Haghdoost, A. A., Alizadeh, S. M., Atapour, J., & Zolala, F. (2010). A comparison of HbA1c and fasting blood sugar tests in general population. *International Journal of Preventive Medicine*, *1*(3).

Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., Griss, J., Sevilla, C., Matthews, L., Gong, C., Deng, C., Varusai, T., Ragueneau, E., Haider, Y., May, B., Shamovsky, V., Weiser, J., Brunson, T., Sanati, N., … D'Eustachio, P. (2022). The reactome pathway knowledgebase 2022. *Nucleic Acids Research*, *50*(D1). https://doi.org/10.1093/nar/gkab1028

Gogola, S., Rejzer, M., & Poppiti, R. (2024). Prostate gland anatomy and hormonal factors contributing to cancer development. In *Therapy Resistance in Prostate Cancer*. https://doi.org/10.1016/b978-0-443-16032-5.00011-7

Gollapalli, P., G, T. S., H, M., Shetty, P., & N, S. K. (2021). Network topology analysis of essential genes interactome of Helicobacter pylori to explore novel therapeutic targets. *Microbial Pathogenesis*, *158*. https://doi.org/10.1016/j.micpath.2021.105059

Gomes, K. P., Jadli, A. S., de Almeida, L. G. N., Ballasy, N. N., Edalat, P., Shandilya, R., Young, D., Belke, D., Shearer, J., Dufour, A., & Patel, V. B. (2022). Proteomic Analysis Suggests Altered Mitochondrial Metabolic Profile Associated With Diabetic Cardiomyopathy. *Frontiers in Cardiovascular Medicine*, *9*. https://doi.org/10.3389/fcvm.2022.791700

Goyal, N., Kesharwani, D., & Datta, M. (2018). Lnc-ing non-coding RNAs with metabolism and diabetes: roles of lncRNAs. In *Cellular and Molecular Life Sciences* (Vol. 75, Issue 10). https://doi.org/10.1007/s00018-018-2760-9

Grossmann, M., Thomas, M. C., Panagiotopoulos, S., Sharpe, K., MacIsaac, R. J., Clarke, S., Zajac, J. D., & Jerums, G. (2008). Low testosterone levels are common and associated with insulin resistance in men with diabetes. *Journal of Clinical Endocrinology and Metabolism*, *93*(5). https://doi.org/10.1210/jc.2007-2177

Gu, P., Chen, X., Xie, R., Xie, W., Huang, L., Dong, W., Han, J., Liu, X., Shen, J., Huang, J., & Lin, T. (2019). A novel AR translational regulator lncRNA LBCS inhibits castration resistance of prostate cancer. *Molecular Cancer*, *18*(1). https://doi.org/10.1186/s12943-019-1037-8

Guerrero-Ochoa, P., Rodríguez-Zapater, S., Anel, A., Esteban, L. M., Camón-Fernández, A.,

Espilez-Ortiz, R., Gil-Sanz, M. J., & Borque-Fernando, Á. (2024). Prostate Cancer and the Mevalonate Pathway. In *International Journal of Molecular Sciences* (Vol. 25, Issue 4). https://doi.org/10.3390/ijms25042152

Gupta, A., Shukla, N., Nehra, M., Gupta, S., Malik, B., Mishra, A. K., Vijay, M., Batra, J., Lohiya, N. K., Sharma, D., & Suravajhala, P. (2020). A Pilot Study on the Whole Exome Sequencing of Prostate Cancer in the Indian Phenotype Reveals Distinct Polymorphisms. *Frontiers in Genetics*. https://doi.org/10.3389/fgene.2020.00874

Ha Chung, B., Horie, S., & Chiong, E. (2019). The incidence, mortality, and risk factors of prostate cancer in Asian men. In *Prostate International* (Vol. 7, Issue 1). https://doi.org/10.1016/j.prnil.2018.11.001

Haddad, M., Eid, S., Harb, F., Massry, M. E. L., Azar, S., Sauleau, E. A., & Eid, A. A. (2022). Activation of 20-HETE Synthase Triggers Oxidative Injury and Peripheral Nerve Damage in Type 2 Diabetic Mice. *Journal of Pain*, *23*(8). https://doi.org/10.1016/j.jpain.2022.02.011

Hammerich, K. H., Ayala, G. E., & Wheeler, T. M. (2008). Anatomy of the prostate gland and surgical pathology of prostate cancer. In *Prostate Cancer*. https://doi.org/10.1017/CBO9780511551994.003

Harrison, S. M., Riggs, E. R., Maglott, D. R., Lee, J. M., Azzariti, D. R., Niehaus, A., Ramos, E. M., Martin, C. L., Landrum, M. J., & Rehm, H. L. (2016). Using ClinVar as a resource to support variant interpretation. *Current Protocols in Human Genetics*, *2016*. https://doi.org/10.1002/0471142905.hg0816s89

Hartley, A., & Ahmad, I. (2023). The role of PPARγ in prostate cancer development and progression. In *British Journal of Cancer* (Vol. 128, Issue 6). https://doi.org/10.1038/s41416-022-02096-8

Heinlein, C. A., & Chang, C. (2004). Androgen receptor in prostate cancer. In *Endocrine Reviews* (Vol. 25, Issue 2, pp. 276–308). https://doi.org/10.1210/er.2002-0032

Hernandez, J., & Thompson, I. M. (2004). Diagnosis and treatment of prostate cancer. In *Medical Clinics of North America* (Vol. 88, Issue 2, pp. 267–279). W.B. Saunders. https://doi.org/10.1016/S0025-7125(03)00144-5

Hong, G., Zhang, W., Li, H., Shen, X., & Guo, Z. (2014). Separate enrichment analysis of pathways for up- And downregulated genes. *Journal of the Royal Society Interface*, *11*(92). https://doi.org/10.1098/rsif.2013.0950

Hopkins, B. D., Goncalves, M. D., & Cantley, L. C. (2016). Obesity and cancer mechanisms: Cancer metabolism. In *Journal of Clinical Oncology* (Vol. 34, Issue 35). https://doi.org/10.1200/JCO.2016.67.9712

Hori, M., & Palmer, M. (2021). Age-specific prostate cancer incidence rate in the world. In *Japanese Journal of Clinical Oncology* (Vol. 51, Issue 1). https://doi.org/10.1093/jjco/hyaa253

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, *82*(11). https://doi.org/10.1016/j.humimm.2021.02.012

Hu, Y., Liu, W., Chen, Y., Zhang, M., Wang, L., Zhou, H., Wu, P., Teng, X., Dong, Y., Zhou, J. W., Xu, H., Zheng, J., Li, S., Tao, T., Hu, Y., & Jia, Y. (2010). Combined use of fasting plasma glucose and glycated hemoglobin A1c in the screening of diabetes and impaired glucose tolerance. *Acta Diabetologica*, *47*(3), 231–236. https://doi.org/10.1007/s00592-009-0143-2

Huang, H., Li, L., & Wen, K. (2021). Interactions between long non-coding RNAs and RNA-binding proteins in cancer (Review). In *Oncology Reports* (Vol. 46, Issue 6). https://doi.org/10.3892/or.2021.8207

Huo, Q., Wang, S., Hou, Y., Gorczynski, R. M., Shen, Y., Wang, B., Ge, H., & Li, T. (2023). Editorial: The relationship between diabetes and cancers and its underlying mechanisms, volume II. In *Frontiers in Endocrinology* (Vol. 14). https://doi.org/10.3389/fendo.2023.1357577

Iczkowski, K. A., & Lucia, M. S. (2011). Current perspectives on Gleason grading of prostate cancer. *Current Urology Reports*. https://doi.org/10.1007/s11934-011-0181-5

Jacob, A., Raj, R., Allison, D. B., & Myint, Z. W. (2021). Androgen receptor signaling in prostate cancer and therapeutic strategies. In *Cancers* (Vol. 13, Issue 21). https://doi.org/10.3390/cancers13215417

Jain, P., Vig, S., Datta, M., Jindel, D., Mathur, A. K., Mathur, S. K., & Sharma, A. (2013). Systems Biology Approach Reveals Genome to Phenome Correlation in Type-2-diabetes. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0053522

Jamaspishvili, T., Berman, D. M., Ross, A. E., Scher, H. I., De Marzo, A. M., Squire, J. A., & Lotan, T. L. (2018). Clinical implications of PTEN loss in prostate cancer. In *Nature Reviews Urology* (Vol. 15, Issue 4). https://doi.org/10.1038/nrurol.2018.9

Jamnagerwalla, J., Howard, L. E., Allott, E. H., Vidal, A. C., Moreira, D. M., Castro-Santamaria, R., Andriole, G. L., Freeman, M. R., & Freedland, S. J. (2018). Serum cholesterol and risk of high-grade prostate cancer: Results from the REDUCE study. *Prostate Cancer and Prostatic Diseases*, *21*(2). https://doi.org/10.1038/s41391-017-0030-9

Jang, M., Kim, S. S., & Lee, J. (2013). Cancer cell metabolism: Implications for therapeutic

targets. In *Experimental and Molecular Medicine* (Vol. 45, Issue 10). https://doi.org/10.1038/emm.2013.85

Kachuri, L., Hoffmann, T. J., Jiang, Y., Berndt, S. I., Shelley, J. P., Schaffer, K. R., Machiela, M. J., Freedman, N. D., Huang, W. Y., Li, S. A., Easterlin, R., Goodman, P. J., Till, C., Thompson, I., Lilja, H., Van Den Eeden, S. K., Chanock, S. J., Haiman, C. A., Conti, D. V., … Witte, J. S. (2023). Genetically adjusted PSA levels for prostate cancer screening. *Nature Medicine*, *29*(6). https://doi.org/10.1038/s41591-023-02277-9

Kakkat, S., Pramanik, P., Singh, S., Singh, A. P., Sarkar, C., & Chakroborty, D. (2023). Cardiovascular Complications in Patients with Prostate Cancer: Potential Molecular Connections. In *International Journal of Molecular Sciences* (Vol. 24, Issue 8). https://doi.org/10.3390/ijms24086984

Kang, Y., Zheng, C., Ye, J., Song, F., Wang, X., Liu, Y., Tian, M., Dong, J., & Lu, S. (2021). Effects of advanced glycation end products on neutrophil migration and aggregation in diabetic wounds. *Aging*, *13*(8). https://doi.org/10.18632/aging.202924

Kapoor, D., Goodwin, E., Channer, K. S., & Jones, T. H. (2006). Testosterone replacement therapy improves insulin resistance, glycaemic control, visceral adiposity and hypercholesterolaemia in hypogonadal men with type-2-diabetes. *European Journal of Endocrinology*, *154*(6). https://doi.org/10.1530/eje.1.02166

Kassem, S. A., Ariel, I., Thornton, P. S., Hussain, K., Smith, V., Lindley, K. J., Aynsley-Green, A., & Glaser, B. (2001). p57KIP2 expression in normal islet cells and in hyperinsulinism of infancy. *Diabetes*, *50*(12). https://doi.org/10.2337/diabetes.50.12.2763

Kawahara, R., Recuero, S., Nogueira, F. C. S., Domont, G. B., Leite, K. R. M., Srougi, M., Thaysen-Andersen, M., & Palmisano, G. (2019). Tissue Proteome Signatures Associated with Five Grades of Prostate Cancer and Benign Prostatic Hyperplasia. *Proteomics*, *19*(21–22). https://doi.org/10.1002/pmic.201900174

Kenfield, S. A., Stampfer, M. J., Chan, J. M., & Giovannucci, E. (2011). Smoking and prostate cancer survival and recurrence. *JAMA*, *305*(24). https://doi.org/10.1001/jama.2011.879

Kenkpen, A. K., Storey, J. J., Olson, E. R., Guden, T. E., Card, T. T., Jensen, A. S., Ahrens, J. L., & Hellmann Whitaker, R. A. (2023). Developing Connections Between LINC00298 RNA and Alzheimer's Disease Through Mapping Its Interactome and Through Biochemical Characterization. *Journal of Alzheimer's Disease*, *95*(2). https://doi.org/10.3233/JAD-230057

Khan, A., Chandra, A., Mahdi, A. A., & Sankhwar, S. N. (2019). PROSTATE CANCER

AND DIABETES LINK: ROLE OF INSULIN AND INSULIN LIKE GROWTH FACTORS. *Era's Journal of Medical Research*, *6*(2). https://doi.org/10.24041/ejmr2019.145

Khandekar, M. J., Cohen, P., & Spiegelman, B. M. (2011). Molecular mechanisms of cancer development in obesity. In *Nature Reviews Cancer* (Vol. 11, Issue 12). https://doi.org/10.1038/nrc3174

Khilwani, B., Kour, B., Shukla, N., Vure, S., Ansari, A. S., Lohiya, N. K., Suravajhala, R., & Suravajhala, P. (n.d.). *In silico approaches to identify the role of lncRNAs in Prostate cancer and Androgen receptor-targeted proteins*. http://www.noncode.org/

Khokhar, M., Roy, D., Modi, A., Agarwal, R., Yadav, D., Purohit, P., & Sharma, P. (2020). Perspectives on the role of PTEN in diabetic nephropathy: an update. In *Critical Reviews in Clinical Laboratory Sciences* (Vol. 57, Issue 7). https://doi.org/10.1080/10408363.2020.1746735

Kimura, T., & Egawa, S. (2018). Epidemiology of prostate cancer in Asian countries. In *International Journal of Urology* (Vol. 25, Issue 6). https://doi.org/10.1111/iju.13593

Kingshott, G., Biernacka, K., Sewell, A., Gwiti, P., Barker, R., Zielinska, H., Gilkes, A., McCarthy, K., Martin, R. M., Lane, J. A., McGeagh, L., Koupparis, A., Rowe, E., Oxley, J., Holly, J. M. P., & Perks, C. M. (2021). Alteration of metabolic conditions impacts the regulation of igf-ii/h19 imprinting status in prostate cancer. *Cancers*. https://doi.org/10.3390/cancers13040825

Klil-Drori, A. J., Azoulay, L., & Pollak, M. N. (2017). Cancer, obesity, diabetes, and antidiabetic drugs: Is the fog clearing? In *Nature Reviews Clinical Oncology* (Vol. 14, Issue 2). https://doi.org/10.1038/nrclinonc.2016.120

Kohl, M., Wiese, S., & Warscheid, B. (2011). Cytoscape: Software for Visualization and Analysis of Biological Networks. In *Methods in Molecular Biology* (Vol. 696). https://doi.org/10.1007/978-1-60761-987-1_18

Koko, M., Abdallah, M. O. E., Amin, M., & Ibrahim, M. (2018). Challenges imposed by minor reference alleles on the identification and reporting of clinical variants from exome data. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-4433-3

Kong, X., Zhao, Q., Xing, X., Zhang, B., Zhang, X., Hong, J., & Yang, W. (2015). Genetic variants associated with lipid profiles in Chinese patients with type-2-diabetes. *PLoS ONE*, *10*(8). https://doi.org/10.1371/journal.pone.0135145

Koochekpour, S. (2010). Androgen receptor signaling and mutations in prostate cancer. In *Asian Journal of Andrology* (Vol. 12, Issue 5). https://doi.org/10.1038/aja.2010.89

Kour, B., Shukla, N., Bhargava, H., Sharma, D., Sharma, A., Singh, A., Valadi, J., Sadasukhi,

T. C., Vuree, S., & Suravajhala, P. (2023). Identification of Plausible Candidates in Prostate Cancer Using Integrated Machine Learning Approaches. *Current Genomics*, *24*(5). https://doi.org/10.2174/0113892029240239231109082805

Kumar, P. R., Bhansali, A., Ravikiran, M., Bhansali, S., Dutta, P., Thakur, J. S., Sachdeva, N., Bhadada, S. K., & Walia, R. (2010). Utility of glycated hemoglobin in diagnosing type-2-diabetes mellitus: A community-based study. *Journal of Clinical Endocrinology and Metabolism*, *95*(6), 2832–2835. https://doi.org/10.1210/jc.2009-2433

Kumar, S., Prajapati, K. S., Singh, A. K., Kushwaha, P. P., Shuaib, M., & Gupta, S. (2021). Long non-coding RNA regulating androgen receptor signaling in breast and prostate cancer. In *Cancer Letters* (Vol. 504). https://doi.org/10.1016/j.canlet.2020.11.039

