

RESEARCH ARTICLE

Identification of Plausible Candidates in Prostate Cancer Using Integrated Machine Learning Approaches



Bhumandeep Kour^{1,2,#}, Nidhi Shukla^{2,3,#}, Harshita Bhargava^{4,#}, Devendra Sharma⁵, Amita Sharma⁴, Anjuvan Singh⁶, Jayaraman Valadi⁷, Trilok Chand Sadasukhi⁸, Sugunakar Vuree^{2,9,*} and Prashanth Suravajhala^{2,10,*}

¹Department of Biotechnology, Lovely Professional University, Jalandhar, Punjab, India; ²Bioclues.org, India; ³Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research, Jaipur, Rajasthan, India; ⁴Department of Computer Science, IIS University, Jaipur, Rajasthan, India; ⁵Urology and Renal Transplant Department of Renal Sciences, Rukmani Birla Hospital, Jaipur, Rajasthan, India; ⁶Department of Biotechnology, School of Bioengineering and Biosciences, Lovely Professional University, Punjab, Phagwara, 144001, India; ⁷Department of Computer Science, FLAME University, Pune, Maharashtra, India; ⁸Department of Urology and Renal Transplant, Mahatma Gandhi University of Medical Sciences and Technology, Jaipur, Rajasthan, India; ⁹MNR Foundation for Research & Innovation, MNR Medical College and Hospital, MNR University, Telangana, India; ¹⁰Amrita School of Biotechnology, Amrita Vishwa Vidyapeetham, Kollam, Kerala, India

Abstract: Background: Currently, prostate-specific antigen (PSA) is commonly used as a prostate cancer (PCa) biomarker. PSA is linked to some factors that frequently lead to erroneous positive results or even needless biopsies of elderly people.

Objectives: In this pilot study, we undermined the potential genes and mutations from several databases and checked whether or not any putative prognostic biomarkers are central to the annotation. The aim of the study was to develop a risk prediction model that could help in clinical decision-making.

Methods: An extensive literature review was conducted, and clinical parameters for related comorbidities, such as diabetes, obesity, as well as PCa, were collected. Such parameters were chosen with the understanding that variations in their threshold values could hasten the complicated process of carcinogenesis, more particularly PCa. The gathered data was converted to semi-binary data (-1, -0.5, 0, 0.5, and 1), on which machine learning (ML) methods were applied. First, we cross-checked various publicly available datasets, some published RNA-seq datasets, and our whole-exome sequencing data to find common role players in PCa, diabetes, and obesity. To narrow down their common interacting partners, interactome networks were analysed using GeneMANIA and visualised using Cytoscape, and later cBioportal was used (to compare expression level based on Z scored values) wherein various types of mutation w.r.t their expression and mRNA expression (RNA seq FPKM) plots are available. The GEPIA 2 tool was used to compare the expression of resulting similarities between the normal tissue and TCGA databases of PCa. Later, top-ranking genes were chosen to demonstrate striking clustering coefficients using the Cytoscape-cytoHubba module, and GEPIA 2 was applied again to ascertain survival plots.

Results: Comparing various publicly available datasets, it was found that BLM is a frequent player in all three diseases, whereas comparing publicly available datasets, GWAS datasets, and published sequencing findings, SPFTPC and PPIMB were found to be the most common. With the assistance of GeneMANIA, TMPO and FOXP1 were found as common interacting partners, and they were also seen participating with BLM.

Conclusion: A probabilistic machine learning model was achieved to identify key candidates between diabetes, obesity, and PCa. This, we believe, would herald precision scale modeling for easy prognosis.

Keywords: cBioPortal, prostate cancer, biomarkers, machine learning, bioinformatics, Prostate-Specific Antigens (PSA).

1. INTRODUCTION

Prostate cancer (PCa) has become the second most common cancer type in men and the fifth major reason for

mortality across the world. As PCa is largely caused by the enlargement of the prostate gland with late onset, the difference in incidence rates in different parts of the world is due to the use of varying diagnostic approaches [1]. One in every nine men above the age of 65 is affected by PCa, which has become the most common cancer type in American men [2]. Around 1.6 million men are positively diagnosed every year, out of which 366,000 die [3]. The prostate-specific antigen (PSA) test has shown an undesirable high false-positive rate, revealing a poor prognosis. Alternatively, it

*Address correspondence to these authors at the MNR Foundation for Research & Innovation, MNR Medical College and Hospital, MNR University, Telangana, India; Bioclues.org, India; E-mail: sugunakarvure@gmail.com; Amrita School of Biotechnology, Amrita Vishwa Vidyapeetham, Kollam, Kerala, India; Bioclues.org, India; Tel/Fax: 9461159780; E-mail: prash@bioclues.org

These authors contributed equally to this work.

has been proposed to lower the PSA threshold to solve the current problem with PSA testing. However, reducing the PSA threshold increases the risk of diagnosing painless disease and unnecessary treatment [4]. Over the years, there have been some developments that have improved the PSA testing and diagnostic accuracy, viz. measuring PSA with different molecular configurations and the rate of increase in PSA. Total PSA (tPSA) refers to the sum of free PSA (unbound) and bound PSA (mainly complexing with α_1 -antichymotrypsin). The increase in tPSA, depicted by the total PSA rate (tPSAV), has now gained great attention in diagnosis and prognosis [4]. PSA seems difficult to completely replace because of its least invasive properties and cost-effectiveness; nevertheless, there is an urgent requirement to supplement PSA with other such biomarkers that could enhance both its specificity and sensitivity during screening tests. Figuring out a panel of this type of diagnostic and prognostic-associated biomarkers could be considered ideal when working with PSA. Even though PCa has a lot of heterogeneity, it is not ideal for forming the screening basis based on a single antigen to provide an accurate analysis report for every patient [4].

Genome-Wide Association Studies (GWAS) serve as a resource providing complete genetic association data to compare and discover significant datasets by taking genes, gene regions, phenotypes, or any trait through GWAS central (<http://www.gwascentral.org> last accessed May 23rd, 2022). It gives more than 67 million P-values for 1600 above studies, making it the world's biggest database of GWAS information at the summary level [5]. On the other hand, the Prostate Cancer Association Group investigating cancer-associated alterations in the genome (PRACTICAL) is an international consortium formed to get precise evaluations of risk related to variants evaluated in large numbers of both cases and controls [6]. The rationale behind using

this meta-centric approach is that there is a dearth of screening candidate genes and mutations overlaying both tools. Therefore, as a mandate of our recently carved out cancer prostate consortium of India [7], we attempt to undermine the candidate genes and mutations from the myriad of databases and examine whether or not any candidate prognostic biomarkers are centric to the annotation.

2. MATERIALS AND METHODS

2.1. Datasets

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. The associated clinical parameters of comorbidities leading to cancer progression, especially for the prostate malignancy, were carefully chosen and were further categorized into unknown significance, likely benign, benign, likely pathogenic, and pathogenic, based on their threshold values with binary/semi-binary scores, viz. -1, -0.5, 0, 0.5, and 1, respectively. While these binary scores are given based on the risk of acquiring PCa that increases with the rising pathogenicity of a variant, the general stratification of low grade, intermediate but less risk, intermediate, moderate risk, and high risk corresponding with -1, -0.5, 0, 0.5, and 1, respectively, was checked and tabulated (Fig. 1 and Supplementary spreadsheet 1). The data was then subjected to five different machine learning algorithms, namely linear regression, multilayer perceptron, random forest, random tree, and REP tree (Table 1), and further compared to predict the chance of developing PCa when there are any change variables from the control population. Similarly, the annotated data of clinical attributes for all three diseases was transformed into semi-binary

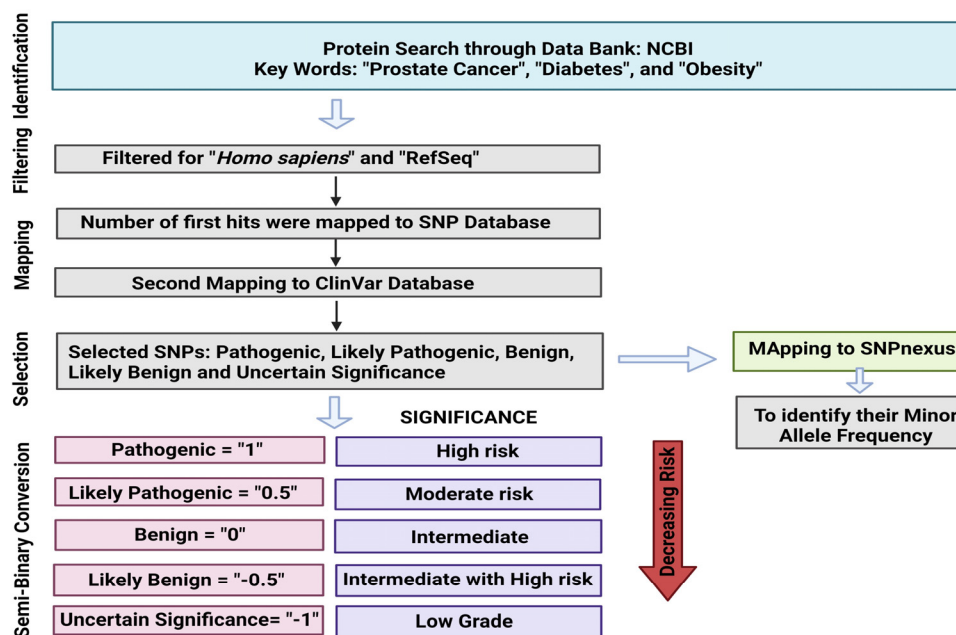


Fig. (1). Flow chart representing identification, filtering, mapping (to SNP database, ClinVar, and SNPnexus), selecting, and finally, semi-binary data conversion (1, 0.5, 0, -0.5, -1) with respective significance. (Figure created through BioRender <https://www.biorender.com/>). (A higher resolution / colour version of this figure is available in the electronic copy of the article).

Table 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.
Multilayer perceptron (MLP)	MLP is an artificial neural network that 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.
Random forest	A random forest algorithm is an ensemble of decision trees, each trained on a random set of features and a random set of samples. This sampling procedure tends to minimize the overfitting and induces diversity in the ensemble. The random forest can be configured using WEKA in terms of the number of trees, the 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.
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.
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 the same or better than the original tree on the validation set, the subtree is discarded.

coding (-1, -0.5, 0, 0.5, 1) based on the different ranges and their corresponding degree of risk for developing PCa (Supplementary Fig. 1).

2.2. Clinical Parameters

2.2.1. Clinical Parameters of PCa

2.2.1.1. 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 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 formed 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 [8]. Modern biopsy 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 tumors [8]. Scoring is the sum of highly frequent common and second-

ary patterns 3+5 [9]. 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 [10].

2.2.1.2. 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 overdiagnosis (false positive or false negative), which can lead to prolonged side effects of treatments [4].

2.2.1.3. 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 screening [11].

2.2.2. Clinical Parameters of Diabetes Associated with Cancer/PCa

2.2.2.1. 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 [12, 13]. 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 [14]. 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 [15, 16]. Its limitations include its association with poor performance in pregnant females, old age, and the chance of overdoing in anemia and genetically predisposed ones [17]. 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, esophagus, lungs, endometrium, pancreas, and kidneys and decreased risk for PCa, suggesting that diabetes and glycemic control is crucial in limiting cancer risk [17].

2.2.2.2. WBC Count Test

WBC count testing alone can predict diabetes even in non-glycemic men. It has been estimated that for every 1000 cells/mm³ rise within the normal range, the chance of diabetes rises by 7.6%. In addition, chronic inflammation increases the likelihood of diabetes even without obesity in autoimmune-aided patients. So, WBC count is considered an independent risk factor for diabetes in young people [18, 19]. 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 to those who did not develop VTE [20].

2.2.2.3. 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 [21]. 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 [22]. Fasting blood sugar testing is more reliable than HbA1c [23].

