Identification of Biomarker using mRNA Seq, miRNA and clinical data of Prostate Cancer using similarity network fusion, ML approach and survival analysis

Report of Major Project

(BT4094D)

Submitted by

Lakavath Rajendar (B191225BT)
Nived S (B190014BT)
Sanandh Kumar M (B191114BT)
Sarath P(B190522BT)

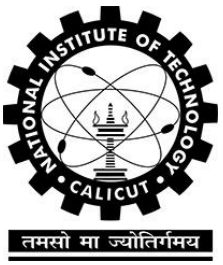
Under the guidance of

Dr. Ravindra Kumar

Assistant Professor

School of Biotechnology

NIT CALICUT



SCHOOL OF BIOTECHNOLOGY NATIONAL
INSTITUTE OF TECHNOLOGY CALICUT NITC
CAMPUS P.O KERALA - 673 601, INDIA

April 2023

SCHOOL OF BIOTECHNOLOGY
National Institute of Technology
Calicut

CERTIFICATE

This is to certify that the bonafide report of the Major project work (BT4094D) entitled **"Identification of Biomarker using mRNA Seq, miRNA and clinical data of Prostate Cancer using similarity network fusion, ML approach and survival analysis"** in requirement for the partial fulfillment of semester VI in Biotechnology from the School of Biotechnology, National Institute of Technology, Calicut by the following students:

1. Lakavath Rajendar (B191225BT)
2. Nived S (B190014BT)
3. Sanandh Kumar M (B191114BT)
4. Sarath P (B190522BT)

Guide	Course Coordinator	Head of the Department
Dr. Ravindra Kumar	**Dr. Md. Anaul Kabir**	**Dr. Rathinasamy K**

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to Dr. Ravindra Kumar, to have taken us in as his mentees and be our guide for this project. From the start, he has helped us steer this project in the right direction. His immense knowledge, depth of concepts, profound experience and professional expertise in the field of biotechnology has assisted us every step of the way in the process. We also wish to extend our feelings of gratitude to Dr. Rathinasamy K, the HOD of the School of Biotechnology, National Institute of Technology, Calicut for providing us with the opportunity to work on this project. We also want to thank Dr. M. Anaul Kabir for demonstrating the research methodology and imparting to us the best practices for conducting a solid research project. We also want to thank Ashitha Washington for helping us understand the research concepts despite her busy doctoral schedule.

1. Lakavath Rajendar (B191225BT)

2. Nived S (B190014BT)

3. Sanandh Kumar M (B191114BT)

4. Sarath P (B190522BT)

Place: NIT, Calicut

Date: 12/04/23

DECLARATION

We, Lakavath Rajendar (B191225BT) ,Nived S (B190014BT), Sanandh Kumar M (B191114BT), Sarath P (B190522BT) certify that, to the best of our knowledge, the current report or the Main Project does not infringe upon any copyright, nor does it violate any ethical regulations.

1. Lakavath Rajendar (B191225BT)

2. Nived S (B190014BT)

3. Sanandh Kumar M (B191114BT)

4. Sarath P (B190522BT)

Place: Calicut

Date: 12/04/23

ABSTRACT

Prostate cancer is a type of cancer that is very common in men. It affects about one in every five men and causes around one in every five cancer-related deaths in men worldwide. There are many factors that can increase the risk of developing prostate cancer, such as smoking, obesity, age, and exposure to chemicals and radiation. However, the exact changes that occur at the molecular level to cause prostate cancer are not yet fully understood. This work involves using data from three sources: miRNA, mRNA, and methylation. We used a tool called TCGABiolinks to prepare the data for analysis. We then used three algorithms, XGBoost, LASSO, and ANOVA, to find the most important features in the data that could predict whether a patient had cancer or not. We tested these important features using three machine learning algorithms: Random forest, K-nearest neighbor, and support vector classifier. As a result, we found 37 genes from mRNA data, 12 MiRNAs from MiRNA expression data, and 38 methylation points from Methylation expression data. 90 target genes were identified From miRDB database and 20 genes were identified from methDB database. Finally, we analyzed these genes to identify potential biomarkers using techniques such as pathway analysis, gene ontology, transcription factor enrichment analysis (TFEA), kinase enrichment analysis (KEA), and survival analysis. In this project, we aimed to find biomarkers that could help in the research and detection of prostate cancer. Biomarkers are biological signs that can indicate the presence or progress of a disease. Prostate cancer tests are usually expensive, and many people cannot afford them. Therefore, finding cost-effective biomarkers can help detect prostate cancer in a more affordable way. This research can benefit a larger segment of the population, who may otherwise not have access to expensive tests.

TABLE OF CONTENTS

INTRODUCTION

Prostate cancer is the most common malignant neoplasm of the urinary tract and most common non cutaneous cancer in men worldwide, with an estimated 1,600,000 cases and 365,000 deaths annually[1] . Even after extensive combination therapy, metastatic prostate cancer remains usually incurable, despite the high long-term survival rate in localized prostate cancer. The peripheral zone of the human prostate is where 60%–70% of human prostate malignancies begin [2]. Human prostates have a pseudostratified epithelium with three types of terminally developed epithelial cells: luminal, basal, and neuroendocrine epithelial cells. The genesis cell's biological and medicinal significance is unclear. According to one study, prostate cancer originating from luminal cells is more aggressive than prostate cancer derived from basal cells. [3]

Male gender, advanced age, positive family history, increased height, obesity, hypertension, lack of exercise, consistently raised testosterone levels, Agent Orange exposure, and ethnicity are all risk factors for prostate cancer.[4] Prostate cancer (PC) is a complex age-related disease that is influenced by demographic factors, race, and genetic predisposition[5]. The increased incidence of prostate cancer in men over 60 years of age confirms a definite link between older age and prostate cancer [6]. The breakdown of glandular homeostasis, which is caused by an imbalance between cell proliferation and death, is the reason for the relation of PC with aging. Prostate cancer, for example, is caused by an imbalance between mitosis and apoptosis. [7].

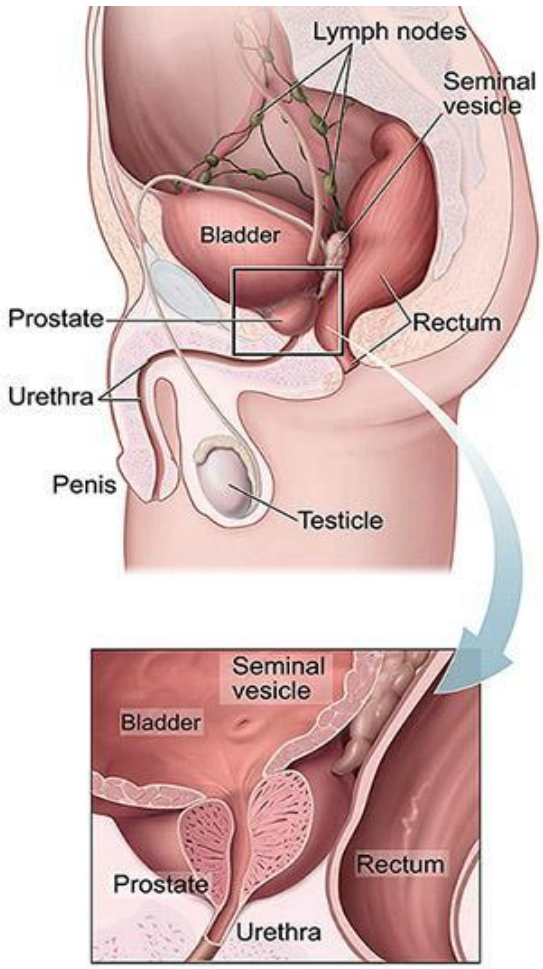


Figure 1: *Anatomy diagram shows the prostate, urethra, penis, testicle, bladder, lymph nodes, seminal vesicle, and rectum are labeled. An inset provides a close-up view of the prostate, urethra, bladder, seminal vesicles, and rectum.. Source: Rawla et al, 2019* [8]

The European Association of Urology has published guidelines for the diagnosis of prostate cancer. This includes a PSA (prostate specific antigen) level measurement and a digital rectal examination (DRE). The PSA's current problem is that it is organ specific, resulting in low specificity. The sensitivity of the DRE (Digital Rectal Exam) is also low. As a result, the ultimate confirmation is obtained through a prostate biopsy to identify the stage and grade of the cancer. However, prostate biopsy has a number of negative consequences for the patient, including lost workdays, excruciating pain and suffering, blood in the sperm, urine, or stool, the risk of infection, acute urine

retention, and sexual issues. Furthermore, one of the biggest drawbacks of prostate biopsy is the moderate false negative rate that comes with it. As a result, novel biomarkers with clinical relevance are urgently needed to aid in the diagnosis of prostate cancer with increased specificity and sensitivity of the test. Our main goal in this project was to uncover several biomarkers that may be explored for efficient prostate cancer research and how these biomarkers can be used to detect prostate cancer. Another key reason we need to conduct research on this topic is because these tests are highly expensive to conduct, and so cannot benefit a larger segment of our country's population who cannot afford them. In 2022, it is expected that 268,490 new cases of prostate cancer would be diagnosed, with 34,500 people dying from the condition.

	Common Types of Cancer	Estimated New Cases 2022	Estimated Deaths 2022
1.	Breast Cancer (Female)	287,850	43,250
2.	Prostate Cancer	268,490	34,500
3.	Lung and Bronchus Cancer	236,740	130,180
4.	Colorectal Cancer	151,030	52,580
5.	Melanoma of the Skin	99,780	7,650
6.	Bladder Cancer	81,180	17,100
7.	Non-Hodgkin Lymphoma	80,470	20,250
8.	Kidney and Renal Pelvis Cancer	79,000	13,920
9.	Uterine Cancer	65,950	12,550
10.	Pancreatic Cancer	62,210	49,830

Figure 2 : *Cancer Statistics(ACS journal), Rebecca L. Siegel MPH et al, 2021*

The trend is similar in India also. The incidence rate of prostate cancer in India is 9-10/100000 population, higher compared to countries like Asia, Africa, lower compared to US, Europe.

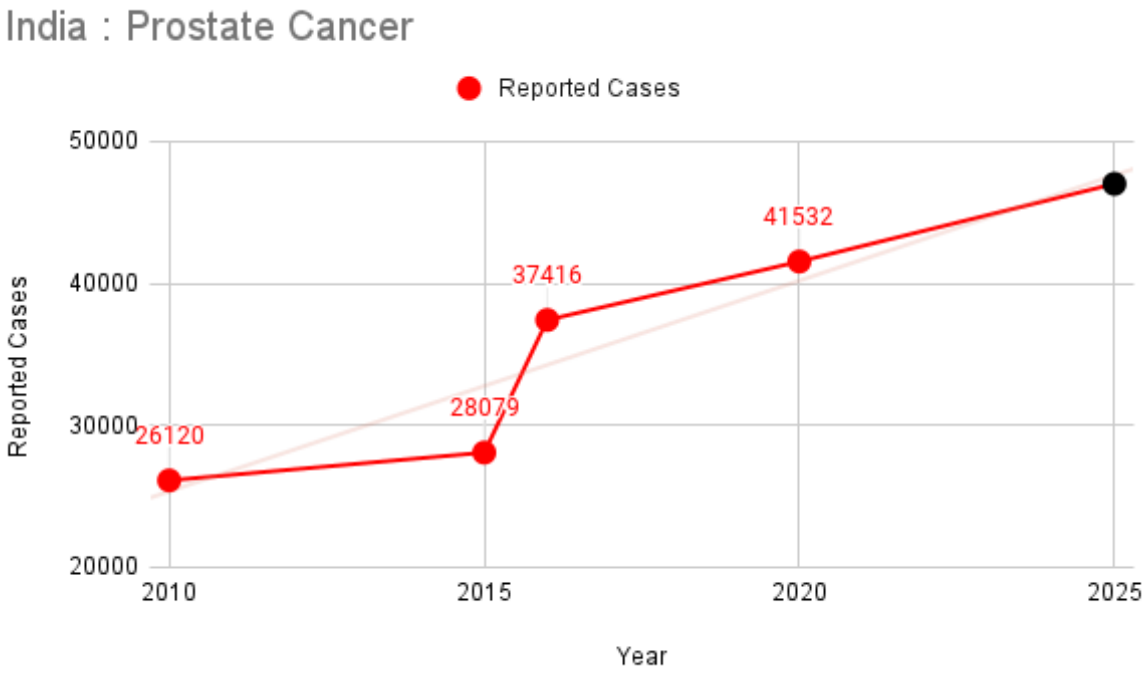


Figure 3 : *Cancer Statistics India (*Based on report of National Cancer Registry Programme (ICMR|NCDIR)

Here's a graph we made using data from the national cancer registry programme. As can be seen, there is a clear positive correlation as well as an increase in the number of cases. We projected the graph further, and 47000 more cases of prostate cancer are expected by 2025.

