An *in-silico* Approach for Recognition of Long non-coding RNA-Associated Competing Endogenous RNA Axes in Prostate Cancer

Mohammad Taheri[1], Arash Safarzadeh[2], Soudeh Ghafouri-Fard[2*], Aria Baniahmad[1*]

1. Institute of Human Genetics, Jena University Hospital, Jena, Germany.
2. Department of Medical Genetics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Corresponding authors: Soudeh Ghafouri-Fard and Aria Baniahmad

s.ghafourifard@sbmu.ac.ir and aria.baniahmad@med.uni-jena.de

Abstract

Prostate cancer is among the most central sources of cancer-related mortalities. In order to find novel candidates for therapeutic strategies in this kind of cancer, we developed an *in-silico* method for identification of competing endogenous RNA network. According to the microarray data analyses between prostate tumor and normal specimens, we attained 1312 differentially expressed (DE)mRNAs, including 778 down-regulated DEmRNAs (such as CXCL13 and BMP5) and 584 up-regulated DEmRNAs (such as OR51E2 and LUZP2), 39 DElncRNAs, including 10 down-regulated DElncRNAs (such as UBXN10-AS1 and FENDRR) and 29 up-regulated DElncRNAs (such as PCA3 and LINC00992) and 10 DEmiRNAs, including 2 down-regulated DEmiRNAs (such as MIR675 and MIR1908) and 8 up-regulated DEmiRNAs (such as MIR6773 and MIR4683). We constructed the ceRNA network between these transcripts. We also evaluated the related signaling pathways and the significance of these RNAs in prediction of survival of patients with prostate cancer. This study provides novel candidates for construction of specific treatment routes for prostate cancer.

Key words: prostate cancer, ceRNA, lncRNA, miRNA

Introduction

Long non-coding RNAs (lncRNAs) are a group of transcripts with sizes more than 200 nt. They have diverse regulatory roles on expression of genes. This kind of epigenetic regulators influence epigenetic marks mainly in the nucleus, thus affecting gene transcription (1). In addition, they can serve as molecular sponges for microRNAs (miRNAs) (2), thus regulating expression of miRNA targets. This mode of action leads to establishment of lncRNA/miRNA/mRNA axes that contribute in the several physiological processes. Dysregulation of lncRNAs can lead to several disorders through induction of imbalances in these molecular axes. LncRNAs that act as molecular sponges are called competitive endogenous RNAs (ceRNAs). Evaluation of the activity of ceRNA networks has practical significance, particularly in unraveling the mechanism of carcinogenesis. Prostate cancer as one of the most important sources of cancer-related mortalities (3), is one of the hot topics in the field of cancer-related ceRNA networks. For instance, Li et al. have constructed a prostate cancer-specific ceRNA network by incorporating lncRNA/miRNA/mRNA interactions based on experimental and *in silico* methods. Their method has led to identification of 42 significant prostate cancer-survival-associated triplets which make a condensed subnetwork consisted of only 25 nodes. The latter finding shows the involvement of some nodes in many triplets. MIR22HG/hsa-mir-21/TGFBR2 and MIR22HG/hsa-mir-21/BCL2 triplets have been recognized as two significantly survival associated triplets with the greatest average degree in the detected subnetwork (4). Similarly, Guo et al. have constructed a prostate cancer-specific core ceRNA network with the capability to be applied as diagnostic and prognostic marker in this type of cancer (5).

The current study aimed at identification of the ceRNA network in prostate cancer using an *in-silico* approach. The related signaling pathways and the significance of these RNAs in prediction of survival of prostate cancer patients have also been evaluated.

Methods

Microarray Data Assessment

The human expression profile of GSE69223, GSE46602, and GSE55945, all with [HG-U133 Plus 2] Affymetrix Human Genome U133 Plus 2.0 Array, which contained 30, 50, and 21 specimens, respectively, were obtained using the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/). We chose 15 prostate tumor and 15 normal tissue specimens from GSE69223, 36 prostate tumor samples from GSE46602 and 13 prostate tumor samples from GSE55945 for additional analyses. This data contained lncRNAs, miRNAs and mRNAs expression profile.

Data processing, meta-analysis and evaluation of data quality

The statistical programming language R was used to analyse and combine all of the microarray data. Data from Affymetrix and Agilent was initially normalized individually for pre-processing using the preprocessCore package's normalizeQuantiles function (version 1.58.0). (https://bioconductor.org/packages/release/bioc/html/preprocessCore.html). With the purpose of exclusion of batch effects (non-biological differences), we used the ComBat function from the R Package Surrogate Variable Analysis (SVA) (version 3.44.0) (6). Batch effect removal was then evaluated. We showed the result of the meta-analysis in a unit expression matrix.

Assessment of differentially expressed transcripts

We used the Limma package (version 3.52.3) (7) in R language to find differentially expressed mRNAs (DEmRNAs), lncRNAs (DElncRNAs) and miRNAs (DEmiRNAs) between prostate tumor and normal specimens. DEmRNAs, DElncRNAs and DEmiRNAs were appraised with the cut-off criteria of false discovery rate (FDR; adjusted p value) < 0.05 and |log2 fold Change (FC)| > 0.5. Subsequently, we identified DElncRNAs and DEmiRNAs using HUGO gene nomenclature.

Two-Way Clustering of DEGs

Expression levels of significant DEmRNAs, DElncRNAs, and DEmiRNAs were obtained and used in the pheatmap package (version 1.0.12) (8) in R language to conduct the two-way clustering based on the Euclidean distance.

Gene Ontology (GO) Enrichment

ClusterProfiler R package (version 4.4.4) (9) was applied to conduct gene ontology (GO) enrichment and investigation of the functions of the significantly up-regulated and down-regulated DEGs. The functional category criteria were established at an adjusted p-value<0.05.

KEGG Pathway Analysis

KEGG pathway analysis of considerably DEGs was performed using the KEGG database (10).

PPI Network Construction

PPI network for DEGs was identified using the STRING database (11). Highest level of confidence was utilized to create the interactions parameter (confidence score >0.9). Protein interactions were visualized using the Cytoscape software v3.9 (12). The top 20 DEGs related to hub genes were lastly detected using the Cytohubba plugin (13) of Cytoscape using the betweenness method.

ceRNA Network and Hub Genes

A ceRNA network was constructed through these steps: 1) Searching the miR2Disease database (http://watson.compbio.iupui.edu:8080/miR2Disease/index.jsp) (14) utilizing the term "prostate cancer" for the prostate cancer (PC)-related miRNAs. 2) measuring the interactions between lncRNAs and miRNAs based on the PC-related miRNAs using miRcode (http://www.mircode.org/); 2) Application of miRDB (http://www.mirdb.org/) (15), miRTarBase (https://mirtarbase.cuhk.edu.cn/) (16), TargetScan (http://www.targetscan.org/) (17) and miRWalk (http://129.206.7.150/) (18) for predicting miRNAs-targeted mRNAs; 3) Discovery of the intersections of the DElncRNAs and DEmRNAs, and formation of lncRNA/mRNA/miRNA ceRNA network using Cytoscape v3.9 and 4) predicting hub genes using cytohubba plugin based on degree approach.

Confirmation of hub genes via expression values

Expression value of hub genes was evaluated using the ualcan database (19).

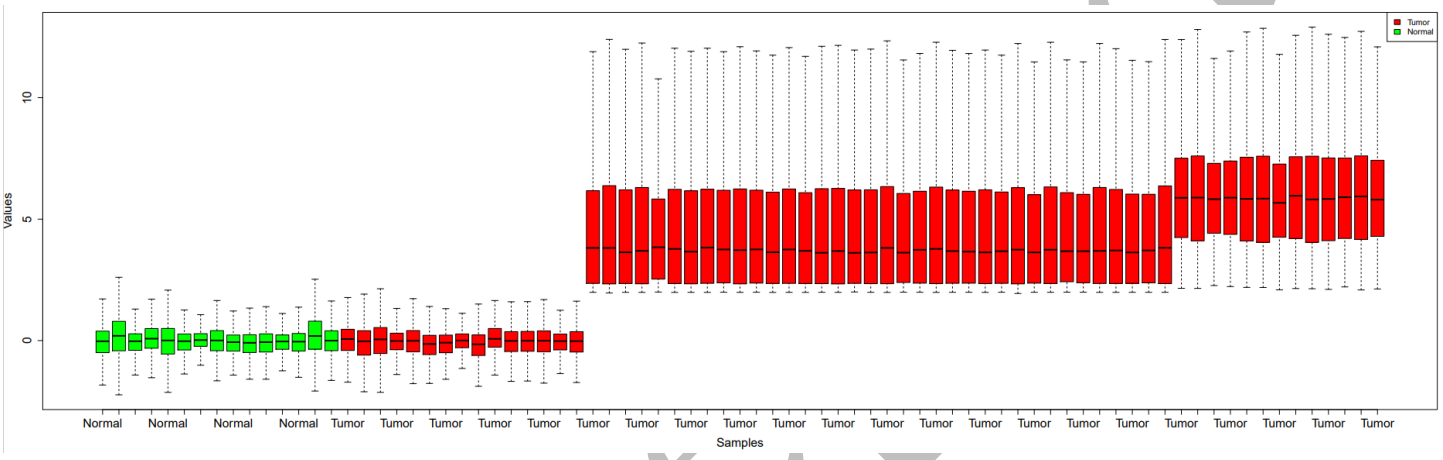
Survival analysis

Survival package (version 3.5.0) (https://CRAN.R-project.org/package=survival) in R was utilized to find survival curves. The clinical data for patients with prostate cancer was obtained from TCGA (PRAD-TCGA). Univariate survival analysis was performed using Kaplan-Meier curves. Statistics were considered significant for P-value<0.05. The start time to the end time in this analysis is from 0 to 5000 days.