2.2.2.4. 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 30 kg per meter square [23]. A number of genetic variants are found common in GWAS, which confirms their associations [24].

2.2.3. Clinical Parameters of Obesity Associated with Cancer/PCa

2.2.3.1. 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 [25]. 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 18.5-25 kg/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 to 25 kg/m². Therefore, it has been highly advised to maintain weight in lower fields only [26]. A systemic pro-inflammatory environment caused by abdominal adiposity might initiate diabetes and cancer [27] 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 [25, 28]. 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 during biopsies [29]. 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 type [30].

2.2.3.2. 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 results in a high rate of PSA-based biopsies and diagnosis, resulting in high cholesterol in men. It can be concluded that high total serum cholesterol or HDL (high-density lipoprotein) is a risk factor for having a more aggressive form of PCa [31].

2.3. Machine Learning Algorithms

We have used the Waikato Environment for Knowledge Analysis (Weka) for implementing the machine learning algorithms. The algorithms used for the study included Linear regression, Multilayer perceptron, Random forest, Random tree, and REPTree for the regression analysis.

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 (Supplementary spreadsheet 2). 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

Table 2. List of different regression algorithms applied in Weka software on datasets of PCa, diabetes mellitus, and obesity and their root mean square error (RMSE) value.

Dataset Used	Algorithm	Root Mean Square Error (RMSE)
Prostate cancer	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

standard deviation (PCa=0.000612372; DM=0.012077094; Obesity=0.00128582 (Table 2) and then the normalized deviation values are calculated and plotted against each other by line graphs.

2.4. Correlational Study

To categorize this, the risk variants were retrieved from ClinVar, published GWAS data (from the PRACTICAL consortium and GWAS central), our own Exome data (Saxena *et al.*, 2021), and RNA seq Diabetes Type 2 Mellitus data [32] (Fig. 2). A cross-check was performed to identify common key players using Venn plots (<https://bioinfogp.cnb.csic.es/tools/venny/>).

2.5. 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, which is then extended to incorporate other genes as interacting partners predicted to share the same function based on their overlapping connections in biological pathways [33].

2.6. CBioportal

We used datasets provided by Armenia *et al.* (2018) [34], wherein they identified 97 significantly mutated genes (SMGs), 70 of which were not earlier involved in PCa, followed by several mutations that were seen in less than 3% of the cases and another study used was of TCGA, Cell 2015 [35].

2.7. 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) [36]. 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 functions like correlation, survival, profiling, plotting, analysis, dimension reduction or differential expression analysis, and detection of a similar gene [37].

2.8. Survival Analysis

For the survival analysis, we used GEPIA 2 again, wherein the log-rank below 0.05 ($p < 0.05$) is referred to as significant [38]. Survival plots are formed with Kaplan-Meier (KM curve), and 2 curves are then compared with the log-rank test [39].

2.9. Cytoscape

For ranking nodes of a network on the basis of network features, a plugin called CytoHubba was added to Cytoscape. It provides a platform for several topological analyses in retrieving subnetworks from the whole protein-protein interaction network. A list of a few nodes was taken from the complete network, and computed topological features were saved as node attributes in the Cytoscape data structure. The clustering coefficient based on the nodal size and

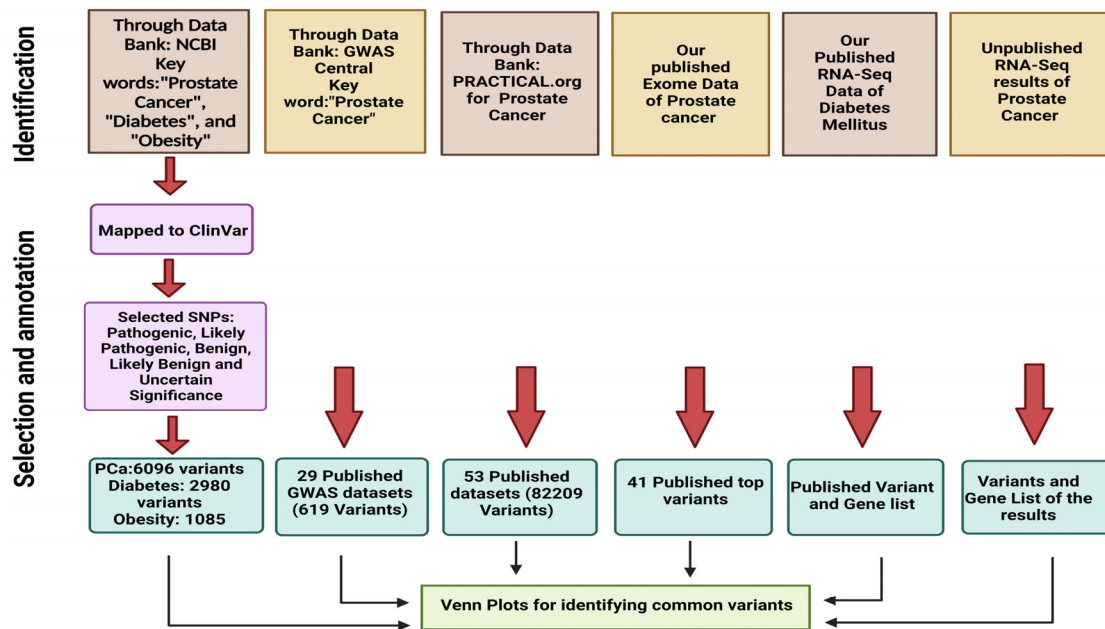


Fig. (2). Flow chart for the correlational study of ClinVar variants (PCa, diabetes, and obesity) and published GWAS data from GWAS Central, PRACTICAL consortium, exome data, and RNA Seq data of both prostate cancer and diabetes mellitus for identifying common variants. (Figure created through BioRender <https://www.biorender.com/>). (A higher resolution / colour version of this figure is available in the electronic copy of the article).

color of the node related to the degree [40] was performed with a PPI network constructed using the STRING database by giving a list of genes in the query, and then its visualization and subnetworks were formed by Cytoscape using the bottleneck method. Later, topological parameters were analysed with the help of a plug-in, viz. network analyzer. It calculates different topological features like connectivity, node number, connecting edges, clustering coefficient, average clustering coefficient, centralization, connectivity degree, etc. [41].

3. RESULTS AND DISCUSSIONS

3.1. Comparison of Vivid Datasets Yields 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 changes. A few common variants were also identified in correlating different GWAS datasets with ClinVar data (Fig. 3).

3.2. The GWAS Central Comparison Yielded rs721048 (EHBPI) 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 were found to be common, i.e., rs145988146, rs80358600, rs276174854, rs276174889, rs771203198, and 1 variant of *BRCA1* (rs28897696) GWAS central was common to both of them (Fig. 4). Whereas the PRACTICAL consortium comparison yielded rs721048 (EHBPI) and rs138213197 (HOXB13), which were earlier

found common to ClinVar PCa and GWAS Central, 17 other variants were 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) (Fig. 5B). No commonality was observed between exome data and obesity (Fig. 5A). One common variant, rs61816761, associated with the *FLAG* gene, was found to be common to obesity and diabetes.

3.3. RNA-Seq Data of both PCa and Diabetes Type 2 Mellitus

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 (Fig. 5C).

3.4. Interactome Networks Using GeneMANIA

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* (Fig. 6A). The interactome network of *PP1MB* and *SFTPC* showed some commonalities in their interacting partners with other datasets. *SFTPC*'s interacting partner *TMEM67* is common in PCa ClinVar data and *FOXP1* in diabetes ClinVar data. *PPMIB*'s interacting partner *PPP2CA* is common in PCa ClinVar data, and *PPARG*

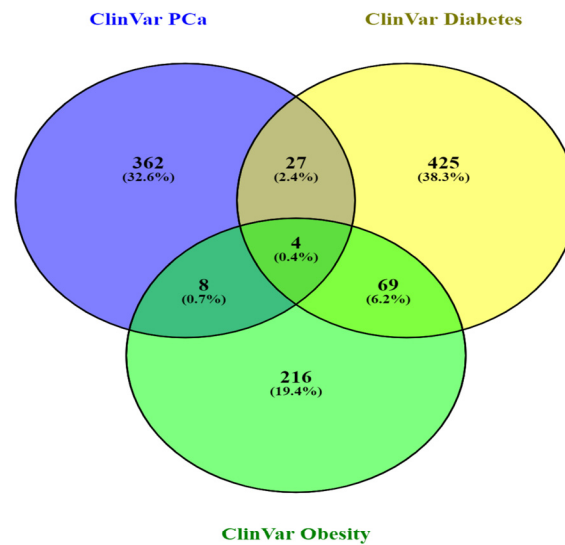


Fig. (3). Venn plot for identifying commonalities between Clinvar datasets of prostate cancer, diabetes, and obesity. It showed 27 common genes between prostate cancer and diabetes, 8 common genes between prostate cancer and obesity, 69 common genes between diabetes and obesity, and 4 among all three 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 NCL* were common in PCa and diabetes. *BRCA2, FANCI, BAP1, AR, and PMS2* were 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, MECPP2, 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* were common in diabetes and obesity with variants. *BLM, TMEM67, RFX6, NR0B2, and NUDC* were found to be common among PCa, diabetes mellitus, and obesity. (Figure created through Venny 2.1 <https://bioinfogp.cnb.csic.es/tools/venny>). (A higher resolution / colour version of this figure is available in the electronic copy of the article).

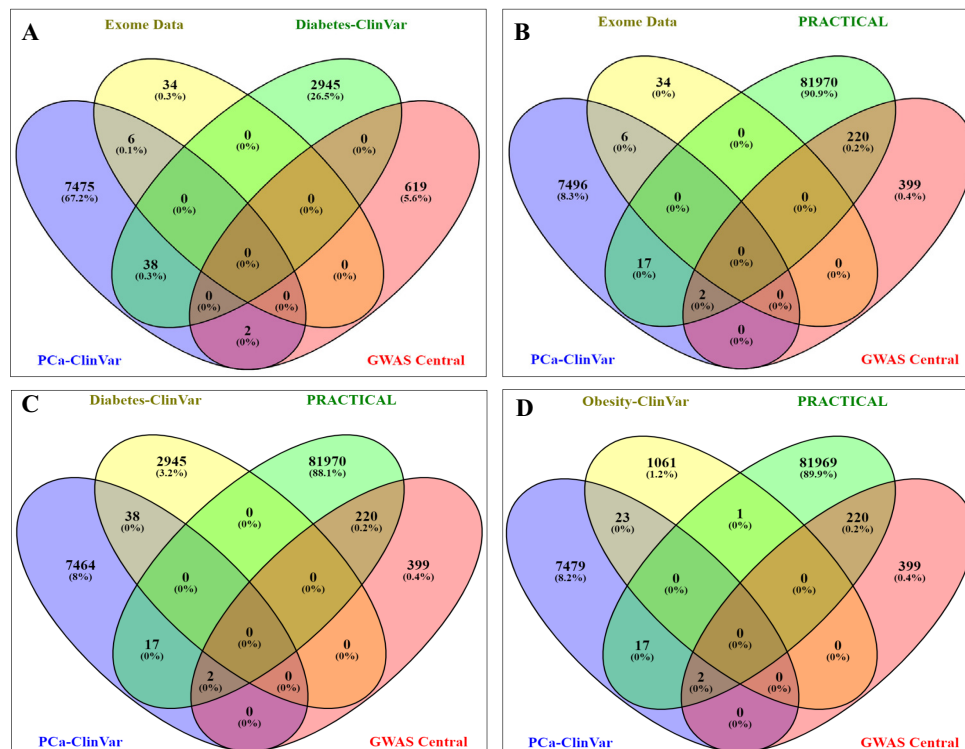


Fig. (4). A, B: Venn plot to identify common variants among ClinVar prostate cancer, diabetes, GWAS central, PRACTICAL consortium for prostate cancer, and prostate cancer exome data; C, D: Venn plot to identify common variants among ClinVar prostate cancer, diabetes, obesity ClinVar, GWAS central, PRACTICAL consortium for prostate cancer and prostate cancer exome data. (Figure created through Venny 2.1 <https://bioinfogp.cnb.csic.es/tools/venny>). (A higher resolution / colour version of this figure is available in the electronic copy of the article).