KUNKLER, C. N., SCHIEFELBEIN, G. E., O'LEARY, N. J., MCCOWN, P. J., & BROWN, J. A. (2022). A single natural RNA modification can destabilize a U A-T-rich RNA DNA-DNA triple helix. *RNA*, *28*(9). https://doi.org/10.1261/rna.079244.122

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., Hoffman, D., Jang, W., Kaur, K., Liu, C., Lyoshin, V., Maddipatla, Z., Maiti, R., Mitchell, J., O'Leary, N., Riley, G. R., Shi, W., Zhou, G., Schneider, V., … Kattman, B. L. (2020). ClinVar: Improvements to accessing data. *Nucleic Acids Research*, *48*(D1). https://doi.org/10.1093/nar/gkz972

Lee, S. H., & Shen, M. M. (2015). Cell types of origin for prostate cancer. In *Current Opinion in Cell Biology* (Vol. 37). https://doi.org/10.1016/j.ceb.2015.10.002

Lee, S., Xu, H., Van Vleck, A., Mawla, A. M., Li, A. M., Ye, J., Huising, M. O., & Annes, J. P. (2022). β-Cell Succinate Dehydrogenase Deficiency Triggers Metabolic Dysfunction and Insulinopenic Diabetes. *Diabetes*, *71*(7). https://doi.org/10.2337/db21-0834

Lee, Y. H. A., Hui, J. M. H., Chung, C. T., Liu, K., Dee, E. C., Ng, K., Tse, G., Chan, J. S. K., & Ng, C. F. (2023). Metformin use and hospital attendance-related resources utilization among diabetic patients with prostate cancer on androgen deprivation therapy: A population-based cohort study. *Cancer Medicine*, *12*(8). https://doi.org/10.1002/cam4.5651

Lemos, A. E. G., Da Rocha Matos, A., Ferreira, L. B., & Gimba, E. R. P. (2019). The long non-coding RNA PCA3: An update of its functions and clinical applications as a biomarker in prostate cancer. In *Oncotarget* (Vol. 10, Issue 61). https://doi.org/10.18632/oncotarget.27284

Li, C., Tang, Z., Zhang, W., Ye, Z., & Liu, F. (2021). GEPIA2021: Integrating multiple deconvolution-based analysis into GEPIA. *Nucleic Acids Research*, *49*(W1). https://doi.org/10.1093/nar/gkab418

Li, J., & Huang, G. (2024). Insulin receptor alternative splicing in breast and prostate cancer. In *Cancer Cell International* (Vol. 24, Issue 1). https://doi.org/10.1186/s12935-024-03252-1

Liang, Y., Ketchum, N. S., Goodman, P. J., Klein, E. A., & Thompson, I. M. (2014). Is there a role for body mass index in the assessment of prostate cancer risk on biopsy? *Journal of Urology*, *192*(4), 1094–1099. https://doi.org/10.1016/j.juro.2014.04.015

Lin, Y., & Jin, X. (2023). Effect of ubiquitin protease system on DNA damage response in prostate cancer (Review). *Experimental and Therapeutic Medicine*, *27*(1). https://doi.org/10.3892/etm.2023.12321

Lindberg, J., Klevebring, D., Liu, W., Neiman, M., Xu, J., Wiklund, P., Wiklund, F., Mills, I. G., Egevad, L., & Grönberg, H. (2013). Exome sequencing of prostate cancer supports the hypothesis of independent tumour origins. *European Urology*, *63*(2). https://doi.org/10.1016/j.eururo.2012.03.050

Lippi, G., & Targher, G. (2010). Glycated hemoglobin (HbA1c): Old dogmas, a new perspective? *Clinical Chemistry and Laboratory Medicine*, *48*(5), 609–614. https://doi.org/10.1515/CCLM.2010.144

Liu, H., Yuan, M., Mitra, R., Zhou, X., Long, M., Lei, W., Zhou, S., Huang, Y. e., Hou, F., Eischen, C. M., & Jiang, W. (2022). CTpathway: a CrossTalk-based pathway enrichment analysis method for cancer research. *Genome Medicine*, *14*(1). https://doi.org/10.1186/s13073-022-01119-6

Liu, J. M., Liu, J. N., Wei, M. T., He, Y. Z., Zhou, Y., Song, X. B., Ying, B. W., & Huang, J. (2013). Effect of IL-18 gene promoter polymorphisms on prostate cancer occurrence and prognosis in Han Chinese population. *Genetics and Molecular Research*, *12*(1). https://doi.org/10.4238/2013.March.15.2

Liu, J. Y., Yao, J., Li, X. M., Song, Y. C., Wang, X. Q., Li, Y. J., Yan, B., & Jiang, Q. (2014). Pathogenic role of lncRNA-MALAT1 in endothelial cell dysfunction in diabetes mellitus. *Cell Death and Disease*, *5*(10). https://doi.org/10.1038/cddis.2014.466

Liu, M. H., Borg, S., Salowsky, R., Padmanaban, A., Saligrama, D., Mcdade Walker, D., Inche, A., Elliott, J., Connelly, M., Boland, D., Padmanabanv, A., Graf, E., Risinger, J., Dickman, J., Toloue, M. M., Kamberov, E., Tesmer, T., Yerramilli, S., Carey, M., … Carroll, M. (2014). Automated Assessment of Next Generation Sequencing Library Preparation Workflow for Quality and Quantity Using the Agilent 2200 TapeStation System Automated RNA Sample Quality Control Rapid DNA-Seq to Achieve High Coverage Libraries from 1ng-1 g in 2 Hours Sequencing Single Human and Bacterial Cells at Low Coverage for Aneuploidy, CNV, and Genotyping Applications. In *ABRF*

2014 POSTER ABSTRACTS S18 JOURNAL OF BIOMOLECULAR TECHNIQUES (Vol. 25).

Liu, R., Duan, T., Yu, L., Tang, Y., Liu, S., Wang, C., & Fang, W. J. (2023). Acid sphingomyelinase promotes diabetic cardiomyopathy via NADPH oxidase 4 mediated apoptosis. *Cardiovascular Diabetology*, *22*(1). https://doi.org/10.1186/s12933-023-01747-1

Liu, S. X., Zheng, F., Xie, K. L., Xie, M. R., Jiang, L. J., & Cai, Y. (2019). Exercise Reduces Insulin Resistance in Type-2-diabetes Mellitus via Mediating the lncRNA MALAT1/MicroRNA-382-3p/Resistin Axis. *Molecular Therapy Nucleic Acids*, *18*. https://doi.org/10.1016/j.omtn.2019.08.002

Liu, X., Tan, W., Wang, W., Feng, T., Wang, C., Wang, L., & Zhou, W. (2023). SEMA4A promotes prostate cancer invasion: involvement of tumor microenvironment. *Journal of Cancer*, *14*(14). https://doi.org/10.7150/jca.86739

Liu, Y., Hu, Y., Zhang, M., Jiang, R., & Liang, C. (2018). Polymorphisms in ERCC2 and ERCC5 and risk of prostate cancer: A meta-analysis and systematic review. *Journal of Cancer*, *9*(16). https://doi.org/10.7150/jca.25356

Liu, Z., Chen, C., Yu, F., Yuan, D., Wang, W., Jiao, K., Yang, S., Zhang, Y., Wang, Y., Liu, L., Xu, H., Zhang, Y., Zhu, G., Hu, B., & Zhu, J. (2021). Association of Total Dietary Intake of Sugars with Prostate-Specific Antigen (PSA) Concentrations: Evidence from the National Health and Nutrition Examination Survey (NHANES), 2003-2010. In *BioMed Research International* (Vol. 2021). https://doi.org/10.1155/2021/4140767

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., … Moore, H. F. (2013). The Genotype-Tissue Expression (GTEx) project. In *Nature Genetics* (Vol. 45, Issue 6). https://doi.org/10.1038/ng.2653

Lu, Q., & Zhu, L. (2020). The role of semaphorins in metabolic disorders. In *International Journal of Molecular Sciences* (Vol. 21, Issue 16). https://doi.org/10.3390/ijms21165641

Luo, R., Li, L., Hu, Y. X., & Xiao, F. (2021). LncRNA H19 inhibits high glucose-induced inflammatory responses of human retinal epithelial cells by targeting miR-19b to increase SIRT1 expression. *Kaohsiung Journal of Medical Sciences*, *37*(2). https://doi.org/10.1002/kjm2.12302

Lygirou, V., Latosinska, A., Makridakis, M., Mullen, W., Delles, C., Schanstra, J. P., Zoidakis, J., Pieske, B., Mischak, H., & Vlahou, A. (2018). Plasma proteomic analysis

reveals altered protein abundances in cardiovascular disease. *Journal of Translational Medicine*, *16*(1). https://doi.org/10.1186/s12967-018-1476-9

Ma, X. Y., Xu, H. Q., Zhao, J. F., Ruan, Y., & Chen, B. (2022). Discovery of a Novel Bloom's Syndrome Protein (BLM) Inhibitor Suppressing Growth and Metastasis of Prostate Cancer. *International Journal of Molecular Sciences*, *23*(23). https://doi.org/10.3390/ijms232314798

Makowska, K. A. A., Hughes, R. E. E., White, K. J. J., Wells, C. M. M., & Peckham, M. (2015). Specific Myosins Control Actin Organization, Cell Morphology, and Migration in Prostate Cancer Cells. *Cell Reports*, *13*(10). https://doi.org/10.1016/j.celrep.2015.11.012

Malik, B., & Feng, F. (2016). Long noncoding RNAs in prostate cancer: Overview and clinical implications. In *Asian Journal of Andrology* (Vol. 18, Issue 4). https://doi.org/10.4103/1008-682X.177123

Mamidi, T. K. K., Wu, J., & Hicks, C. (2019). Integrating germline and somatic variation information using genomic data for the discovery of biomarkers in prostate cancer. *BMC Cancer*, *19*(1). https://doi.org/10.1186/s12885-019-5440-8

Marchetti, C., Patriarca, P., Solero, G. Pietro, Baralle, F. E., & Romano, M. (2004). Genetic Characterization of Myeloperoxidase Deficiency in Italy. *Human Mutation*, *23*(5). https://doi.org/10.1002/humu.20027

Marín-Aguilera, M., Codony-Servat, J., Kalko, S. G., Fernández, P. L., Bermudo, R., Buxo, E., Ribal, M. J., Gascón, P., & Mellado, B. (2012). Identification of docetaxel resistance genes in castration-resistant prostate cancer. *Molecular Cancer Therapeutics*, *11*(2). https://doi.org/10.1158/1535-7163.MCT-11-0289

Mathiyalagan, P., Okabe, J., Chang, L., Su, Y., Du, X. J., & El-Osta, A. (2014). The primary microRNA-208b interacts with Polycomb-group protein, Ezh2, to regulate gene expression in the heart. *Nucleic Acids Research*, *42*(2). https://doi.org/10.1093/nar/gkt896

McNeal, J. E. (1981). The zonal anatomy of the prostate. *The Prostate*, *2*(1). https://doi.org/10.1002/pros.2990020105

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, *47*(D1). https://doi.org/10.1093/nar/gky1038

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2016). PANTHER version 10: Expanded protein families and functions, and analysis tools. *Nucleic Acids Research*, *44*(D1). https://doi.org/10.1093/nar/gkv1194

Minoretti, P., Falcone, C., Aldeghi, A., Olivieri, V., Mori, F., Emanuele, E., Calcagnino, M., & Geroldi, D. (2006). A novel Val734Ile variant in the ABCC9 gene associated with myocardial infarction. *Clinica Chimica Acta*, *370*(1–2). https://doi.org/10.1016/j.cca.2006.02.007

Misawa, A., Takayama, K. I., & Inoue, S. (2017). Long non-coding RNAs and prostate cancer. In *Cancer Science* (Vol. 108, Issue 11). https://doi.org/10.1111/cas.13352

Mistry, T., Digby, J. E., Desai, K. M., & Randeva, H. S. (2007). Obesity and Prostate Cancer: A Role for Adipokines. In *European Urology* (Vol. 52, Issue 1). https://doi.org/10.1016/j.eururo.2007.03.054

Mittal, R. D., Mittal, T., Singh, A. K., & Mandal, R. K. (2012). Association of caspases with an increased prostate cancer risk in north indian population. *DNA and Cell Biology*, *31*(1). https://doi.org/10.1089/dna.2011.1285

Miyamoto, H., Messing, E. M., & Chang, C. (2004). Androgen deprivation therapy for prostate cancer: Current status and future prospects. *Prostate*, *61*(4). https://doi.org/10.1002/pros.20115

Monterisi, S., Lobo, M. J., Livie, C., Castle, J. C., Weinberger, M., Baillie, G., Surdo, N. C., Musheshe, N., Stangherlin, A., Gottlieb, E., Maizels, R., Bortolozzi, M., Micaroni, M., & Zaccolo, M. (2017). PDE2A2 regulates mitochondria morphology and apoptotic cell death via local modulation of cAMP/PKA signalling. *ELife*, *6*. https://doi.org/10.7554/eLife.21374

Mubeen, S., Tom Kodamullil, A. T., Hofmann-Apitius, M., & Domingo-Fernández, D. (2022). On the influence of several factors on pathway enrichment analysis. In *Briefings in Bioinformatics* (Vol. 23, Issue 3). https://doi.org/10.1093/bib/bbac143

Munkley, J., Rajan, P., Lafferty, N. P., Dalgliesh, C., Jackson, R. M., Robson, C. N., Leung, H. Y., & Elliott, D. J. (2014). A novel androgen-regulated isoform of the TSC2 tumour suppressor gene increases cell proliferation. *Oncotarget*, *5*(1). https://doi.org/10.18632/oncotarget.1405

Murray, N. H., Lewis, A., Rincon Pabon, J. P., Gross, M. L., Henzler-Wildman, K., & Pagliarini, D. J. (2022). 2-Propylphenol Allosterically Modulates COQ8A to Enhance ATPase Activity. *ACS Chemical Biology*, *17*(8). https://doi.org/10.1021/acschembio.2c00434

Murray, P. ., & Clayton, P. . (2000). Disorders of Growth Hormone in Childhood. In *Endotext*.

Nagamani, S., Sahoo, R., Muneeswaran, G., & Narahari Sastry, G. (2019). Data Science Driven Drug Repurposing for Metabolic Disorders. In *In Silico Drug Design:*

Repurposing Techniques and Methodologies. https://doi.org/10.1016/B978-0-12-816125-8.00007-9

Nair, S. V. . M. T. G. et al. a. (2020). Next-Generation Sequencing in Cancer. *Journal of Maxillofacial and Oral Surgery*, *20*, 340–344.

Neeb, A., Herranz, N., Arce-Gallego, S., Miranda, S., Buroni, L., Yuan, W., Athie, A., Casals, T., Carmichael, J., Rodrigues, D. N., Gurel, B., Rescigno, P., Rekowski, J., Welti, J., Riisnaes, R., Gil, V., Ning, J., Wagner, V., Casanova-Salas, I., … de Bono, J. S. (2021). Advanced Prostate Cancer with ATM Loss: PARP and ATR Inhibitors. *European Urology*, *79*(2). https://doi.org/10.1016/j.eururo.2020.10.029

Nencioni, A., Caffa, I., Cortellino, S., & Longo, V. D. (2018). Fasting and cancer: molecular mechanisms and clinical application. *Nature Reviews Cancer*, *18*(11), 707–719. https://doi.org/10.1038/s41568-018-0061-0

Nkyimbeng-Takwi, E. H., & Chapoval, S. P. (2011). Biology and function of neuroimmune semaphorins 4A and 4D. In *Immunologic Research* (Vol. 50, Issue 1). https://doi.org/10.1007/s12026-010-8201-y

Noto, H., Goto, A., Tsujimoto, T., Osame, K., & Noda, M. (2013). Latest insights into the risk of cancer in diabetes. In *Journal of Diabetes Investigation* (Vol. 4, Issue 3). https://doi.org/10.1111/jdi.12068

Oderda, M., Dematteis, A., Calleris, G., Conti, A., D'Agate, D., Falcone, M., Marquis, A., Montefusco, G., Marra, G., & Gontero, P. (2023). Predictors of Prostate Cancer at Fusion Biopsy: The Role of Positive Family History, Hypertension, Diabetes, and Body Mass Index. *Current Oncology*, *30*(5). https://doi.org/10.3390/curroncol30050374

Ogata, H., Goto, S., Fujibuchi, W., & Kanehisa, M. (1998). Computation with the KEGG pathway database. *BioSystems*, *47*(1–2). https://doi.org/10.1016/S0303-2647(98)00017-3

Oh, M., Alkhushaym, N., Fallatah, S., Althagafi, A., Aljadeed, R., Alsowaida, Y., Jeter, J., Martin, J. R., Babiker, H. M., McBride, A., & Abraham, I. (2019). The association of BRCA1 and BRCA2 mutations with prostate cancer risk, frequency, and mortality: A meta-analysis. *Prostate*, *79*(8). https://doi.org/10.1002/pros.23795

Ohmoto, A., Morizane, C., Kubo, E., Takai, E., Hosoi, H., Sakamoto, Y., Kondo, S., Ueno, H., Shimada, K., Yachida, S., & Okusaka, T. (2018). Germline variants in pancreatic cancer patients with a personal or family history of cancer fulfilling the revised Bethesda guidelines. *Journal of Gastroenterology*, *53*(10). https://doi.org/10.1007/s00535-018-1466-y

Okotie, O. T., Roehl, K. A., Han, M., Loeb, S., Gashti, S. N., & Catalona, W. J. (2007).