REVIEW OF LITERATURE

A. GENES

Prostate cancer risk has been linked to over 100 single nucleotide polymorphisms (SNPs) and other genes . The most prevalent genetic changes in prostate cancer are translocations affecting the ETS family of transcription factors, such as the ERG and ETV genes [9]. The first translocation discovered was a recurring gene fusion of TMPRSS2's 5′ untranslated region to ERG (TMPRSS2:ERG) [10]. The TMPRSS2:ERG fusion is found in 50% of localized prostate tumors [11], and recurrent gene fusions are also found between TMPRSS2 and ETV1, ETV4, and ETV5. ETS2 deletion is seen in around one-third of deadly mCRPCs, most typically through TMPRSS2:ERG fusions[12]. Overexpression of oncogenes like myristoylated AKT1 (myrAKT1), which is constitutively active, changes normal human prostate epithelial cells into prostate cancer cells with adenocarcinoma and squamous cell carcinoma phenotypes. Genes such as Smad4, Pten and Trp53 in both basal cells and luminal cells (ARR2PB-Cre), in basal cells (CK14-CreER), and in luminal cells (CK8-CreER) are tumor suppressor genes which are conditionally inactivated resulted the formation of prostate adenocarcinoma [13]**.** Pten, Rb1, and Trp53 inactivation also resulted in the development of NEPC (neuroendocrine prostate cancer). BRCA1 and BRCA2 mutations are linked to an increased risk of prostate cancer, as well as a more aggressive phenotype and deadly results [14]**.** Several independent genomic investigations have found that 15%–35% of mCRPC patients had DNA repair abnormalities in the BRCA1/2, ATM, ATR, and RAD51 genes [15]. P53 mutations are uncommon in primary prostate cancer, although they are more common in metastatic illness. As a result, p53 mutations in prostate cancer are often regarded as a late and grave finding [16]**.**

B. DIAGNOSIS

Prostate cancer is diagnosed primarily through prostate-specific antigen (PSA) testing and transrectal ultrasound-guided (TRUS) prostate tissue biopsies, while PSA testing for screening is debatable and TRUS is ineffective for staging [17]**.** Staging is done using magnetic resonance imaging or computed tomography, which has a low accuracy when compared to pathological specimens. Detection of lymph nodes and bone metastases is also a problem with current approaches. Free and total PSA levels, PCA3 urine testing, Prostate Health Index scoring (PHI), the "4K" test, exosome testing, genomic analysis, MRI imaging, PIRADS scoring, and MRI-TRUS fusion guided biopsies are some of the more recent diagnostic approaches[18]**.**

C. TREATMENT

It is considered confined and potentially treatable when the cancer is limited to the prostate. [19]**.** Pain relievers, bisphosphonates, rank ligand inhibitors, hormone therapy, chemotherapy, radiopharmaceuticals, immunotherapy, focused radiation, and other targeted therapies can be utilized if the illness has progressed to the bones or elsewhere beyond the prostate [20]**.** Chemotherapeutic medicines such as mitoxantrone, doxorubicin, vinblastine, paclitaxel, docetaxel, and others are used to treat advanced prostate cancer [21]**.**

D. BIOMARKERS

Cancer Epidemiology and Biomarkers	
Contributing Factors	**Three Major Kinds of Biomarkers**
Genetic (family history) Lifestyle Age Environment - Radiation - Toxic Agents - Mutagens - Infectious Agents Drugs for Treatment Race	1.Epigenetic 2. Genetic 3.Proteomic **Utilization of Biomarker** Toxic Agents Cancer Screening Cancer Detection Cancer Diagnosis Cancer Survival and Outcome Cancer Prognosis

Ideal Biomarker	Factors that Impact Sensitivity and Specificity
❖ Capable of distinguishing between healthy and sick people ❖ Should be expressed early in the course of the disease ❖ It is simple to test, less expensive, and should aid in the reduction of cancer-related deaths. ❖ It produces repeatable results and allows for multiplexing.	❖ Sample type (biofluid vs. tissue) ❖ The sample's stability and the time it takes to assay the biomarker ❖ Negative controls should be used correctly. ❖ Background profiling

Genetic biomarkers are frequently connected to overexpression of a gene. ERG and ETV1 are overexpressed in prostate cancer, but they also fuse with TMPRSS2, triggering tumor formation [22]. This TMPRSS2-ERG gene fusion rearrangement can be detected in urine and may aid in the prediction of prostate cancer development. PCA3 (prostate cancer antigen 3) is a gene that produces non coding mRNA that is particular to the prostate. PCA3 has a higher specificity than PSA and can differentiate between prostate cancer and benign diseases, making it a better biomarker for prostate cancer screening. [23]. Golgi phosphoprotein 2 is encoded by GOLPH2, a Golgi membrane antigen. This gene is increased in around 90% of prostate cancer patients, leading in overexpression of the gene . Epigenetic modifications, which do not include changes in nucleotide sequence, have a role in a range of biological processes, such as transcription, DNA repair, and differentiation, and their alterations have

been related to cancer . In prostate cancer cells, hypermethylation of the PDLIM4 gene lowers both mRNA and protein expression. [24]. PDLIM4 may operate as a tumor suppressor, limiting cell proliferation and predicting recurrence in prostate cancer. PSA is a proteomic biomarker that was one of the first to be completely recognised and is likely one of the most commonly used in the clinic. Fat metabolism is aided by the isomerase AMACR, also known as alpha-methylacyl-CoA racemase. It's overexpressed in prostate cancer tissue and works as a growth promoter for the disease. PSCA, or prostate stem cell antigen, is a glycoprotein present on the surface of cells in the prostate. PSCA expression has been associated with advanced prostate cancer stages, and there is a relationship between PSCA and prostate cancer risk.

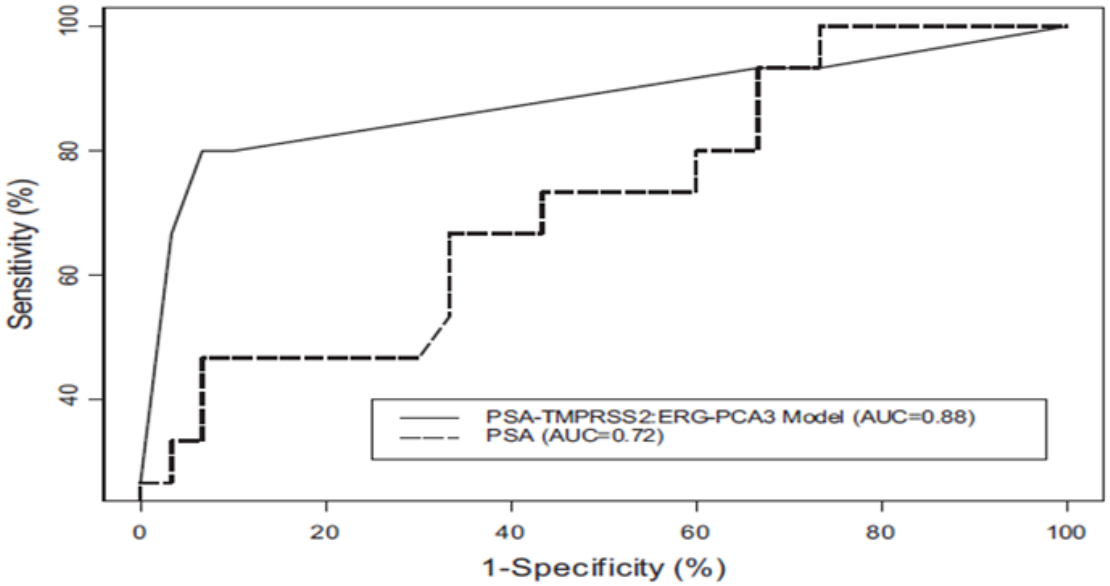


Figure 4 : *To predict prostate cancer on future biopsy, an algorithm combining serum PSA, urine TMPRSS2:ERG fusion, and PCA3 was given. Improvements in cancer prediction (AUC=0.88; specificity 90%; sensitivity 80%)*

Source: Simpa S Salami et al, 2013

METHODOLOGY

Research Workflow

1. **Data Preprocessing** : miRNA & mRNA data is taken from TCGA
2. **Feature Importance** : Machine learning feature importance models are used to identify important genes.
3. **Machine Learning Classification** : ML classification models are used to verify accuracy of feature selection.
4. **Pathway Analysis & Gene Ontology** : To translate the list of genes that are differentially expressed across the given phenotypes.
5. **KEA & TFEA** : To connect a set of genes with Transcription factors and Kinases

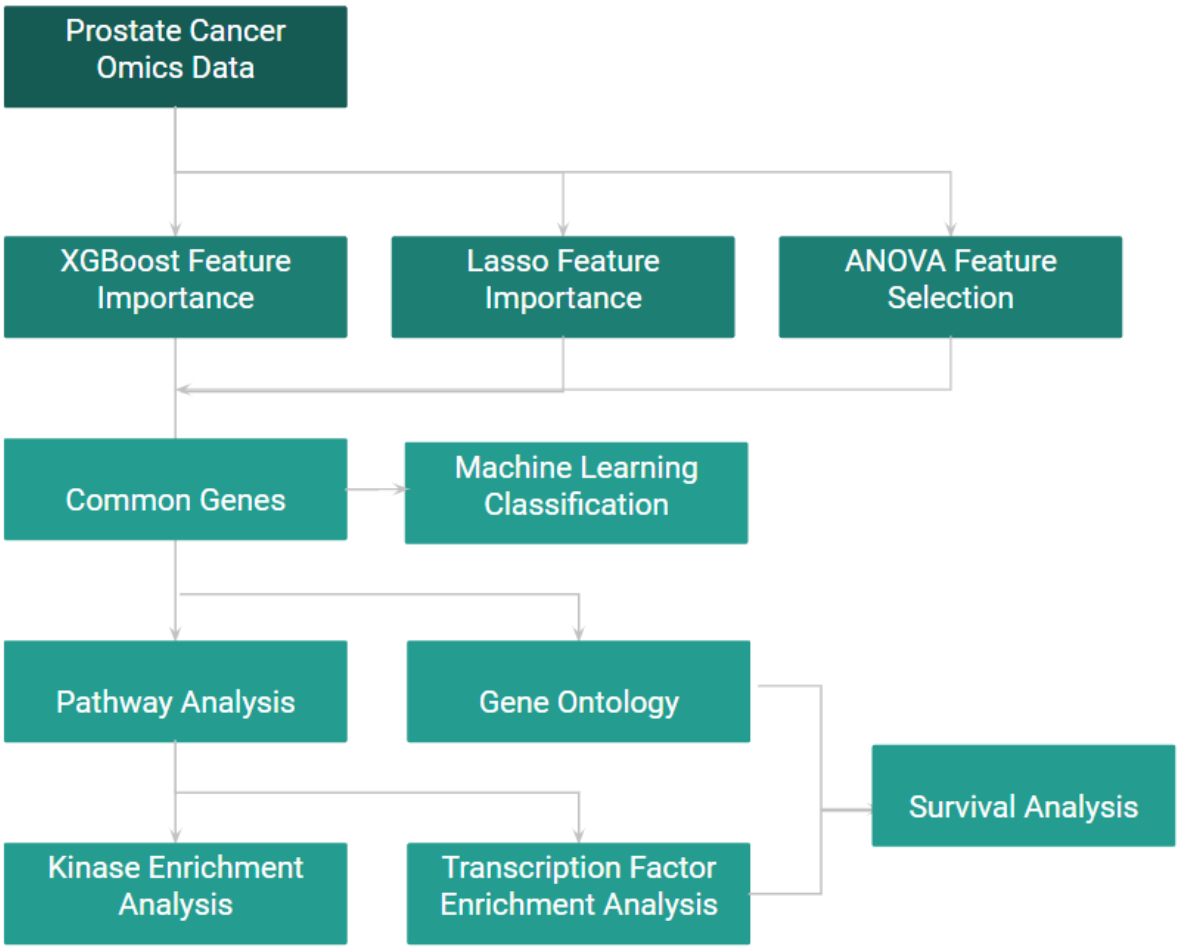


Figure 5: *Workflow of methodology.*

1.Data Preprocessing

The Cancer Genome Atlas (TCGA) is used to obtain prostate cancer omics data, which includes miRNA expression data, mRNA expression data, Methylation expression data, and patient clinical data. Because the data is raw and unprocessed, some data preprocessing steps are required to transform the raw data into a useful and efficient format. In the case of mRNA data, the raw data contains 20112 gene expression data from 550 individuals. 497 of these samples are from patients with primary solid tumors, one from a patient with a metastatic tumor, and 52 from healthy people. Because only one sample of metastatic tumor was available, we combined metastatic and primary tumor into one tumor class. In addition, the data contains 331 missing values spread across 53 gene data. As part of the data preprocessing, we removed the 53 gene data. The raw data is normalized as the final data preparation step to transform features to be on a similar scale. This improves the model's performance and training stability. These steps are repeated for miRNA data as well.

2.Feature Importance

Feature importance refers to techniques that assign a score to input features based on how useful they are at predicting a target variable. We have developed a feature importance layer, which consists of three algorithms: XGBoost feature importance [25], Lasso feature importance [26], and ANOVA feature importance [27]. mRNA expression data, miRNA expression data, and methylation data were fed into this layer separately and output features are noted.

The first step in the feature importance layer was to use the XGBoost algorithm to identify important features from the mRNA expression data. The XGBoost algorithm is a gradient boosting machine learning algorithm that is commonly used for feature selection. The algorithm identifies important features by iteratively building decision trees and determining which features

have the most impact on the classification accuracy of the model. The second algorithm used was the Lasso feature importance algorithm, which is a regularization method that shrinks the coefficients of less important features to zero. This method is commonly used in high-dimensional data, such as omics data, where the number of features is much larger than the number of samples. The Lasso algorithm helps to identify important features that have non-zero coefficients. The third algorithm used in the feature importance layer was ANOVA feature importance. ANOVA (analysis of variance) is a statistical method that is used to compare the means of two or more groups. In this case, the ANOVA algorithm was used to identify genes that have significantly different expression levels between cancerous and non-cancerous samples.