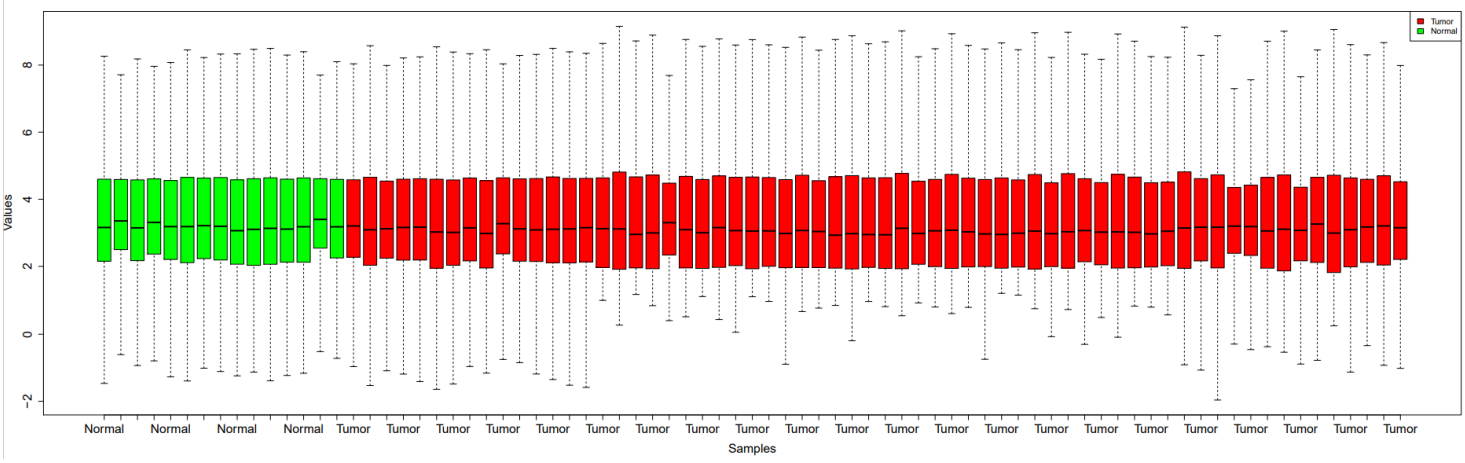
Results

Microarray Data Processing

Figure 1 depicts the boxplots of raw data, normalized data after batch effect removal and quantile normalization. These plots show the reliability of the quality of the expression data. Moreover, the boxplot of the preprocessed data had good normalization. Figure 2 shows the Euclidean distances between the samples after batch effect removal. In the PCA plot (Figure 3), 101 specimens are shown in the 2D plane traversed by their first two principal components (PC1 and PC2) According to this plot, the samples had a good dispersion following removal of batch effect.



A



B

Figure 1. Boxplots after combining datasets. A) First box plot shows the combination of datasets B) The second boxplot shows the merged datasets after removing the batch effect removal.

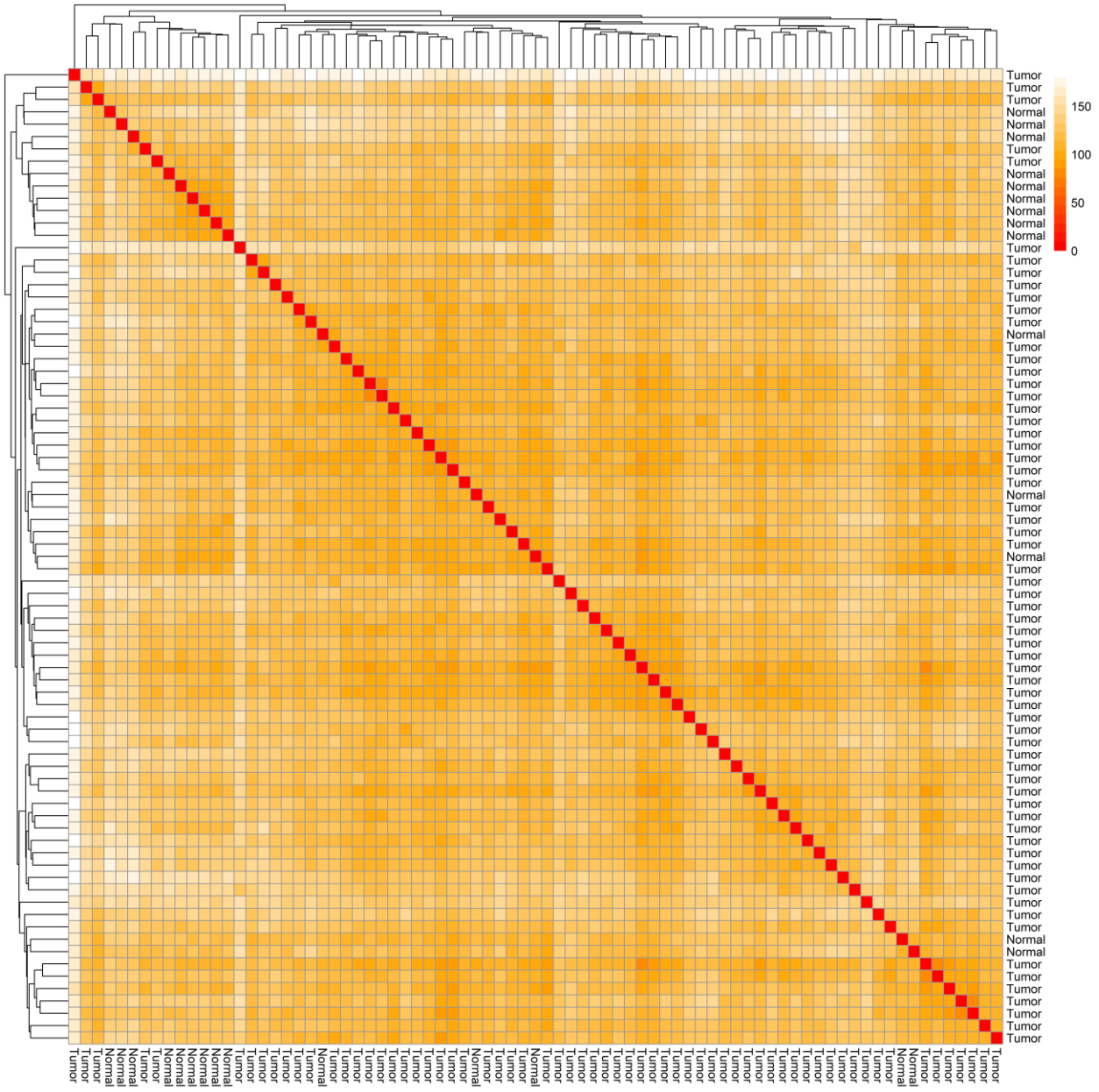


Figure 2. The Euclidean distances between samples. Based on the Euclidean distance, hierarchical clustering between the samples has been established; Legend shows the distance value between samples.

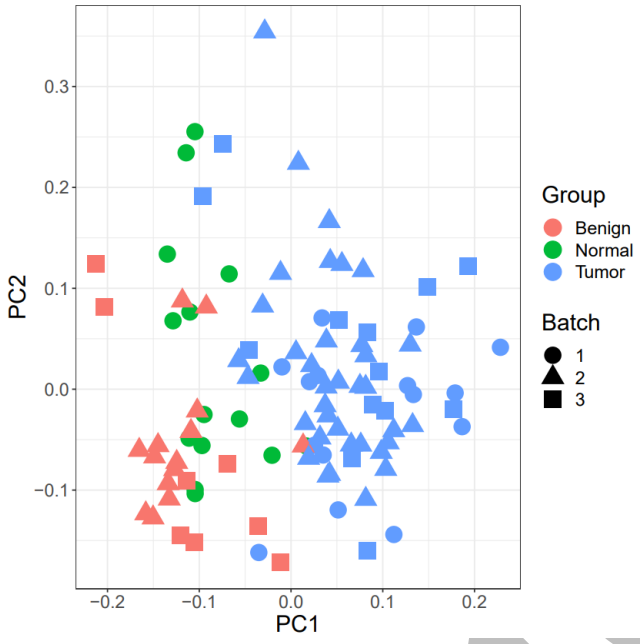


Figure 3. PCA plot. The Batch implies that the data includes three platforms. Also, healthy benign and tumor samples were divided into three groups.

DEGs Analysis

Based on the microarray data analysis between prostate tumor and normal samples by Limma, we analyzed differentially expressed mRNA, lncRNA and miRNAs and obtained 1312 DEmRNAs, including 778 downregulated DEmRNAs (such as CXCL13 and BMP5) and 584 upregulated DEmRNAs (such as OR51E2 and LUZP2), 39 DElncRNAs, including 10 downregulated DElncRNAs (such as UBXN10-AS1 and FENDRR) and 29 upregulated DElncRNAs (such as PCA3 and LINC00992) and 10 DEmiRNAs, including 2 downregulated DEmiRNAs (such as MIR675 and MIR1908) and 8 upregulated DEmiRNAs (such as MIR6773 and MIR4683). The most significantly upregulated and downregulated DEmRNAs, DElncRNAs, and DEmiRNAs are shown in Tables 1-3, respectively.

Table 1. The top 10 up- and downregulated DEmRNAs between prostate tumor and normal samples.