Characteristics of Prostate Cancer Detected by Digital Rectal Examination Only. *Urology*. https://doi.org/10.1016/j.urology.2007.07.019

Oscanoa, J., Sivapalan, L., Gadaleta, E., Dayem Ullah, A. Z., Lemoine, N. R., & Chelala, C. (2020). SNPnexus: A web server for functional annotation of human genome sequence variation (2020 update). *Nucleic Acids Research*, *48*(W1). https://doi.org/10.1093/NAR/GKAA420

Oseni, S. O., Naar, C., Pavlović, M., Asghar, W., Hartmann, J. X., Fields, G. B., Esiobu, N., & Kumi-Diaka, J. (2023). The Molecular Basis and Clinical Consequences of Chronic Inflammation in Prostatic Diseases: Prostatitis, Benign Prostatic Hyperplasia, and Prostate Cancer. In *Cancers* (Vol. 15, Issue 12). https://doi.org/10.3390/cancers15123110

Pandiyan, A., Murugesan, S., Ramadoss, K., & Mackraj, I. (2021). *A Systems Biological Approach to the Identification of Exosomal NTRK1 as a Potent Biomarker for Prolonged type II Diabetes-Induced Alzheimer's Disease* (Vol. 25, Issue 2). http://annalsofrscb.ro

Patel, A. R., & Klein, E. A. (2009). Risk factors for prostate cancer. *Nature Clinical Practice Urology*, *6*(2), 87–95. https://doi.org/10.1038/ncpuro1290

Pathak, N., Anu, R. I., Kulkarni, P., & Patel, A. (2024). A Survey on Unmet Need for Uniform Next-Generation Sequencing Reporting in India. *Indian Journal of Medical and Paediatric Oncology*, *45*(2). https://doi.org/10.1055/s-0043-1770936

Pearson-Stuttard, J., Buckley, J., Cicek, M., & Gregg, E. W. (2021). The Changing Nature of Mortality and Morbidity in Patients with Diabetes. In *Endocrinology and Metabolism Clinics of North America* (Vol. 50, Issue 3). https://doi.org/10.1016/j.ecl.2021.05.001

Peila, R., & Rohan, T. E. (2020). Diabetes, glycated hemoglobin, and risk of cancer in the UK Biobank Study. *Cancer Epidemiology Biomarkers and Prevention*, *29*(6), 1107–1119. https://doi.org/10.1158/1055-9965.EPI-19-1623

Peters, K. E., Davis, W. A., Ito, J., Winfield, K., Stoll, T., Bringans, S. D., Lipscombe, R. J., & Davis, T. M. E. (2017). Identification of novel circulating biomarkers predicting rapid decline in renal function in type-2-diabetes: The fremantle diabetes study phase II. *Diabetes Care*, *40*(11). https://doi.org/10.2337/dc17-0911

Phan, N. N., Wang, C. Y., Chen, C. F., Sun, Z., Lai, M. D., & Lin, Y. C. (2017). Voltage-gated calcium channels: Novel targets for cancer therapy. *Oncology Letters*, *14*(2). https://doi.org/10.3892/ol.2017.6457

Pidsley, R., Lam, D., Qu, W., Peters, T. J., Luu, P., Korbie, D., Stirzaker, C., Daly, R. J., Stricker, P., Kench, J. G., Horvath, L. G., & Clark, S. J. (2022). Comprehensive

methylome sequencing reveals prognostic epigenetic biomarkers for prostate cancer mortality. *Clinical and Translational Medicine*, *12*(10). https://doi.org/10.1002/ctm2.1030

Pierotti, M. A., & Greco, A. (2006). Oncogenic rearrangements of the NTRK1/NGF receptor. *Cancer Letters*, *232*(1). https://doi.org/10.1016/j.canlet.2005.07.043

Poloz, Y., & Stambolic, V. (2015). Obesity and cancer, a case for insulin signaling. In *Cell Death and Disease* (Vol. 6, Issue 12). Springer Nature. https://doi.org/10.1038/CDDIS.2015.381

Post, P. N., Kil, P. J. M., Hendrikx, A. J. M., Janssen-Heijnen, M. L. G., Crommelins, M. A., & Coebergh, J. W. W. (1999). Comorbidity in patients with prostate cancer and its relevance to treatment choice. *BJU International*, *84*(6). https://doi.org/10.1046/j.1464-410x.1999.00279.x

Prokopenko, D., Morgan, S. L., Mullin, K., Hofmann, O., Chapman, B., Kirchner, R., Amberkar, S., Wohlers, I., Lange, C., Hide, W., Bertram, L., & Tanzi, R. E. (2021). Whole-genome sequencing reveals new Alzheimer's disease–associated rare variants in loci related to synaptic function and neuronal development. *Alzheimer's and Dementia*, *17*(9). https://doi.org/10.1002/alz.12319

Radišauskas, R., Kuzmickiene, I., Milinavičiene, E., & Everatt, R. (2016). Hypertension, serum lipids and cancer risk: A review of epidemiological evidence. In *Medicina (Lithuania)* (Vol. 52, Issue 2). https://doi.org/10.1016/j.medici.2016.03.002

Ramesh, S., & Kosalram, K. (2023). The burden of non-communicable diseases: A scoping review focus on the context of India. In *Journal of Education and Health Promotion*. https://doi.org/10.4103/jehp.jehp_1113_22

Ramnarine, V. R., Kobelev, M., Gibb, E. A., Nouri, M., Lin, D., Wang, Y., Buttyan, R., Davicioni, E., Zoubeidi, A., & Collins, C. C. (2019). The evolution of long noncoding RNA acceptance in prostate cancer initiation, progression, and its clinical utility in disease management. In *European Urology* (Vol. 76, Issue 5). https://doi.org/10.1016/j.eururo.2019.07.040

Ravi, C., Sanjeevan, K., Thomas, A., & Pooleri, G. (2021). Development of an Indian nomogram for predicting extracapsular extension in prostate cancer. *Indian Journal of Urology*, *37*(1). https://doi.org/10.4103/iju.IJU_200_20

Rebbeck, T. R., Devesa, S. S., Chang, B.-L., Bunker, C. H., Cheng, I., Cooney, K., Eeles, R., Fernandez, P., Giri, V. N., Gueye, S. M., Haiman, C. A., Henderson, B. E., Heyns, C. F., Hu, J. J., Ingles, S. A., Isaacs, W., Jalloh, M., John, E. M., Kibel, A. S., … Zeigler-Johnson, C. M. (2013). Global Patterns of Prostate Cancer Incidence, Aggressiveness,

and Mortality in Men of African Descent. *Prostate Cancer*, *2013*. https://doi.org/10.1155/2013/560857

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., Wadi, L., Meyer, M., Wong, J., Xu, C., Merico, D., & Bader, G. D. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nature Protocols*, *14*(2). https://doi.org/10.1038/s41596-018-0103-9

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, *47*(D1). https://doi.org/10.1093/nar/gky1016

Rothfels, K., Milacic, M., Matthews, L., Haw, R., Sevilla, C., Gillespie, M., Stephan, R., Gong, C., Ragueneau, E., May, B., Shamovsky, V., Wright, A., Weiser, J., Beavers, D., Conley, P., Tiwari, K., Jassal, B., Griss, J., Senff-Ribeiro, A., … Stein, L. (2023). Using the Reactome Database. *Current Protocols*, *3*(4). https://doi.org/10.1002/cpz1.722

Rout, M., Kour, B., Vuree, S., Lulu, S. S., Medicherla, K. M., & Suravajhala, P. (2022). Diabetes mellitus susceptibility with varied diseased phenotypes and its comparison with phenome interactome networks. *World Journal of Clinical Cases*, *10*(18). https://doi.org/10.12998/wjcc.v10.i18.5957

Sakkiah, S., Wang, T., Zou, W., Wang, Y., Pan, B., Tong, W., & Hong, H. (2018). Endocrine disrupting chemicals mediated through binding androgen receptor are associated with diabetes mellitus. In *International Journal of Environmental Research and Public Health* (Vol. 15, Issue 1). https://doi.org/10.3390/ijerph15010025

Salciccia, S., Capriotti, A. L., Laganà, A., Fais, S., Logozzi, M., De Berardinis, E., Busetto, G. M., Di Pierro, G. B., Ricciuti, G. P., Del Giudice, F., Sciarra, A., Carroll, P. R., Cooperberg, M. R., Sciarra, B., & Maggi, M. (2021). Biomarkers in prostate cancer diagnosis: From current knowledge to the role of metabolomics and exosomes. In *International Journal of Molecular Sciences* (Vol. 22, Issue 9). https://doi.org/10.3390/ijms22094367

Sant'anna-Silva, A. C. B., Perez-Valencia, J. A., Sciacovelli, M., Lalou, C., Sarlak, S., Tronci, L., Nikitopoulou, E., Meszaros, A. T., Frezza, C., Rossignol, R., Gnaiger, E., & Klocker, H. (2021). Succinate anaplerosis has an onco-driving potential in prostate cancer cells. *Cancers*, *13*(7). https://doi.org/10.3390/cancers13071727

Sarkar, S., Gogoi, M., Mahato, M., Joshi, A. B., Baruah, A. J., Kodgire, P., & Boruah, P. (2022). Biosensors for detection of prostate cancer: a review. In *Biomedical Microdevices* (Vol. 24, Issue 4). https://doi.org/10.1007/s10544-022-00631-1

Sathishkumar, C., Prabu, P., Mohan, V., & Balasubramanyam, M. (2018). Linking a role of lncRNAs (long non-coding RNAs) with insulin resistance, accelerated senescence, and inflammation in patients with type-2-diabetes. *Human Genomics*, *12*(1). https://doi.org/10.1186/s40246-018-0173-3

Saxena, A., Mathur, N., Tiwari, P., & Mathur, S. K. (2021). Whole transcriptome RNA-seq reveals key regulatory factors involved in type-2-diabetes pathology in peripheral fat of Asian Indians. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-90148-z

Schnedl, W. J., Liebminger, A., Roller, R. E., Lipp, R. W., & Krejs, G. J. (2001). Hemoglobin variants and determination of glycated hemoglobin (HbA1c). In *Diabetes/Metabolism Research and Reviews* (Vol. 17, Issue 2, pp. 94–98). https://doi.org/10.1002/dmrr.186

Sciacca, L., Vigneri, R., Tumminia, A., Frasca, F., Squatrito, S., Frittitta, L., & Vigneri, P. (2013). Clinical and molecular mechanisms favoring cancer initiation and progression in diabetic patients. In *Nutrition, Metabolism and Cardiovascular Diseases* (Vol. 23, Issue 9). https://doi.org/10.1016/j.numecd.2013.05.006

Segura-Moreno, Y. Y., Sanabria-Salas, M. C., Varela, R., Mesa, J. A., & Serrano, M. L. (2021). Decoding the heterogeneous landscape in the development prostate cancer (Review). In *Oncology Letters* (Vol. 21, Issue 5). https://doi.org/10.3892/OL.2021.12637

Sekhoacha, M., Riet, K., Motloung, P., Gumenku, L., Adegoke, A., & Mashele, S. (2022). Prostate Cancer Review: Genetics, Diagnosis, Treatment Options, and Alternative Approaches. In *Molecules* (Vol. 27, Issue 17). https://doi.org/10.3390/molecules27175730

Shah, P., Wang, X., Yang, W., Eshghi, S. T., Sun, S., Hoti, N., Chen, L., Yang, S., Pasay, J., Rubin, A., & Zhang, H. (2015). Integrated proteomic and glycoproteomic analyses of prostate cancer cells reveal glycoprotein alteration in protein abundance and glycosylation. *Molecular and Cellular Proteomics*, *14*(10). https://doi.org/10.1074/mcp.M115.047928

Shahid, M., Kim, M., Lee, M. Y., Yeon, A., You, S., Kim, H. L., & Kim, J. (2019). Downregulation of CENPF Remodels Prostate Cancer Cells and Alters Cellular Metabolism. *Proteomics*, *19*(11). https://doi.org/10.1002/pmic.201900038

Shahid, R. K., Ahmed, S., Le, D., & Yadav, S. (2021). Diabetes and cancer: Risk, challenges, management and outcomes. In *Cancers* (Vol. 13, Issue 22). https://doi.org/10.3390/cancers13225735

Shaker, O. G., Abdelaleem, O. O., Mahmoud, R. H., Abdelghaffar, N. K., Ahmed, T. I., Said, O. M., & Zaki, O. M. (2019). Diagnostic and prognostic role of serum miR-20b, miR-

17-3p, HOTAIR, and MALAT1 in diabetic retinopathy. *IUBMB Life*, *71*(3). https://doi.org/10.1002/iub.1970

Shan, T., Liu, J., Xu, Z., & Wang, Y. (2019). Roles of phosphatase and tensin homolog in skeletal muscle. In *Journal of Cellular Physiology* (Vol. 234, Issue 4). https://doi.org/10.1002/jcp.26820

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11). https://doi.org/10.1101/gr.1239303

Sharifi, N., Gulley, J. L., & Dahut, W. L. (2005). Androgen deprivation therapy for prostate cancer. In *JAMA* (Vol. 294, Issue 2). https://doi.org/10.1001/jama.294.2.238

Sharma, D., Someshwar, S., Kour, B., Shukla, N., Khilwani, B., Vijay, M., Gupta, A., Ansari, A., Vuree, S., Kumar, A., Singh, S., Ravi, A., Mathur, P., Kumar Mishra, A., Ramaswamy, G., Suravajhala, R., Sadasukhi, N., Valadi, J., Mohan Medicherla, K., … Suravajhala, P. (n.d.). The CAPCI network: A CAncer Prostate Consortium of India for conducting next-generation genomic sequencing studies Equal contributing authors. *Cancer Health Disparities*, *7*, 1–13. https://doi.org/10.9777/chd.2023.1001

Sharma, N., Arora, S., Saurav, S., & Motiani, R. K. (2020). Pathophysiological Significance of Calcium Signaling at Mitochondria-Associated Endoplasmic Reticulum Membranes (MAMs). In *Current Opinion in Physiology* (Vol. 17). https://doi.org/10.1016/j.cophys.2020.08.012

Sharma, R., Abbastabar, H., Abdulah, D. M., Abidi, H., Abolhassani, H., Abrehdari-Tafreshi, Z., Absalan, A., Ali, H. A., Abu-Gharbieh, E., Acuna, J. M., Adib, N., Sakilah Adnani, Q. E., Aghaei, A., Ahmad, A., Ahmad, S., Ahmadi, A., Ahmadi, S., Ahmed, L. A., Ajami, M., … Zoladl, M. (2024). Temporal patterns of cancer burden in Asia, 1990–2019: a systematic examination for the Global Burden of Disease 2019 study. *The Lancet Regional Health - Southeast Asia*, *21*. https://doi.org/10.1016/j.lansea.2023.100333

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). DbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, *29*(1). https://doi.org/10.1093/nar/29.1.308

Shukla, N., Kour, B., Sharma, D., Vijayvargiya, M., Sadasukhi, T. C., Medicherla, K. M., Malik, B., Bissa, B., Vuree, S., Lohiya, N. K., & Suravajhala, P. (2023). Towards Understanding the Key Signature Pathways Associated from Differentially Expressed Gene Analysis in an Indian Prostate Cancer Cohort. *Diseases*, *11*(2).