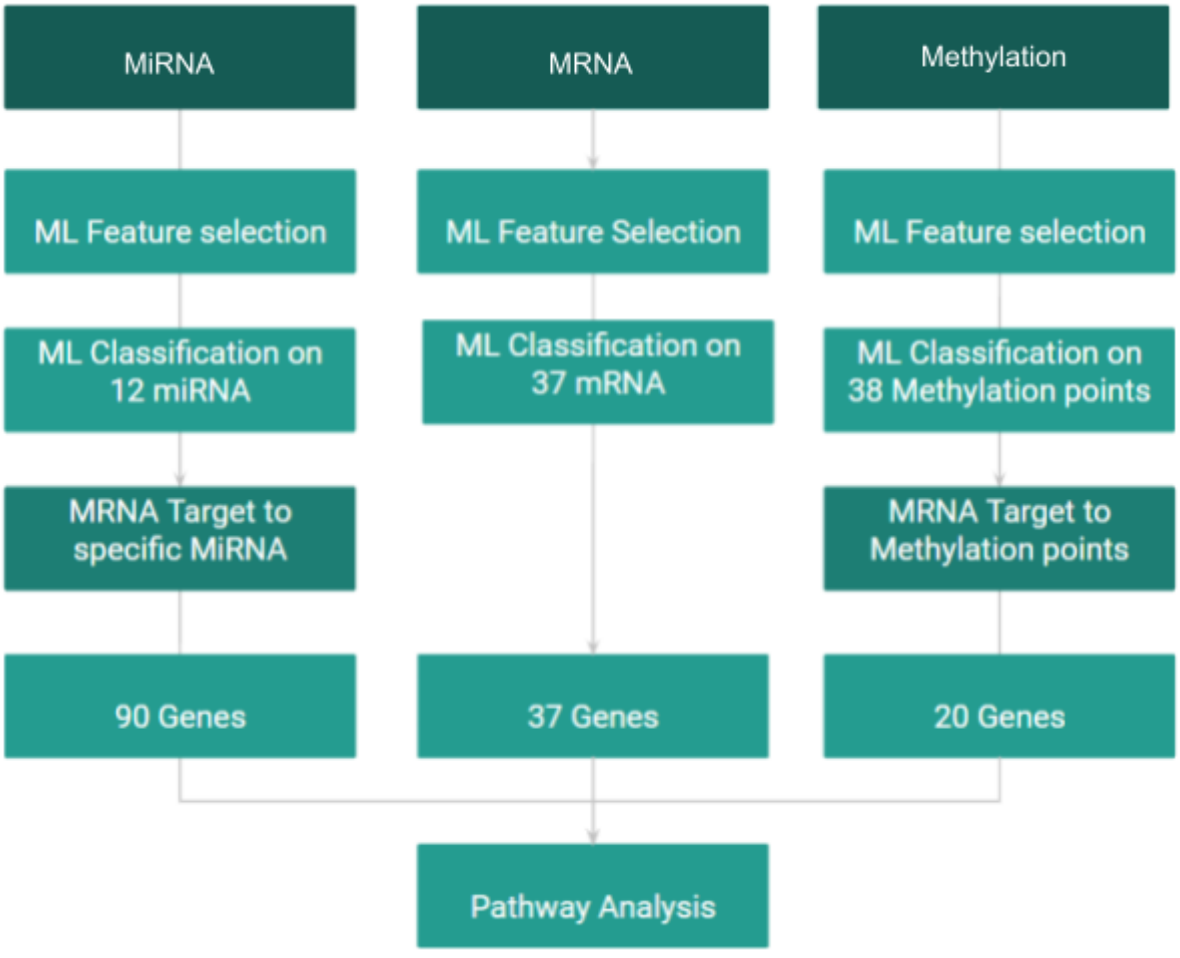


Figure 6: *Workflow of Feature Importance.*

After running these three algorithms on the mRNA expression data, miRNA expression data, and methylation data, the results were filtered to include only features that appeared in at least two of the algorithms. This process helped to identify the most important features that were consistently selected by multiple algorithms. In the case of mRNA data, the significant mRNA was directly selected from the feature importance part. In the case of miRNA data, the output features were miRNAs, which were then used to identify their target mRNA. Similarly, in the case of methylation data, the target mRNA was identified. Finally, the selected genes from all three omics data were used for machine learning classification and further studies. These genes can serve as potential biomarkers for prostate cancer detection and could be used to develop more accurate diagnostic tests.

In conclusion, the feature importance layer developed in this project provides a powerful tool for identifying important genes from multiple omics data types. By combining the results of multiple algorithms, researchers can identify the most important features consistently selected by different methods. The selected genes can serve as potential biomarkers for prostate cancer detection and may ultimately help to improve patient outcomes.

3.Machine Learning Classification

Machine Learning Classification is one of the most common tasks in machine learning, where the goal is to predict a categorical label or class for a given set of input features. In this case, a feature importance layer was applied to the data to filter out the most significant features. These features were selected based on their ability to differentiate between cancer patients and normal patients. We refer to these selected features as "significant features". To test the accuracy of these features, we used a machine learning layer that included three popular algorithms: random forest [28], K-nearest neighbors

[29], and support vector classifier [30]. We assessed the performance of these algorithms on all three omics datasets.

Random Forest is an ensemble learning algorithm that is used for classification and regression tasks. It creates multiple decision trees and aggregates the results to produce a final prediction. Each decision tree is created by randomly selecting a subset of features and data points. Random Forest is used to handle high-dimensional data and can handle missing data effectively. K-Nearest Neighbors (KNN) is a non-parametric machine learning algorithm used for classification and regression tasks. KNN uses the distance metric to identify the K nearest data points and assigns the class based on the majority vote. KNN is a simple and effective algorithm but can be computationally expensive for large datasets. Support Vector Classifier (SVC) is a supervised learning algorithm used for classification tasks. SVC tries to find the hyperplane that maximizes the margin between the classes. Our classification layer consists of these three machine learning algorithms. We then trained the three classification algorithms on the selected features using a train-test split strategy. Specifically, we randomly divided the data into two parts, a training set, and a test set. The training set was used to train the classification models, while the test set was used to evaluate their performance.

We have used f1 score as a classification matrix for machine learning classification. F1 score is chosen because of its performance to combine precision and recall, and also works for cases where the datasets are imbalanced as it requires both precision and recall to have a reasonable value [31,32]. Table 1.1 and Table 1.2 summarizes the accuracy of the four models for each dataset.

Table 1.1 : F1 score of training data on the three models

	Random Forest	K-Nearest Neighbour	Support Vector Classifier
mRNA	98.68	87.65	100
MiRNA	98.68	88.97	87.66
Methylation	98.64	93.72	93.60

Table 1.2 : F1 score of testing data on the three models

	Random Forest	K-Nearest Neighbour	Support Vector Classifier
mRNA	89.28	92.85	88.76
MiRNA	82.14	81.62	82.14
Methylation	91.79	89.75	90.26

Our study demonstrates that the three classification algorithms consistently achieved high performance across all three omics datasets. This suggests that the significant features selected through the feature importance layer are effective in filtering cancer patient data from patient omics data. Random Forest consistently achieved the highest F1 score for all three datasets, but the difference in F1 score between the three models was small, indicating that the features are suitable for further analysis.

4.Pathway Analysis & Gene Ontology

Following the completion of Machine Learning classification, a comprehensive Pathway Analysis and Gene Ontology analysis were conducted to further elucidate the functional characteristics of the identified set of Differentially Expressed Genes (DEGs). This rigorous analysis aimed to validate and strengthen the findings by identifying classes of genes that

exhibited overexpression and could potentially be associated with the disease under investigation. This integrative approach provided a robust and multi-dimensional understanding of the biological pathways and functional annotations that may be implicated in the disease pathogenesis, thereby bolstering the significance of the DEGs identified through the ML modeling process.

5.KEA & TFEA

Transcription Factor Enrichment Analysis (TFEA) predicts transcription factors that are predicted to regulate the input gene list by performing gene set enrichment analysis using different transcription factor gene set libraries, for example, integrated target genes for transcription factors as determined by ChIP-seq experiments (ChEA).

Protein kinases that are probably the controllers of the extended protein-protein interaction network are predicted by KEA. The list of proteins from the subnetwork is subjected to enrichment analysis by KEA utilizing gene set libraries from kinase-substrate interaction databases.

6. Survival Analysis

Survival analysis is a powerful statistical technique that allows for the estimation of the probability of survival over time following a diagnosis of prostate cancer.

Similarity Network Fusion

In addition to using XGBoost, LASSO, and ANOVA feature selection algorithms, we also utilized a method called similarity network fusion in this project. This approach involved combining all of the available data, including mRNA, miRNA, and methylation data, into a single network known as a similarity network. Instead of analyzing the data separately, we grouped the omics data of each patient into a single formatted data set to use for further

analysis. To accomplish this, each omics data was formatted into a uniform format using affinity network, which created a network based on the number of patients. The affinity network was then combined to obtain a single network known as a fused similarity network. This network was then grouped into two clusters, and Kaplan-Meier plots were plotted and analyzed to determine any patterns or trends in the data. By using this approach, we were able to obtain a more comprehensive understanding of the underlying molecular mechanisms of prostate cancer and identify potential biomarkers for the disease.

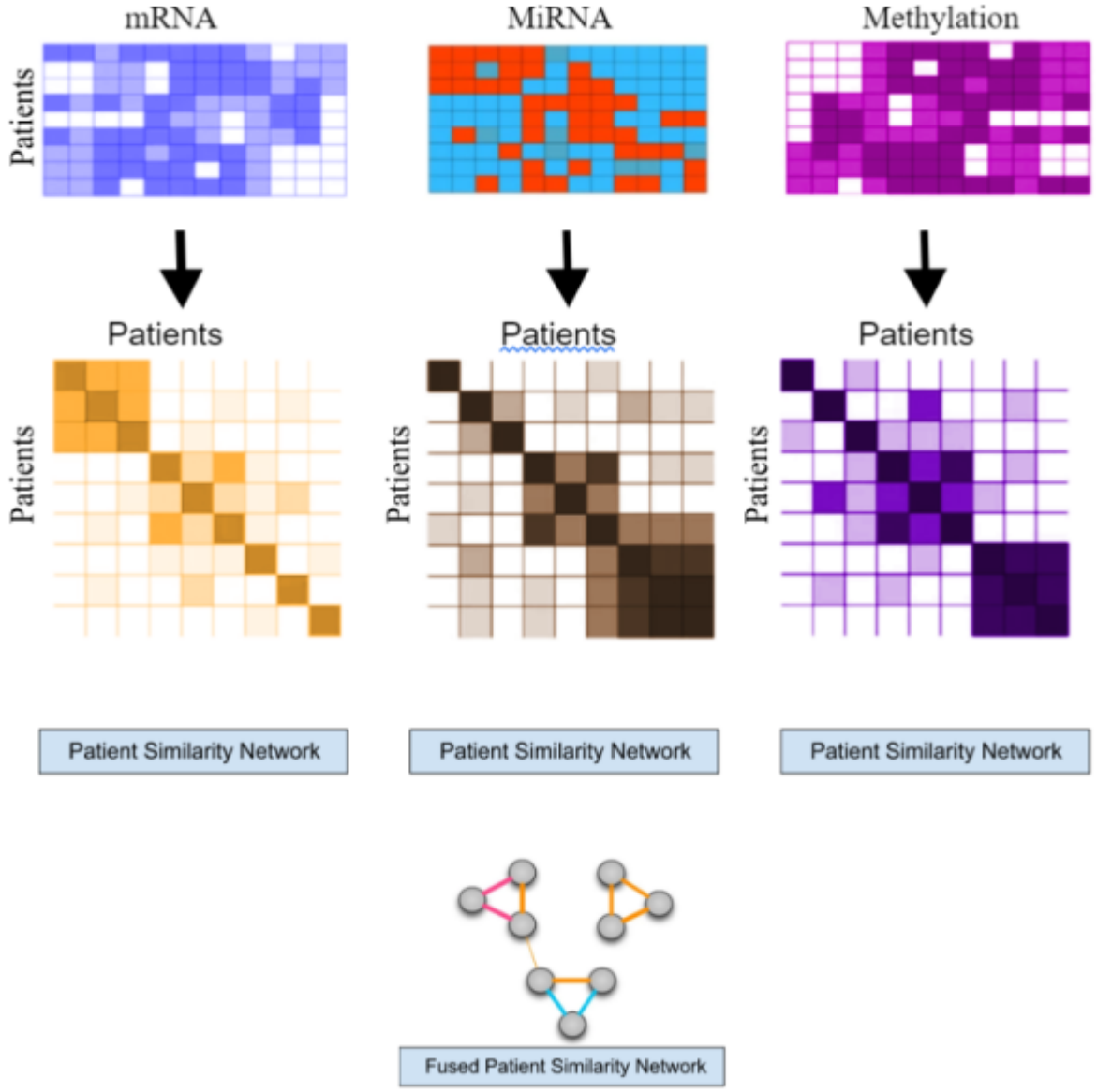


Figure 7: *Similarity Network Fusion Workflow (Plots only for representation)*

RESULTS AND DISCUSSION

After the Machine Learning modeling was performed, we obtained a total of 90 differentially expressed genes from 12 distinct miRNA. We used the mirRDB6 database for correlating mRNA and miRNA. One gene is shared by these two techniques when 90 genes from miRNA are analyzed with 37 genes from mRNA. PPARGC1A - The prostate tumor suppressor PPAR coactivator 1 alpha (PGC1) regulates the ratio of anabolism to catabolism. Metastasis development in prostate cancer is causally linked to PGC1A downregulation. To do pathway analysis, these 90 genes derived from miRNA and 37 genes derived from mRNA are merged. After that, we turned to GeneMANIA to forecast the gene's function that was shown to be differentially expressed in all the statutes listed**(Figure-8)**. Utilizing a massive collection of functional association data, GeneMANIA discovers additional genes that are connected to a set of input genes. Protein and genetic relationships, pathways, co-expression, co-localization, and protein domain similarity are all examples of association data[33].