Down-regulated			**Up-regulated**		
DEmRNA	**Log FC**	**Adjusted *P* value**	**DEmRNA**	**Log FC**	**Adjusted P value**
CXCL13	-2.914284	0.0001	**OR51E2**	2.410149	0.002
BMP5	-2.549856	0.0002	**LUZP2**	2.205251	0.0005
WIF1	-2.453527	0.0001	**HOXC6**	2.178773	0.0001
NELL2	-2.383551	0.00003	**HPN**	2.004699	0.000000002
SLC14A1	-2.214827	0.00004	**C15orf48**	1.984213	0.01
DAPL1	-2.038041	0.002	**TRPM4**	1.971099	0.0000005
KRT23	-2.011459	0.00004	**B3GAT1**	1.834909	0.0002
LGR6	-1.843800	0.00001	**PRCAT47**	1.807160	0.01
CFD	-1.792887	0.00002	**THBS4**	1.804980	0.0001
PTGS1	-1.750086	0.00001	**DLX1**	1.804387	0.0001

Table 2. The up- and downregulated DElncRNAs between prostate tumor and normal samples.

Down-regulated			**Up-regulated**		
DElncRNA	**Log FC**	**Adjusted *P* value**	**DElncRNA**	**Log FC**	**Adjusted P value**
UBXN10-AS1	-1.085041069	0.0002	**PCA3**	2.194085974	0.001
FENDRR	-1.001246896	0.0002	**LINC00992**	2.131068688	0.00008
MAGI2-AS3	-0.951283196	0.00003	**C1QTNF3-AMACR**	2.129359564	0.0001
MAGI2-IT1	-0.877464251	0.0002	**PCAT18**	1.444563038	0.03
BOLA3-AS1	-0.818242331	0.00004	**ERVH48-1**	1.384998066	0.02
ADAMTS9-AS2	-0.79674974	0.00003	**DRAIC**	1.351053821	0.006
HCG11	-0.765278673	0.0001	**FOXP4-AS1**	1.343377023	0.0009
TBX5-AS1	-0.74698408	0.00001	**LINC00842**	1.213140476	0.004
RBMS3-AS3	-0.714422829	0.0001	**LINC00920**	1.030649901	0.04
MEG3	-0.707122105	0.0002	**DANCR**	1.026630806	0.0005
			PRRT3-AS1	0.916404009	0.03
			SNHG19	0.900564908	0.0001
			PCAT7	0.827071849	0.02
			C8orf34-AS1	0.824415749	0.01
			CRNDE	0.733537617	0.01
			SNHG9	0.711240397	0.04
			LINC01351	0.708756468	0.03
			ENO1-AS1	0.689974397	0.01
			ZNF793-AS1	0.68123815	0.009
			MCF2L-AS1	0.645850215	0.007
			PRKAG2-AS1	0.640111087	0.02
			PCAT6	0.622871409	0.02
			POU6F2-AS2	0.618162182	0.03
			LINC00862	0.596493068	0.02
			RPARP-AS1	0.595574259	0.004
			LEF1-AS1	0.592353189	0.02
			LINC00665	0.538956003	0.002
			LINC00973	0.520799038	0.02

			LINC01128	0.507755709	0.005

Table 3. The significantly up- and downregulated DEmiRNAs between prostate tumor and normal samples.

Down-regulated			**Up-regulated**		
DEmiRNA	**Log FC**	**Adjusted *P* value**	**DEmiRNA**	**Log FC**	**Adjusted P value**
MIR675	-1.461212788	0.003	**MIR6773**	1.110887917	0.0000363
MIR1908	-0.809060479	0.005	**MIR4683**	0.903634366	0.0003
			MIR7110	0.875949754	0.001
			MIR3658	0.746514424	0.0001
			MIR3185	0.670389146	0.01
			MIR6824	0.617549387	0.02
			MIR4647	0.599044810	0.004
			MIR4784	0.549181183	0.01

Volcano plot was depicted with the EnhancedVolcano package (version 1.14.0) (20) in R to compare the variation in miRNA, lncRNA, and mRNA expression between prostate tumor and normal samples (Figure 4). Moreover, the two-way clustering showed 20 clearly distinct DEmRNA expression patterns between prostate tumor and normal samples (Figure 5a). The expression of DElncRNAs and DEmiRNAs is demonstrated in two heatmaps (Figure 5b).