https://doi.org/10.3390/diseases11020072

Shukla, S., Cyrta, J., Murphy, D. A., Walczak, E. G., Ran, L., Agrawal, P., Xie, Y., Chen, Y., Wang, S., Zhan, Y., Li, D., Wong, E. W. P., Sboner, A., Beltran, H., Mosquera, J. M., Sher, J., Cao, Z., Wongvipat, J., Koche, R. P., … Chen, Y. (2017). Aberrant Activation of a Gastrointestinal Transcriptional Circuit in Prostate Cancer Mediates Castration Resistance. *Cancer Cell*, *32*(6). https://doi.org/10.1016/j.ccell.2017.10.008

Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, *40*(W1). https://doi.org/10.1093/nar/gks539

Skopelitou, D., Miao, B., Srivastava, A., Kumar, A., Kuswick, M., Dymerska, D., Paramasivam, N., Schlesner, M., Lubinski, J., Hemminki, K., Försti, A., & Bandapalli, O. R. (2021). Whole exome sequencing identifies apcdd1 and hdac5 genes as potentially cancer predisposing in familial colorectal cancer. *International Journal of Molecular Sciences*, *22*(4). https://doi.org/10.3390/ijms22041837

Smith, R. D., & Mallath, M. K. (2019). History of the growing burden of cancer in India: From antiquity to the 21st century. In *Journal of Global Oncology* (Vol. 2019, Issue 5). https://doi.org/10.1200/JGO.19.00048

Smith, S. L., Pitt, A. R., & Spickett, C. M. (2021). Approaches to Investigating the Protein Interactome of PTEN. In *Journal of Proteome Research* (Vol. 20, Issue 1). https://doi.org/10.1021/acs.jproteome.0c00570

Sottnik, J. L., & Theodorescu, D. (2016). CD44: A metastasis driver and therapeutic target. In *Oncoscience* (Vol. 3, Issues 11–12). https://doi.org/10.18632/oncoscience.335

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. In *Nature Reviews Molecular Cell Biology* (Vol. 22, Issue 2). https://doi.org/10.1038/s41580-020-00315-9

Stein, M. S., Kalff, V., Williams, S. G., Murphy, D. G., Colman, P. G., & Hofman, M. S. (2023). The GLP-1 receptor is expressed in vivo by human metastatic prostate cancer. *Endocrine Oncology*, *4*(1). https://doi.org/10.1530/eo-23-0015

Suh, S., & Kim, K. W. (2011). Diabetes and cancer: Is diabetes causally related to cancer? In *Diabetes and Metabolism Journal* (Vol. 35, Issue 3). https://doi.org/10.4093/dmj.2011.35.3.193

Sun, H., Wu, X., Wu, F., Li, Y., Yu, Z., Chen, X., Chen, Y., & Yang, W. (2015). Associations of genetic variants in the PSCA, MUC1 and PLCE1 genes with stomach cancer susceptibility in a Chinese population. *PLoS ONE*, *10*(2). https://doi.org/10.1371/journal.pone.0117576

Sun, X., Kour, B. D., Wang, X., Hu, X., Wang, Y., & Vuree, S. (2021). Insights into association between urolithiasis and prostate cancer. *Journal of Men's Health*, *17*(4). https://doi.org/10.31083/jomh.2021.070

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3). https://doi.org/10.3322/caac.21660

Suravajhala, P., Kogelman, L. J. A., Mazzoni, G., & Kadarmideen, H. N. (2015). Potential role of lncRNA cyp2c91-protein interactions on diseases of the immune system. *Frontiers in Genetics*, *6*(JUL). https://doi.org/10.3389/fgene.2015.00255

Szablewski, L. (2024). Insulin Resistance: The Increased Risk of Cancers. In *Current Oncology* (Vol. 31, Issue 2). https://doi.org/10.3390/curroncol31020075

Taghizadeh, N., Boezen, H. M., Schouten, J. P., Schröder, C. P., De Vries, E. G. E., & Vonk, J. M. (2015). BMI and lifetime changes in BMI and cancer mortality risk. *PLoS ONE*, *10*(4). https://doi.org/10.1371/journal.pone.0125261

Taguchi, K., Yasui, T., Milliner, D. S., Hoppe, B., & Chi, T. (2017). Genetic Risk Factors for Idiopathic Urolithiasis: A Systematic Review of the Literature and Causal Network Analysis. In *European Urology Focus* (Vol. 3, Issue 1). https://doi.org/10.1016/j.euf.2017.04.010

Tamguney, T., & Stokoe, D. (2007). New insights into PTEN. In *Journal of Cell Science* (Vol. 120, Issue 23). https://doi.org/10.1242/jcs.015230

Tang, T., Tan, X., Wang, Z., Wang, S., Wang, Y., Xu, J., Wei, X., Zhang, D., Liu, Q., & Jiang, J. (2022). Germline Mutations in Patients With Early-Onset Prostate Cancer. *Frontiers in Oncology*, *12*. https://doi.org/10.3389/fonc.2022.826778

Tang, X., Hu, X., Yang, X., Fan, Y., Li, Y., Hu, W., Liao, Y., Zheng, M. cai, Peng, W., & Gao, L. (2016). Predicting diabetes mellitus genes via protein-protein interaction and protein subcellular localization information. *BMC Genomics*. https://doi.org/10.1186/s12864-016-2795-y

Tang, Z., Li, C., Kang, B., Gao, G., Li, C., & Zhang, Z. (2017). GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research*, *45*(W1). https://doi.org/10.1093/nar/gkx247

Testa, U., Castelli, G., & Pelosi, E. (2019). Cellular and Molecular Mechanisms Underlying Prostate Cancer Development: Therapeutic Implications. *Medicines*, *6*(3). https://doi.org/10.3390/medicines6030082

Tindale, L. C. (2018). *Healthy at 100: Genetic factors and phenotypes associated with*

healthy aging and longevity.

Tiruye, T., O'Callaghan, M., FitzGerald, L. M., Moretti, K., Jay, A., Higgs, B., Kichenadasse, G., Caughey, G., Roder, D., & Beckmann, K. (2024). Medication-based Comorbidity Measures and Prostate Cancer Treatment Selection. *Clinical Genitourinary Cancer*, *22*(2). https://doi.org/10.1016/j.clgc.2024.01.018

Tiwari, P., Saxena, A., Gupta, N., Medicherla, K. M., Suravajhala, P., & Mathur, S. K. (2019). Systems genomics of thigh adipose tissue from asian indian type-2 diabetics revealed distinct protein interaction hubs. *Frontiers in Genetics*. https://doi.org/10.3389/fgene.2018.00679

Twig, G., Afek, A., Shamiss, A., Derazne, E., Tzur, D., Gordon, B., & Tirosh, A. (2013). White blood cells count and incidence of type-2-diabetes in young men. *Diabetes Care*, *36*(2), 276–282. https://doi.org/10.2337/dc11-2298

Tyagi, S., Tyagi, N., Choudhury, A., Gupta, G., Zahra, M. M. A., & Rahin, S. A. (2022). Identification and Classification of Prostate Cancer Identification and Classification Based on Improved Convolution Neural Network. In *BioMed Research International* (Vol. 2022). https://doi.org/10.1155/2022/9112587

Ulgen, E., Ozisik, O., & Sezerman, O. U. (2018). pathfindR: An R Package for Pathway Enrichment Analysis Utilizing Active Subnetworks. *BioRxiv*.

Unnikrishnan, Rahul, Senan, S., Vijayakumar, P., & Paul, G. (2021). Relevance of prostate-specific antigen screening in elderly individuals: A cross-sectional analysis from a tertiary care referral teaching hospital in India. *Hamdan Medical Journal*, *14*(2). https://doi.org/10.4103/hmj.hmj_54_20

Unnikrishnan, Ranjit, Rani, C. S. S., Anjana, R. M., Uthra, S. C., Vidya, J., Sankari, G. U., Venkatesan, U., Rani, S. J., & Mohan, V. (2016). Postpartum development of type 1 diabetes in Asian Indian women with gestational diabetes. *Indian Journal of Endocrinology and Metabolism*. https://doi.org/10.4103/2230-8210.190562

Valkovicova, T., Skopkova, M., Stanik, J., & Gasperikova, D. (2019). Novel insights into genetics and clinics of the HNF1A-MODY. *Endocrine Regulations*, *53*(2). https://doi.org/10.2478/enr-2019-0013

van Nostrand, J. L., Hellberg, K., Luo, E. C., van Nostrand, E. L., Dayn, A., Yu, J., Shokhirev, M. N., Dayn, Y., Yeo, G. W., & Shaw, R. J. (2020). AMPK regulation of Raptor and TSC2 mediate metformin effects on transcriptional control of anabolism and inflammation. *Genes and Development*, *34*(19–20). https://doi.org/10.1101/gad.339895.120

Vancura, P., Oebel, L., Spohn, S., Frederiksen, U., Schäfer, K., Sticht, C., & Spessert, R.

(2021). Evidence for a dysfunction and disease-promoting role of the circadian clock in the diabetic retina. *Experimental Eye Research*, *211*. https://doi.org/10.1016/j.exer.2021.108751

Vanniya. S, P., Chandru, J., Jeffrey, J. M., Rabinowitz, T., Brownstein, Z., Krishnamoorthy, M., Avraham, K. B., Cheng, L., Shomron, N., & Srisailapathy, C. R. S. (2022). PNPT1, MYO15A, PTPRQ, and SLC12A2-associated genetic and phenotypic heterogeneity among hearing impaired assortative mating families in Southern India. *Annals of Human Genetics*, *86*(1). https://doi.org/10.1111/ahg.12442

Velonas, V. M., Woo, H. H., dos Remedios, C. G., & Assinder, S. J. (2013). Current status of biomarkers for prostate cancer. In *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms140611034

Verboom, K., Van Loocke, W., Volders, P. J., Decaesteker, B., Avila Cobos, F., Bornschein, S., De Bock, C. E., Kalender Atak, Z., Clappier, E., Aerts, S., Cools, J., Soulier, J., Taghon, T., Van Vlierberghe, P., Vandesompele, J., Speleman, F., & Durinck, K. (2018). A comprehensive inventory of TLX1 controlled long non-coding RNAs in T-cell acute lymphoblastic leukemia through polyA+ and total RNA sequencing. In *Haematologica* (Vol. 103, Issue 12). https://doi.org/10.3324/haematol.2018.190587

Vietri, M. T., D'elia, G., Caliendo, G., Resse, M., Casamassimi, A., Passariello, L., Albanese, L., Cioffi, M., & Molinari, A. M. (2021). Hereditary prostate cancer: Genes related, target therapy and prevention. In *International Journal of Molecular Sciences* (Vol. 22, Issue 7). https://doi.org/10.3390/ijms22073753

Vigneri, P., Frasca, F., Sciacca, L., Pandini, G., & Vigneri, R. (2009). Diabetes and cancer. In *Endocrine-Related Cancer* (Vol. 16, Issue 4, pp. 1103–1123). https://doi.org/10.1677/ERC-09-0087

von Toerne, C., Huth, C., de las Heras Gala, T., Kronenberg, F., Herder, C., Koenig, W., Meisinger, C., Rathmann, W., Waldenberger, M., Roden, M., Peters, A., Thorand, B., & Hauck, S. M. (2016). MASP1, THBS1, GPLD1 and ApoA-IV are novel biomarkers associated with prediabetes: the KORA F4 study. *Diabetologia*, *59*(9). https://doi.org/10.1007/s00125-016-4024-2

Vozarova, B., Weyer, C., Lindsay, R. S., Pratley, R. E., Bogardus, C., & Tataranni, P. A. (n.d.). *High White Blood Cell Count Is Associated With a Worsening of Insulin Sensitivity and Predicts the Development of Type-2-diabetes*.

Wang, G., Zhao, D., Spring, D. J., & Depinho, R. A. (2018). Genetics and biology of prostate cancer. In *Genes and Development* (Vol. 32, Issues 17–18). https://doi.org/10.1101/gad.315739.118

Wang, L., Zheng, Z., Feng, X., Zang, X., Ding, W., Wu, F., & Zhao, Q. (2019). CircRNA/lncRNA-miRNA-mRNA Network in Oxidized, Low-Density, Lipoprotein-Induced Foam Cells. *DNA and Cell Biology*, *38*(12). https://doi.org/10.1089/dna.2019.4865

Wang, M., Yang, Y., & Liao, Z. (2020). Diabetes and cancer: Epidemiological and biological links. *World Journal of Diabetes*, *11*(6). https://doi.org/10.4239/wjd.v11.i6.227

Wang, H. L. V., & Chekanova, J. A. (2019). An Overview of Methodologies in Studying lncRNAs in the High-Throughput Era: When Acronyms ATTACK! In *Methods in Molecular Biology* (Vol. 1933). https://doi.org/10.1007/978-1-4939-9045-0_1

Wei, L., Tian, Y., Chen, Y., Wei, Q., Chen, F., Cao, B., Wu, Y., Zhao, B., Chen, X., Xie, C., Xi, C., Yu, X., Wang, J., Lv, X., Du, J., Wang, Y., Shen, L., Wang, X., Shen, B., … Wang, K. (2019). Identification of TYW3/CRYZ and FGD4 as susceptibility genes for amyotrophic lateral sclerosis. *Neurology: Genetics*, *5*(6). https://doi.org/10.1212/NXG.0000000000000375

Wendt, F. R., Pathak, G. A., Tylee, D. S., Goswami, A., & Polimanti, R. (2020). Heterogeneity and Polygenicity in Psychiatric Disorders: A Genome-Wide Perspective. In *Chronic Stress* (Vol. 4). https://doi.org/10.1177/2470547020924844

Wilcox, G. (2005). Insulin and Insulin Resistance. In *Clin Biochem Rev* (Vol. 26).

Wilkinson, E. J., Woodworth, A. M., Parker, M., Phillips, J. L., Malley, R. C., Dickinson, J. L., & Holloway, A. F. (2020). Epigenetic regulation of the ITGB4 gene in prostate cancer. *Experimental Cell Research*, *392*(2). https://doi.org/10.1016/j.yexcr.2020.112055

Wilson, R. L., Taaffe, D. R., Newton, R. U., Hart, N. H., Lyons-Wall, P., & Galvão, D. A. (2022). Obesity and prostate cancer: A narrative review. In *Critical Reviews in Oncology/Hematology* (Vol. 169). https://doi.org/10.1016/j.critrevonc.2021.103543

Wise, H. M., Hermida, M. A., & Leslie, N. R. (2017). Prostate cancer, PI3K, PTEN and prognosis. *Clinical Science*, *131*(3). https://doi.org/10.1042/CS20160026

Witts, W. K., Amu, H., Dowou, R. K., Kwafo, F. O., & Bain, L. E. (2024). Health-related quality of life among adults living with chronic non-communicable diseases in the Ho Municipality of Ghana: a health facility-based cross-sectional study. *BMC Public Health*. https://doi.org/10.1186/s12889-024-18143-3

Wright, J. L., Plymate, S. R., Porter, M. P., Gore, J. L., Lin, D. W., Hu, E., & Zeliadt, S. B. (2013). Hyperglycemia and prostate cancer recurrence in men treated for localized prostate cancer. *Prostate Cancer and Prostatic Diseases*, *16*(2). https://doi.org/10.1038/pcan.2013.5

Wu, S. Y., Fang, S. C., Hwang, O. R., Shih, H. J., & Shao, Y. H. J. (2020). Influence of baseline cardiovascular comorbidities on mortality after androgen deprivation therapy for metastatic prostate cancer. *Cancers*, *12*(1). https://doi.org/10.3390/cancers12010189

Xu, X., Kharazmi, E., Tian, Y., Mukama, T., Sundquist, K., Sundquist, J., Brenner, H., & Fallah, M. (2021). Risk of prostate cancer in relatives of prostate cancer patients in Sweden: A nationwide cohort study. *PLoS Medicine*, *18*(6). https://doi.org/10.1371/journal.pmed.1003616

Yan, R., Wang, Y., Shi, M., Xiao, Y., Liu, L., Liu, L., & Guo, B. (2019). Regulation of PTEN/AKT/FAK pathways by PPARγ impacts on fibrosis in diabetic nephropathy. *Journal of Cellular Biochemistry*, *120*(5). https://doi.org/10.1002/jcb.27937

Yang, G., Lu, X., & Yuan, L. (2014). LncRNA: A link between RNA and cancer. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* (Vol. 1839, Issue 11). https://doi.org/10.1016/j.bbagrm.2014.08.012

Yang, H., Robinson, P. N., & Wang, K. (2015). Phenolyzer: Phenotype-based prioritization of candidate genes for human diseases. *Nature Methods*, *12*(9). https://doi.org/10.1038/nmeth.3484

Yang, S., Sun, Z., Zhou, Q., Wang, W., Wang, G., Song, J., Li, Z., Zhang, Z., Chang, Y., Xia, K., Liu, J., & Yuan, W. (2018). microRNAs, long noncoding RNAs, and circular RNAs: Potential tumor biomarkers and targets for colorectal cancer. In *Cancer Management and Research* (Vol. 10). https://doi.org/10.2147/CMAR.S166308

Yang, Ying, Jia, J., Sun, Z., Liu, C., Li, Z., Xiao, Y., Yu, J., Du, F., Shi, Y., Sun, J., Shui, J., & Zhang, X. (2021). Polymorphism of FGD4 and myelosuppression in patients with esophageal squamous cell carcinoma. *Future Oncology*, *17*(18). https://doi.org/10.2217/fon-2020-1191

Yang, Yongyong, Liu, K. Y., Liu, Q., & Cao, Q. (2021). Androgen Receptor-Related Non-coding RNAs in Prostate Cancer. In *Frontiers in Cell and Developmental Biology* (Vol. 9). https://doi.org/10.3389/fcell.2021.660853

Yao, Q., & Chen, T. (2020). LINC01128 regulates the development of osteosarcoma by sponging miR-299-3p to mediate MMP2 expression and activating Wnt/β-catenin signalling pathway. *Journal of Cellular and Molecular Medicine*, *24*(24). https://doi.org/10.1111/jcmm.16046

Yee, S. W., Chen, L., & Giacomini, K. M. (2012). The role of ATM in response to metformin treatment and activation of AMPK. In *Nature Genetics* (Vol. 44, Issue 4). https://doi.org/10.1038/ng.2236

Yeh, Y. A., Yang, S., Constantinescu, M., Chaudoir, C., Tanner, A., Henry, M., Anderson, S.,

Saldivar, J.-S., Serkin, F., Fazili, T., Lurie, A. A., & Yu, X. (2019). Prostatic adenocarcinoma with novel NTRK3 gene fusion: a case report. *American Journal of Clinical and Experimental Urology*, *7*(5).