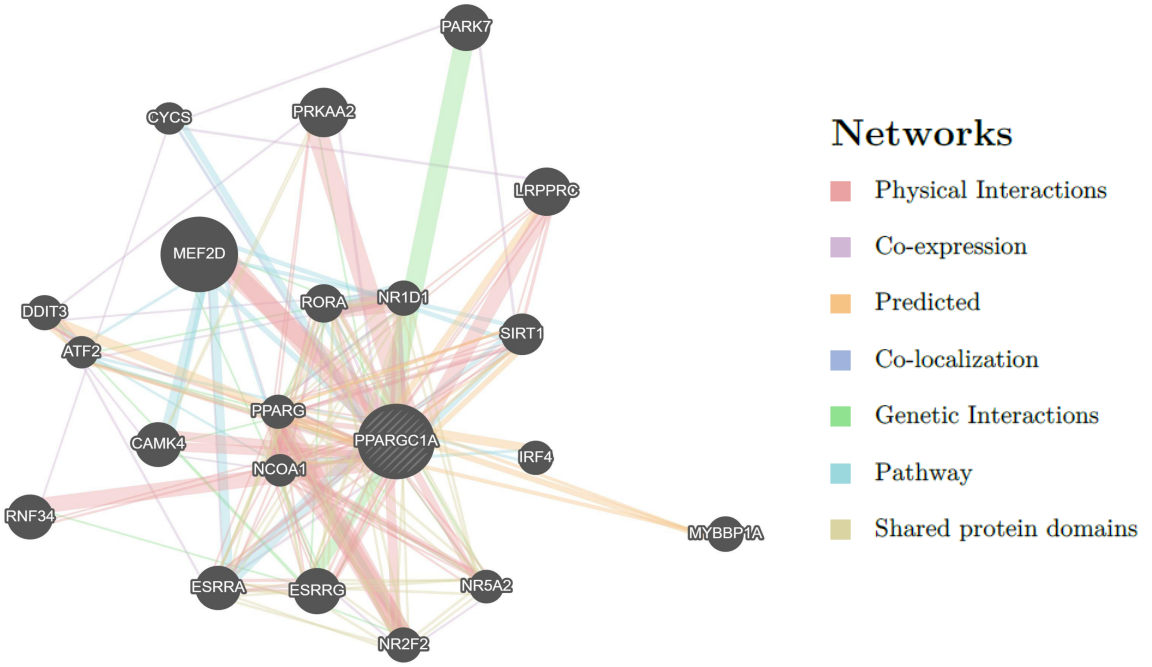


Figure 8: GeneMANIA network showing network of differentially expressed genes related to differentially expressed PPARGC1A gene.

The network included 8.01 percent co-expression (denoted in lilac), 77.6 percent physical interactions (denoted in pink), 3.63 percent co-localization (denoted in purple), 5.37 percent predicted values (denoted in orange), 1.8 percent pathways (denoted in blue), 0.6 percent shared protein domains (denoted in yellow), and 2.87 percent genetic interactions (denoted in green).

EnrichR[34] is a platform that integrates knowledge from a variety of well-known initiatives in a novel way to offer synthesized data about mammalian genes and gene sets. The platform offers a number of techniques to compute gene set enrichment, and the outcomes are shown in a number of engaging ways.The results from the EnrichR(**Figure-9**) showed that Chemokine and Cytokine signaling pathway and CCKR signaling pathway were most differentially expressed. Cytokines are a diverse family of low-molecular weight proteins involved in the mediation of communication between cells. Cytokines interact with members of the immunoglobulin superfamily, type I, type II, TNF, G-protein coupled (chemokine), TGF, and IL-17 receptor families to produce their effects.The canonical mechanism triggered after cytokine-receptor contact is the Janus kinases (JAK)—signal transducers and activators (STAT) pathway.Chemokines are a wide class of chemotactic cytokines that influence immune cell placement and movement. They work by coupling to G-protein coupled receptors, which are seven-transmembrane protein receptors (GPCRs).Tumor cells go through gene alterations during the first stage of prostate cancer metastasis, which enables decreased cell-cell and cell-ECM adhesion as well as improved invasiveness and migratory ability. As cytokines and chemokines are produced into the tumor microenvironment (TME), the ECM and basement membrane degrade, immune cells chemotactically move, adhesion protein expression is altered, and the primary tumor site is remodeled. Numerous Cytokines, such as vascular endothelial growth factor (VEGF), CXCL8, IL-6, and TGF-, promote tumor-associated angiogenesis.

Cholecystokinin receptors CCK1R and CCK2R, respectively, are the biological conduits via which the digestive peptide hormones gastrin and cholecystokinin operate.Through a number of processes, the expression of the gastrin gene is increased in a range of pre-cancerous diseases as well as in advanced cancer. Differential processing of the polypeptide product results in the synthesis of distinct physiologically active peptides, depending on the tissue where it is expressed and the level of expression. These peptides then activate signaling pathways that change the activation of downstream genes that affect cell survival, angiogenesis, and invasion by binding to the classical gastrin cholecystokinin B receptor CCK-BR, its isoforms, and alternative receptor.

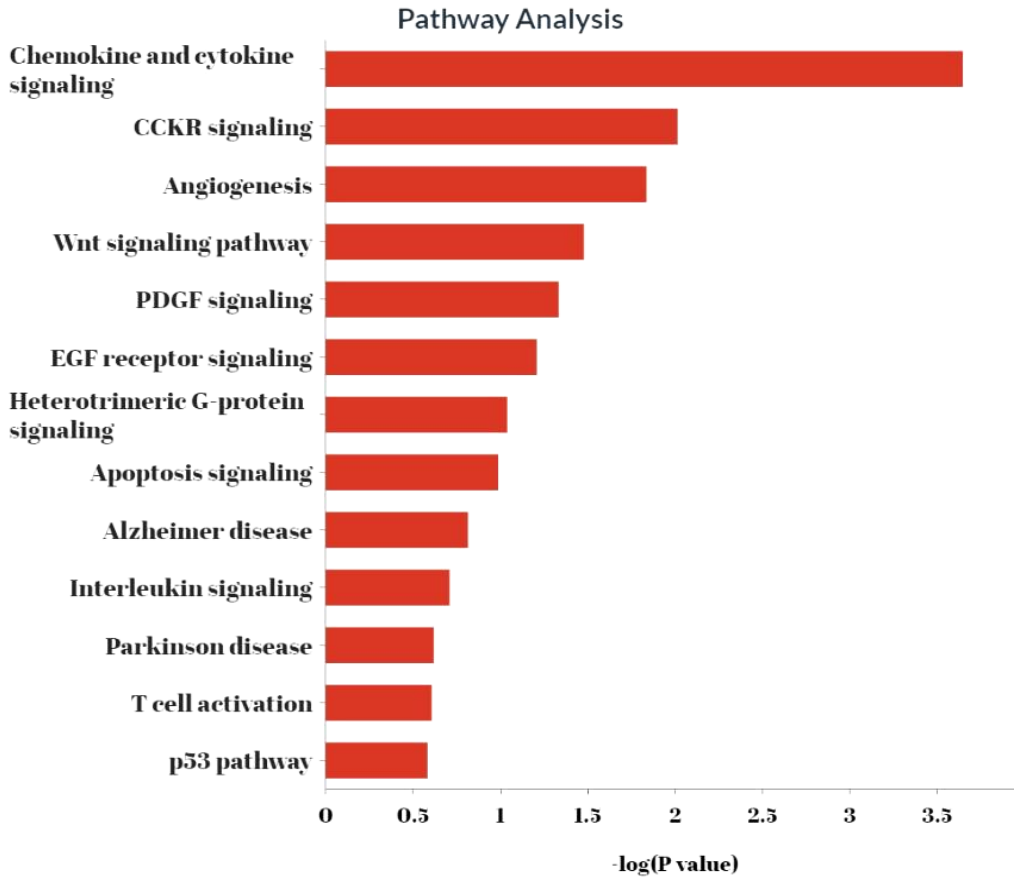


Figure 9: Pathway analysis report indicating significant pathway related to given DEGs

Pathway	Genes
Inflammation mediated by chemokine and cytokine signaling pathway Homo sapiens P00031	PAK3;UBE2D2;EIF4G2;HSP90B1
CCKR signaling map ST Homo sapiens P06959	PAK3;RGL1;NR2C2
Angiogenesis Homo sapiens P00005	COL4A6;COL17A1;SEMA3A;RPS6KA6;IGF1;ESR1;GAB1
Wnt signaling pathway Homo sapiens P00057	ESR1;LGR6;FZD4;CDH11; DACT2
PDGF signaling pathway Homo sapiens P00047	EPS8;GAB1;CDH11;RPS6KA6
EGF receptor signaling pathway Homo sapiens P00018	RRM2;GAB1;RPS6KA6;EPS8;PLSCR1;IGF1;EEA1A;SEMA6D;EIF4G2;SLC8A1
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway Homo sapiens P00027	POLR1D;GARNL3;GNRH1;PDE4B
Apoptosis signaling pathway Homo sapiens P00006	PLK2;IGF1;TOP1;H6PA6
Alzheimer disease-presenilin pathway Homo sapiens P00004	RRM2;FZD4;IGF1
Interleukin signaling pathway Homo sapiens P00036	IGF1;RPS6KA6

Table 1 : List of Genes in Pathway Analysis

We performed survival analysis on the significant genes identified from the top two pathways(**Figure 10)**. The following genes, namely PAK3, UBE2D2, EIF4G2, HSP90B1, RGL1, and NR2C2, were subjected to Survival Analysis, revealing that HSP90B1 and UBE2D2 exhibit potential as cancer-promoting genes, while PAK3 and RGL1 function as tumor suppressor genes.

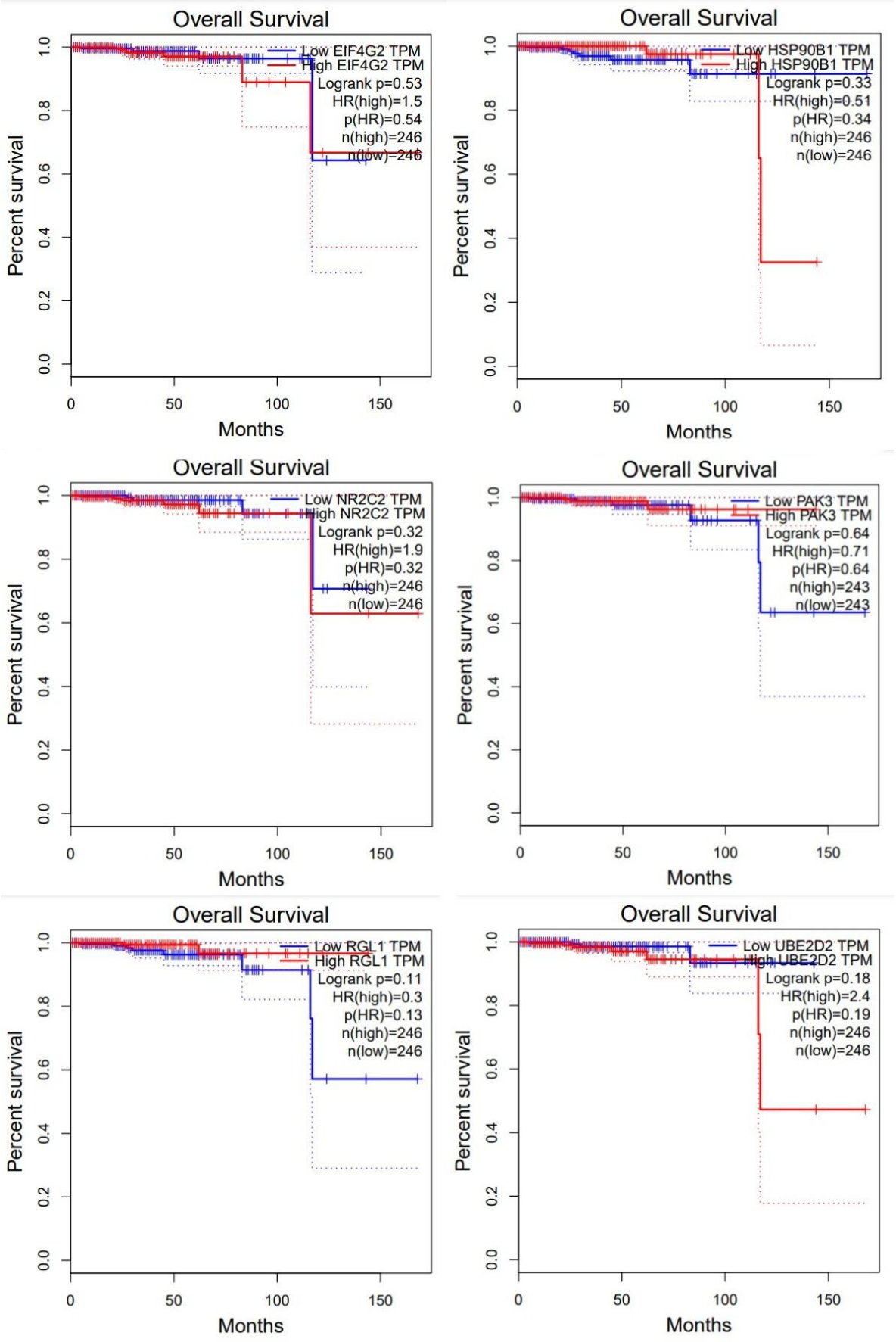


Figure 10: Survival analysis report indicating survival probability on given Gene.

We performed survival analysis on the significant genes identified from the top two pathways(**Figure 10)**. The following genes, namely PAK3, UBE2D2, EIF4G2, HSP90B1, RGL1, and NR2C2, were subjected to Survival Analysis, revealing that HSP90B1 and UBE2D2 exhibit potential as cancer-promoting genes, while PAK3 and RGL1 function as tumor suppressor genes.