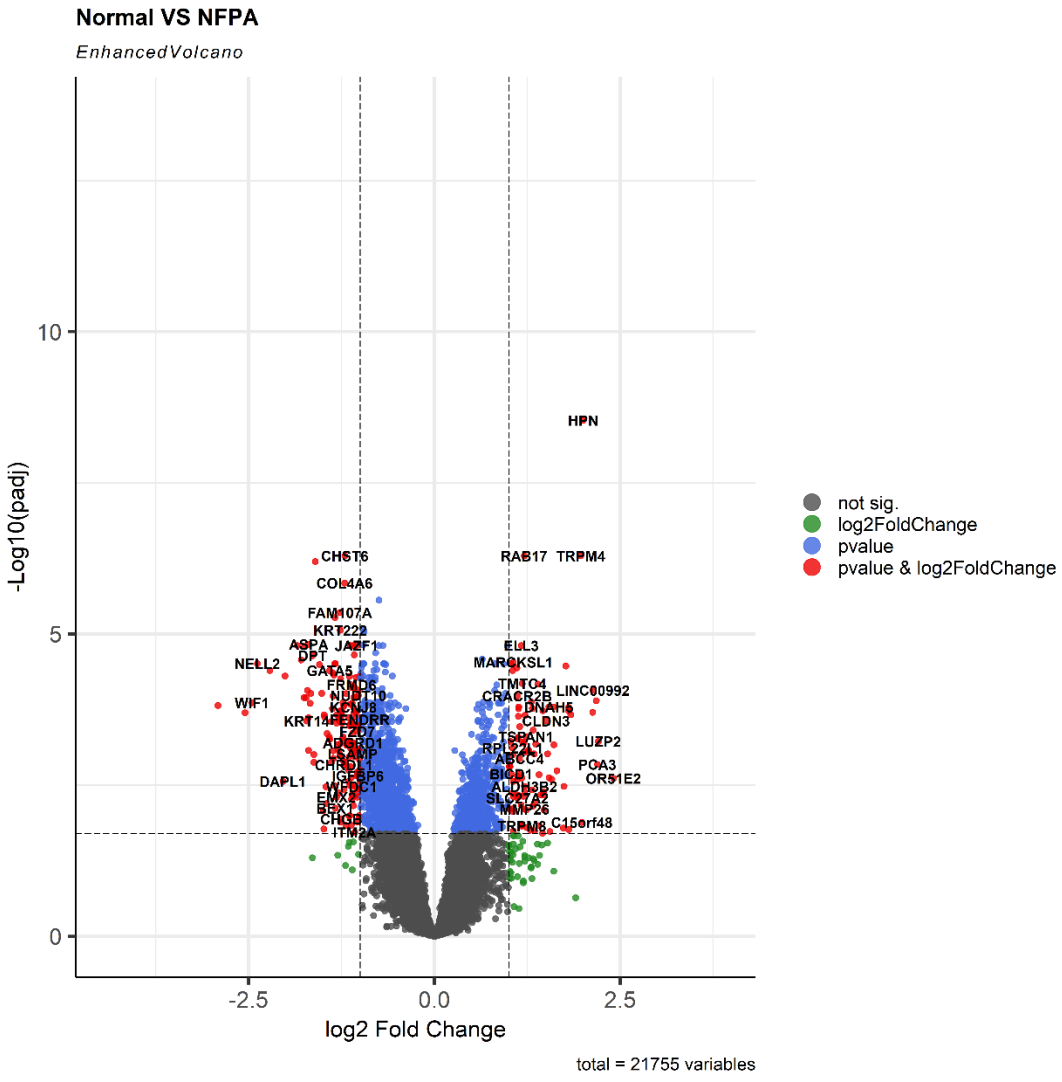
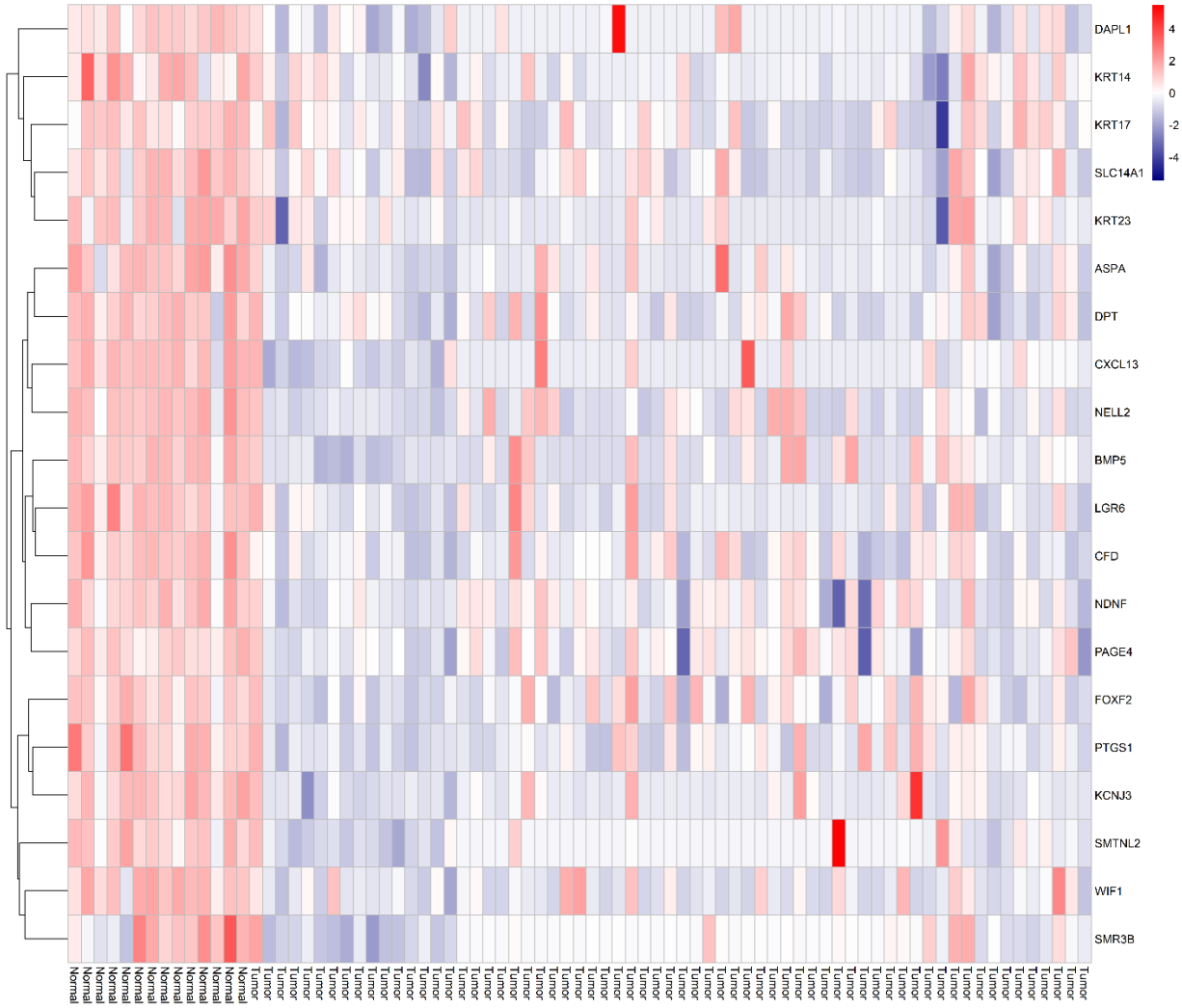
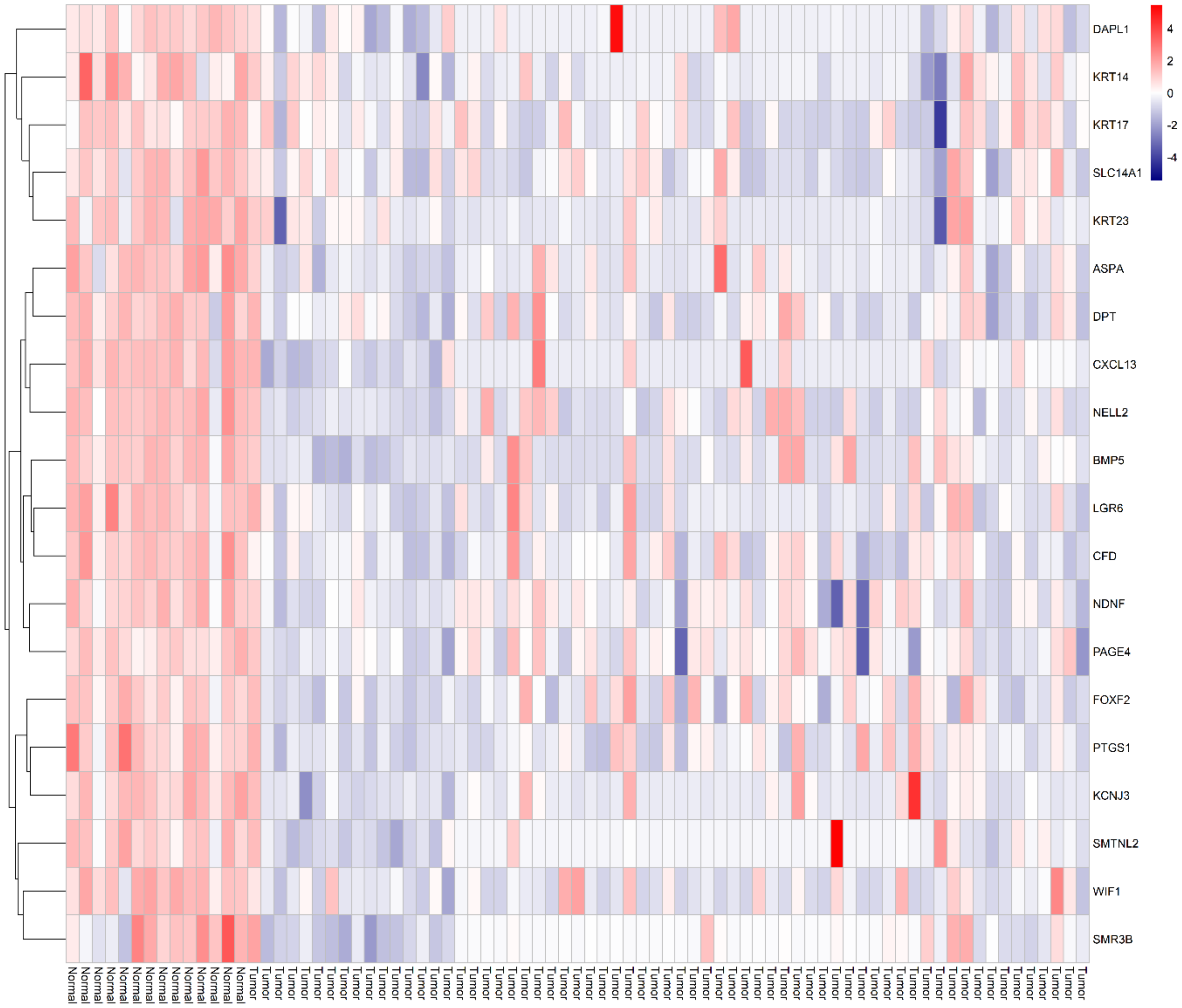
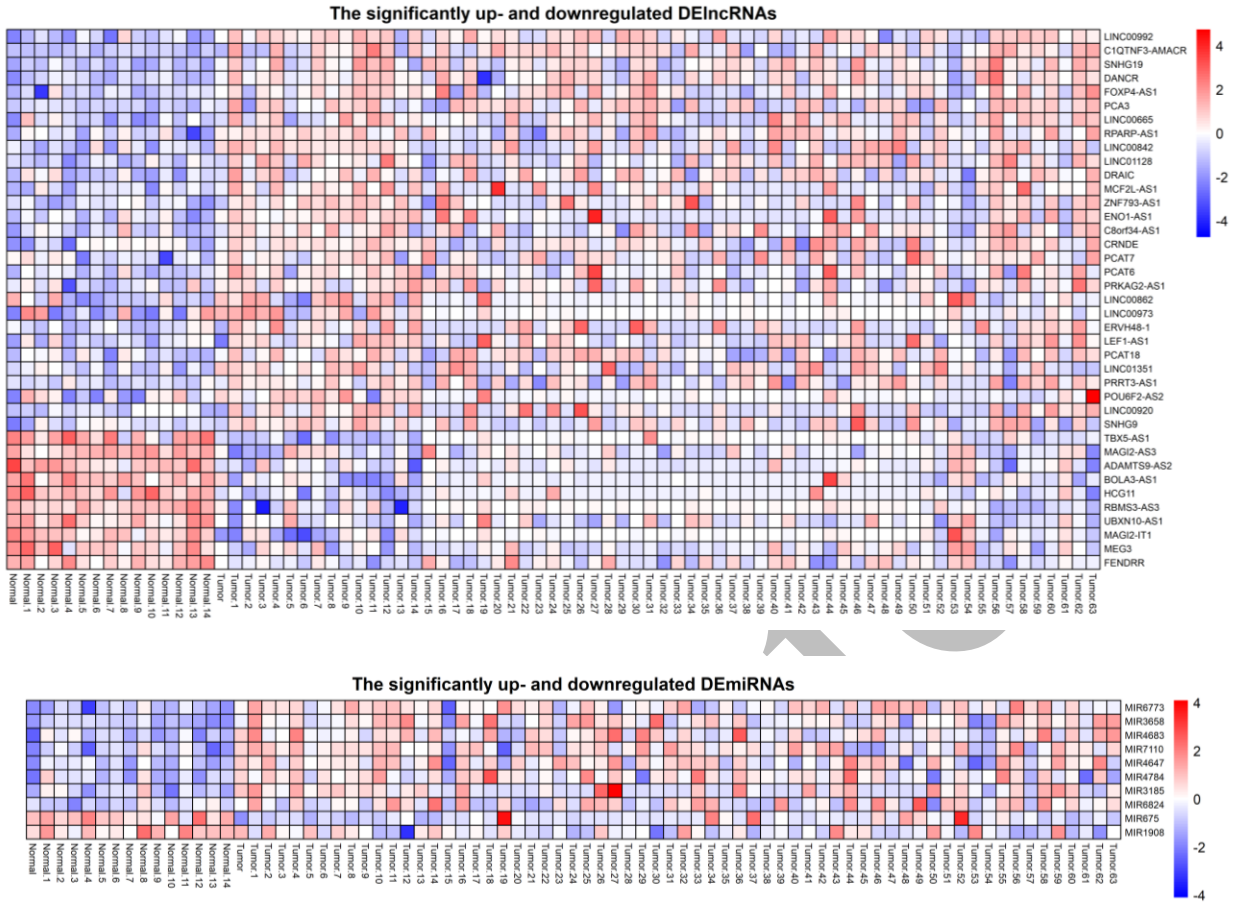


Figure 4. The volcano plot of differentially expressed genes (DEGs); horizontal axis, $\log_2$(FC); vertical axis, -$\log_{10}$(adjusted P value).


DAPL1
KRT14
KRT17
SLC14A1
KRT23
ASPA
DPT
CXCL13
NELL2
BMP5
LGR6
CFD
NDNF
PAGE4
FOXF2
PTGS1
KCNJ3
SMTNL2
WIF1
SMR3B
4
2
0
-2
-4
Normal
Tumor


DAPL1
KRT14
KRT17
SLC14A1
KRT23
ASPA
DPT
CXCL13
NELL2
BMP5
LGR6
CFD
NDNF
PAGE4
FOXF2
PTGS1
KCNJ3
SMTNL2
WIF1
SMR3B
4
2
0
-2
-4
Normal
Tumor

(A)



(B)

Figure 5. A. The two-way clustering of DEmRNAs between prostate tumor samples and normal samples; horizontal axis, the samples; vertical axis, DEmRNAs. B. Two heatmaps depicting expression of DElncRNAs and DEmiRNAs.