Yin, S., Liu, L., & Gan, W. (2021). The roles of post-translational modifications on MTOR signaling. In *International Journal of Molecular Sciences* (Vol. 22, Issue 4). https://doi.org/10.3390/ijms22041784

Yin, X., Lin, H., Lin, L., Miao, L., He, J., & Zhuo, Z. (2022). LncRNAs and CircRNAs in cancer. In *MedComm* (Vol. 3, Issue 2). https://doi.org/10.1002/mco2.141

Yu, X., Liu, R., Gao, W., Wang, X., & Zhang, Y. (2023). Single-cell omics traces the heterogeneity of prostate cancer cells and the tumor microenvironment. In *Cellular and Molecular Biology Letters* (Vol. 28, Issue 1). https://doi.org/10.1186/s11658-023-00450-z

Yu, Y. P., Liu, S., Ren, B. G., Nelson, J., Jarrard, D., Brooks, J. D., Michalopoulos, G., Tseng, G., & Luo, J. H. (2023). Fusion Gene Detection in Prostate Cancer Samples Enhances the Prediction of Prostate Cancer Clinical Outcomes from Radical Prostatectomy through Machine Learning in a Multi-Institutional Analysis. *American Journal of Pathology*, *193*(4). https://doi.org/10.1016/j.ajpath.2022.12.013

Yun, S. J., Kim, S. K., Kim, J., Cha, E. J., Kim, J. S., Kim, S. J., Ha, Y. S., Kim, Y. H., Jeong, P., Kang, H. W., Kim, J. H., Park, J. L., Choi, Y. K., Moon, S. K., Choi, Y. H., Kim, S. Y., & Kim, W. J. (2017). Transcriptomic features of primary prostate cancer and their prognostic relevance to castration-resistant prostate cancer. *Oncotarget*, *8*(70). https://doi.org/10.18632/oncotarget.22296

Zaidi, S., Gandhi, J., Joshi, G., Smith, N. L., & Khan, S. A. (2019). The anticancer potential of metformin on prostate cancer. In *Prostate Cancer and Prostatic Diseases* (Vol. 22, Issue 3). https://doi.org/10.1038/s41391-018-0085-2

Zarzour, A., Morgans, A. K., Palmbos, P., Jovanovic, B., Tomlins, S. A., VanderWeele, D. J., Yang, X. J., Schaeffer, E. M., Sharifi, R., Meeks, J. J., Abdulkadir, S., & Hussain, M. H. A. (2020). Evaluating the clinical, environmental, genetic, and genomic profile of men with early-onset aggressive prostate cancer (PCa). *Journal of Clinical Oncology*, *38*(15_suppl). https://doi.org/10.1200/jco.2020.38.15_suppl.e17517

Zhang, A., Zhang, J., Kaipainen, A., Lucas, J. M., & Yang, H. (2016). Long non-coding RNA: A newly deciphered “code” in prostate cancer. In *Cancer Letters* (Vol. 375, Issue 2). https://doi.org/10.1016/j.canlet.2016.03.003

Zhang, D., Qin, H., Leng, Y., Li, X., Zhang, L., Bai, D., Meng, Y., & Wang, J. (2018). LncRNA MEG3 overexpression inhibits the development of diabetic retinopathy by

regulating TGF-β1 and VEGF. *Experimental and Therapeutic Medicine*, *16*(3). https://doi.org/10.3892/etm.2018.6451

Zhang, K., Han, Y., Hu, Z., Zhang, Z., Shao, S., Yao, Q., Zheng, L., Wang, J., Han, X., Zhang, Y., Chen, T., Yao, Z., Han, T., & Hong, W. (2019). SCARNA10, a nuclear-retained long non-coding RNA, promotes liver fibrosis and serves as a potential biomarker. *Theranostics*, *9*(12). https://doi.org/10.7150/thno.32935

Zhang, W., Wang, T., Wang, Y., Zhu, F., Shi, H., Zhang, J., Wang, Z., Qu, M., Zhang, H., Wang, T., Qian, Y., Yang, J., Gao, X., & Li, J. (2022). Intratumor heterogeneity and clonal evolution revealed in castration-resistant prostate cancer by longitudinal genomic analysis. *Translational Oncology*, *16*. https://doi.org/10.1016/j.tranon.2021.101311

Zhang, Y. H., Zeng, T., Chen, L., Huang, T., & Cai, Y. D. (2021). Determining protein–protein functional associations by functional rules based on gene ontology and KEGG pathway. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1869*(6). https://doi.org/10.1016/j.bbapap.2021.140621

Zhao, G. Y., Ning, Z. F., & Wang, R. (2021). LncRNA SNHG19 Promotes the Development of Non-Small Cell Lung Cancer via Mediating miR-137/E2F7 Axis. In *Frontiers in Oncology* (Vol. 11). https://doi.org/10.3389/fonc.2021.630241

Zhao, W., Liang, J., Chen, Z., Diao, Y., & Miao, G. (2020). Combined analysis of circRNA and mRNA profiles and interactions in patients with Diabetic Foot and Diabetes Mellitus. *International Wound Journal*, *17*(5). https://doi.org/10.1111/iwj.13420

Zhou, C. K., Check, D. P., Lortet-Tieulent, J., Laversanne, M., Jemal, A., Ferlay, J., Bray, F., Cook, M. B., & Devesa, S. S. (2016). Prostate cancer incidence in 43 populations worldwide: An analysis of time trends overall and by age group. *International Journal of Cancer*, *138*(6). https://doi.org/10.1002/ijc.29894

Zhou, X., Ao, X., Jia, Z., Li, Y., Kuang, S., Du, C., Zhang, J., Wang, J., & Liu, Y. (2022). Non-coding RNA in cancer drug resistance: Underlying mechanisms and clinical applications. In *Frontiers in Oncology* (Vol. 12). https://doi.org/10.3389/fonc.2022.951864

Zhu, Xiaolei, & Wu, S. (2019). Increased Risk of Hypertension with Enzalutamide in Prostate Cancer: A Meta-Analysis. *Cancer Investigation*, *37*(9). https://doi.org/10.1080/07357907.2019.1670203

Zhu, Xiaolin, Petrovski, S., Xie, P., Ruzzo, E. K., Lu, Y. F., McSweeney, K. M., Ben-Zeev, B., Nissenkorn, A., Anikster, Y., Oz-Levi, D., Dhindsa, R. S., Hitomi, Y., Schoch, K., Spillmann, R. C., Heimer, G., Marek-Yagel, D., Tzadok, M., Han, Y., Worley, G., … Goldstein, D. B. (2015). Whole-exome sequencing in undiagnosed genetic diseases:

Interpreting 119 trios. *Genetics in Medicine*, *17*(10). https://doi.org/10.1038/gim.2014.191

Zhu, Y., Zhang, F., Zhang, S., & Yi, M. (2023). Predicting latent lncRNA and cancer metastatic event associations via variational graph auto-encoder. *Methods*, *211*. https://doi.org/10.1016/j.ymeth.2023.01.006

Zi, F., Zi, H., Li, Y., He, J., Shi, Q., & Cai, Z. (2018). Metformin and cancer: An existing drug for cancer prevention and therapy (review). In *Oncology Letters* (Vol. 15, Issue 1). https://doi.org/10.3892/ol.2017.7412

URL

1. Global Cancer Observatory. Available from:https://gco.iarc.who.int. [last accessed on 21 April 2024].
2. Batch Entrez. Available from: https://www.ncbi.nlm.nih.gov/sites/batchentrez [last accessed on 12 April 2024].
3. National Center for Biotechnology Information. Available from:https://www.ncbi.nlm.nih.gov/snp/. [last accessed on 12 April 2024].
4. ClinVar. Available from: https://www.ncbi.nlm.nih.gov/clinvar/. [last accessed on 12 April 2024].
5. KEGG Mapper. Available from:https://www.genome.jp/kegg/mapper/. [Last accessed on 12 April 2024].
6. Reactome. Available from: https://reactome.org/. [Last accessed on 12 April 2024].
7. PantherDB. Available from: https://pantherdb.org/ [Last accessed on 12 April 2024].

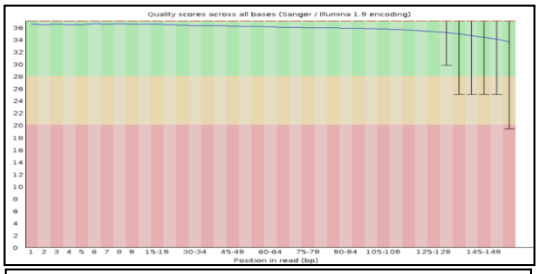
APPENDICES:

Supplementary Figures of FastQC files of Clinical samples:

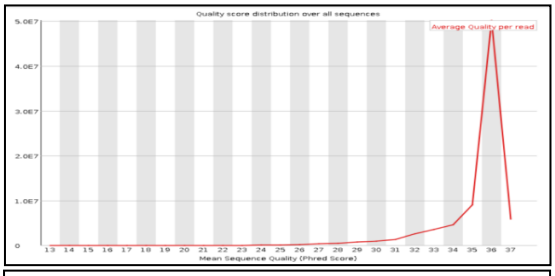
S1_R1.fastqc

Measure	Value
Filename	S1_R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	81209072
Sequences flagged as poor quality	0
Sequence length	159
%GC	51

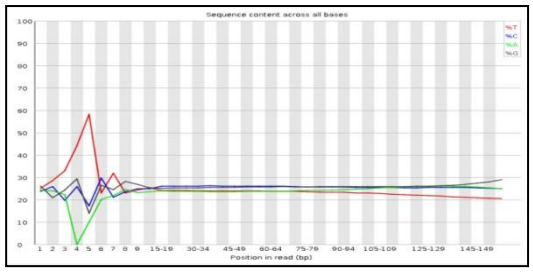
Basic Statistic



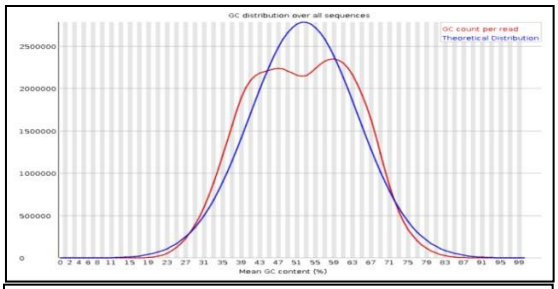
Per base sequence quality



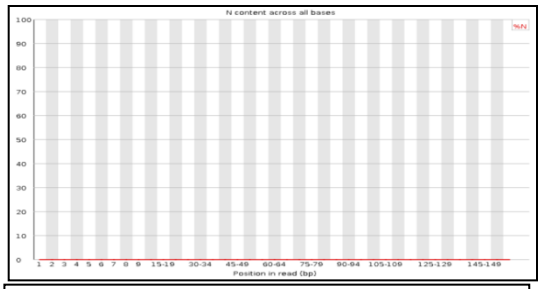
Per sequence quality score



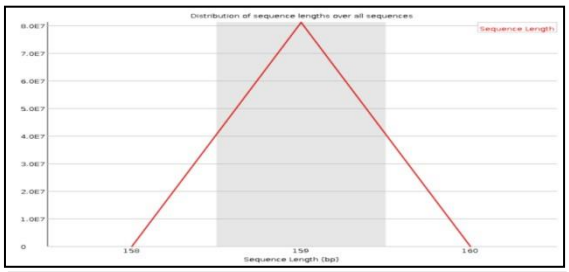
Per base sequence content



Per Sequence GC content



Per base N content



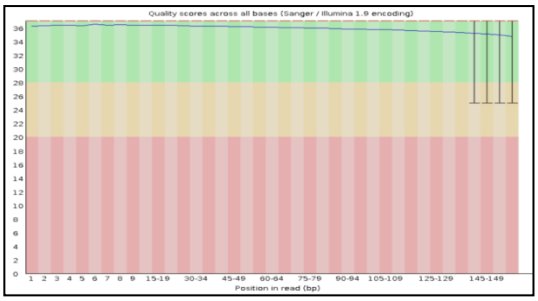
Sequence length Distribution

Supplementary Figure 1: Fastq reports for quality values of the bases in forward reads of sample 1 where red line in the centre of graphs depicts the median value, average quality is represented by blue line and green line shows the very high quality. Y-axis shows quality scores and X-axis shows read position

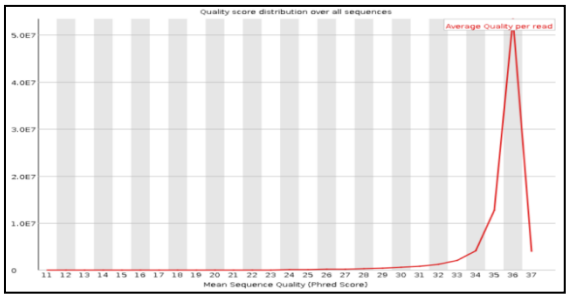
S1_R2.fastqc

Measure	Value
Filename	S1_R2.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	81209072
Sequences flagged as poor quality	0
Sequence length	159
%GC	51