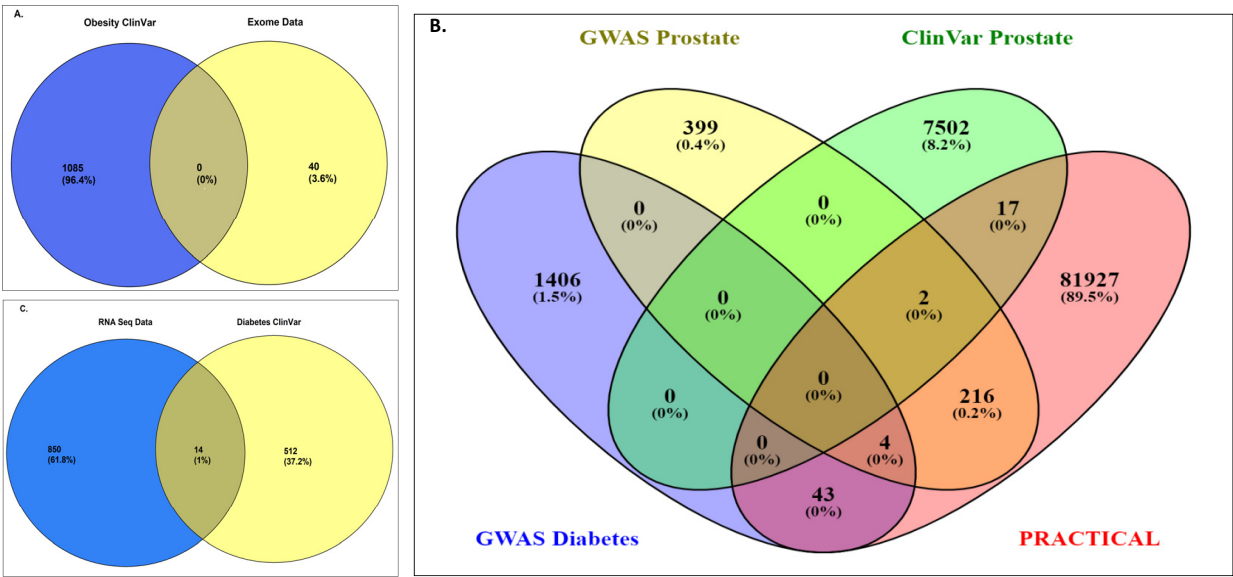


Fig. (5). **A:** Venn plot showing no commonality between exome data and obesity ClinVar data; **B:** Venn plot for GWAS central data of both type 2 diabetes and prostate cancer, PRACTICAL Consortium, and ClinVar Prostate cancer; **C:** Venn plot showing some commonalities between RNAseq results of three prostate cancer samples and ClinVar data of diabetes. (Figure created through Venny 2.1 <https://bioinfogp.cnb.csic.es/tools/venny>). (A higher resolution / colour version of this figure is available in the electronic copy of the article).

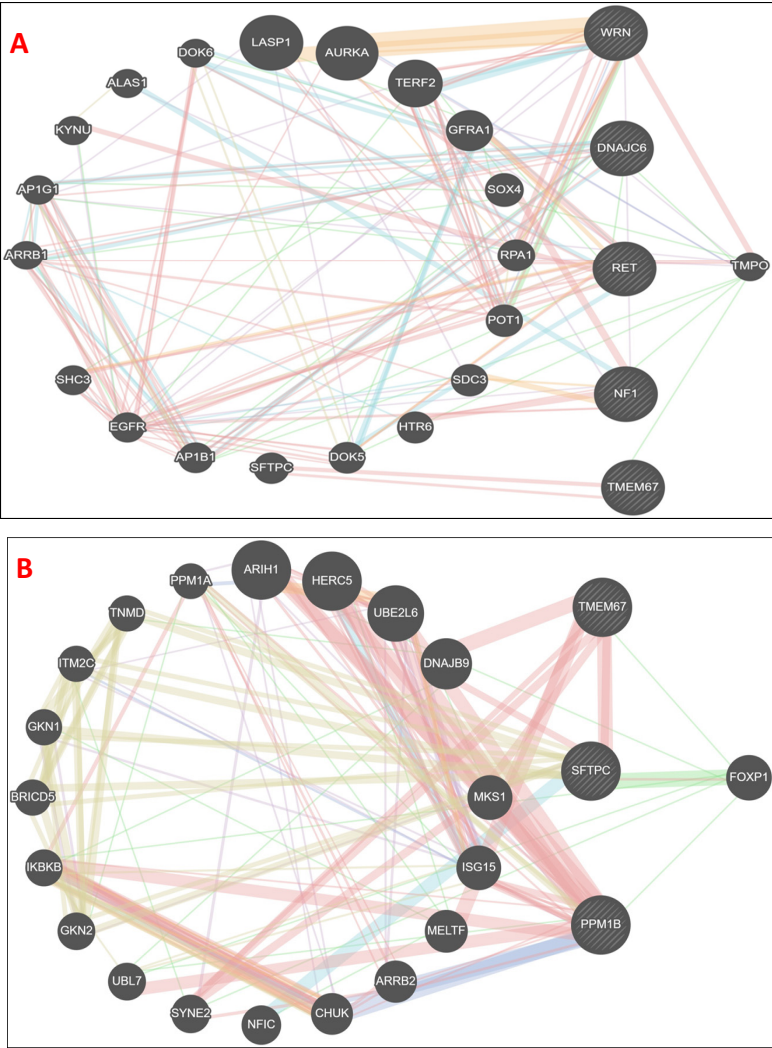


Fig. (6). contd....

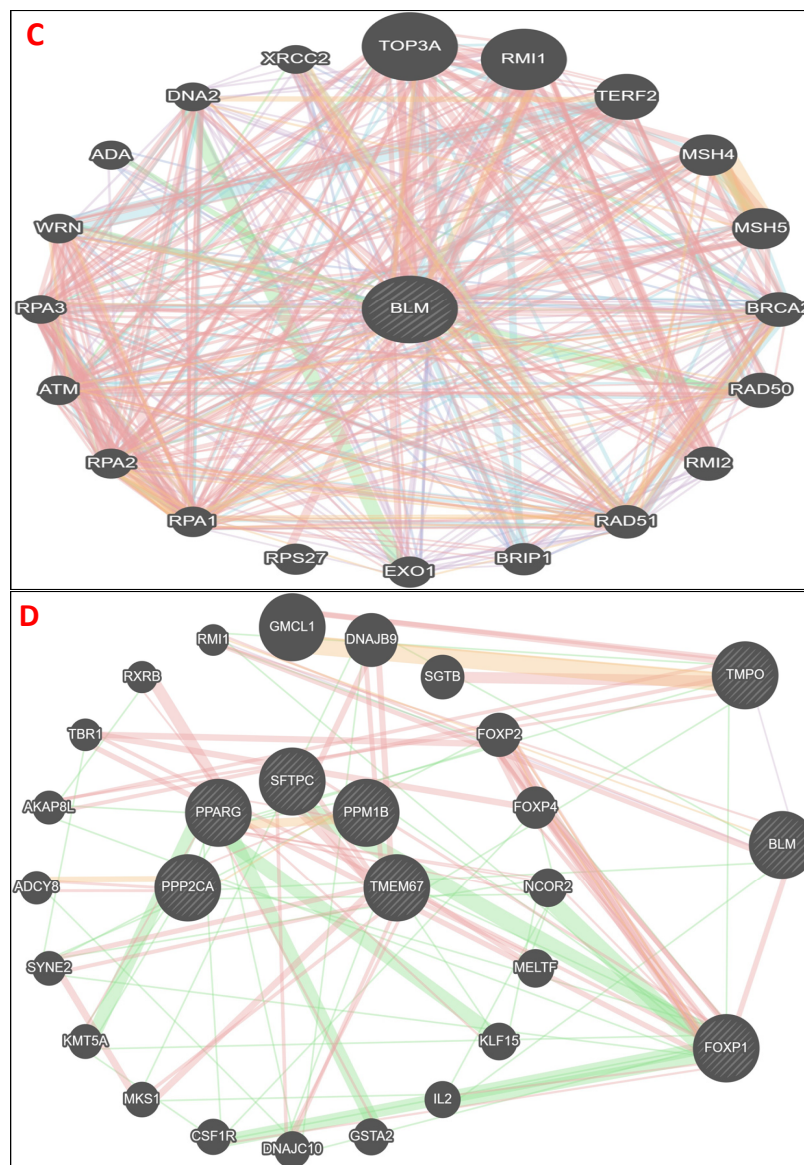


Fig. (6). **A.** Interactome network of commonly associated genes (WRN, DNAJC6, RET, NF1AND TMEM67) of prostate cancer and diabetes from ClinVar data; TMPO is shown as a common interacting partner; **B.** Interactome network of TMEM67, SFTPC, and PPMIB showing mainly genetic interactions (green) with one common FOXP1; **C.** Interactome network of BLM gene found as the only one common gene among prostate cancer, diabetes, and obesity that interacted 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.** Source: <https://genemania.org> (A higher resolution / colour version of this figure is available in the electronic copy of the article).

is common in diabetes ClinVar. *TMEM67* is also common among the ClinVar data sets and *SFTPC*, whereas *FOXP1* was found as the common interacting partner of *PP1MB*, *SFTPC*, and *TMEM67* (Fig. 6). *FOXP1* showed genetic interactions with both *SFTPC* and *PPM1B* (Fig. 6). 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 25th, 2022) (Tables 3 and 4). On the other hand, the BLM gene, which was found as the only common gene between prostate cancer and diabetes, interacts with both the earlier identified *FOXP1* (physical interactions shown in red color) and TMPO (co-expression shown in purple color) (Fig. 6C and D).

3.5. Machine Learning Result

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 prostate cancer dataset. The linearq12 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.

3.6. cBioPortal Results

On putting different queries of individual genes, it summarizes the genomic alterations across the whole sample,

Table 3. Functions and pathways involved in the common genes among ClinVar PCa and ClinVar diabetes from GeneCards.

Gene	Function	Pathways Involved
<i>WRN</i>	Magnesium and ATP-dependent DNA multifunctional helicase enzyme, 3'→5' exonuclease (ds DNA) at 5' overhangs. It 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. It also 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
<i>DNAJC6</i>	It promotes the 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
<i>RET</i>	It plays a role in neuronal navigation, cell proliferation, cell differentiation, and cell migration upon binding with glial cell-derived neurotrophic factor family ligands, PTK2/FAK1 phosphorylation, and regulates the balance between both cell death and survival. It is active without ligand and triggers apoptosis by a process that needs receptor intracellular caspase cleavage. It behaves as a dependence receptor in the presence of the ligand GDNF in somatotrophs, promotes survival, and downregulates growth hormone (GH) production, and if GDNF is absent, it triggers apoptosis.	• RET signaling • Developmental biology • Cytokine signaling in the immune system • Innate immune system • Signaling by GPCR • Tyrosine kinases / Adaptors • VEGF pathway (Tocris) • G-protein signaling_H-RAS regulation pathway • Aryl hydrocarbon receptor • Dopaminergic neurogenesis • Signaling events regulated by Ret tyrosine kinase • Sudden Infant Death Syndrome (SIDS) susceptibility pathways
<i>NF1</i>	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

(Table 3) contd....

Gene	Function	Pathways Involved
<i>TMPO</i>	It 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 a lamin binding domain, whereas beta and gamma isoforms are 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 (depolymerization of the nuclear lamina) - Apoptosis and autophagy

Table 4. Details of common genes retrieved using geneCards.