GO Enrichment Analysis of DEGs

The substantially DEGs were enriched in 497 GO terms. Clusterprofiler package was used for analysis. For performing this analysis, the all genes listed in the database in clusterProfiler package have been used as background. In GO functional enrichment analysis, 497 GO entries fulfil Adjusted P value<0.05, the majority of which are biological processes, followed by cellular component and molecular function. The first 30 entries are collagen-containing extracellular

matrix (CC), extracellular matrix structural constituent (MF), extracellular matrix organization (BP), extracellular structure organization (BP), cell-cell junction (CC), basement membrane (CC), cell junction assembly (BP), cell-substrate adhesion (BP), collagen trimer (CC), sarcolemma (CC), muscle contraction (BP), muscle system process (BP), urogenital system development (BP), morphogenesis of a branching structure (BP), I band (CC), endoplasmic reticulum lumen (CC), extracellular matrix structural constituent conferring tensile strength (MF), contractile fiber (CC), regulation of cell-substrate adhesion (BP), Z disc (CC), mesenchyme development (BP), mesenchymal cell differentiation (BP), respiratory tube development (BP), myofibril (CC), membrane raft (CC), membrane microdomain (CC), gland morphogenesis (BP), renal system development (BP), glycosaminoglycan binding (MF) and sarcomere (CC). Figure 6 shows the barplots of top 10 enriched functions.

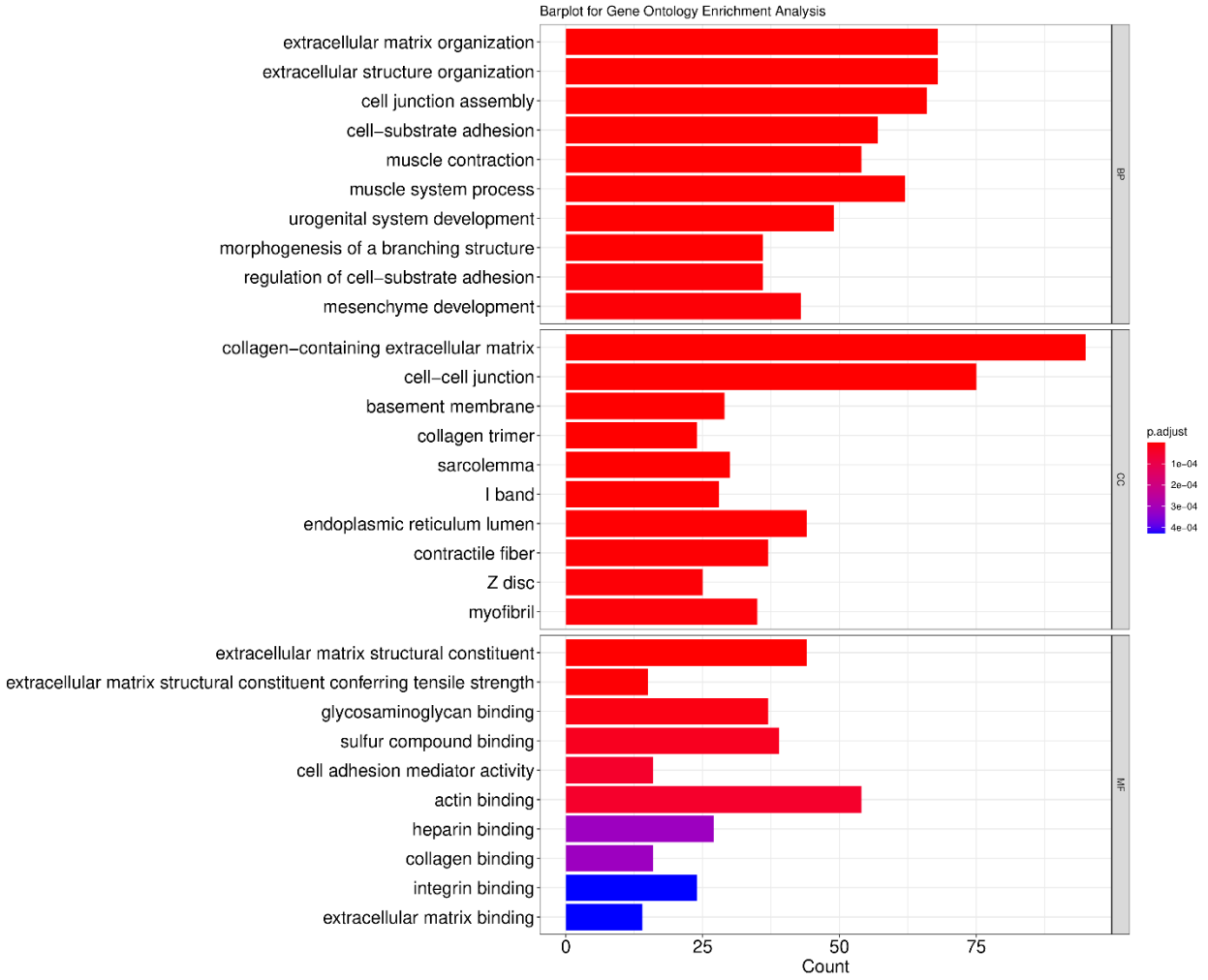


Figure 6. The barplots of top 10 enriched functions. BP (biological process), CC (cellular component) and MF (molecular function). X axis displays the count of geneset; Y axis displays the geneset function; Bar color signifies the adjusted P.value, ranging from red (most significant) to blue (least significant).

Figures 7 and 8 (supplementary file) show the dotplots of top 10 enriched functions and enriched GO induced graph, respectively.

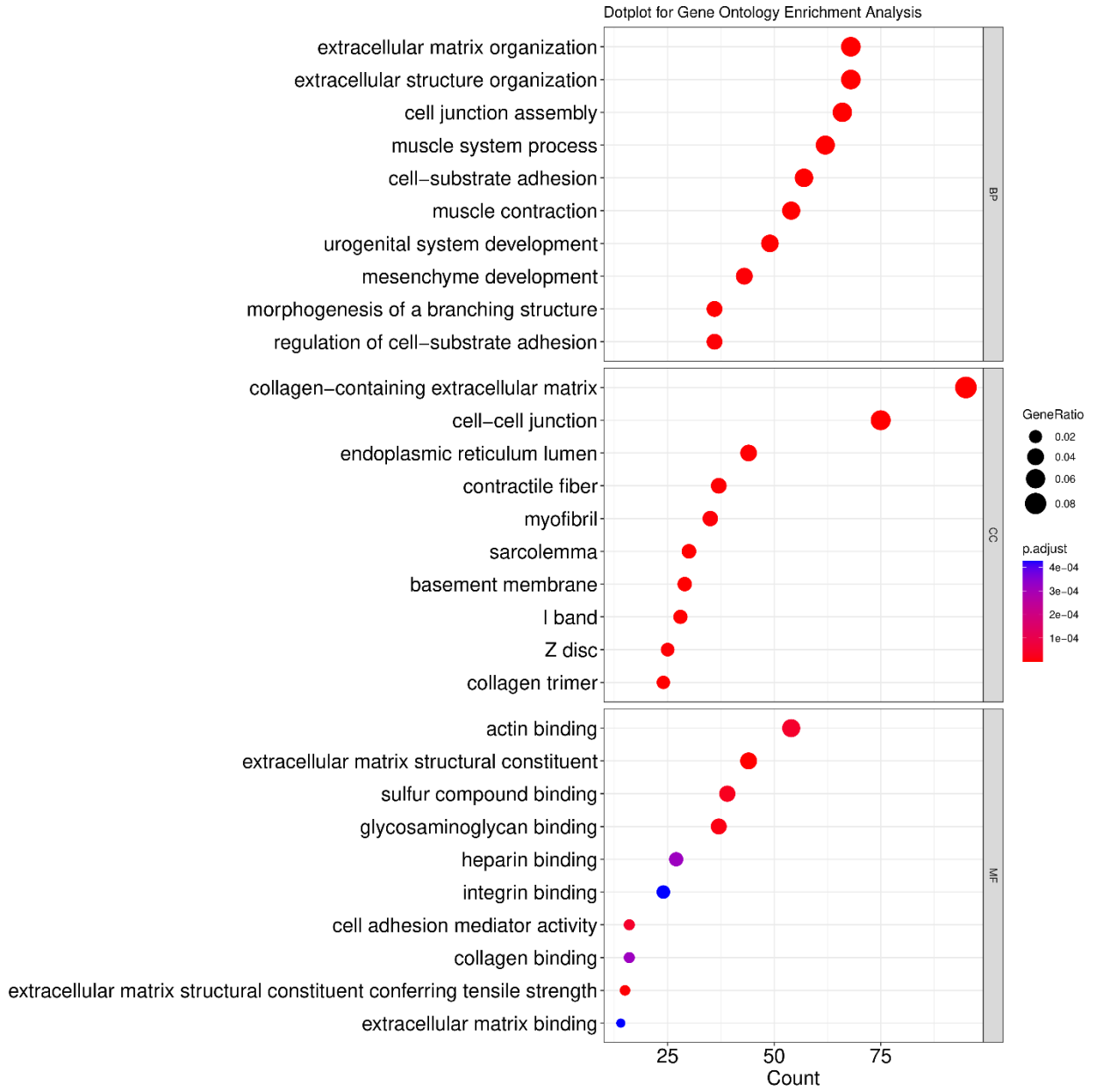


Figure 7. The dotplots of top 10 enriched functions. X axis displays the count of geneset; Y axis displays the geneset function; Dot color represents the adjusted P.value, ranging from dark blue (most significant) to red (least significant). Dot size represents the GeneRatio and The larger the size of the dot, the higher the value of the gene ratio.