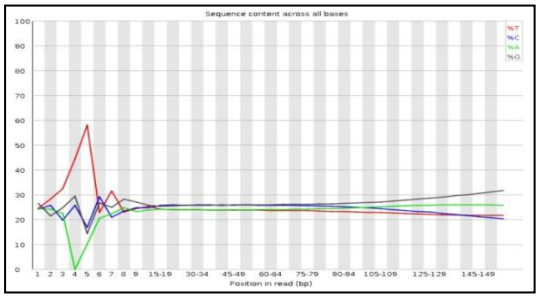
Basic Statistic



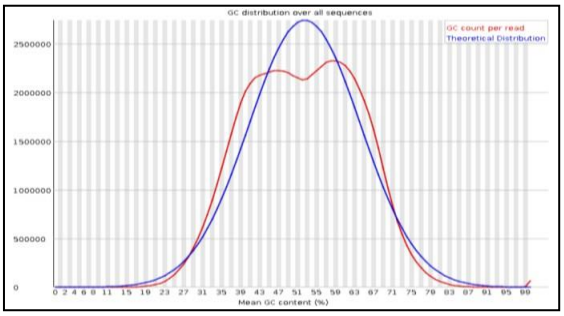
Per base sequence quality



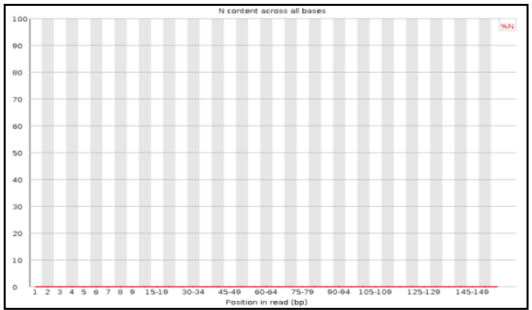
Per sequence quality score



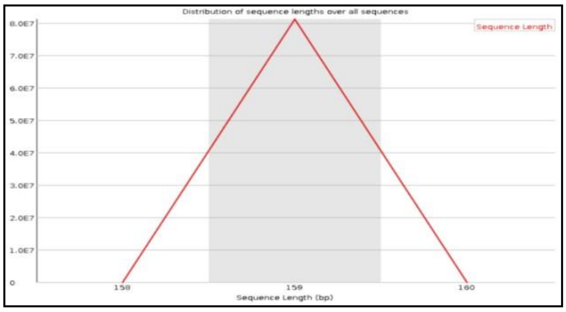
Per base sequence content

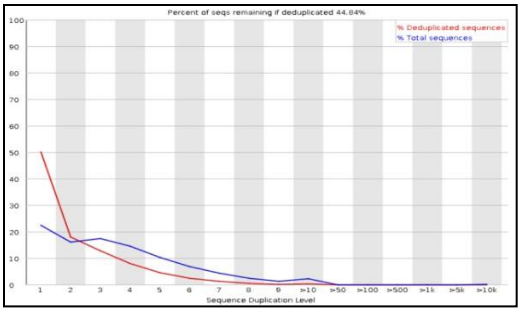


Per Sequence GC content



Per base N content



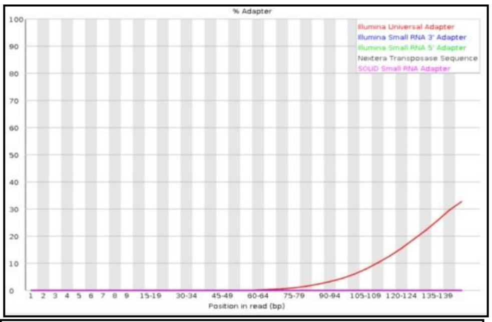


Sequence length Distribution

Sequence Duplication Levels

Sequence	Count	Percentage	Possible Source
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	144311	0.17770305268357212	No Hit

Overrepresented sequences

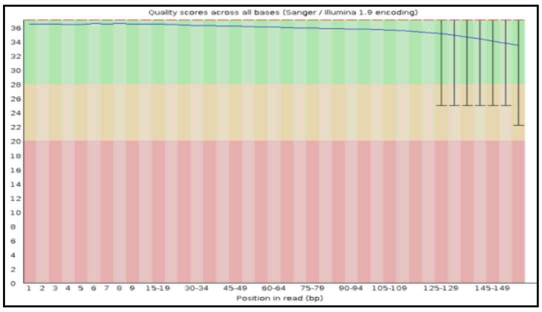


Adapter Content

Supplementary Figure 2: Fastq graph plot reports for quality values of the bases in reverse read of sample 1

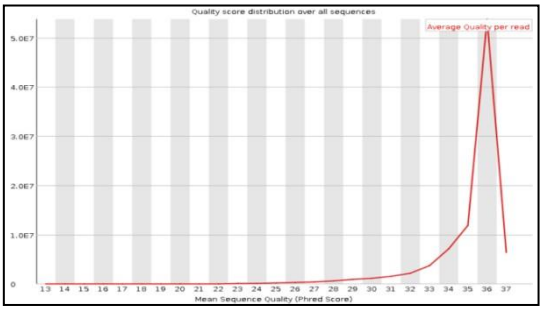
S2_R1_fatsqc

Measure	Value
Filename	S2_R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	91315062
Sequences flagged as poor quality	0
Sequence length	159
%GC	51

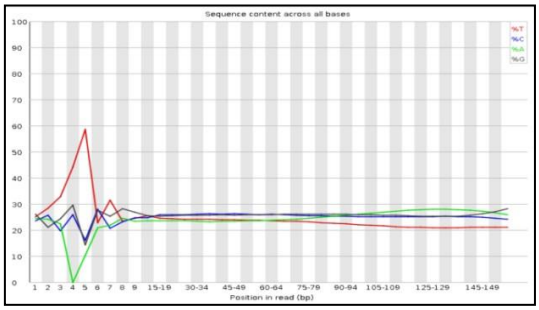


Basic Statistic

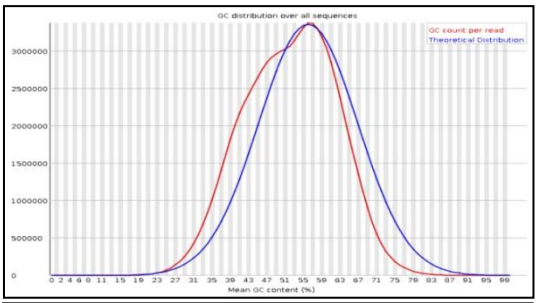
Per base sequence quality



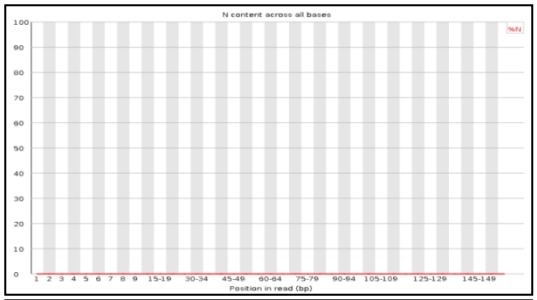
Per sequence quality score



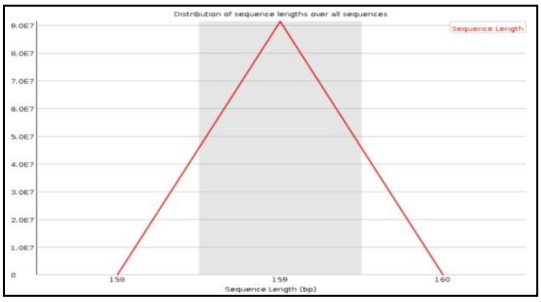
Per base sequence content



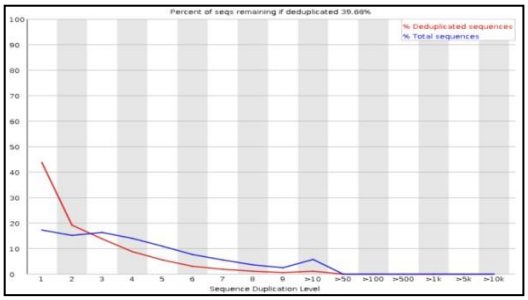
Per Sequence GC content



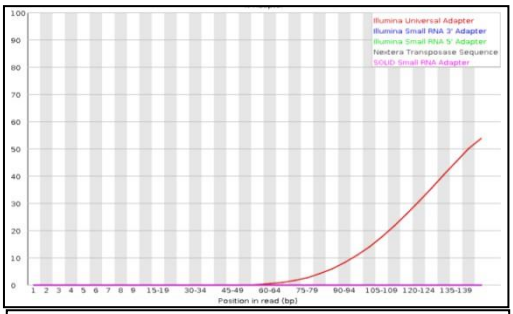
Per base N content



Sequence length Distribution



Sequence Duplication Level



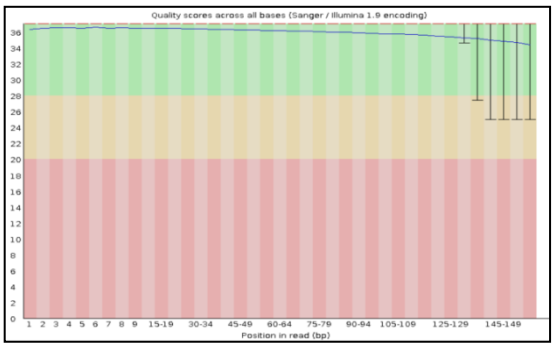
Adapter Content

Supplementary Figure 3: Fastq graph plot reports for quality values of the bases in forward read of sample-2

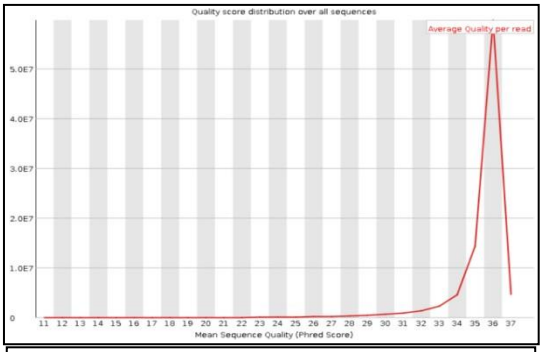
S2_R2_Qc

Measure	Value
Filename	S2_R2.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	91315062
Sequences flagged as poor quality	0
Sequence length	159
%GC	51

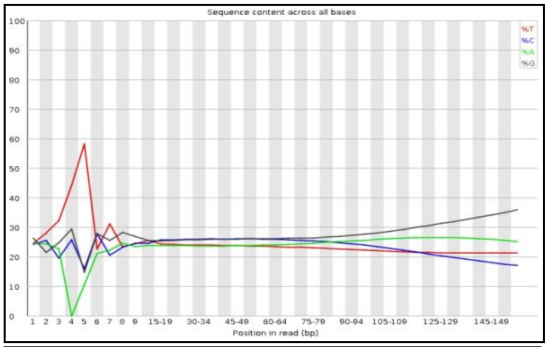
Basic Statistic



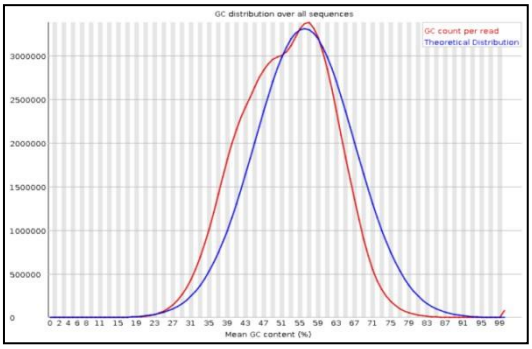
Per base sequence quality



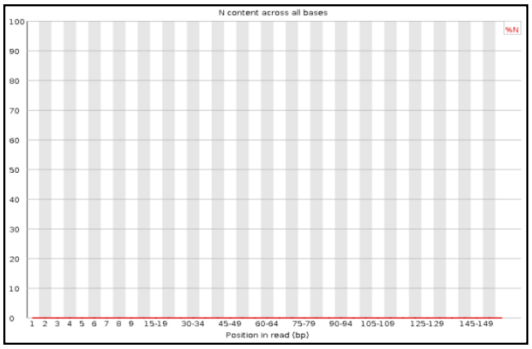
Per sequence quality score



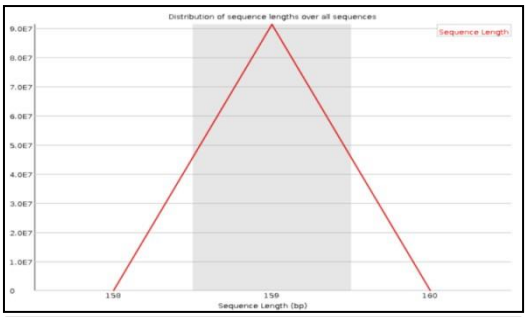
Per base sequence content



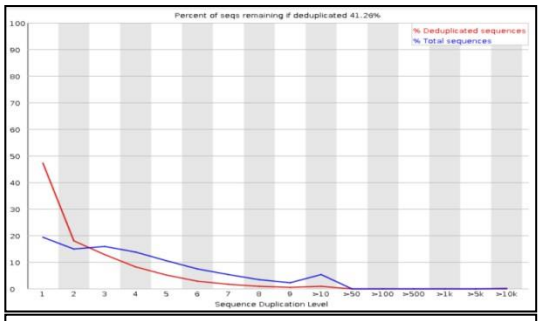
Per Sequence GC content



Per base N content



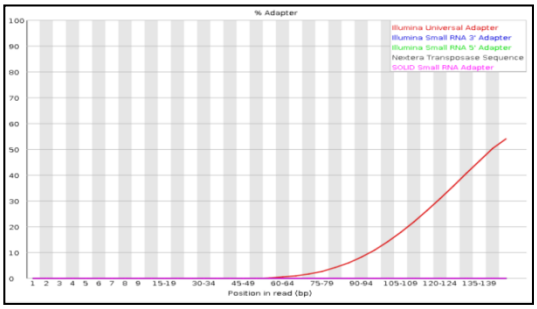
Sequence length Distribution



Sequence Duplication Level

Sequence	Count	Percentage	Possible Source
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	178874	0.19588663259079864	No Hit

Overrepresented sequences



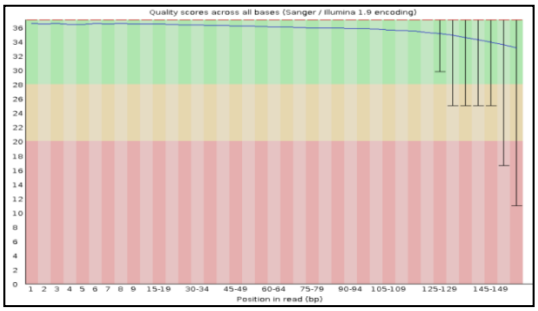
Adapter Content

Supplementary Figure 4: Fastq graph plot reports for quality values of the bases in reverse read of sample-2

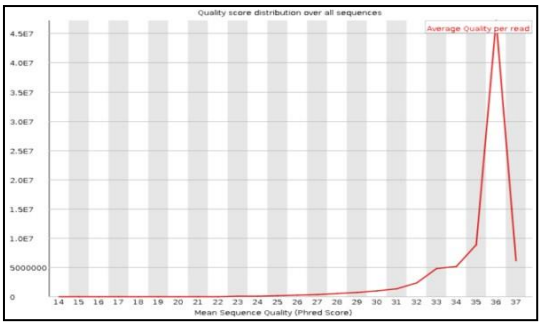
S3_R1_Qc

Measure	Value
Filename	S3_R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	79963266
Sequences flagged as poor quality	0
Sequence length	159
%GC	52