Gene	Function (UniProtKB)
<i>SFTPC</i>	Elevates alveolar stability by reducing surface tension at the air-liquid interface in the peripheral air spaces.
<i>PPM1B</i>	It encodes an enzyme that has a large specificity. This enzyme can dephosphorylate PRKAA1 and PRKAA2, CDK2 and CDK6 <i>in vitro</i> . Its dephosphorylation at 'Ser-17' in can inhibit TBK1-mediated antiviral signaling. It has an important role in terminating TNF-alpha-mediated NF-kappa-B activation by dephosphorylating and inactivating IKKB/IKKB.
<i>PPP2CA</i>	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.
<i>PPARG</i>	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.
<i>TMEM67</i>	Important for ciliary structure and function; may regulate ciliary membrane composition; during early ciliogenesis, it helps in centrosome migration to the apical cell surface; plays a role in maintaining cilia length and appropriate number through the control of centrosome duplication; needed for cell branching morphology; important in endoplasmic reticulum-associated degradation (ERAD) of surfactant protein C (SFTPC).
<i>FOXP1</i>	Transcriptional repressor, acts with CTBP1 and synergistically represses transcription; plays role in the specification and differentiation of lung epithelium; regulates fate of lung secretory epithelial cell and regeneration by restricting the goblet cell lineage program; important regulator of B-cell development in transcription; regulates proliferation of cardiac muscle cell; helps in columnar organization of spinal motor neurons and promoting both lateral motor neuron column (LMC) and preganglionic motor column (PGC) formation and is needed for appropriate motor axon projections; regulates PITX3 promoter activity; may aid in identity of midbrain in embryonic stem cell-derived dopamine neurons by regulating PITX3; down-regulates T follicular helper cells T(FH)s differentiation; maintains hair follicle stem cell quiescence; represses several pro-apoptotic genes transcription and work together with NF-kappa B-signaling to promote B-cell expansion through inhibition of caspase-dependent apoptosis.

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 show *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) (Fig. 7) in a prostate adenocarcinoma study by Abeshouse *et al.* (2018) [35].

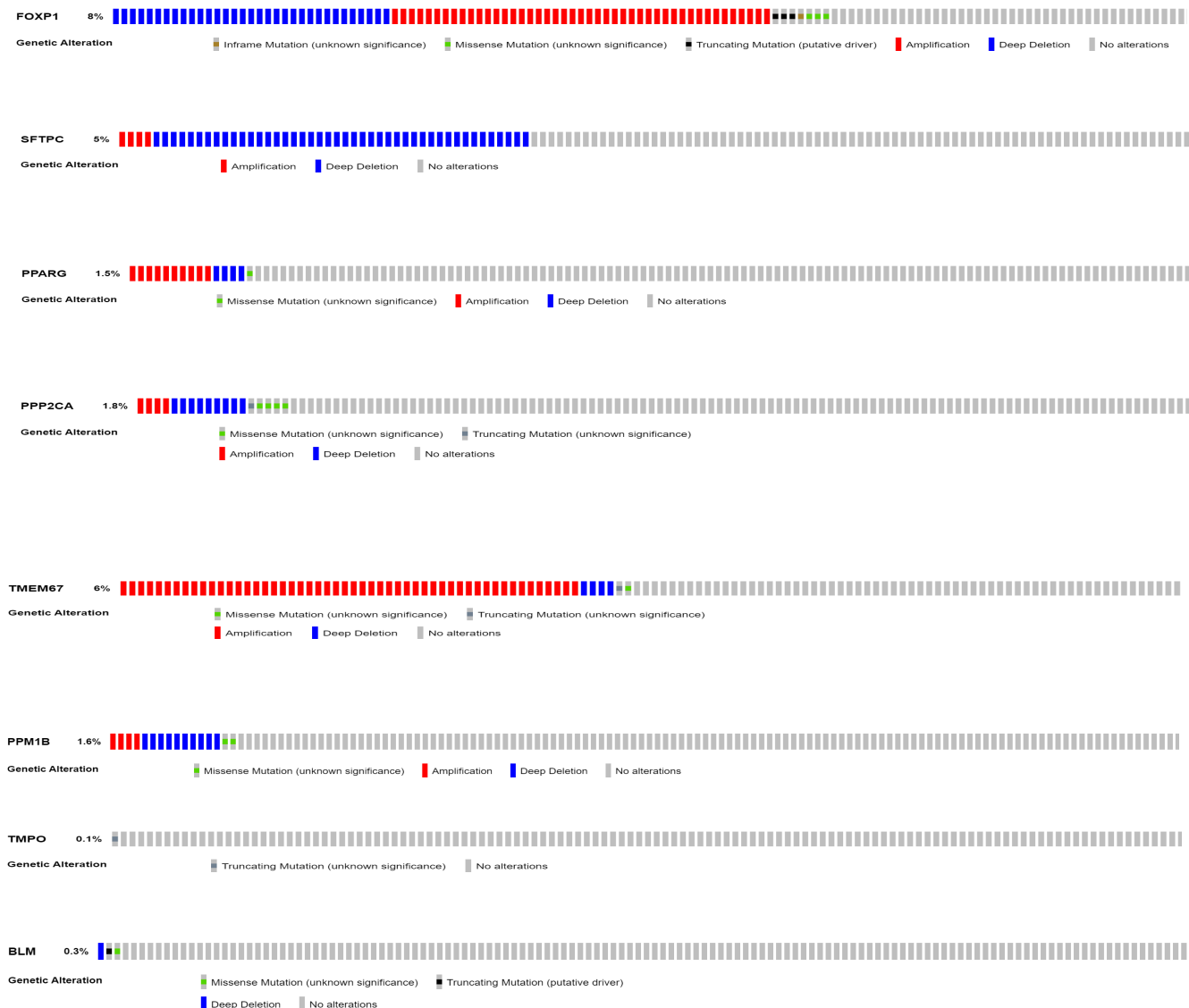


Fig. (7). Overall percentage and type of genetic mutations in *FOXPI*, *SFTPC*, *PPP2CA*, *PPARG*, *PPM1B*, *TMEM67*, *TMPO*, and *BLM* related to prostate adenocarcinoma. (Source: www.cbioportal.org/) (A higher resolution / colour version of this figure is available in the electronic copy of the article).

3.7. Gene Expression Patterns were Viewed Using GEPIA 2

Using normal expression profile graphs (TPM) and box plots with the help of GEPIA 2, we performed a comparative expression analysis. 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, a deviation limit is set; beyond that is known for outlier regions (abundant expression) (Fig. 8a). 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.

Comparing only tumor tissue expressions and matching TCGA normal and GTEx data, we found that the *PPP2CA* (5.5) and *FOXPI* (4.9) are highly expressing genes as compared to others, whereas *TMEM67* and *PPARG* are low expressing, and *TMPO*, *SFTC*, and *PPM1B* are least expressing genes (Fig. 8b). *TMPO* showed a very small change in expression profile from the normal but with a slightly higher deviation from the median expression in the tumor (Fig. 8c), and the most outlied expression was seen for the *PPARG* gene (Fig. 8a).

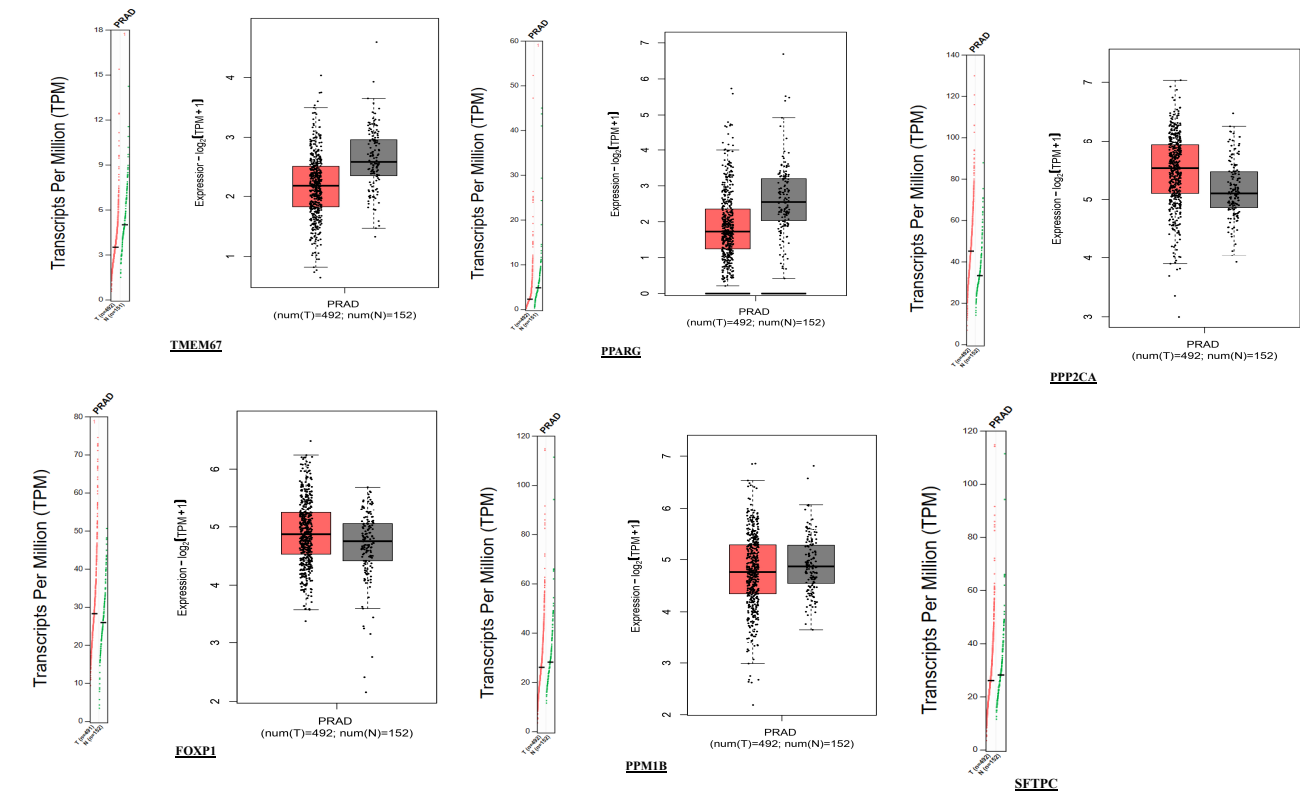


Fig. (8a). Box plots for expression of common genes compared to TCGA tumor-normal datasets of prostate adenocarcinomas (PRAD). Source: <http://gepia.cancer-pku.cn/> (A higher resolution / colour version of this figure is available in the electronic copy of the article).