Figure 9 indicates the gene-concept network of top 5 GO terms (Cell-substrate adhesion, cell junction assembly, extracellular matrix organization, extracellular structure organization and muscle contraction).

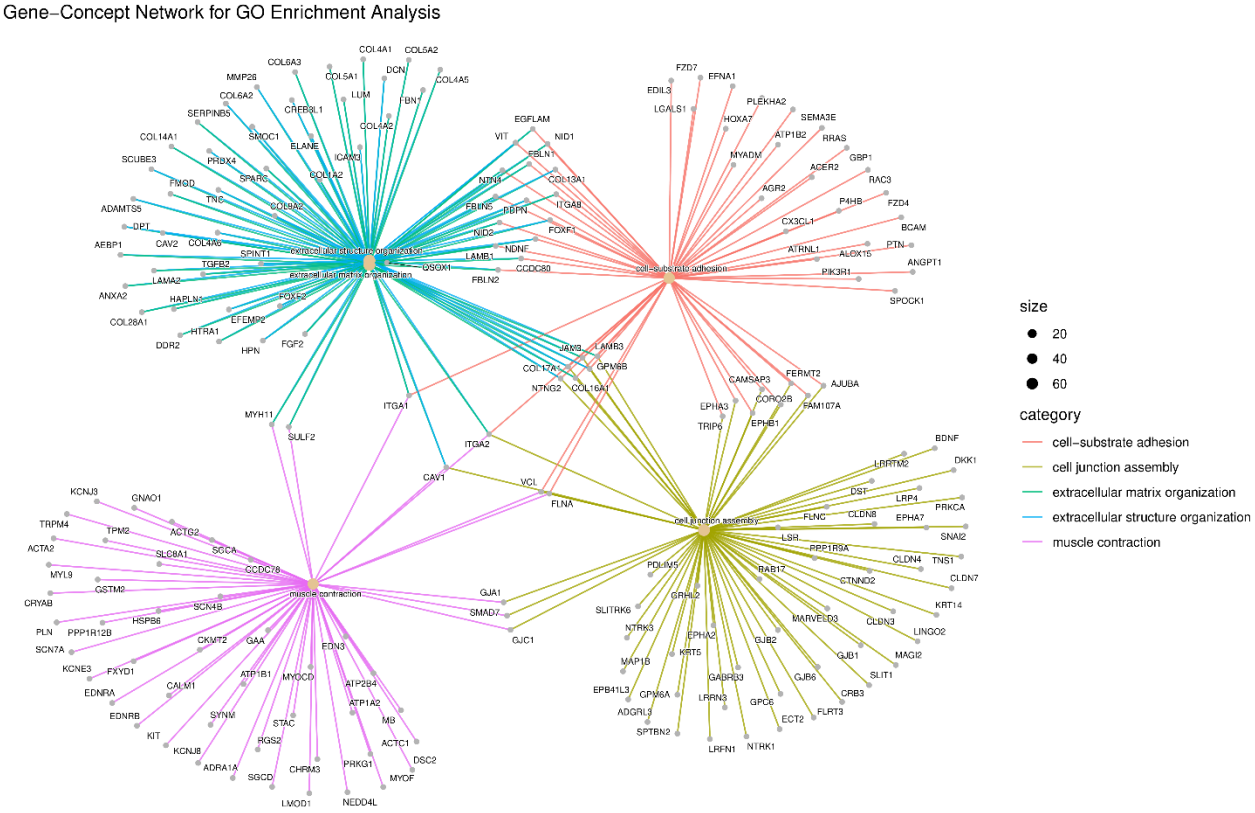


Figure 9. Network plot of top 5 GO terms. GO terms are linked with genes. There are more genes for a specific GO term if the dot relating to it is bigger.

In figure 10, the UpSet plot visualized the intersection between top 10 GO terms. It highlights the gene overlap between several gene sets.

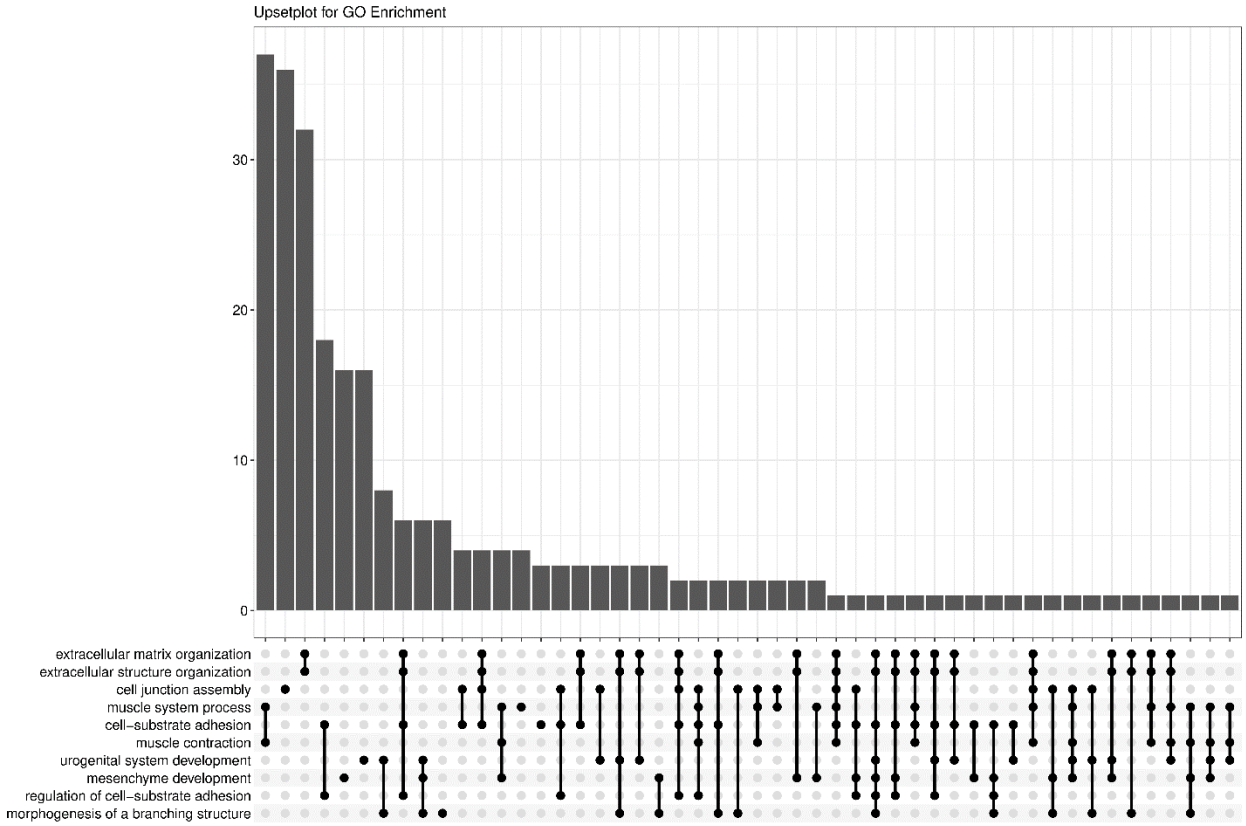


Figure 10. UpsetPlot of 10 GO terms.