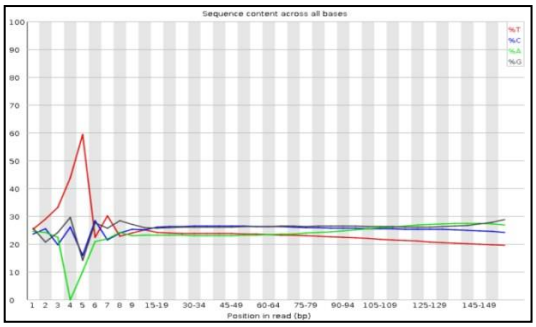
Adapter Content



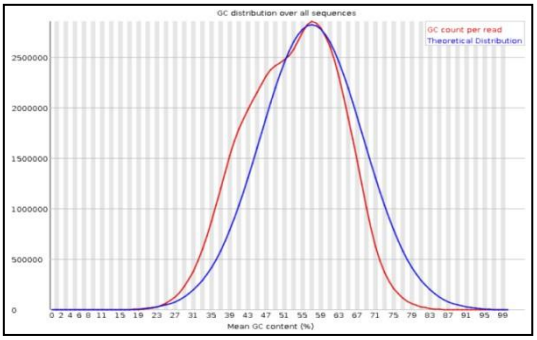
Per base sequence quality



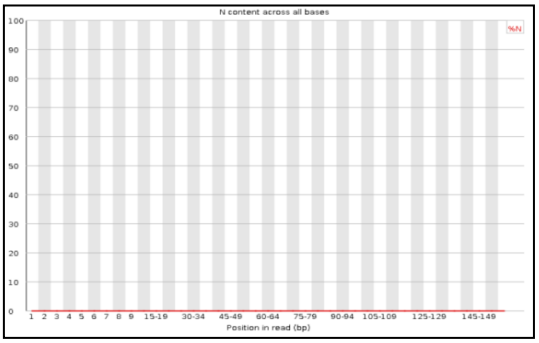
Per sequence quality score



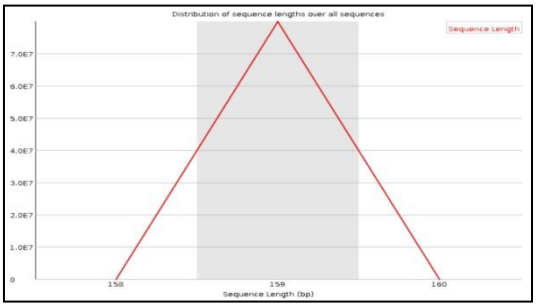
Per base sequence content



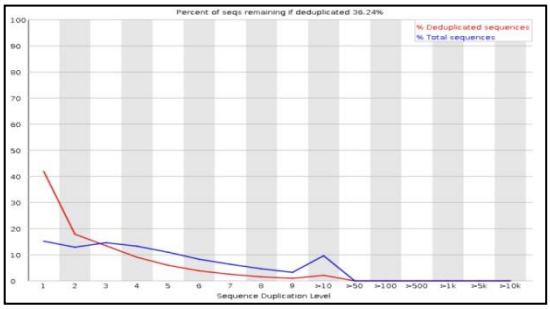
Per Sequence GC content



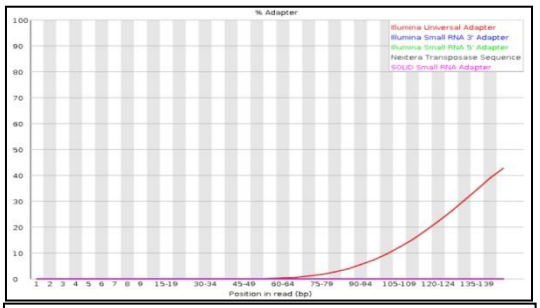
Per base N content



Sequence length Distribution



Sequence l Sequence Duplication Level



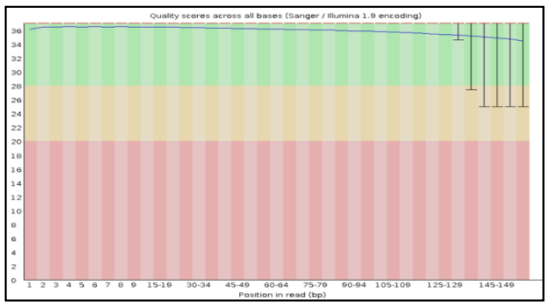
Sequence length Distribution

Supplementary Figure 5: Fastq graph plot reports for quality values of the bases in forward read of sample-3

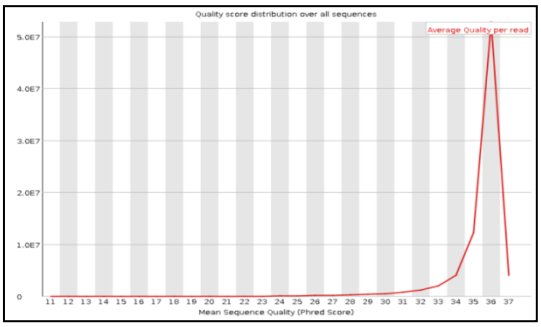
S3_R2_Qc

Measure	Value
Filename	S3_R2.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	79963266
Sequences flagged as poor quality	0
Sequence length	159
%GC	52

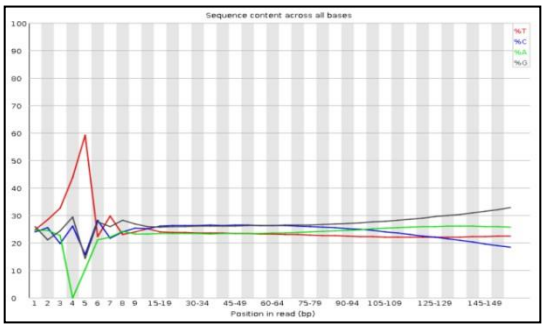
Basic Statistic



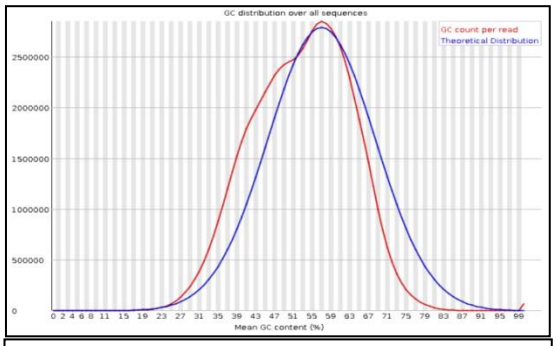
Per base sequence quality



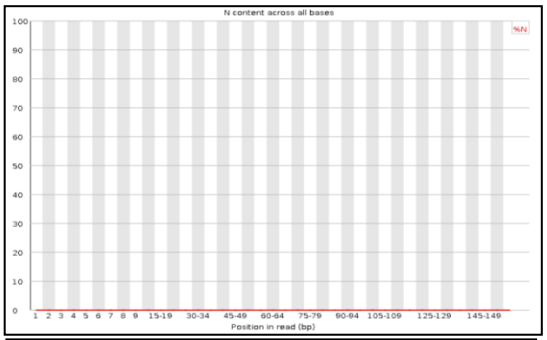
Per sequence quality score



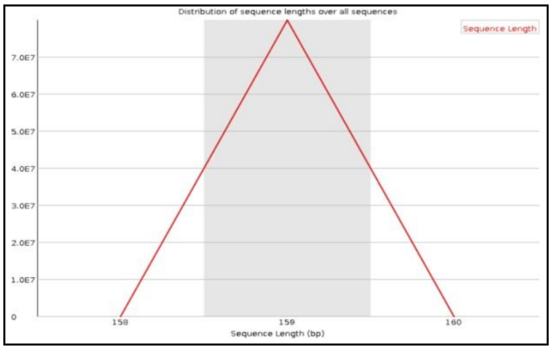
Per base sequence content



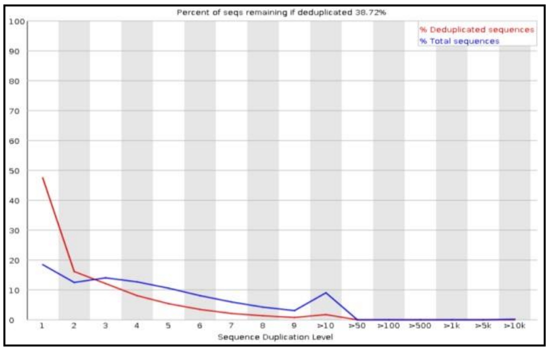
Per Sequence GC content



Per base N content



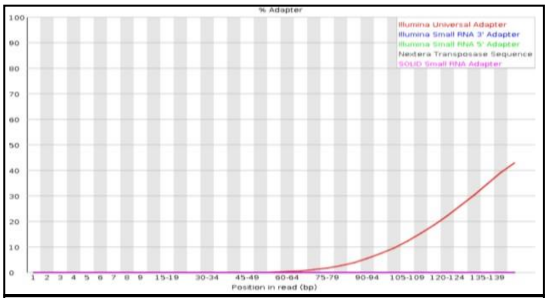
Sequence length Distribution



Sequence Duplication Level

Sequence	Count	Percentage	Possible Source
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	149356	0.18678076505779542	No Hit

Overrepresented sequences



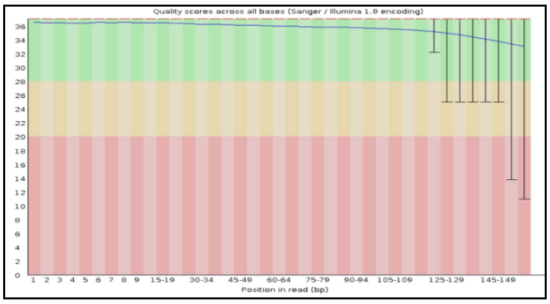
Adapter Content

Supplementary Figure 6: Fastq graph plot reports for quality values of the bases in reverse read of sample-3

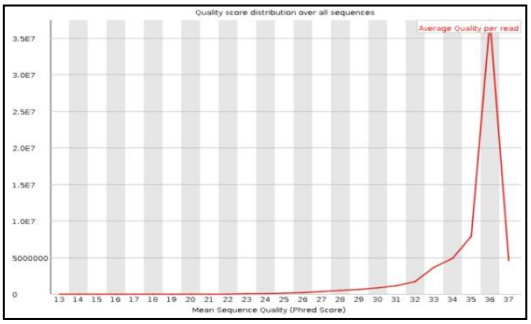
S4_R1_Qc

Measure	Value
Filename	S4_R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	64931810
Sequences flagged as poor quality	0
Sequence length	159
%GC	51

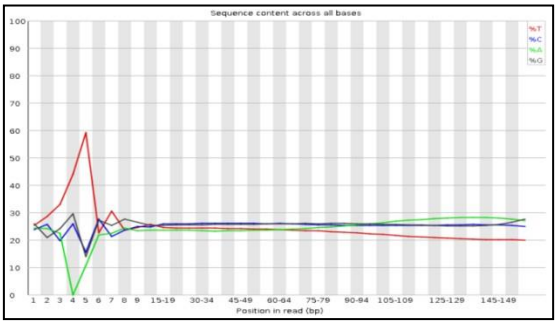
Basic Statistic



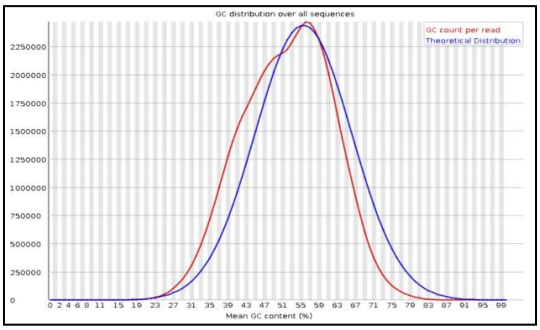
Per base sequence quality



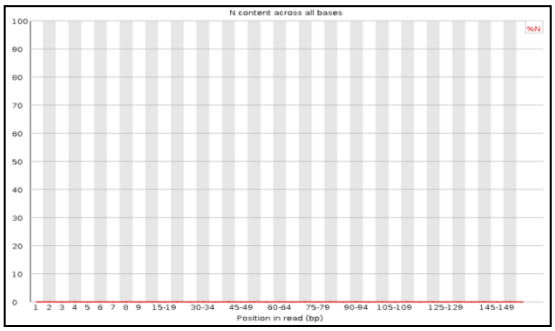
Per sequence quality score



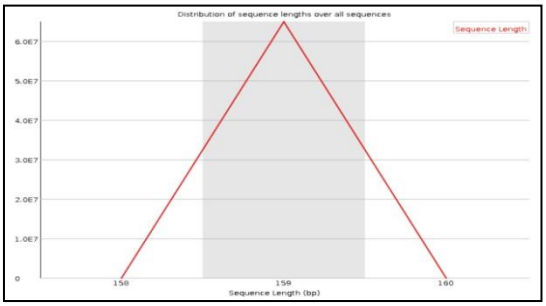
Per base sequence content



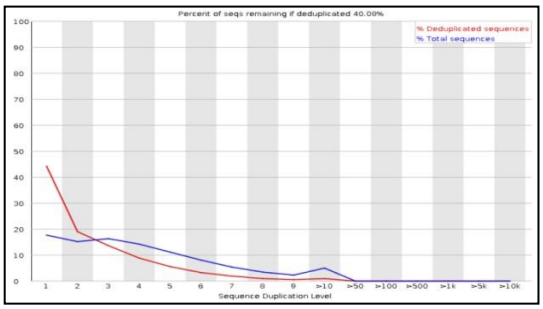
Per Sequence GC content



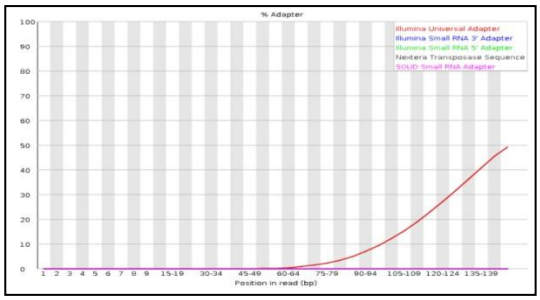
Per base N content



Sequence length Distribution



Sequence Duplication Level



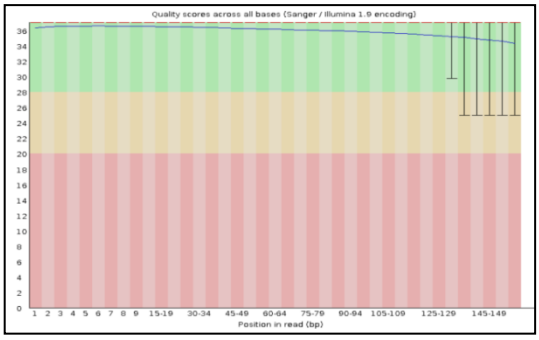
Adapter content

Supplementary Figure 7: Fastq graph plot reports for quality values of the bases in forwards read of sample 4

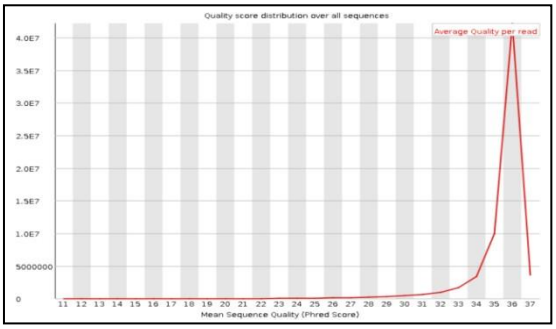
S4_R2_QC

Measure	Value
Filename	S4_R2.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	64931810
Sequences flagged as poor quality	0
Sequence length	159
%GC	51

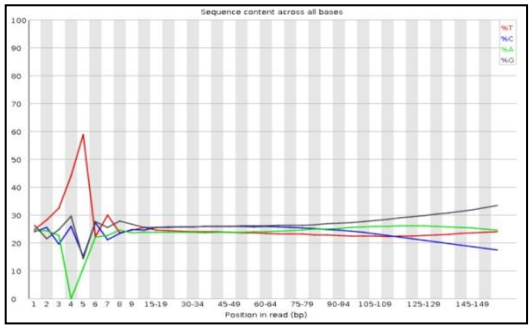
Basic Statistic



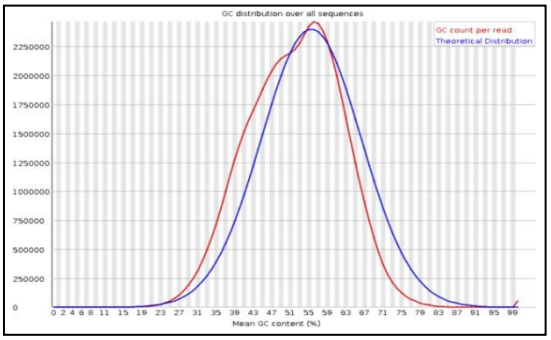
Per base sequence quality



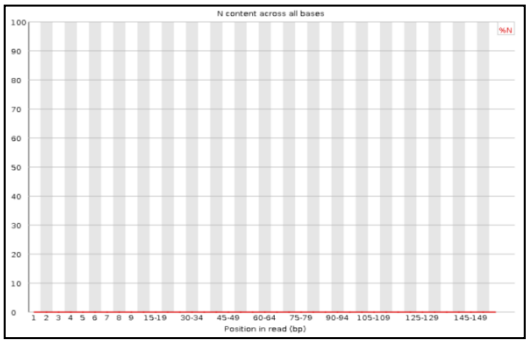
Per sequence quality score



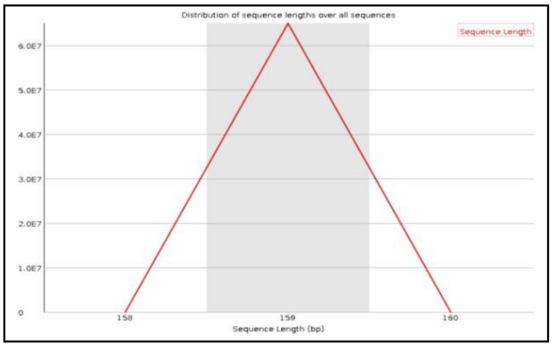
Per base sequence content



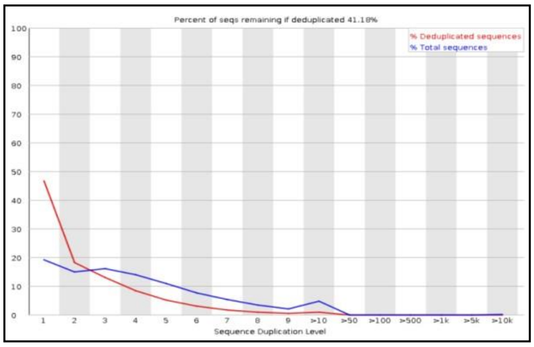
Per Sequence GC content



Per base N content



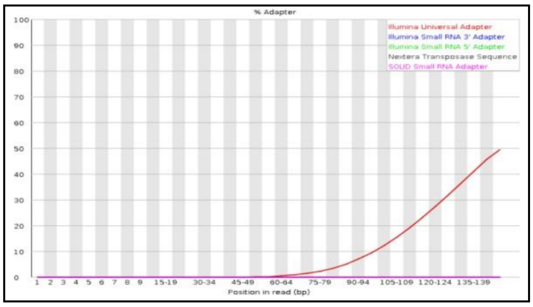
Sequence length Distribution



Sequence Duplication Level

Sequence	Count	Percentage	Possible Source
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	123980	0.19093877099683498	No Hit