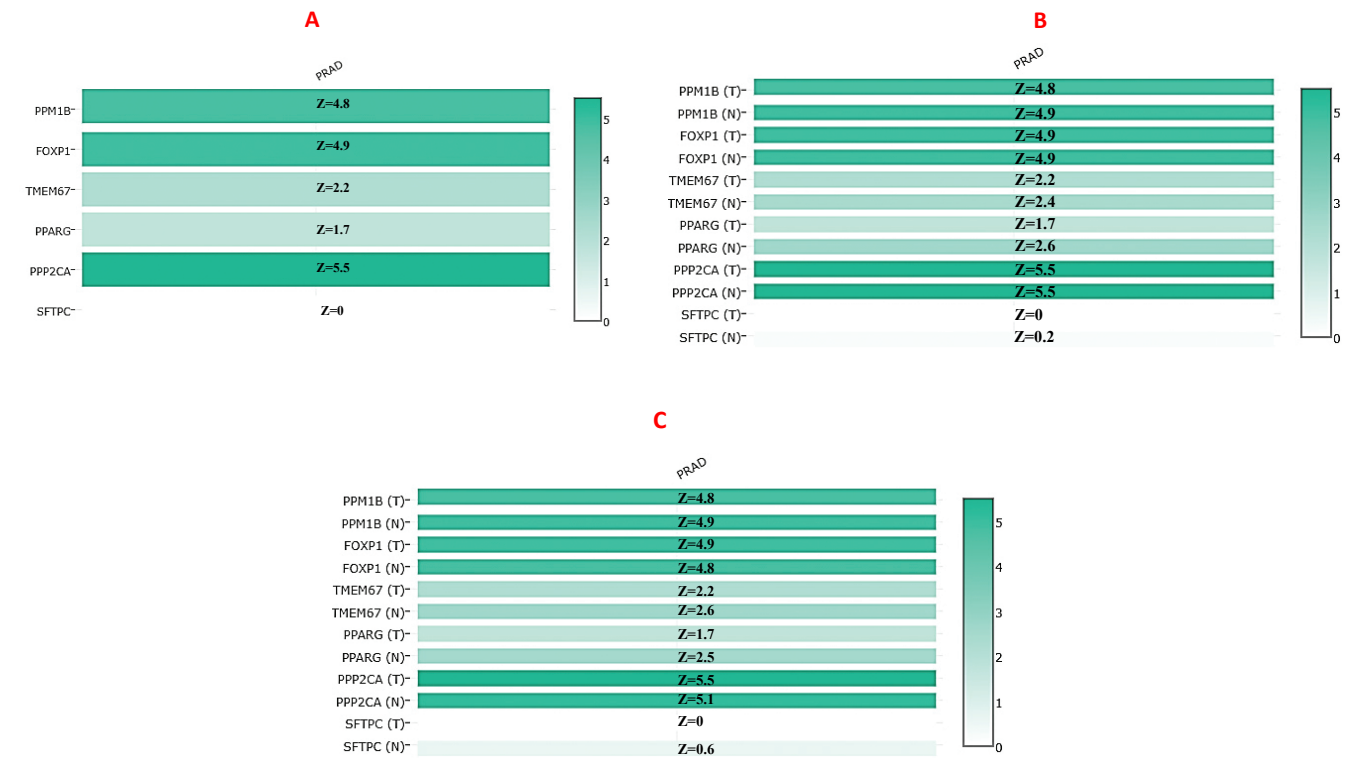


Fig. (8b). Box plots for expression of common genes in comparison to TCGA tumor-normal datasets of prostate adenocarcinomas (PRAD) and GTEx data (multiple genes based on Z scores) A. Only tumor tissue expression; B. Match TCGA normal data; C. Match TCGA normal and GTEx data. Source: <http://gepia.cancer-pku.cn/> (A higher resolution / colour version of this figure is available in the electronic copy of the article).

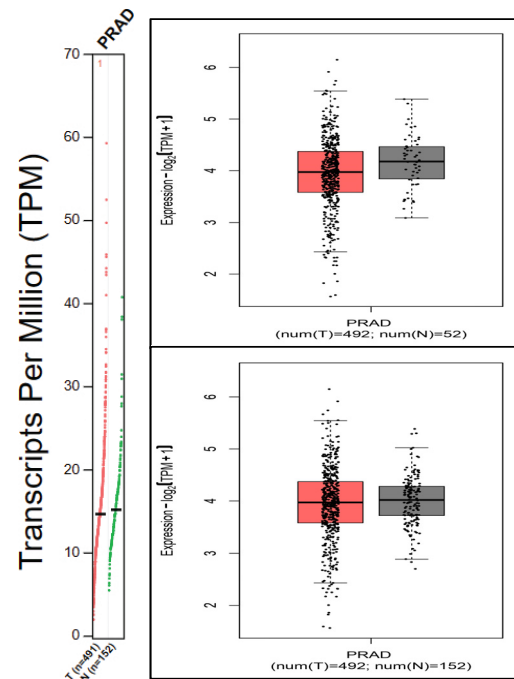


Fig. (8c). Expression in comparison to TCGA tumor-normal dataset of prostate adenocarcinomas (PRAD) dataset and GTEx data for TMPO gene. Source: <http://gepia.cancer-pku.cn/> (A higher resolution / colour version of this figure is available in the electronic copy of the article).

3.8. Top-ranking Genes Showed Vivid Clustering Coefficients for Ascertaining Survival Plots

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 (Fig. 9).

On analysing, we found that the disease-free survival for *FOXP1* is log-rank $p=0.005$, and *BLM* is log-rank with $p=0.00065$ (Fig. 10). In a comparison of overall survival and disease-free survival in reference to the significance, only *BLM* was observed to have a significant p-value in disease-free survival. The insignificance with respect to other genes of p-value may be because of the individual genetic variability. Each study exhibits clinical heterogeneity. Therefore, we need to include higher sample size studies. Comparison of other genes' overall survival (OS) and disease-free survival (DFS), *TMPO*, *TMEM67*, *TMPO*, *PPP2CA*, *PPM1B*, and *PPARG* using GEPIA 2 tool with PRAD dataset for both TCGA normalized data and GTEx data are shown in Supplementary Fig. (2).

Prostate cancer is age-linked and the second most common type of cancer among men. Most commonly, it is found in people above the age of 63 years. *PSA* is a widely used diagnostic biomarker with a high rate of false-positive results. Considering the heterogeneity of prostate cancer only, it does not seem justified to depend on a single biomarker for its screening and diagnosis for every patient. However, it seems difficult to replace such a biomarker as *PSA*, which has the least invasiveness and cost-effectiveness properties. A PRACTICAL consortium was established to get precise evaluations of risk related to different genetic variants and assessment of their combined associations of such variations by evaluating variants in large numbers of both cases and controls, and their GWAS datasets were published. Another resource called GWAS central (<http://www.gwascentral.org>)

can provide complete genetic association data at a summary level, which is designed to serve its maximum utility and safe open approach. Our study categorically was divided into two parts, wherein the former part is the use of machine learning algorithms to predict the chance of developing prostate cancer, and the later part is completely the correlational study on both publically available datasets and published RNA-seq and WES results of prostate cancer and diabetes to identify the common associations among them, giving a *vice versa* relationship to help in early diagnosis of PCa. Later, different bioinformatic tools were used to study their interaction networks and their expression patterns.

For the risk prediction model, individual sets of clinically verified variants (Clinvar) were first prepared for PCa, diabetes, and obesity from the NCBI. The datasets were retrieved by using keywords, viz. "*Prostate Cancer*", "*Diabetes*", and "*Obesity*". The data was later provided and labelled with binary/semi-binary scores (-1, -0.5, 0, 0.5, and 1) based on the clinical significance attributes as unknown significance, likely benign, benign, likely pathogenic, and pathogenic, considering the lowest to the highest risk ranges. The weak results predicted the chance of developing PCA when there are any changes in variables from the normal included in the machine learning. The prediction was done on the basis of the root mean square error standard deviation (PCa=0.000612372; DM=0.012077094; Obesity=0.00128582). As the data was normalized and plotted, we observed that the linear regression and tree-based algorithms were shown to have a lower RMSE than the multi-layer perceptron algorithm for the PCa dataset. As the linear regression has the lowest RMSE compared to multilayer perceptron and tree-based algorithms for the diabetes dataset, the tree-based algorithms also have the lowest RMSE as compared to linear regression and multilayer perceptron algorithms for the obesity dataset. We argue that this prediction invariably would help in clinical decision-making.

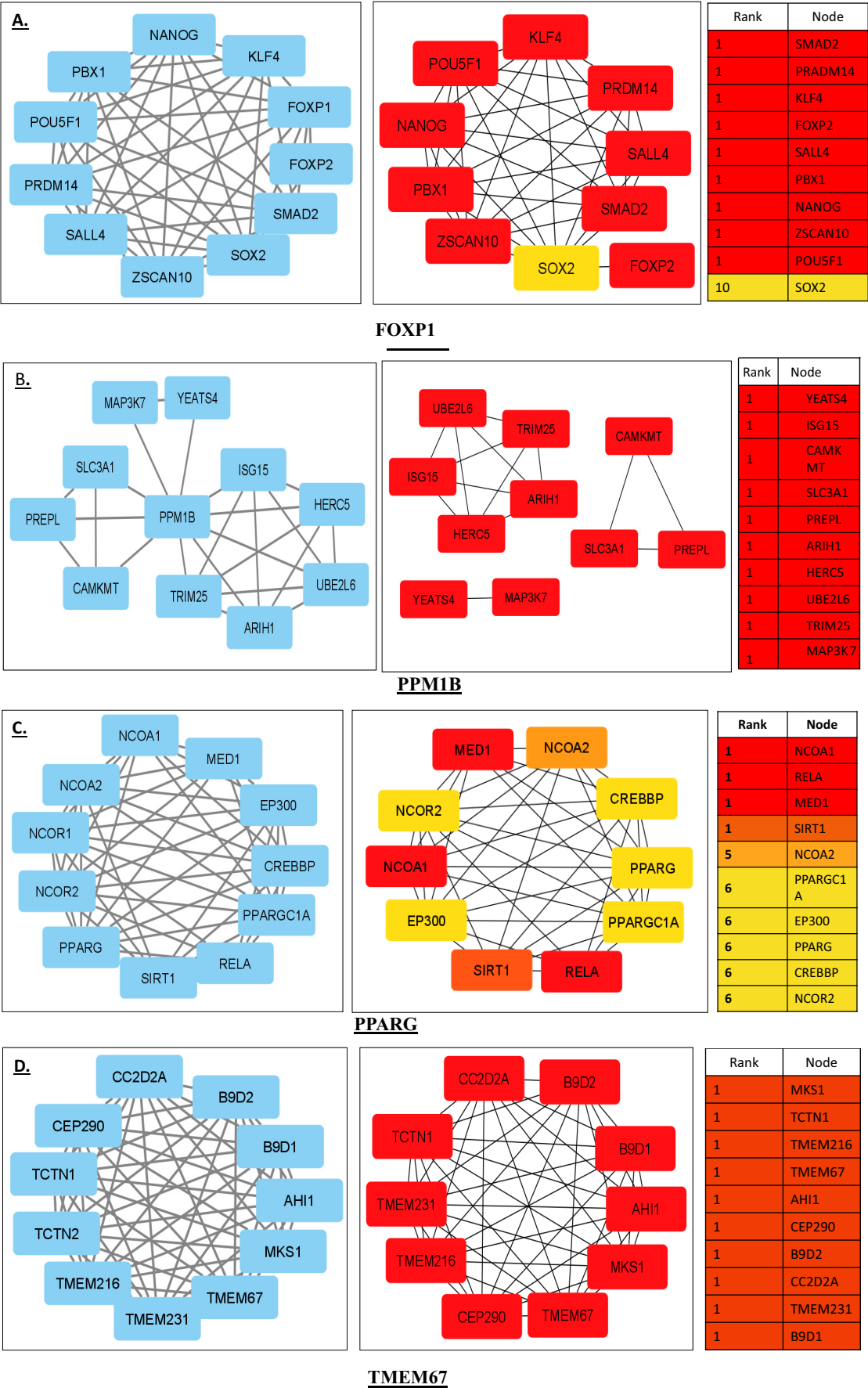


Fig. (9). contd....