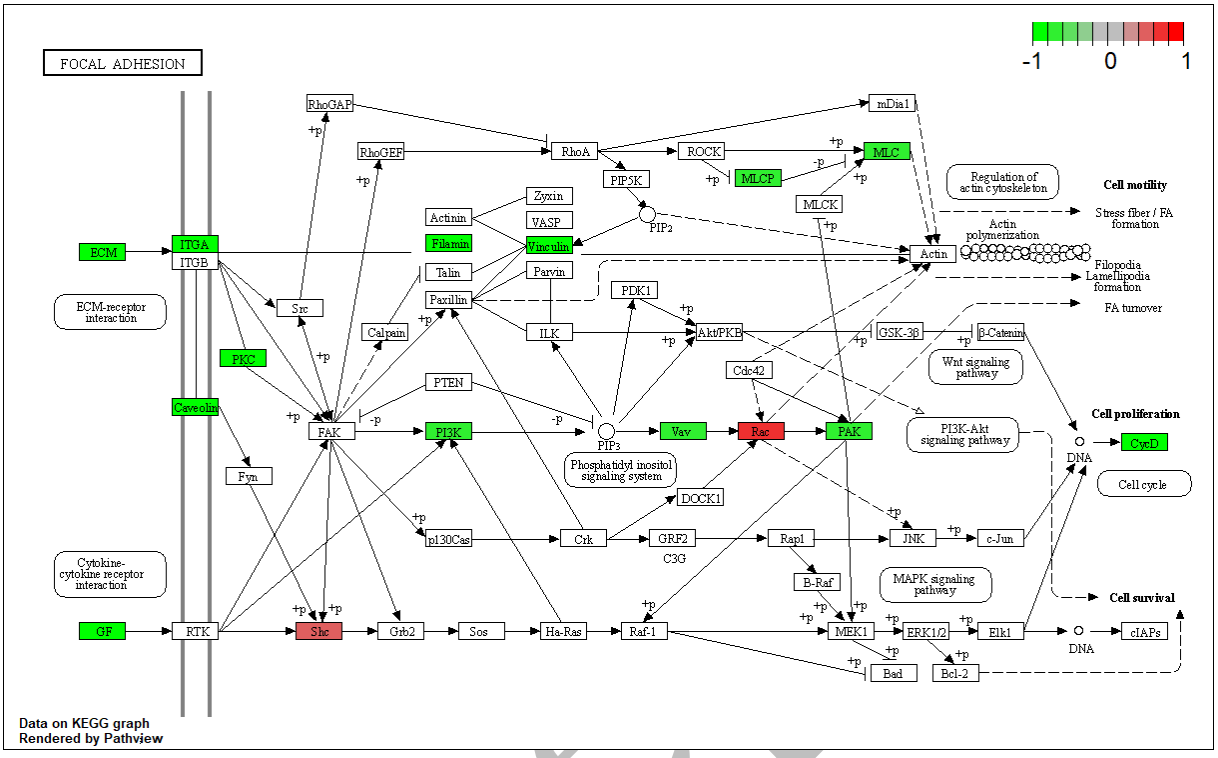
Pathway Analysis

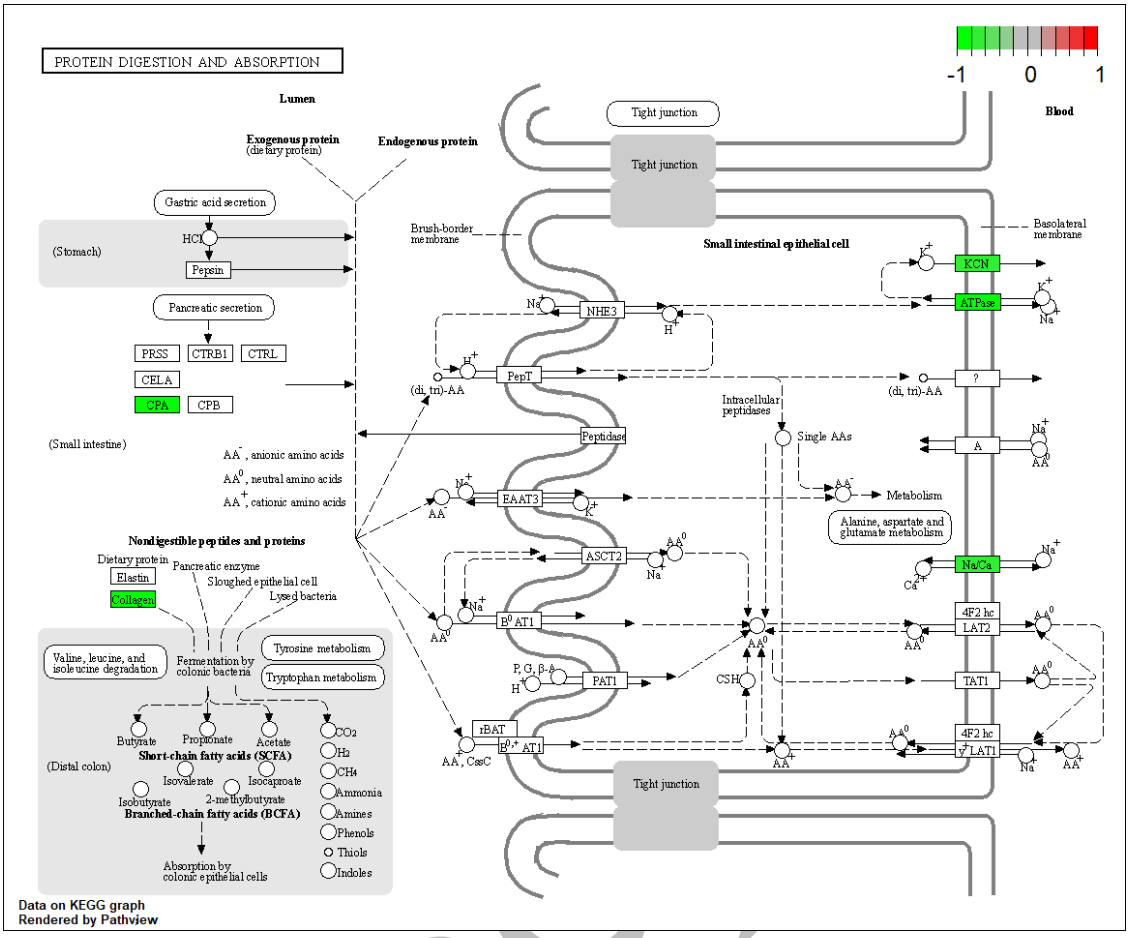
Using Pathview (version 1.36.1) (21) and gage (version 2.46.1) (22) packages in R, KEGG pathway analyses of 177 down-regulated and 177 up-regulated DEGs were performed to identify the potential functional genes (Table 4 and Figure 11). Figure 10 shows the schematic visualization of 3 pathways (1 up-regulated and 2 down-regulated pathways).

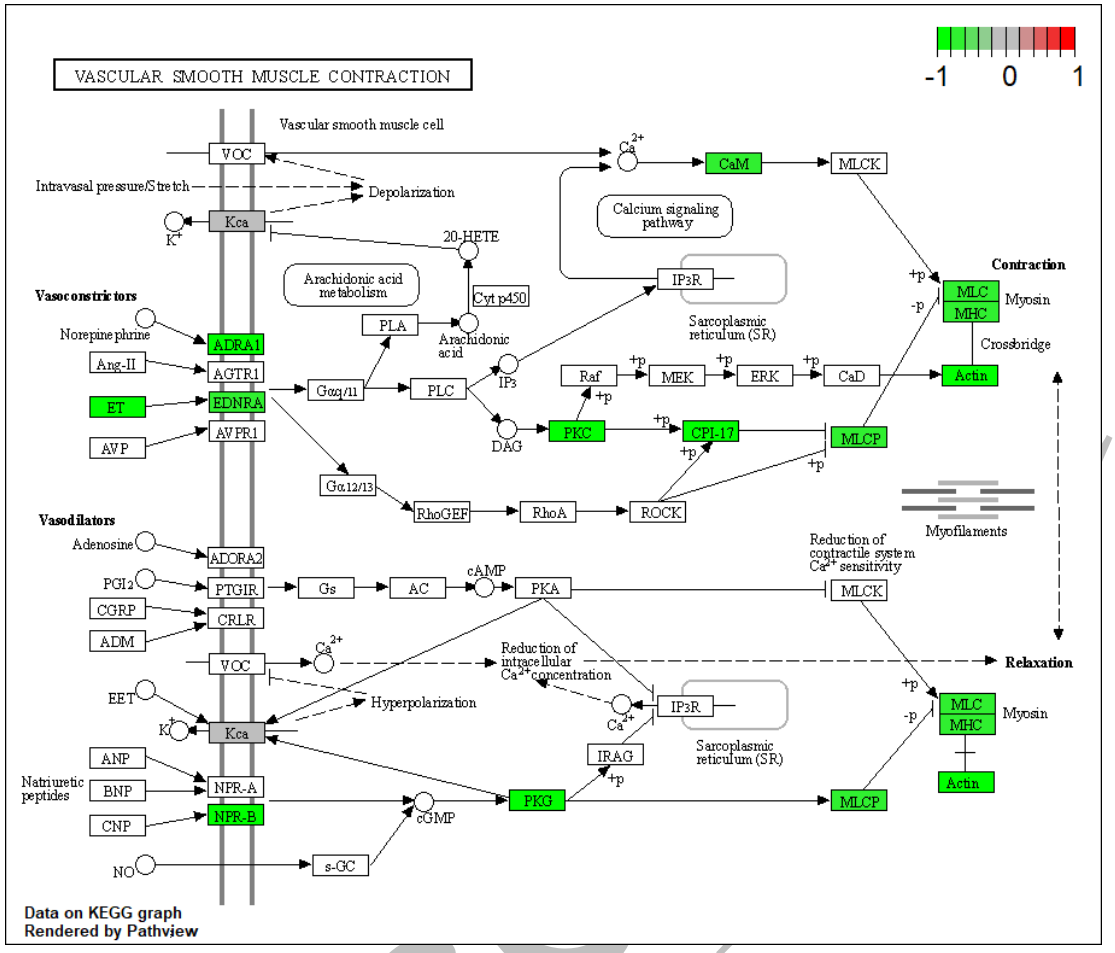
Table 4. Down-regulated and Up-regulated Pathways