To deduce the underlying biological process from the DEGs, we performed a Gene Ontology analysis. Both regulation and negative regulation of G0 to G1 transition are the most enriched biological function which shows the presence of Oncogene and Tumor suppressor genes in DEG

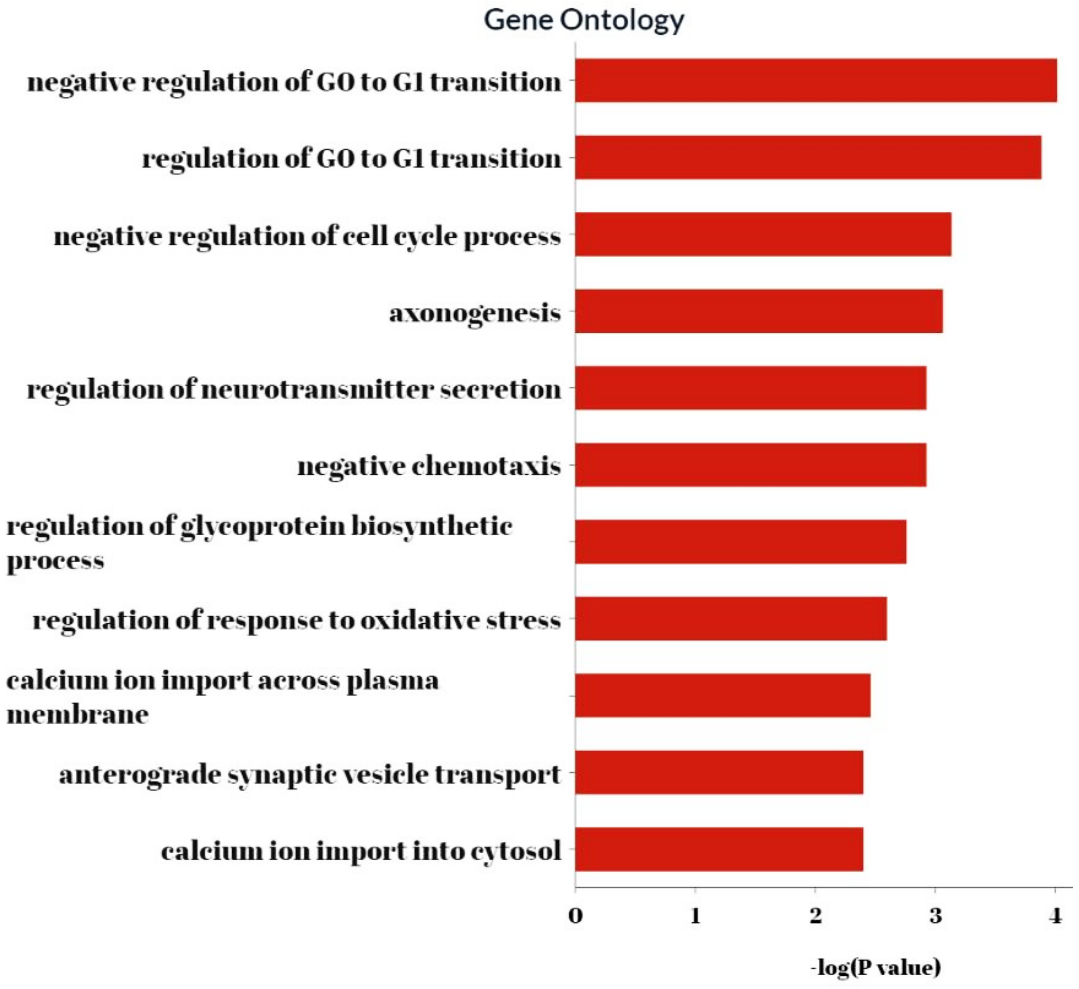


Figure 11: Gene Ontology report indicating significant biological process related to given DEGs.

Term	Genes
negative regulation of G0 to G1 transition (GO:0070317)	RRM2;MGA;EPC1;PHC3
regulation of G0 to G1 transition (GO:0070316)	RRM2;MGA;EPC1;PHC3
negative regulation of cell cycle process (GO:0010948)	RRM2;MGA;EPC1;PHC3
axonogenesis (GO:0007409)	RAB10;EPHA10;SEMA6D;SEMA3A;GAB1;PAK3;LGR6
regulation of neurotransmitter secretion (GO:0046928)	MCTP1;DTNBP1;MCTP2
negative chemotaxis (GO:0050919)	SEMA3A;SEMA6D;LGR6
regulation of glycoprotein biosynthetic process (GO:0010559)	NECAB1;IGF1
regulation of response to oxidative stress (GO:1902882)	MCTP1;ACOX2
calcium ion import across plasma membrane (GO:0098703)	TRPV3;SLC8A1

Table 2 : List of Genes in Ontology

We followed EnrichR with X2 Kweb where we carried out Transcription Factor Enrichment Analysis (TFEA), Kinase Enrichment Analysis (KEA), and built an eXpression2Kinases (X2K) network. From the fingerprints of differentially expressed genes, X2K Web [35] deduces upstream regulatory networks. X2K Web generates inferred networks of transcription factors, proteins, and kinases

anticipated to govern the expression of the inputted gene list by integrating transcription factor enrichment analysis, protein-protein interaction network expansion, and kinase enrichment analysis. The following process in the X2K pipeline uses the enhanced transcription factors as its input. The TFEA**(Figure-12)** projected that AR was the transcription factor with the highest significance for our group since it scored 1.76e-3, the highest value on the -log10 negative scale (p-value).AR is a ligand-activated transcription factor that controls genes essential for the development and differentiation of male sexuality.

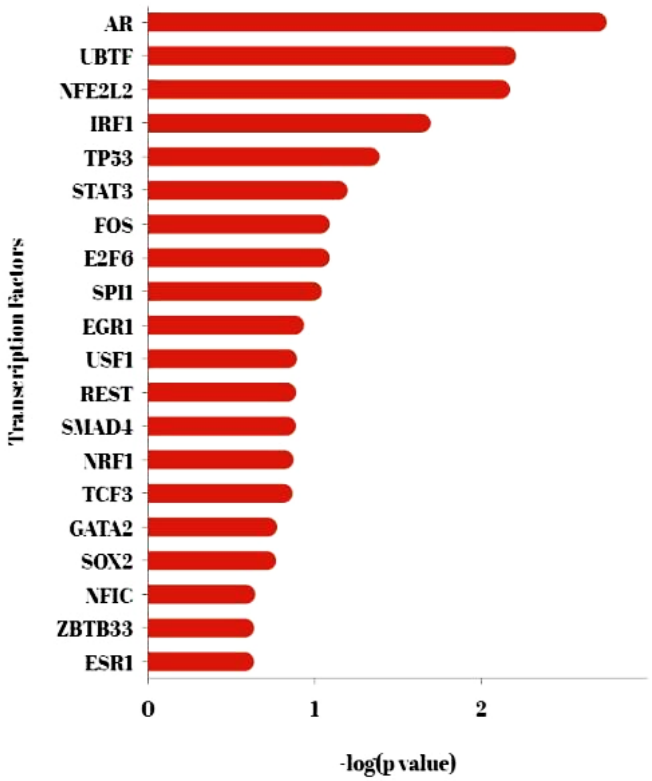


Figure 12: Transcription factor enrichment analysis report indicating significant transcription factors related to given DEGs.

From the gene list, TFEA also inferred that AR targets 15 genes, including RFX3, MTMR2, LGR6, ATXN7, GRIA4, NKX3-1, IGF1, GAB1, FUT9, CPEB3, ELMOD2, SLC8A1, SEMA6D, and RFC3.

Transcription Factor	Enriched Targets
AR	15 genes (RFX3, MTMR2, LGR6, ATXN7, GRIA4, NKX3-1, IGF1, GAB1, FUT9, CPEB3, ELMOD2, SLC8A1, SEMA6D, RFC3, HMCN1)
NFE2L2	13 genes (C5ORF24, GSTM3, EEA1, KCNA4, KIAA1522, IGF1, PHYHIPL, SEMA6D, CEP170, RAB10, ESR1, CDH11 POLQ)
UBTF	17 genes (TOP1, ARFGEF1, PHC3, TNPO1, RAB10, ASXL2, NR2C2, NCAPG, EEA1, ZNF236, UBE2D2, ATXN7, AFF4, CPEB3, EIF4G2, ELAC1, AUH)
IRF1	5 genes (ZNF204P, APOBEC3C, PLSCR1, KDSR, RAB10)
TP53	5 genes (GCC2, PLK2, NKX3-1, CDH11, RRM2)
STAT3	8 genes (PLSCR1, ANKRD13C, RAB30, HSP90B1, KDSR, AFF4, KDM4C, RRM2)
FOS	7 genes (ANKRD12, HSP90B1, POLR1D, DCUN1D4, AFF4, KDM4C, RRM2)
E2F6	25 genes (PLSCR1, TOP1, RRM2, ASXL2, MTMR2, EEA1, UBE2D2, MGA, FZD4, DTNBP1, HSP90B1, EPC1, AKAP11, ANKRD13C, RFC3, DCUN1D4, KDM4C, C5ORF24, RFX3, NCAPG, RGL1, GXYLT1, AFF4, LPCAT2, ELAC1)

Table 3 : List of Genes from TFEA

The next phase was Kinase Enrichment Analysis (KEA), which forecasts the protein kinases that are probably the regulators of the larger protein-protein interaction network from the preceding step **(Figure-13)**. The list of proteins from the subnetwork is subjected to enrichment analysis by KEA utilizing gene set libraries from kinase-substrate interaction databases.

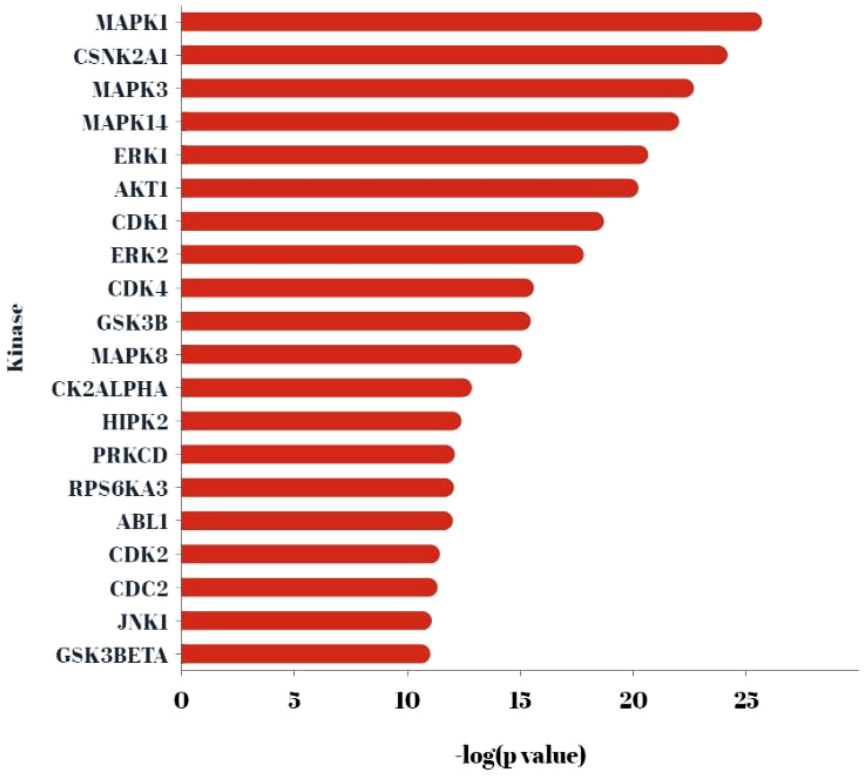


Figure 13: Bar plot of Kinase Enrichment Analysis report indicating most enriched protein kinases

Since they scored 2.06e-25 and 1.18e-24, respectively, which is the highest value on the negative scale of -log10 for our cohort, MAPK14 and MAPK1 were projected to be the two protein kinases that were the most significant (p-value).