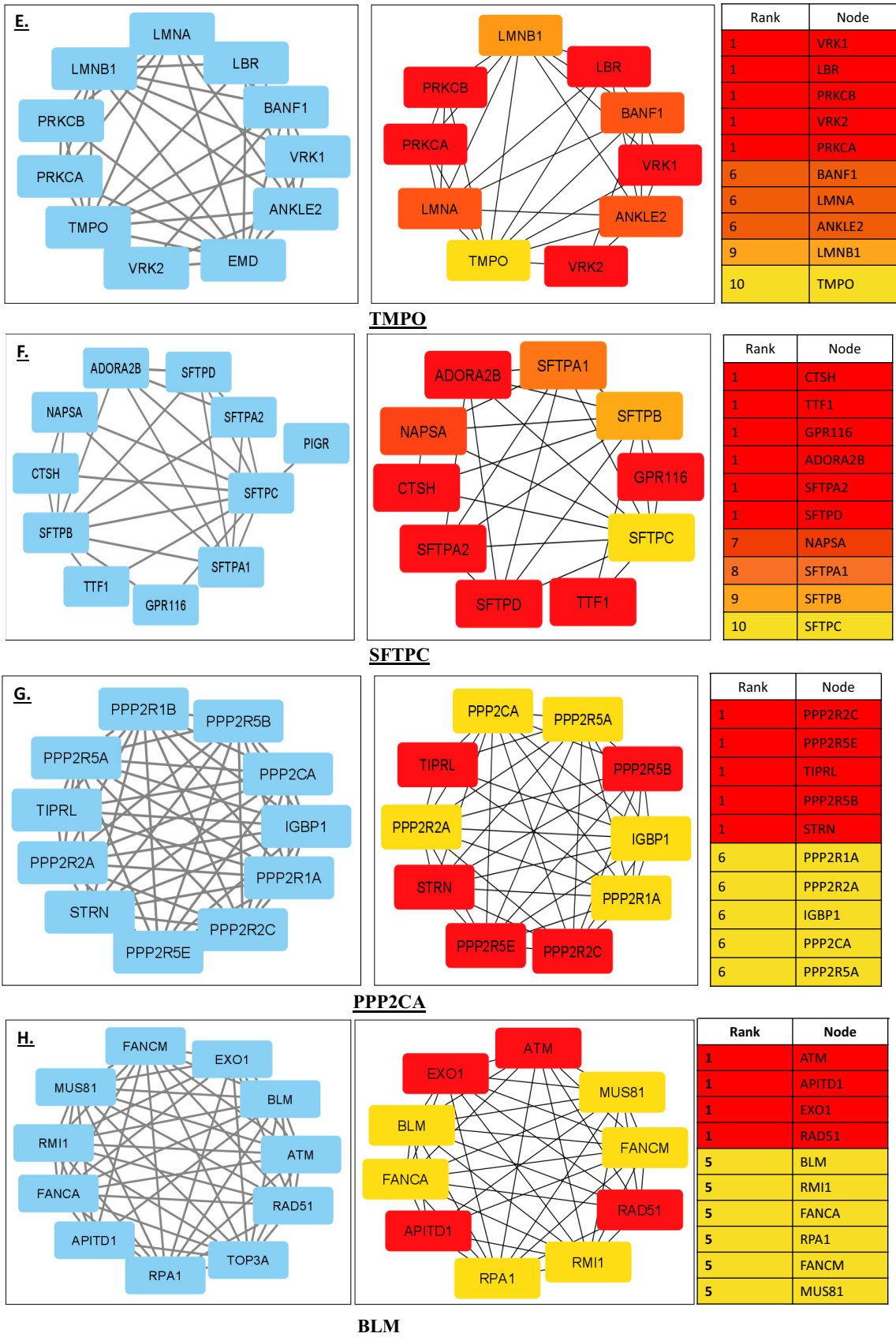


Fig. (9). Clustering coefficient (CC) networks of all the identified genes (A. *FOXPI*, B. *PPM1B*, C. *PPARG*, D. *TMEM67*, E. *TMPO*, F. *SFTPC*, G. *PPP2CA* and H. *BLM*) formed by cytoHubba plugin in Cytoscape to segregate and analyse highly interacting gene groups that might be participating in same biological function. The degree of clustering coefficient is represented in these networks through a colour scale ranging from red to yellow. The highest (top-ranked) CC is shown in red, moderate are in orange, and least are in yellow (low rank). Source: Software Cytoscape. (A higher resolution / colour version of this figure is available in the electronic copy of the article).

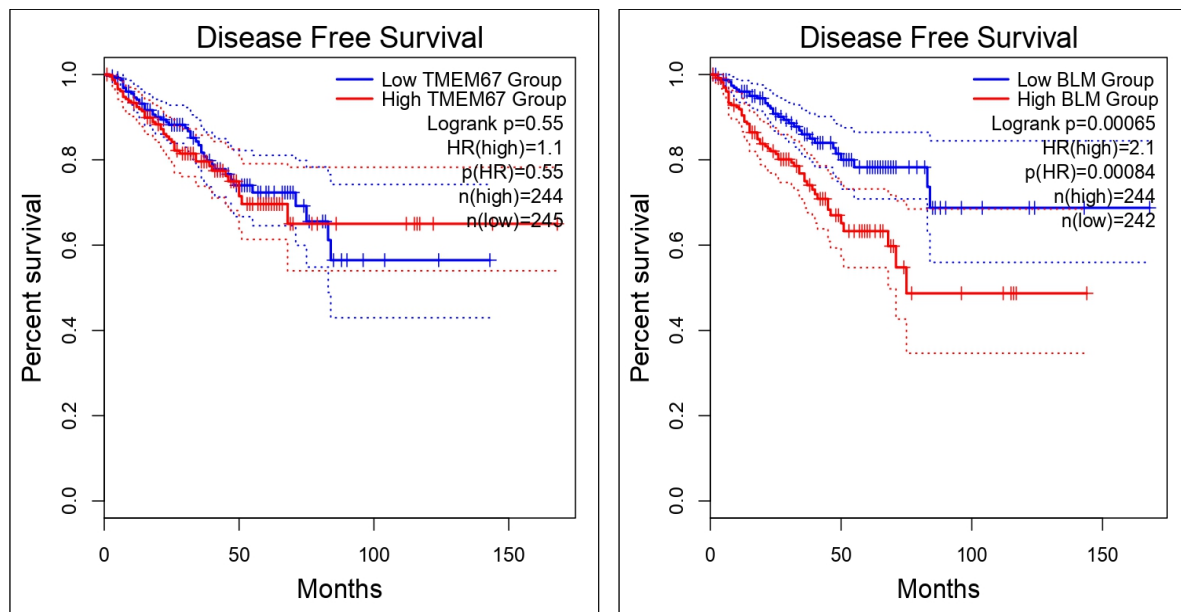


Fig. (10). Comparison of overall survival (OS) and disease-free survival (DFS) of *FOXPI* and *BLM* using the GEPIA 2 tool with *PRAD* dataset for both TCGA normalized data and GTEx data. Source: <http://gepia.cancer-pku.cn/> (A higher resolution / colour version of this figure is available in the electronic copy of the article).

CONCLUSION

In 1980, prostatic acid phosphatase (PAP) was the first described biomarker of PCa progression. Its levels were found to be high in metastatic patients. Later, this was replaced by PSA, a serine protease secreted by epithelial cells inside the prostate gland, and its level was also found to be increased compared to normal in PCa patients [42]. Although the management and survival rate of this disease have been improved with PSA screening since 1980, its limitations have kept the study of more precise biomarkers for prostate cancer evolving fast. The recent genomic and proteomic technologies have helped to understand the biology of PCa in better ways to contribute to biomarker discovery [42]. The PSA's role is controversial in screening asymptomatic men because of overdiagnosis and consequent overtreatment without any lethal ailment. However, it is not only broadly used as a biomarker for screening but also in therapy response monitoring and risk stratification for relapse [43, 44]. Identification of potential biomarkers could certainly improve screening, diagnosis, and prognosis; therefore, apart from other biomarkers, isoform assays of PSA are more concentrated in current studies, and those biomarkers could predict aggressiveness and put forward better treatment approaches [44, 42]. As per the approval of the US Food and Drug Administration or Clinical Laboratory Improvement Amendments-based laboratory-developed tests, recent markers can be identified from urine and serum or can be tissue-based. Various tests include *TMPRSS2-ERG* gene fusion test, Mi-Prostate score test, Oncotype DX test, ProMark test, ConfirmMDx test, Prolaris test, Prostate Core Mitomic test, 4K score test, Prostarix test, Decipher test. Other include α -Methylacyl coenzyme A racemase (AMACR). *PTEN* gene deletions are not approved by the US FDA but are commercially developed and available as Clinical Laboratory Improvement Amendments-based laboratory-developed tests. Only a few have been approved

yet by the US FDA, including *PSA*, *PHI*, and *PCA3*. Emerging biomarkers like tumor cells, microRNAs, and exosomes are still in the infant stage due to certain reasons like their flawed preclinical trials, inapt statistical analysis, etc. Hence, vigilant validation of each biomarker could only help to resolve prevailing or unmet challenges and lead to better diagnosis by clinicians, and the goal of personalized medicine could be achieved [42].

The ML heuristics has set a precedent to bring a transformation in cancer with data-driven pipelines for understanding possible causal relationships. What we sought to achieve through this pilot is to check whether candidate genes could be considered for precision scale modeling, and therefore, we employed 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 *FOXPI*. While *FOXPI* 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 among PCa, diabetes mellitus, and obesity, interacting with both *FOXPI* 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. Nevertheless, our work has limitations, where we lack clinical interpretation as we set a hypothesis that the differentially expressed genes (DEGs) harbour certain mutations that tend to be pathogenic. To achieve this, we need a large number of datasets to be screened. Our CAPCI and Systems Genomics Lab are working towards these goals.

On a granular level, this analysis has allowed us to bring insights into ascertaining three different phenotypes in the form of diabetes, PCa, and obesity. There is room for better ML-based integration, and time will reveal where we stand.

AUTHORS' CONTRIBUTIONS

BK, NS, and HB contributed equally. BK wrote the first draft. All other authors participated in lateral sections. PS proofread the manuscript before all authors agreed to the final version.

LIST OF ABBREVIATIONS

ADA	=	American Diabetes Association
DEGs	=	Differentially Expressed Genes
DFS	=	Disease-free Survival
GTEX	=	Genotype-Tissue Expression
GWAS	=	Genome-Wide Association Studies
HCC	=	Hepatocellular Carcinoma
HDL	=	High-density Lipoprotein
IR	=	Insulin Resistance
ML	=	Machine Learning
OS	=	Overall Survival
PAP	=	Prostatic Acid Phosphatase
PCa	=	Prostate Cancer
PRAD	=	Prostate Adenocarcinoma
PSA	=	Prostate-specific Antigen
SMGs	=	Significantly Mutated Genes
TCGA	=	The Cancer Genome Atlas
VTE	=	Venous Thromboembolism

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals/humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The authors confirm that the data supporting the findings of this research are available within the article.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