Overrepresented sequences



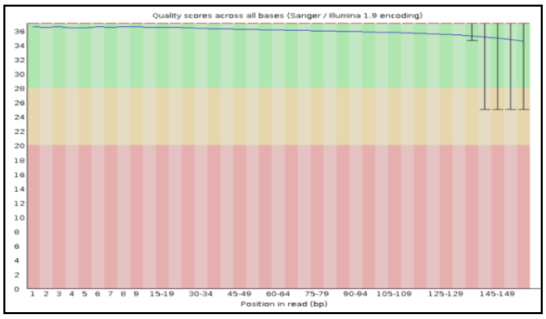
Adapter content

Supplementary Figure 8: Fastq graph plot reports for quality values of the bases in reverse read of sample-4

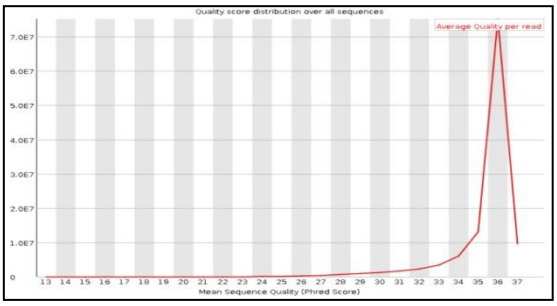
S5_R1_QC

Measure	Value
Filename	S5_R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	116462674
Sequences flagged as poor quality	0
Sequence length	159
%GC	52

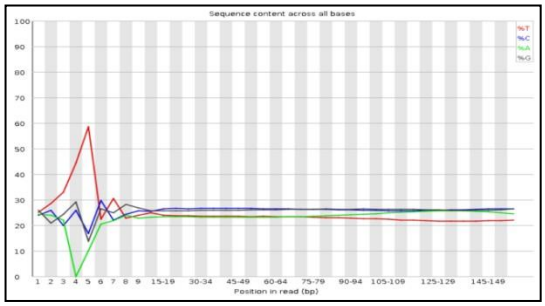
Basic Statistic



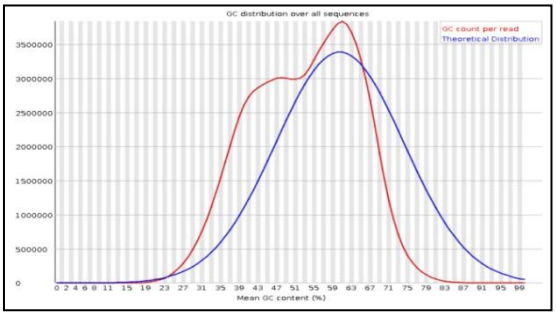
Per base sequence quality



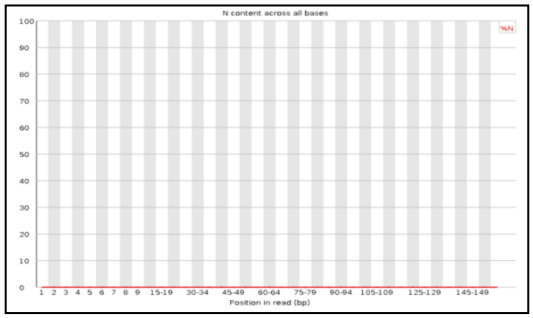
Per sequence quality score



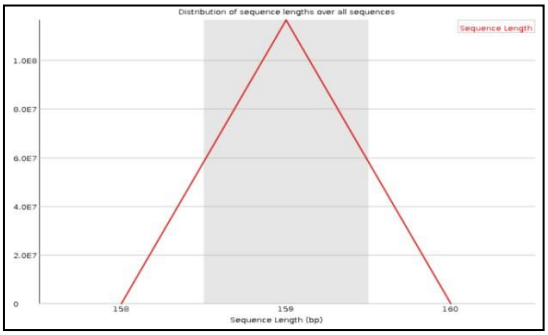
Per base sequence content



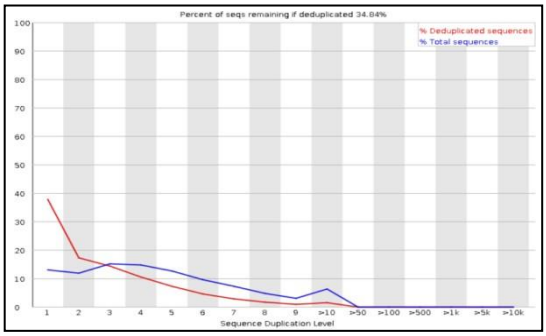
Per Sequence GC content



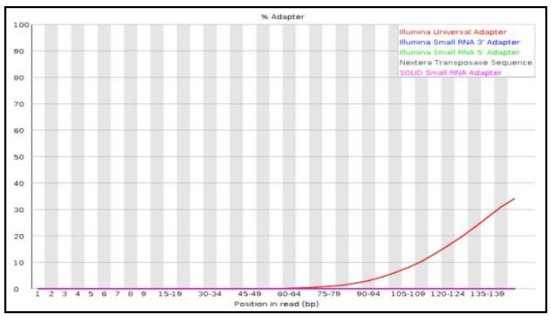
Per base N content



Sequence length Distribution



Sequence Duplication Level



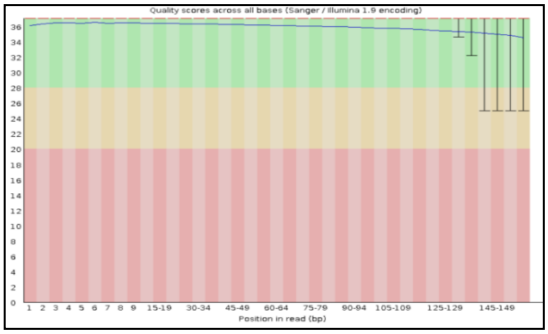
Adapter content

Supplementary Figure 9: Fastq graph plot reports for quality values of the bases in forwards read of sample-4

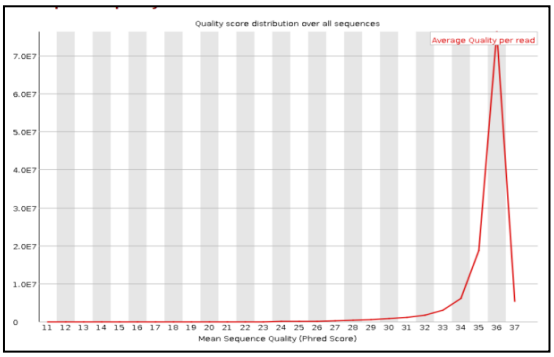
S5_R2_QC

Measure	Value
Filename	S5_R2.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	116462674
Sequences flagged as poor quality	0
Sequence length	159
%GC	52

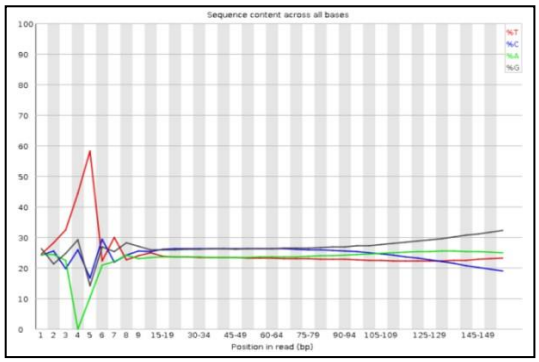
Basic Statistic



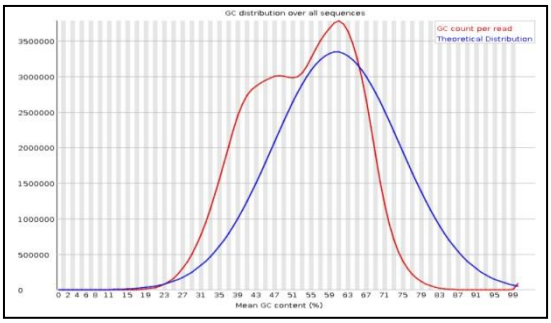
Per base sequence quality



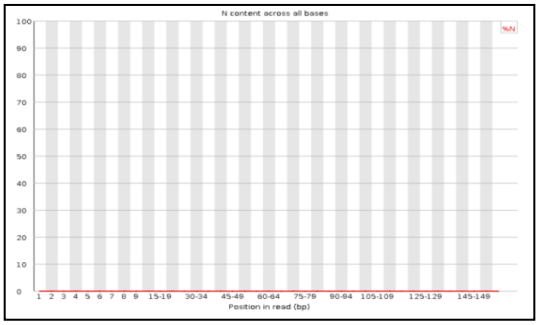
Per sequence quality score



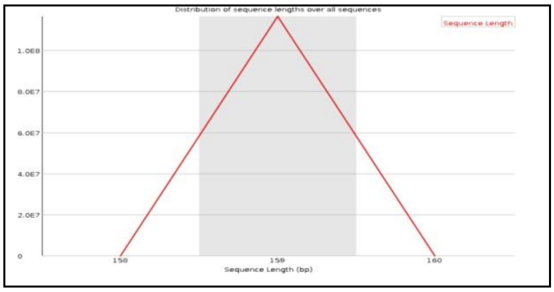
Per base sequence content



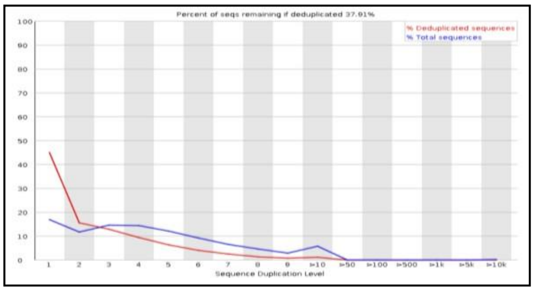
Per Sequence GC content



Per base N content



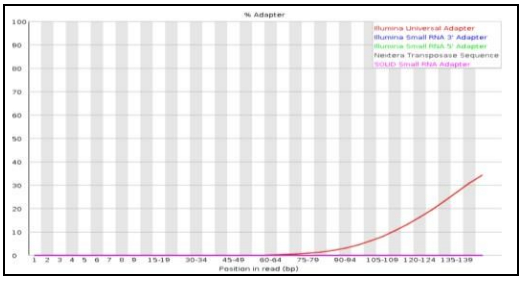
Sequence length Distribution



Sequence Duplication Level

Sequence	Count	Percentage	Possible Source
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	204210	0.17534373287702462	No Hit

Overrepresented sequences



Adapter content

Supplementary Figure 10: Fastq graph plot reports for quality values of the bases in reverse read of sample-5

LIST OF PUBLICATIONS:

Research articles:

A. Published:

1. **Kour**, **Bhumandeep**, Nidhi Shukla, Harshita Bhargava, Devendra Sharma, Amita Sharma, Anjuvan Singh, Jayaraman Valadi, T. S. Sadasukhi, Sugunakar Vuree, and Prashanth Suravajhala. "Identification of plausible candidates in Prostate Cancer using integrated machine learning approaches."*Current Genomics 24, no. 5 (2023); (287-306)*
2. Shukla, Nidhi, **Bhumandeep Kour**, Devendra Sharma, Maneesh Vijayvargiya, T. C. Sadasukhi, Krishna Mohan Medicherla, Babita Malik et al. "Towards Understanding the Key Signature Pathways Associated from Differentially Expressed Gene Analysis in an Indian Prostate Cancer Cohort." *Diseases* 11, no. 2 (2023): 72.

B. Communicated:

3. **Kour, Bhumandeep,** Anjuvan Singh, Nirmal Kumar Lohiya, Devendra Sharma, Nidhi Shukla, Santosh Kumari Duppala, Sugunakar Vuree, Prashanth Suravajhala, "Genome to Phenome Pathway Functional Annotation Using Whole Exome Sequencing of Prostate Cancer-Specific to India" **(Communicated to Human Gene) (Under Review)**
4. Khilwani, Barkha, **Bhumandeep Kour**, Nidhi Shukla, Sugunakar Vuree, Abdul S. Ansari, Nirmal K. Lohiya, Renuka Suravajhala, and Prashanth Suravajhala. "In silico approaches to identify the role of lncRNAs in Prostate cancer and Androgen receptor-targeted proteins." *bioRxiv* (2023): 2023-10 **(Communicated to Biochemistry and Biophysics Reports) (Under Review)**

Review papers:

5. Rout, Madhusmita, **Bhumandeep Kour**, Sugunakar Vuree, Sajitha S. Lulu, Krishna Mohan Medicherla, and Prashanth Suravajhala. "Diabetes mellitus susceptibility with varied diseased phenotypes and its comparison with phenome interactome networks." *World Journal of Clinical Cases* 10, no. 18 (2022): 5957.
6. Sun, Xiaolu, **Bhuman Deep Kour**, Xiangtao Wang, Xiufang Hu, Yufen Wang, and Vuree Sugunakar. "Insights into association between urolithiasis and prostate cancer." *Journal of Men's Health* 17, no. 4 (2021): 52-61.

7. Sharma, Devendra, Saloni Someshwar, **Bhumandeep Kour**, Nidhi Shukla, Barkha Khilwani, Maneesh Vijay, Ayam Gupta et al. "The CAPCI network: A CAncer Prostate Consortium of India for conducting next-generation genomic sequencing studies." *Cancer Health Disparities* (2019).
8. Bhargavi, R., Barkha Khilwani, **Bhumandeep Kour**, Nidhi Shukla, Rajaguru Aradhya, Devendra Sharma, Maneesh Vijayvargiya et al. "Prostate cancer in India: Current perspectives and the way forward." *Journal of Reproductive Healthcare and Medicine* 4, no. 8 (2023): 1.

LIST OF CONFRENCES:

1. Poster Presentation in "48th Annual Meeting and International Conference of the **INDIAN SOCIETY OF HUMAN GENETICS 2024"**, Institute of Human Genetics, Ahmedabad, India, held on 21st to 24rth January 2024

2. Poster Presentation in **"International Conference on Reproductive Health with Emphasis on Innovations in Reproductive Sciences and Technology: Hope, Risk and Responsibilities",** Ravenshaw University, Cuttack, Odisha, India held on 24th to 26th February, 2023

3. Oral Presentation in **"International Conference on Frontiers in Nutrition, Medical Genomics, and Drug Discovery, InBix 2022",** Vignan University, Guntur, Andhra Pradesh, India held on October 31 to November 2nd 2022

4. Participated (video abstract) in virtual conference "**Indian conference on Bioinformatics, InBix 2020"**, CSIR-North-East Institute of Science and Technology Jorhat, Assam, India, November 11-13, 2021

5. Attended **"International Conference on Biosciences and Biotechnology (ICBB-2019)"** held at Lovely Professional University, Punjab, India on 4th-5th November 2019

LIST OF WORKSHOPS:

1. Participated in a one-week international workshop cum training on **"System Bioinformatics and Biopython Modules"** held during 4th-8th Feb, 2020, jointly organized by Birla Institute of Scientific Research, Jaipur and Perdana University, Malaysia.

2. Participated in a virtual workshop on **"Essential topics in Bioinformatics for intermediates"** from 21st to 30th September 2020, jointly organized by Birla Institute of Scientific Research, Jaipur and Perdana University, Malaysia.