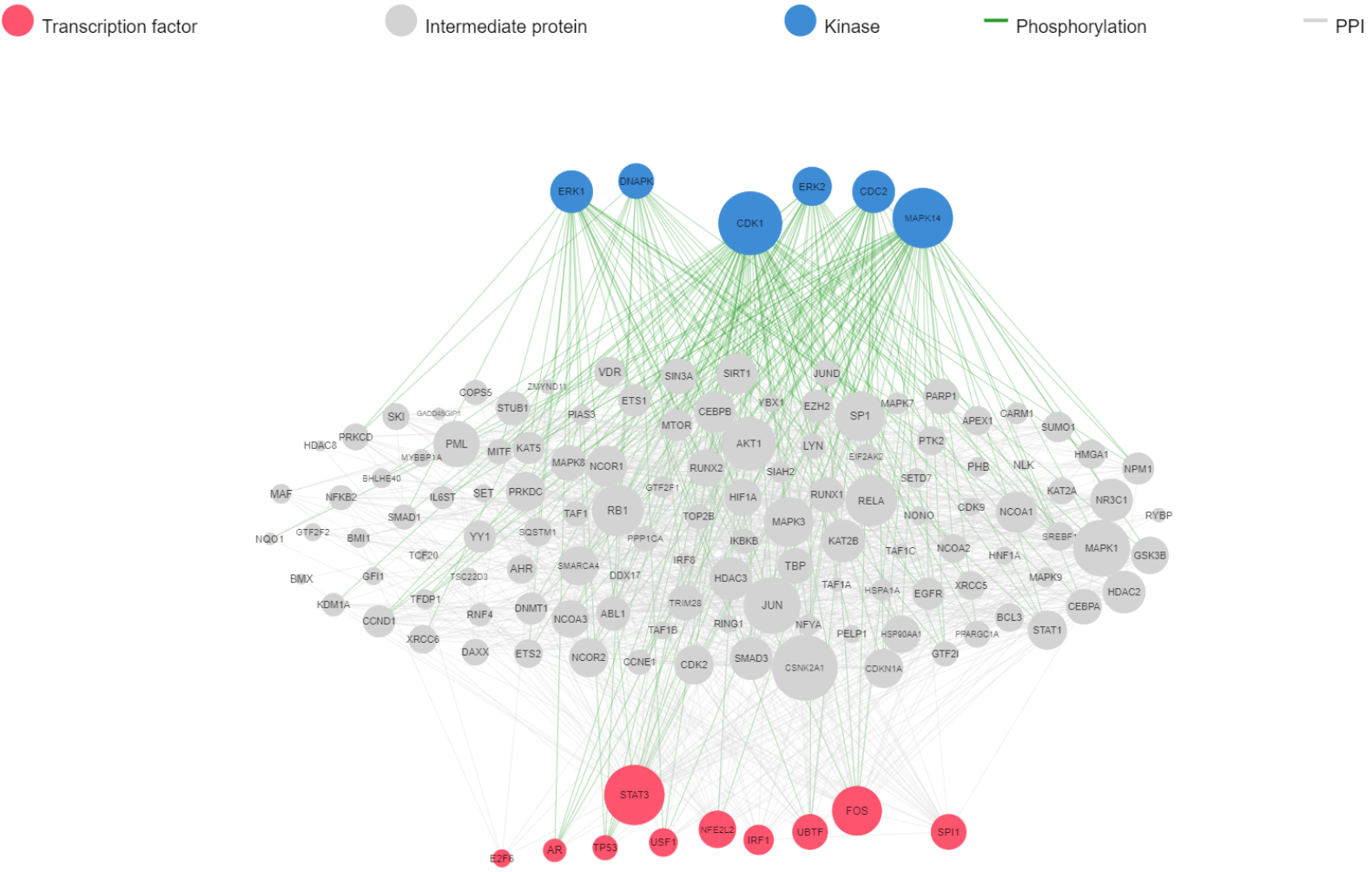


Figure 14: eXpression2Kinases (X2K) network..

TFEA and KEA were then followed by the eXpression2Kinases (X2K) network, which displays the inferred upstream regulatory network predicted to regulate the input list of genes by integrating the results from the TFEA and the kinase enrichment.

From TFEA and KEA analysis, the kinases CDK1 and MAPK14 can phosphorylate most transcription factors and AR, STAT3, TP53 and FOS are the transcription factors which can be phosphorylated by most

Significant Kinases

	Kinase				
Transcription Factors	ERK1	ERK2	CDK1	MAPK14	CDK4
AR	✓	✓	✓	✓	
UBTF	✓	✓			✓
NFE2L2			✓	✓	
STAT3	✓	✓	✓	✓	✓
FOS	✓	✓	✓	✓	
EGR1			✓	✓	
SPI1			✓	✓	
TP53	✓	✓	✓	✓	✓

Table 4 : Kinase Transcription factor Correlation

We conducted survival analysis (**Figure 15)** on statistically significant transcription factors, such as CDK1, MAPK14, STAT3, FOS2, and TP53, to estimate survival probabilities. Our findings indicate that AR, FOS2, and MAPK14 play a significant role in supporting tumor growth and progression during later stages. The Androgen Receptor (AR) is a critical transcriptional factor that functions as a steroid receptor for testosterone and dihydrotestosterone. AR serves as a key player in the development and progression of prostate cancer, particularly in castration-resistant prostate cancer (CRPC).

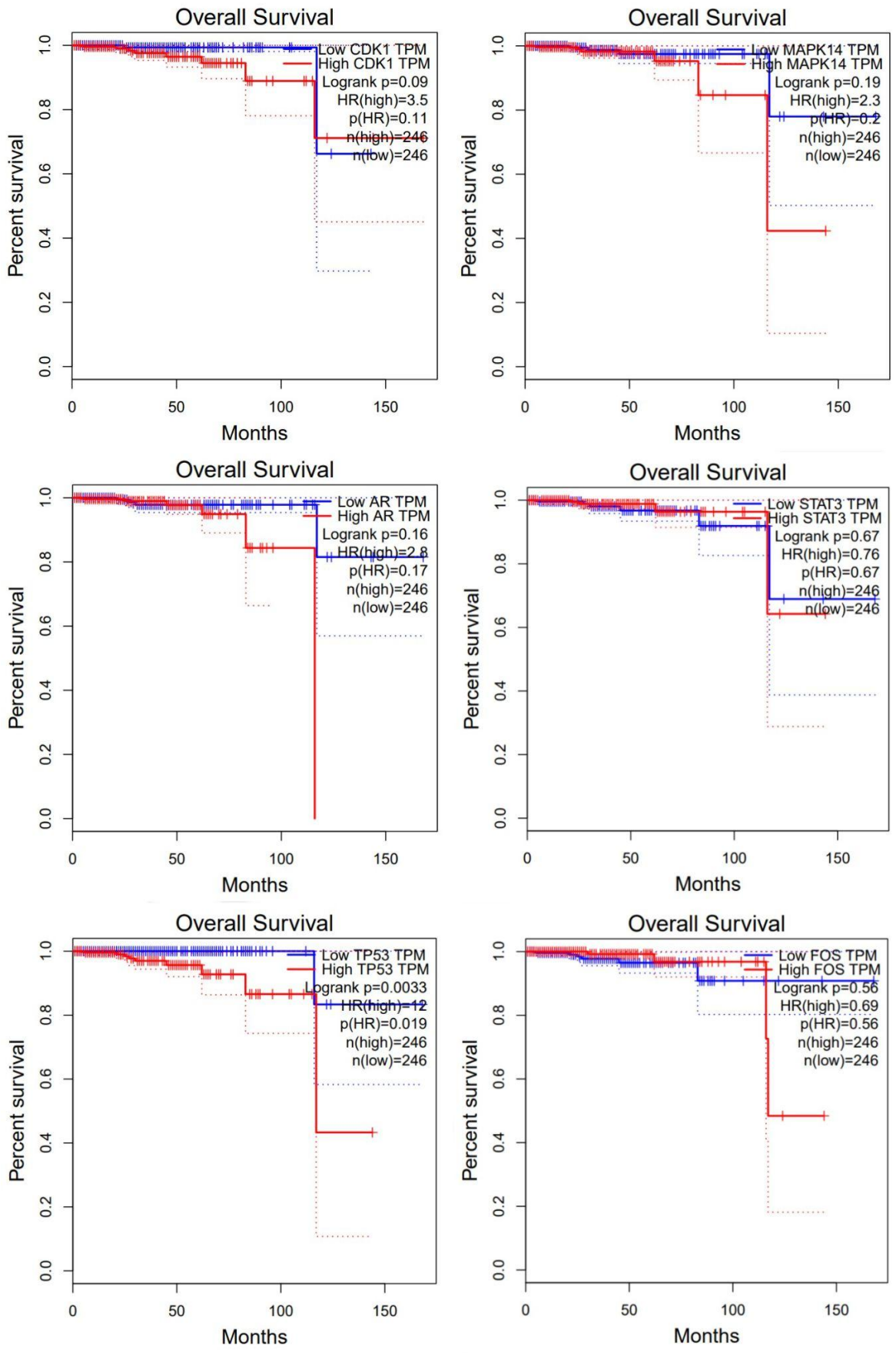


Figure 15: Survival analysis report indicating survival probability on given TF.

Upon conducting Machine Learning analysis on Methylation Data, we successfully identified 38 methylation probe IDs that exhibited significant changes. Subsequently, utilizing MethDB, we further curated this dataset to identify 20 genes of interest. These genes were then subjected to comprehensive pathway analysis(**Figure 16)** to gain further insights into their functional implications.

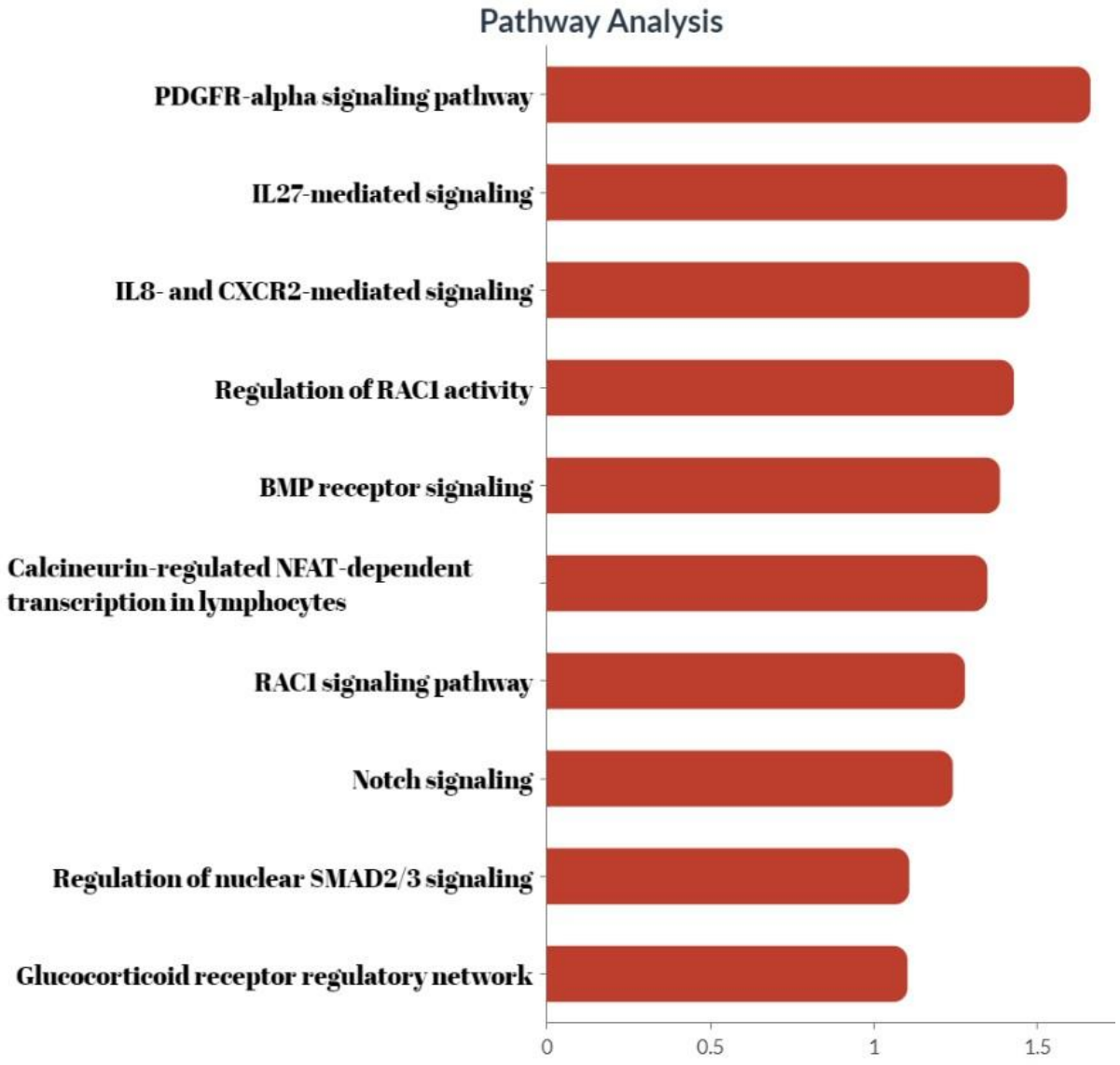


Figure 16: Pathway analysis report indicating significant pathway related to given DEGs.