- [1] Rawla, P. Epidemiology of prostate cancer. *World J. Oncol.*, **2019**, *10*(2), 63-89.
<http://dx.doi.org/10.14740/wjon1191> PMID: 31068988
- [2] Abate-Shen, C.; Shen, M.M. Molecular genetics of prostate cancer. *Genes Dev.*, **2000**, *14*(19), 2410-2434.
<http://dx.doi.org/10.1101/gad.819500> PMID: 11018010
- [3] Pernar, C.H.; Ebot, E.M.; Wilson, K.M.; Mucci, L.A. The epidemiology of prostate cancer. *Cold Spring Harb. Perspect. Med.*, **2018**, *8*(12), a030361.
<http://dx.doi.org/10.1101/cshperspect.a030361> PMID: 29311132
- [4] Velonas, V.; Woo, H.; Remedios, C.; Assinder, S. Current status of biomarkers for prostate cancer. *Int. J. Mol. Sci.*, **2013**, *14*(6), 11034-11060.
<http://dx.doi.org/10.3390/ijms140611034> PMID: 23708103
- [5] Beck, T.; Hastings, R.K.; Gollapudi, S.; Free, R.C.; Brookes, A.J. GWAS Central: A comprehensive resource for the comparison and interrogation of genome-wide association studies. *Eur. J. Hum. Genet.*, **2014**, *22*(7), 949-952.
<http://dx.doi.org/10.1038/ejhg.2013.274> PMID: 24301061
- [6] Kote-Jarai, Z.; Easton, D.F.; Stanford, J.L.; Ostrander, E.A.; Schleutker, J.; Ingles, S.A.; Schaid, D.; Thibodeau, S.; Dörk, T.; Neal, D.; Cox, A.; Maier, C.; Vogel, W.; Guy, M.; Muir, K.; Lophatananon, A.; Kedda, M.-A.; Spurdle, A.; Steginga, S.; John, E.M.; Giles, G.; Hopper, J.; Chappuis, P.O.; Hutter, P.; Foulkes, W.D.; Hamel, N.; Salinas, C.A.; Koopmeiners, J.S.; Karyadi, D.M.; Johanneson, B.; Wahlfors, T.; Tammela, T.L.; Stern, M.C.; Corral, R.; McDonnell, S.K.; Schürmann, P.; Meyer, A.; Kuefer, R.; Leongamornlert, D.A.; Tymrakiewicz, M.; Liu, J.; O'Mara, T.; Gardiner, R.A.F.; Aitken, J.; Joshi, A.D.; Severi, G.; English, D.R.; Southey, M.; Edwards, S.M.; Al Olama, A.A.; Eeles, R.A.; Al Olama, A.A.; Eeles, R.A. Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRAC-TICAL Consortium. *Cancer Epidemiol. Biomarkers Prev.*, **2008**, *17*(8), 2052-2061.
<http://dx.doi.org/10.1158/1055-9965.EPI-08-0317> PMID: 18708398
- [7] Sharma, D.; Someshwar, S.; Kour, B.; Shukla, N.; Khilwani, B.; Vijay, M.; Gupta, A.; Ansari, A.; Vuree, S.; Kumar, A.; Singh, S. The CAPCI network: A CAncer Prostate Consortium of India for conducting next-generation genomic sequencing studies. *Cancer Health Disparities*, **2019**.
- [8] Chen, N.; Zhou, Q. The evolving Gleason grading system. *Chin. J. Cancer Res.*, **2016**, *28*(1), 58-64.
PMID: 27041927
- [9] 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. Automated deep-learning system for Gleason grading

- of prostate cancer using biopsies: A diagnostic study. *Lancet Oncol.*, **2020**, *21*(2), 233-241.
http://dx.doi.org/10.1016/S1470-2045(19)30739-9 PMID: 31926805
- [10] Iczkowski, K.A.; Lucia, M.S. Current perspectives on Gleason grading of prostate cancer. *Curr. Urol. Rep.*, **2011**, *12*(3), 216-222.
http://dx.doi.org/10.1007/s11934-011-0181-5 PMID: 21424766
- [11] Okotie, O.T.; Roehl, K.A.; Han, M.; Loeb, S.; Gashti, S.N.; Catalona, W.J. Characteristics of prostate cancer detected by digital rectal examination only. *Urology*, **2007**, *70*(6), 1117-1120.
http://dx.doi.org/10.1016/j.urol.2007.07.019 PMID: 18158030
- [12] Hu, Y.; Liu, W.; Chen, Y.; Zhang, M.; Wang, L.; Zhou, H.; Wu, P.; Teng, X.; Dong, Y.; Zhou, J.; Xu, H.; Zheng, J.; Li, S.; Tao, T.; Hu, Y.; Jia, Y. Combined use of fasting plasma glucose and glycated hemoglobin A1c in the screening of diabetes and impaired glucose tolerance. *Acta Diabetol.*, **2010**, *47*(3), 231-236.
http://dx.doi.org/10.1007/s00592-009-0143-2 PMID: 19760291
- [13] Schnedl, W.J.; Liebming, A.; Roller, R.E.; Lipp, R.W.; Krejs, G.J. Hemoglobin variants and determination of glycated hemoglobin (HbA1c). *Diabetes Metab. Res. Rev.*, **2001**, *17*(2), 94-98.
http://dx.doi.org/10.1002/dmrr.186 PMID: 11307174
- [14] Donadon, V.; Balbi, M.; Valent, F.; Avogaro, A. Glycated hemoglobin and antidiabetic strategies as risk factors for hepatocellular carcinoma. *World J. Gastroenterol.*, **2010**, *16*(24), 3025-3032.
http://dx.doi.org/10.3748/wjg.v16.i24.3025 PMID: 20572306
- [15] Kumar, P.R.; Bhansali, A.; Ravikiran, M.; Bhansali, S.; Dutta, P.; Thakur, J.S.; Sachdeva, N.; Bhadada, S.K.; Walia, R. Utility of glycated hemoglobin in diagnosing type 2 diabetes mellitus: A community-based study. *J. Clin. Endocrinol. Metab.*, **2010**, *95*(6), 2832-2835.
http://dx.doi.org/10.1210/jc.2009-2433 PMID: 20371663
- [16] Lippi, G.; Targher, G. Glycated hemoglobin (HbA1c): Old dogmas, a new perspective? *cclm*, **2010**, *48*(5), 609-614.
http://dx.doi.org/10.1515/CCLM.2010.144 PMID: 20464776
- [17] Peila, R.; Rohan, T.E. Diabetes, glycated hemoglobin, and risk of cancer in the UK biobank study. *Cancer Epidemiol. Biomarkers Prev.*, **2020**, *29*(6), 1107-1119.
http://dx.doi.org/10.1158/1055-9965.EPI-19-1623 PMID: 32179703
- [18] Twig, G.; Afek, A.; Shamiss, A.; Derazne, E.; Tzur, D.; Gordon, B.; Tirosch, A. White blood cells count and incidence of type 2 diabetes in young men. *Diabetes Care*, **2013**, *36*(2), 276-282.
http://dx.doi.org/10.2337/dc11-2298 PMID: 22961572
- [19] Vozarova, B.; Weyer, C.; Lindsay, R.S.; Pratley, R.E.; Bogardus, C.; Tataranni, P.A. High white blood cell count is associated with a worsening of insulin sensitivity and predicts the development of type 2 diabetes. *Diabetes*, **2002**, *51*(2), 455-461.
http://dx.doi.org/10.2337/diabetes.51.2.455 PMID: 11812755
- [20] Blix, K.; Jensvoll, H.; Brækkan, S.K.; Hansen, J.B. White blood cell count measured prior to cancer development is associated with future risk of venous thromboembolism--the Tromsø study. *PLoS One*, **2013**, *8*(9), e73447.
http://dx.doi.org/10.1371/journal.pone.0073447 PMID: 24023876
- [21] Nencioni, A.; Caffa, I.; Cortellino, S.; Longo, V.D. Fasting and cancer: Molecular mechanisms and clinical application. *Nat. Rev. Cancer*, **2018**, *18*(11), 707-719.
http://dx.doi.org/10.1038/s41568-018-0061-0 PMID: 30327499
- [22] Jayedi, A.; Djafarian, K.; Rezagholizadeh, F.; Mirzababaei, A.; Hajimohammadi, M.; Shab-Bidar, S. Fasting blood glucose and risk of prostate cancer: A systematic review and meta-analysis of dose-response. *Diabetes Metab.*, **2018**, *44*(4), 320-327.
http://dx.doi.org/10.1016/j.diabet.2017.09.004 PMID: 29074328
- [23] Ghazanfari, Z.; Haghdost, A.A.; Alizadeh, S.M.; Atapour, J.; Zolala, F. A comparison of HbA1c and fasting blood sugar tests in general population. *Int. J. Prev. Med.*, **2010**, *1*(3), 187-194.
PMID: 21566790
- [24] Cheng, L.; Zhuang, H.; Ju, H.; Yang, S.; Han, J.; Tan, R.; Hu, Y. Exposing the causal effect of body mass index on the risk of type 2 diabetes mellitus: a mendelian randomization study. *Front. Genet.*, **2019**, *10*, 94.
http://dx.doi.org/10.3389/fgene.2019.00094 PMID: 30891058
- [25] Khandekar, M.J.; Cohen, P.; Spiegelman, B.M. Molecular mechanisms of cancer development in obesity. *Nat. Rev. Cancer*, **2011**, *11*(12), 886-895.
http://dx.doi.org/10.1038/nrc3174 PMID: 22113164
- [26] Bianchini, F.; Kaaks, R.; Vainio, H. Overweight, obesity, and cancer risk. *Lancet Oncol.*, **2002**, *3*(9), 565-574.
http://dx.doi.org/10.1016/S1470-2045(02)00849-5 PMID: 12217794
- [27] Garg, S.K.; Maurer, H.; Reed, K.; Selagamsetty, R. Diabetes and cancer: Two diseases with obesity as a common risk factor. *Diabetes Obes. Metab.*, **2014**, *16*(2), 97-110.
http://dx.doi.org/10.1111/dom.12124 PMID: 23668396
- [28] Hopkins, B.D.; Goncalves, M.D.; Cantley, L.C. Obesity and cancer mechanisms: Cancer metabolism. *J. Clin. Oncol.*, **2016**, *34*(35), 4277-4283.
http://dx.doi.org/10.1200/JCO.2016.67.9712 PMID: 27903152
- [29] Liang, Y.; Ketchum, N.S.; Goodman, P.J.; Klein, E.A.; Thompson, I.M., Jr Is there a role for body mass index in the assessment of prostate cancer risk on biopsy? *J. Urol.*, **2014**, *192*(4), 1094-1099.
http://dx.doi.org/10.1016/j.juro.2014.04.015 PMID: 24747090
- [30] Taghizadeh, N.; Boezen, H.M.; Schouten, J.P.; Schröder, C.P.; Vries, E.G.E.; Vonk, J.M. BMI and lifetime changes in BMI and cancer mortality risk. *PLoS One*, **2015**, *10*(4), e0125261.
http://dx.doi.org/10.1371/journal.pone.0125261 PMID: 25881129
- [31] 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. Serum cholesterol and risk of high-grade prostate cancer: Results from the REDUCE study. *Prostate Cancer Prostatic Dis.*, **2018**, *21*(2), 252-259.
http://dx.doi.org/10.1038/s41391-017-0030-9 PMID: 29282360
- [32] Saxena, A.; Mathur, N.; Tiwari, P.; Mathur, S.K. Whole transcriptome RNA-seq reveals key regulatory factors involved in type 2 diabetes pathology in peripheral fat of Asian Indians. *Sci. Rep.*, **2021**, *11*(1), 10632.
http://dx.doi.org/10.1038/s41598-021-90148-z PMID: 34017037
- [33] Warde-Farley, D.; Donaldson, S.L.; Comes, O.; Zuberi, K.; Badrawi, R.; Chao, P.; Franz, M.; Grouios, C.; Kazi, F.; Lopes, C.T.; Maitland, A.; Mostafavi, Z.; Montojo, J.; Shao, Q.; Wright, G.; Bader, G.D.; Morris, Q. The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **2010**, *38*(S2), W214-W220.
http://dx.doi.org/10.1093/nar/gkq537 PMID: 20576703
- [34] Armenia, J.; Wankowicz, S.A.M.; Liu, D.; Gao, J.; Kundra, R.; Reznik, E.; Chatila, W.K.; Chakravarty, D.; Han, G.C.; Coleman, I.; Montgomery, B.; Pritchard, C.; Morrissey, C.; Barbieri, C.E.; Beltran, H.; Shoner, A.; Zafeiriou, Z.; Miranda, J.; Bielski, C.M.; Penson, A.V.; Tolonen, C.; Huang, F.W.; Robinson, D.; Wu, Y.M.; Lonigro, R.; Garraway, L.A.; Demicheli, F.; Kantoff, P.W.; Taplin, M.E.; Abida, W.; Taylor, B.S.; Scher, H.I.; Nelson, P.S.; de Bono, J.S.; Rubin, M.A.; Sawyers, C.L.; Chinnaiyan, A.M.; Schultz, N.; Van Allen, E.M. The long tail of oncogenic drivers in prostate cancer. *Nat. Genet.*, **2018**, *50*(5), 645-651.
http://dx.doi.org/10.1038/s41588-018-0078-z PMID: 29610475
- [35] Abeshouse, A.; Ahn, J.; Akbani, R.; Ally, A.; Amin, S.; Andry, C.D.; Annala, M.; Aprikian, A.; Armenia, J.; Arora, A.; Auman, J.T.; Balasundaram, M.; Balu, S.; Barbieri, C.E.; Bauer, T.; Benz, C.C.; Bergeron, A.; Beroukhi, R.; Berrios, M.; Bivoli, A.; Bodenheimer, T.; Boice, L.; Bootwalla, M.S.; Borges dos Reis, R.; Boutros, P.C.; Bowen, J.; Bowlby, R.; Boyd, J.; Bradley, R.K.; Breggia, A.; Brimo, F.; Bristow, C.A.; Brooks, D.; Broom, B.M.; Bryce, A.H.; Bubley, G.; Burks, E.; Butterfield, Y.S.N.; Button, M.; Canes, D.; Carlotti, C.G.; Carlsen, R.; Carmel, M.; Carroll, P.R.; Carter, S.L.; Cartun, R.; Carver, B.S.; Chan, J.M.; Chang, M.T.; Chen, Y.; Cherniack, A.D.; Chevalier, S.; Chin, L.; Cho, J.; Chu, A.; Chuah, E.; Chudamani, S.; Cibulskis, K.; Ciriello, G.; Clarke, A.; Cooperberg, M.R.; Corcoran, N.M.; Costello, A.J.; Cowan, J.; Crain, D.; Curley, E.; David, K.; Demchok, J.A.; Demicheli, F.; Dhalla, N.; Dhir, R.; Douek, A.; Drake, B.; Dvinge, H.; Dyakova, N.; Felau, I.; Ferguson, M.L.; Frazer, S.; Freedland, S.; Fu, Y.; Gabriel, S.B.; Gao, J.; Gardner, J.; Gastier-Foster, J.M.; Gehlenborg, N.; Gerken, M.; Gerstein, M.B.; Getz, G.; Godwin, A.K.; Gopalan, A.; Graefen, M.; Graim, K.; Gribbin, T.; Guin, R.; Gupta, M.; Hadjipanayis, A.; Haider, S.; Hamel, L.; Hayes, D.N.; Heiman, D.I.; Hess, J.; Hoadley, K.A.; Holbrook, A.H.; Holt, R.A.; Holway, A.; Hovens, C.M.; Hoyle, A.P.; Huang, M.; Hutter, C.M.; Ittmann, M.; Iype, L.; Jefferys, S.R.; Jones, C.D.; Jones, S.J.M.; Juhl, H.; Kahles, A.; Kane, C.J.; Kasaian, K.; Kerger, M.; Khurana, E.; Kim, J.; Klein, R.J.; Kucherlapati, R.; Lacombe, L.; Ladanyi, M.; Lai, P.H.; Laird, P.W.; Lander, E.S.; Latour, M.; Lawrence, M.S.; Lau, K.; LeBien, T.; Lee, D.; Lee, S.; Lehmann,