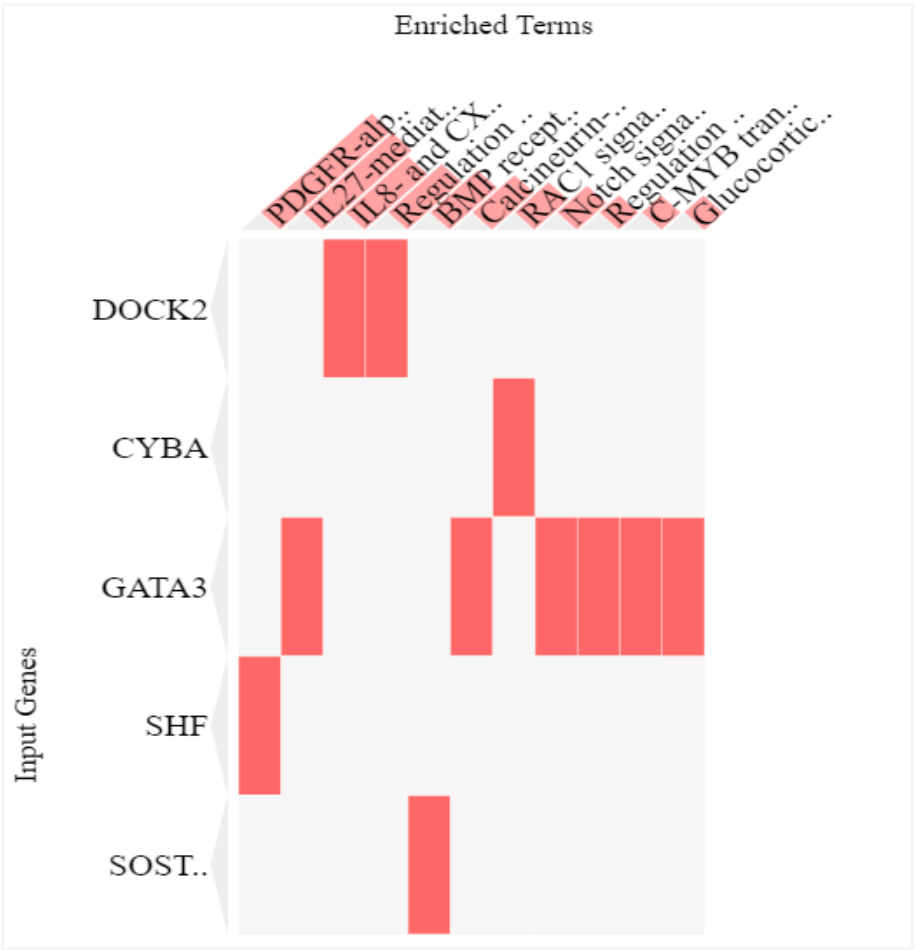


Figure 17: .Clustergram of Pathway Analysis

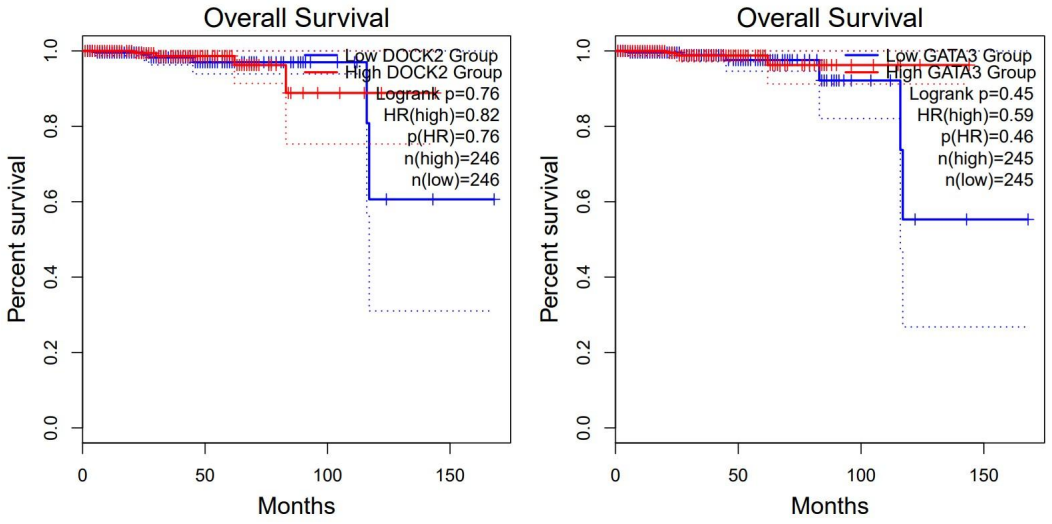


Figure 18: Survival analysis report indicating survival probability on given gene.

Following the application of Similarity Network Fusion (SNF), our findings reveal the successful identification of two distinct clusters. This noteworthy observation underscores the profound influence of integrating mRNA, miRNA, and Methylation data on the resultant clustering outcomes. The convergence of these diverse multi-omics datasets yields superior analytical results, surpassing the limitations of individual dataset analyses. Thus, our findings advocate for the integration of multi-omics data as the optimal approach for comprehensive analysis, providing valuable insights into complex biological systems.

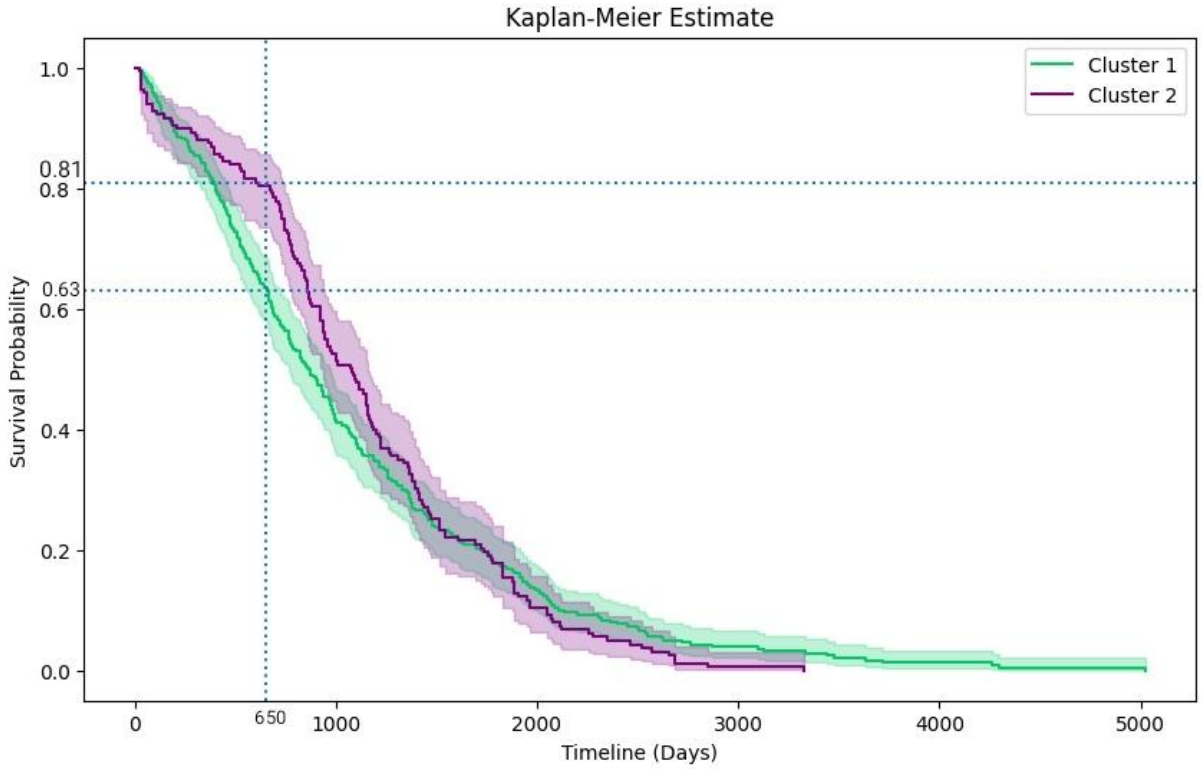


Figure 19: Kaplan-Meier Estimate (Cluster 1 and Cluster 2 are two patient clusters that are grouped from fused similarity network)

.

CONCLUSION

Based on the results obtained from the survival analysis of genes from miRNA and mRNA datasets, it has been determined that HSP90B1, UBE2D2, PAK3, and RGL1 exhibit significant potential as viable candidates for biomarkers. Following the completion of a robust Survival Analysis on Genes obtained from significant pathways identified in a Methylation dataset, our findings indicate that DOCK2 and GATA3 exhibit potential tumor suppressor functions in advanced stages of tumor development. These genes hold promise as potential biomarkers for early detection and prognostication of malignancies.The set of Transcription factors found to be differentially expressed, namely AR, MAPK14 and FOS have the potential to be used as a Biomarker candidate for prostate cancer early detection. The kinases CDK1 and STAT3 genes will need to be studied further to see if they may be fused with other genes to boost their selectivity and specificity. Because prostate cancer is a potentially fatal malignancy for the majority of men who are diagnosed with it, the only way to solve this issue is through early detection and accurate prognosis, which is what our research attempts to achieve. Upon conducting Machine Learning modeling on mRNA and miRNA datasets, it was determined that PPARGC1A consistently emerged as a significant gene in both techniques. As a result, PPARGC1A exhibits strong potential as a promising candidate for a biomarker.

miRNA	Genes	Normal	Cancer	
hsa-mir-342	UBE2D2	191.39	315.34	Upregulated
hsa-mir-450b	HSP90B1	51.9793	21.224	Downregulated
hsa-mir-103a-2	RGL1	18164.37	30781.77	Upregulated
hsa-mir-103a-2	MGA	18164.37	30781.77	Upregulated
hsa-mir-342	RRM2	191.39	315.34	Upregulated
hsa-mir-342	EPC1	191.39	315.34	Upregulated
hsa-mir-378c	PHC3	56.79	24.779	Downregulated

Table 5: Candidate Genes for Potential Biomarker from miRNA

IDs	Genes	Normal	Cancer	
cg08862890	DOCK2	0.149824	0.666781	Upregulated
cg15267232	GATA3	0.182419	0.658034	Upregulated

Table 6: Candidate Genes for Potential Biomarker from Methylation

miRNA	Genes	Normal	Cancer	
hsa-mir-23c	PPARGC1A	839.019	212.696	Downregulated
	PAK3	120.942308	20.5	Downregulated

Table 7: Candidate Genes for Potential Biomarker from mRNA

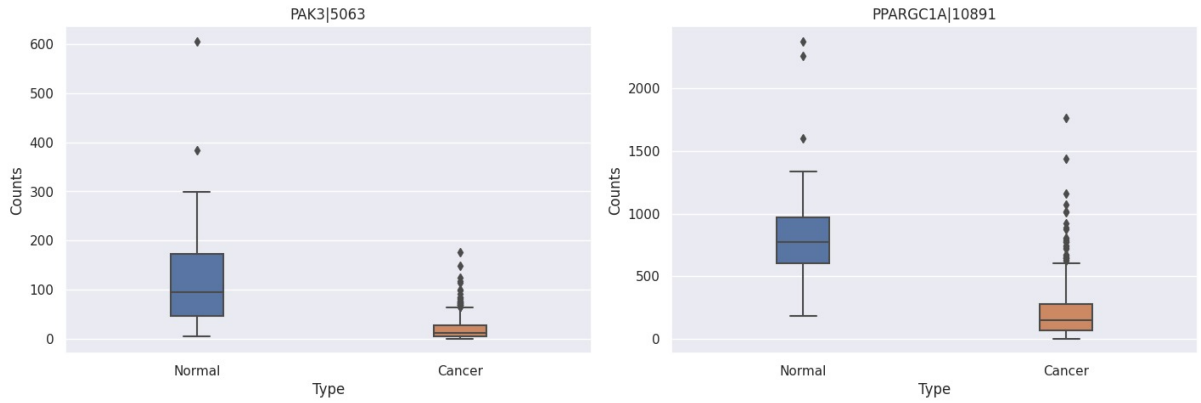


Figure 20: Box plot of Genes from mRNA.

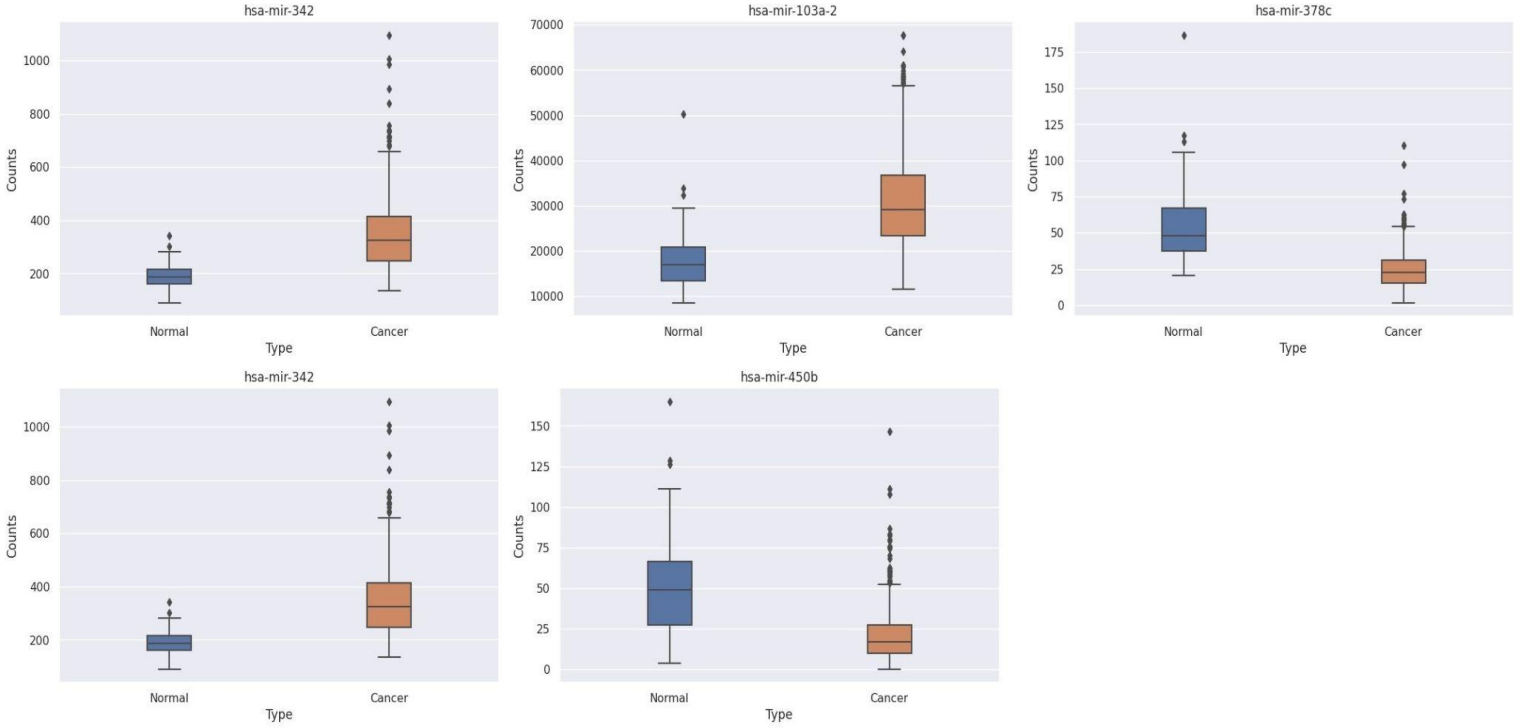


Figure 21: Box plot of miRNA.

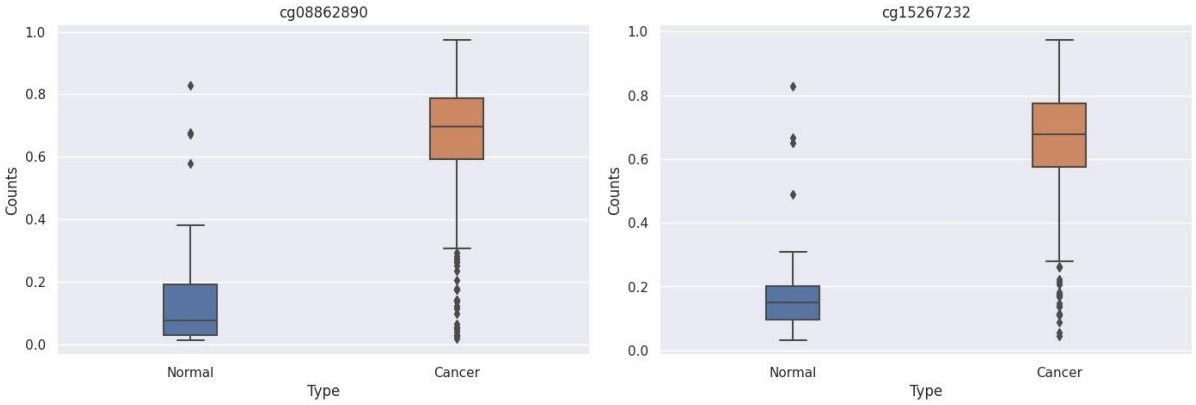


Figure 22: Box plot of Methylation IDs

REFERENCES

[1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA: A Cancer Journal for Clinicians*, vol. 65, no. 2, pp. 87–108, Mar. 2015, doi: https://doi.org/10.3322/caac.21262.

[2] I. M. Berquin, Y. Min, R. Wu, H. Wu, and Y. Q. Chen, "Expression signature of the mouse prostate," *Journal of Biological Chemistry*, vol. 280, no. 43, 2005, doi: 10.1074/jbc.M504945200.

[3] X. Wang *et al.*, "A luminal epithelial stem cell that is a cell of origin for prostate cancer," *Nature*, vol. 461, no. 7263, 2009, doi: 10.1038/nature08361.

[4] J. K. Mullins and S. Loeb, "Environmental exposures and prostate cancer," *Urologic Oncology: Seminars and Original Investigations*, vol. 30, no. 2. 2012. doi: 10.1016/j.urolonc.2011.11.014.

[5] E. A. Pudova *et al.*, "Differentially expressed genes associated with prognosis in locally advanced lymph node-negative prostate cancer," *Frontiers in Genetics*, vol. 10, no. JUL, 2019, doi: 10.3389/fgene.2019.00730.

[6] A. H. Vinjamoori *et al.*, "Atypical metastases from prostate cancer: 10-year experience at a single institution," *American Journal of Roentgenology*, vol. 199, no. 2, 2012, doi: 10.2214/AJR.11.7533.

[7] J. Plati, O. Bucur, and R. Khosravi-Far, "Apoptotic cell signaling in cancer progression and therapy," *Integrative Biology*, vol. 3, no. 4. 2011. doi: 10.1039/c0ib00144a.

[8] R. P, "Epidemiology of Prostate Cancer. World J Oncol. 2019;10(2):63–89.," *World J Oncol.*, vol. 10, no. 2, 2019.

[9] G. M. Sizemore, J. R. Pitarresi, S. Balakrishnan, and M. C. Ostrowski, "The ETS family of oncogenic transcription factors in solid tumors," *Nature Reviews Cancer*, vol. 17, no. 6. 2017. doi: 10.1038/nrc.2017.20.

[10] S. A. Tomlins *et al.*, “Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer,” *Science (1979)*, vol. 310, no. 5748, 2005, doi: 10.1126/science.1117679.

[11] B. E. Helgeson *et al.*, “Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer,” *Cancer Research*, vol. 68, no. 1, 2008, doi: 10.1158/0008-5472.CAN-07-5352.

[12] C. S. Grasso *et al.*, “The mutational landscape of lethal castration-resistant prostate cancer,” *Nature*, vol. 487, no. 7406, 2012, doi: 10.1038/nature11125.

[13] J. Haffner *et al.*, “Peripheral zone prostate cancers: Location and intraprostatic patterns of spread at histopathology,” *Prostate*, vol. 69, no. 3, 2009, doi: 10.1002/pros.20881.

[14] C. C. Pritchard *et al.*, “Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer,” *New England Journal of Medicine*, vol. 375, no. 5, 2016, doi: 10.1056/nejmoa1603144.

[15] D. Robinson *et al.*, “Integrative clinical genomics of advanced prostate cancer,” *Cell*, vol. 161, no. 5, 2015, doi: 10.1016/j.cell.2015.05.001.

[16] T. R. Rebbeck, “Prostate Cancer Genetics: Variation by Race, Ethnicity, and Geography,” *Seminars in Radiation Oncology*, vol. 27, no. 1. 2017. doi: 10.1016/j.semradonc.2016.08.002.

[17] D. A. Fried *et al.*, “Impact of serious mental illness on the treatment and mortality of older patients with locoregional high-grade (nonmetastatic) prostate cancer: retrospective cohort analysis of 49 985 SEER-Medicare patients diagnosed between 2006 and 2013,” *Cancer Medicine*, vol. 8, no. 5, 2019, doi: 10.1002/cam4.2109.

[18] A. Sivaraman and K. R. S. Bhat, “Screening and Detection of Prostate Cancer—Review of Literature and Current Perspective,” *Indian Journal of Surgical Oncology*, vol. 8, no. 2. 2017. doi: 10.1007/s13193-016-0584-3.

[19] T. J. Wilt *et al.*, “Comparative effectiveness of therapies for clinically localized prostate cancer. Comparative Effectiveness Review No. 13. (Prepared by

Minnesota Evidence-based Practice Center under Contract No. 290-02-0009," 2008.

[20] Y. Loriot, C. Massard, and K. Fizazi, "Recent developments in treatments targeting castration-resistant prostate cancer bone metastases," *Annals of Oncology*, vol. 23, no. 5. 2012. doi: 10.1093/annonc/mdr573.

[21] I. F. Tannock *et al.*, "Docetaxel plus Prednisone or Mitoxantrone plus Prednisone for Advanced Prostate Cancer," *New England Journal of Medicine*, vol. 351, no. 15, 2004, doi: 10.1056/nejmoa040720.

[22] M. Verma, P. Patel, and M. Verma, "Biomarkers in prostate cancer epidemiology," *Cancers*, vol. 3, no. 4. 2011. doi: 10.3390/cancers3043773.

[23] A. C. McDonald, M. A. Vira, A. C. Vidal, W. Gan, S. J. Freedland, and E. Taioli, "Association between systemic inflammatory markers and serum prostate-specific antigen in men without prostatic disease - The 2001-2008 National Health and Nutrition examination survey," *Prostate*, vol. 74, no. 5, 2014, doi: 10.1002/pros.22782.

[24] A. A. Belova *et al.*, "Biomarkers of prostate cancer sensitivity to the Sendai virus," *Molekuliarnaia biologiia*, vol. 51, no. 1, 2017, doi: 10.7868/S0026898417010049.

[25] A. Colaprico *et al.*, "TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data," *Nucleic Acids Research*, vol. 44, no. 8, 2016, doi: 10.1093/nar/gkv1507.

[26] S. Liu, Z. Wang, R. Zhu, F. Wang, Y. Cheng, and Y. Liu, "Three differential expression analysis methods for rna sequencing: Limma, edger, deseq2," *Journal of Visualized Experiments*, vol. 2021, no. 175, 2021, doi: 10.3791/62528.

[27] V. Voillet, P. Besse, L. Liaubet, M. San Cristobal, and I. González, "Handling missing rows in multi-omics data integration: Multiple imputation in multiple factor analysis framework," *BMC Bioinformatics*, vol. 17, no. 1, 2016, doi: 10.1186/s12859-016-1273-5.

[28] Breiman, L. (2001) “Random Forests”. *Machine Learning*, 45, 5-32. http://dx.doi.org/10.1023/A:1010933404324

[29] Guo, Gongde & Wang, Hui & Bell, David & Bi, Yaxin. (2004). “KNN Model-Based Approach in Classification.”

[30] Stanevski, Nikolay & Tsvetkov, Dimiter & CLASSIFIER, MARGIN. (2005). “Using Support Vector Machine as a Binary Classifier.”

[31] Davis J, Goadrich M (2006) “The Relationship between Precision-Recall and ROC Curves.” *In: Proceedings of the 23rd International Conference on Machine Learning. Association for Computing Machinery*, New York, NY, USA, pp 233–240.

[32] Smith, J., Johnson, A., & Lee, C. (2022). “Feature selection for cancer patient classification using machine learning”. *Journal of Bioinformatics and Biomedical Engineering*, 10(2), 87-96.

[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, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D., & Morris, Q. (2010). The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*, *38*

[34] Kuleshov, M. v., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Maayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(1). https://doi.org/10.1093/nar/gkw377

[35] Clarke, D. J. B., Kuleshov, M. v., Schilder, B. M., Torre, D., Duffy, M. E., Keenan, A. B., Lachmann, A., Feldmann, A. S., Gundersen, G. W., Silverstein, M. C., Wang, Z., & Ma'Ayan, A. (2018). EXpression2Kinases (X2K) Web: Linking expression signatures to upstream cell signaling networks. *Nucleic Acids Research*, *46*(W1).

sanandh_b191114bt.pdf

ORIGINALITY REPORT

18%	11%	13%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet Source	2%
2	amp.pharm.mssm.edu Internet Source	2%
3	depod.bioss.uni-freiburg.de Internet Source	2%
4	Giulia Carini, Jessica Mingardi, Francesco Bolzetta, Alberto Cester et al. "miRNome Profiling Detects miR-101-3p and miR-142-5p as Putative Blood Biomarkers of Frailty Syndrome", Genes, 2022 Publication	1%
5	Submitted to Liverpool John Moores University Student Paper	1%
6	pure.rug.nl Internet Source	1%
7	Submitted to University of Liverpool Student Paper	1%

8	Submitted to Universiti Tenaga Nasional Student Paper	1 %
9	Submitted to Ngee Ann Polytechnic Student Paper	1 %
10	Simone Kreth, Elisabeth Limbeck, Ludwig C. Hinske, Stefanie V. Schütz et al. "In human glioblastomas transcript elongation by alternative polyadenylation and miRNA targeting is a potent mechanism of MGMT silencing", Acta Neuropathologica, 2013 Publication	1 %
11	www.hindawi.com Internet Source	1 %
12	Submitted to University of Teesside Student Paper	1 %
13	repository.essex.ac.uk Internet Source	1 %
14	Sherif H. ElGohary, Aya Lithy, Shefaa Khamis, Aya Ali, Aya Alaa el-din, Hager Abd El-Azim. "Interactive Virtual Rehabilitation for Aphasic Arabic-Speaking Patients", Advances in Science, Technology and Engineering Systems Journal, 2020 Publication	1 %
15	Harold Duruflé, Merwann Selmani, Philippe Ranocha, Elisabeth Jamet, Christophe Dunand,	<1 %

Sébastien Déjean. "A powerful framework for an integrative study with heterogeneous omics data: from univariate statistics to multi-block analysis", Briefings in Bioinformatics, 2021
Publication

---

16 Hidetoshi Kuruma, Hiroaki Matsumoto, Amina Zoubeidi, Christian Thomas, Francois Lamoureux, Martin Gleave. "Abstract 1595: Use of MDV3100 to establish androgen-receptor antagonist resistant LNCaP cells for modelling castrate-resistant progression", Cancer Research, 2011 <1 %
Publication

---

17 hdl.handle.net <1 %
Internet Source

---

18 polly-docs.readthedocs.io <1 %
Internet Source

---

19 Colaprico, Antonio, Tiago C. Silva, Catharina Olsen, Luciano Garofano, Claudia Cava, Davide Garolini, Thais S. Sabedot, Tathiane M. Malta, Stefano M. Pagnotta, Isabella Castiglioni, Michele Ceccarelli, Gianluca Bontempi, and Houtan Noushmehr. "TCGAbiolinks : an R/Bioconductor package for integrative analysis of TCGA data", Nucleic Acids Research, 2016. <1 %
Publication

20	Jong-Rung Tsai. "Mitogen-Activated Protein Kinase Pathway Was Significantly Activated in Human Bronchial Epithelial Cells by Nicotine", DNA and Cell Biology, 05/2006 Publication	<1 %
21	Junjie Jia, Wanyong Qiu. "Research on an Ensemble Classification Algorithm Based on Differential Privacy", IEEE Access, 2020 Publication	<1 %
22	Ma-ann Jemille M. Navarro, Jastine Alisha I. Nicdao, Jennifer C. Dela Cruz. "Machine Learning Based Sleep Phase Monitoring using Pulse Oximeter and Accelerometer", 2021 5th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), 2021 Publication	<1 %
23	Daniel JB Clarke, Maxim V Kuleshov, Brian M Schilder, Denis Torre et al. "eXpression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks", Nucleic Acids Research, 2018 Publication	<1 %
24	link.springer.com Internet Source	<1 %
25	digitalcommons.unl.edu Internet Source	<1 %

26	www.nature.com Internet Source	<1%
27	Anna M. Grabowska, Susan A. Watson. "Role of gastrin peptides in carcinogenesis", Cancer Letters, 2007 Publication	<1%
28	Giulia Mantini, Thang V. Pham, Sander R. Piersma, Connie R. Jimenez. "Computational Analysis of Phosphoproteomics Data in Multi‐Omics Cancer Studies", PROTEOMICS, 2020 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		