- K-V.; Leraas, K.M.; Leshchiner, I.; Leung, R.; Libertino, J.A.; Lichtenberg, T.M.; Lin, P.; Linehan, W.M.; Ling, S.; Lippman, S.M.; Liu, J.; Liu, W.; Lochovsky, L.; Loda, M.; Logothetis, C.; Lolla, L.; Longacre, T.; Lu, Y.; Luo, J.; Ma, Y.; Mahadeshwar, H.S.; Mallery, D.; Mariamidze, A.; Marra, M.A.; Mayo, M.; McCall, S.; McKercher, G.; Meng, S.; Mes-Masson, A.-M.; Merino, M.J.; Meyerson, M.; Mieczkowski, P.A.; Mills, G.B.; Shaw, K.R.M.; Minner, S.; Moinzadeh, A.; Moore, R.A.; Morris, S.; Morrison, C.; Mose, L.E.; Mungall, A.J.; Murray, B.A.; Myers, J.B.; Naresh, R.; Nelson, J.; Nelson, M.A.; Nelson, P.S.; Newton, Y.; Noble, M.S.; Noushmehr, H.; Nykter, M.; Pantazi, A.; Parfenov, M.; Park, P.J.; Parker, J.S.; Paulauskis, J.; Penny, R.; Perou, C.M.; Piché, A.; Pihl, T.; Pinto, P.A.; Prandi, D.; Protopopov, A.; Ramirez, N.C.; Rao, A.; Rathmell, W.K.; Rättsch, G.; Ren, X.; Reuter, V.E.; Reynolds, S.M.; Rhie, S.K.; Rieger-Christ, K.; Roach, J.; Robertson, A.G.; Robinson, B.; Rubin, M.A.; Saad, F.; Sadeghi, S.; Saksena, G.; Saller, C.; Salner, A.; Sanchez-Vega, F.; Sander, C.; Sandusky, G.; Sauter, G.; Sboner, A.; Scardino, P.; Scarlata, E.; Schein, J.E.; Schlomm, T.; Schmidt, L.S.; Schultz, N.; Schumacher, S.E.; Seidman, J.; Neder, L.; Seth, S.; Sharp, A.; Shelton, C.; Shelton, T.; Shen, H.; Shen, R.; Sherman, M.; Sheth, M.; Shi, Y.; Shih, J.; Shmulevich, I.; Simko, J.; Simon, R.; Simons, J.V.; Sipahimalani, P.; Skelly, T.; Sofia, H.J.; Soloway, M.G.; Song, X.; Sorcini, A.; Sougnez, C.; Stepa, S.; Stewart, C.; Stewart, J.; Stuart, J.M.; Sullivan, T.B.; Sun, C.; Sun, H.; Tam, A.; Tan, D.; Tang, J.; Tarnuzzer, R.; Tarvin, K.; Taylor, B.S.; Tebbag, P.; Tenggara, I.; Têtu, B.; Tewari, A.; Thiessen, N.; Thompson, T.; Thorne, L.B.; Tirapelli, D.P.; Tomlins, S.A.; Trevisan, F.A.; Troncoso, P.; True, L.D.; Tsourlakis, M.C.; Tyekucheva, S.; Van Allen, E.; Van Den Berg, D.J.; Veluvolu, U.; Verhaak, R.; Vocke, C.D.; Voet, D.; Wan, Y.; Wang, Q.; Wang, W.; Wang, Z.; Weinhold, N.; Weinstein, J.N.; Weisenberger, D.J.; Wilkerson, M.D.; Wise, L.; Witte, J.; Wu, C.-C.; Wu, J.; Wu, Y.; Xu, A.W.; Yadav, S.S.; Yang, L.; Yang, L.; Yau, C.; Ye, H.; Yena, P.; Zeng, T.; Zenklusen, J.C.; Zhang, H.; Zhang, J.; Zhang, J.; Zhang, W.; Zhong, Y.; Zhu, K.; Zmuda, E. The molecular taxonomy of primary prostate cancer. *Cell*, **2015**, *163*(4), 1011-1025.
<http://dx.doi.org/10.1016/j.cell.2015.10.025> PMID: 26544944
- [36] 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.; Fleming, J.; Siminoff, L.; Traino, H.; Mosavel, M.; Barker, L.; Jewell, S.; Rohrer, D.; Maxim, D.; Filkins, D.; Harbach, P.; Cortadillo, E.; Berghuis, B.; Turner, L.; Hudson, E.; Feenstra, K.; Sobin, L.; Robb, J.; Branton, P.; Korzeniewski, G.; Shive, C.; Tabor, D.; Qi, L.; Groch, K.; Nampally, S.; Buia, S.; Zimmerman, A.; Smith, A.; Burges, R.; Robinson, K.; Valentino, K.; Bradbury, D.; Cosentino, M.; Diaz-Mayoral, N.; Kennedy, M.; Engel, T.; Williams, P.; Erickson, K.; Ardlie, K.; Winckler, W.; Getz, G.; DeLuca, D.; MacArthur, D.; Kellis, M.; Thomson, A.; Young, T.; Gelfand, E.; Donovan, M.; Meng, Y.; Grant, G.; Mash, D.; Marcus, Y.; Basile, M.; Liu, J.; Zhu, J.; Tu, Z.; Cox, N.J.; Nicolae, D.L.; Gamazon, E.R.; Im, H.K.; Konkashbaev, A.; Pritchard, J.; Stevens, M.; Flutre, T.; Wen, X.; Dermitzakis, E.T.; Lappalainen, T.; Guigo, R.; Monlong, J.; Sammeth, M.; Koller, D.; Battle, A.; Mostafavi, S.; McCarthy, M.; Rivas, M.; Maller, J.; Rusyn, I.; Nobel, A.; Wright, F.; Shabalina, A.; Feolo, M.; Sharopova, N.; Sturcke, A.; Paschal, J.; Anderson, J.M.; Wilder, E.L.; Derr, L.K.; Green, E.D.; Struwing, J.P.; Temple, G.; Volpi, S.; Boyer, J.T.; Thomson, E.J.; Guyer, M.S.; Ng, C.; Abdallah, A.; Colantuoni, D.; Insel, T.R.; Koester, S.E.; Little, A.R.; Bender, P.K.; Lehner, T.; Yao, Y.; Compton, C.C.; Vaught, J.B.; Sawyer, S.; Lockhart, N.C.; Demchok, J.; Moore, H.F. The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **2013**, *45*(6), 580-585.
<http://dx.doi.org/10.1038/ng.2653> PMID: 23715323
- [37] Tang, Z.; Li, C.; Kang, B.; Gao, G.; Li, C.; Zhang, Z. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.*, **2017**, *45*(W1), W98-W102.
<http://dx.doi.org/10.1093/nar/gkx247> PMID: 28407145
- [38] Shen, Y.; Liu, J.; Zhang, L.; Dong, S.; Zhang, J.; Liu, Y.; Zhou, H.; Dong, W. Identification of potential biomarkers and survival analysis for head and neck squamous cell carcinoma using bioinformatics strategy: A study based on TCGA and GEO datasets. *BioMed Res. Int.*, **2019**, *2019*, 1-14.
<http://dx.doi.org/10.1155/2019/7376034> PMID: 31485443
- [39] Li, C.; Tang, Z.; Zhang, W.; Ye, Z.; Liu, F. GEPIA2021: integrating multiple deconvolution-based analysis into GEPIA. *Nucleic Acids Res.*, **2021**, *49*(W1), W242-W246.
<http://dx.doi.org/10.1093/nar/gkab418> PMID: 34050758
- [40] Assenov, Y.; Ramirez, F.; Schelhorn, S.E.; Lengauer, T.; Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics*, **2008**, *24*(2), 282-284.
<http://dx.doi.org/10.1093/bioinformatics/btm554> PMID: 18006545
- [41] Gollapalli, P.; G, T.S.; H, M.; Shetty, P.; N, S.K. Network topology analysis of essential genes interactome of *Helicobacter pylori* to explore novel therapeutic targets. *Microb. Pathog.*, **2021**, *158*, 105059.
<http://dx.doi.org/10.1016/j.micpath.2021.105059> PMID: 34157412
- [42] Saini, S. PSA and beyond: Alternative prostate cancer biomarkers. *Cell Oncol.*, **2016**, *39*(2), 97-106.
<http://dx.doi.org/10.1007/s13402-016-0268-6> PMID: 26790878
- [43] Duffy, M.J. Biomarkers for prostate cancer: Prostate-specific antigen and beyond. *Clin. Chem. Lab. Med.*, **2020**, *58*(3), 326-339.
<http://dx.doi.org/10.1515/cclm-2019-0693> PMID: 31714881
- [44] Hatakeyama, S.; Yoneyama, T.; Tobisawa, Y.; Ohya, C. Recent progress and perspectives on prostate cancer biomarkers. *Int. J. Clin. Oncol.*, **2017**, *22*(2), 214-221.
<http://dx.doi.org/10.1007/s10147-016-1049-y> PMID: 27730440