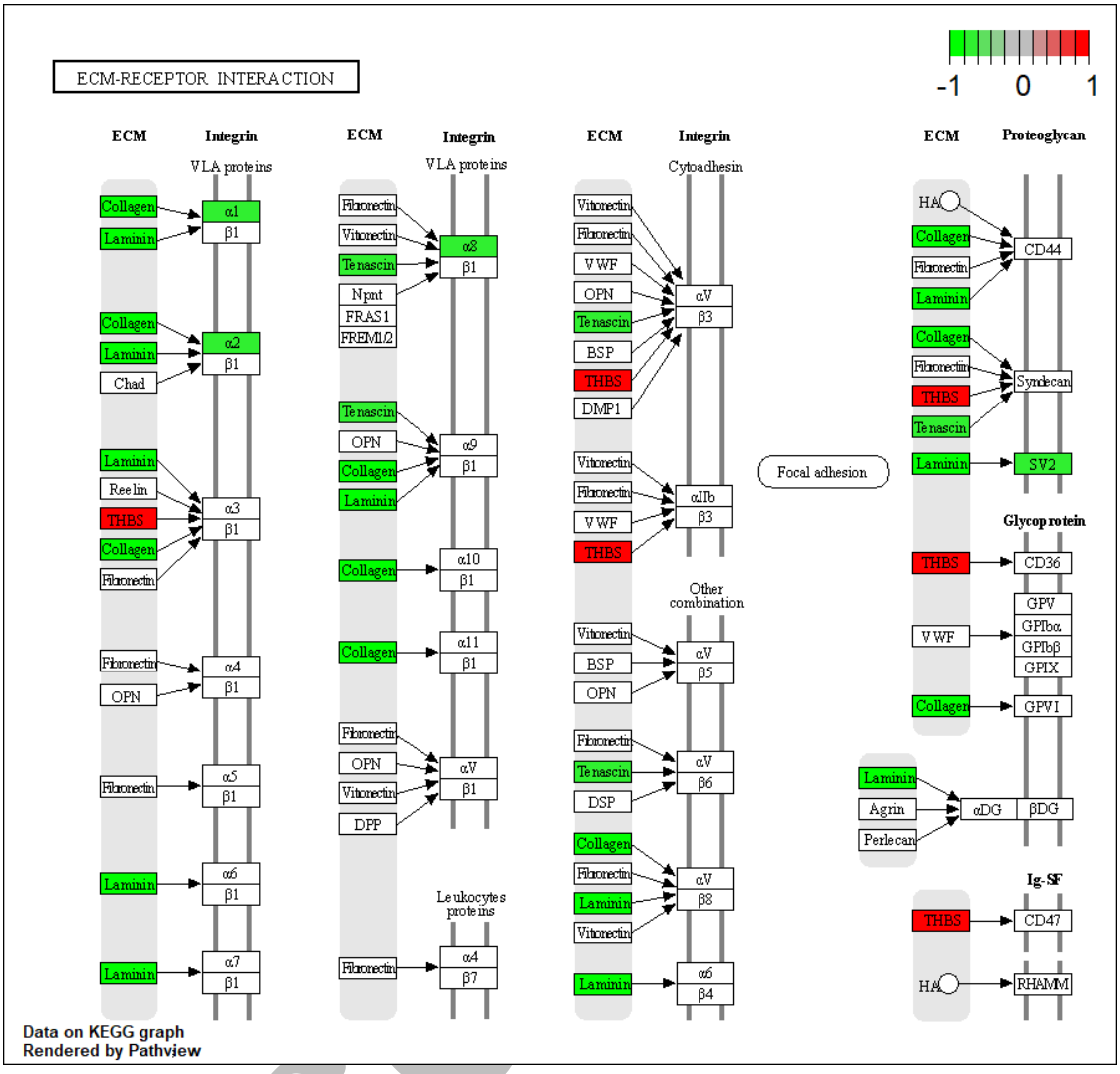
Down-regulated		**Up-regulated**	
Pathway	**P value**	**Pathway**	**P value**
Focal adhesion	0.007574712	**Purine metabolism**	0.04602828
Protein digestion and absorption	0.014154194		
Vascular smooth muscle contraction	0.023808509		
ECM-receptor interaction	0.031880880		

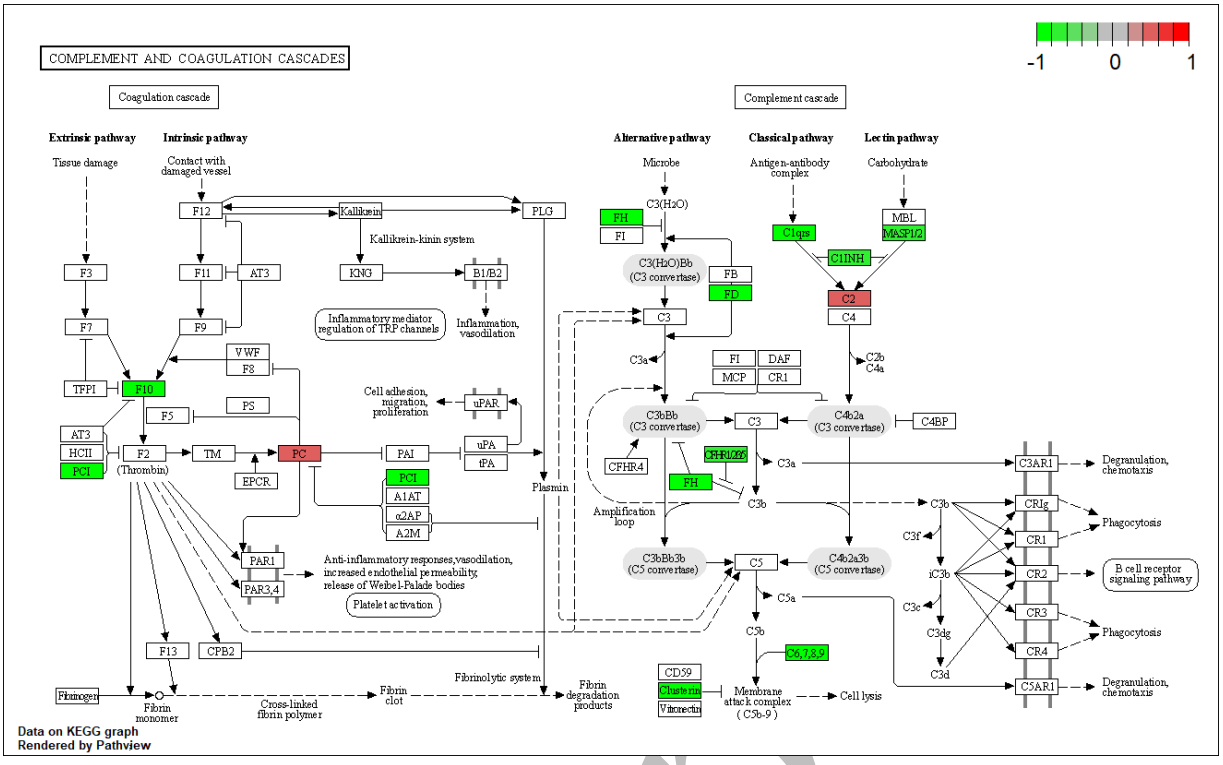
Complement and coagulation cascades	0.049110016		




PROTEIN DIGESTION AND ABSORPTION
-1
0
1
Lumen
Blood
Exogenous protein
(dietary protein)
Endogenous protein
Tight junction
Tight junction
Gastric acid secretion
(Stomach)
HCl
Pepsin
Brush-border membrane
Small intestinal epithelial cell
Basolateral membrane
Pancreatic secretion
PRSS
CTRB1
CTRL
CELA
CPA
CPB
(Small intestine)
NHE3
KCN
ATPase
PepT
(di, tri)-AA
(di, tri)-AA
?
Intracellular peptidases
Peptidase
Single AAs
A
AA-, anionic amino acids
AA0, neutral amino acids
AA+, cationic amino acids
EAAT3
Metabolism
Alanine, aspartate and glutamate metabolism
Nondigestible peptides and proteins
Dietary protein
Elastin
Collagen
Pancreatic enzyme
Sloughed epithelial cell
Lysed bacteria
ASCT2
Na/Ca
B0AT1
4F2 hc
LAT2
Valine, leucine, and isoleucine degradation
Fermentation by colonic bacteria
Tyrosine metabolism
Tryptophan metabolism
PAT1
TAT1
rBAT
B0,+AT1
4F2 hc
y+LAT1
Butyrate
Propionate
Acetate
Short-chain fatty acids (SCFA)
(Distal colon)
Isovalerate
Isocaproate
Isobutyrate
2-methylbutyrate
Branched-chain fatty acids (BCFA)
Absorption by colonic epithelial cells
CO2
H2
CH4
Ammonia
Amines
Phenols
Thiols
Indoles
Tight junction
Data on KEGG graph
Rendered by Pathview


VASCULAR SMOOTH MUSCLE CONTRACTION
-1
0
1
Vascular smooth muscle cell
VOC
Intravasal pressure/Stretch
Depolarization
Kca
K+
20-HETE
Arachidonic acid metabolism
Cyt p450
PLA
Arachidonic acid
Vasoconstrictors
Norepinephrine
ADRA1
Ang-II
AGTR1
ET
EDNRA
AVP
AVPR1
Gαq/11
PLC
IP3
DAG
Gα12/13
RhoGEF
RhoA
ROCK
Ca2+
CaM
MLCK
Calcium signaling pathway
IP3R
Sarcoplasmic reticulum (SR)
Raf
MEK
ERK
CaD
PKC
CPI-17
MLCP
Contraction
MLC
MHC
Myosin
Crossbridge
Actin
Myofilaments
Reduction of contractile system Ca2+ sensitivity
Vasodilators
Adenosine
ADORA2
PGI2
PTGIR
CGRP
CRLR
ADM
Gs
AC
cAMP
PKA
Reduction of intracellular Ca2+ concentration
Relaxation
Hyperpolarization
EET
ANP
Natriuretic peptides
BNP
CNP
NPR-A
NPR-B
cGMP
PKG
IRAG
NO
s-GC
Data on KEGG graph
Rendered by Pathview


ECM-RECEPTOR INTERACTION
-1
0
1
ECM
Integrin
VLA proteins
Collagen
Laminin
α1
β1
Collagen
Laminin
Chad
α2
β1
Laminin
Reelin
THBS
Collagen
Fibronectin
α3
β1
Fibronectin
OPN
α4
β1
Fibronectin
α5
β1
Laminin
α6
β1
Laminin
α7
β1
ECM
Integrin
VLA proteins
Fibronectin
Vitronectin
Tenascin
Npnt
FRAS1
FREM1/2
α8
β1
Tenascin
OPN
Collagen
Laminin
α9
β1
Collagen
α10
β1
Collagen
α11
β1
Fibronectin
OPN
Vitronectin
DPP
αV
β1
Leukocytes proteins
Fibronectin
α4
β7
ECM
Integrin
Cytoadhesin
Vitronectin
Fibronectin
VWF
OPN
Tenascin
BSP
THBS
DMP1
αV
β3
Vitronectin
Fibronectin
VWF
THBS
αIIb
β3
Other combination
Vitronectin
BSP
OPN
αV
β5
Fibronectin
Tenascin
DSP
αV
β6
Collagen
Fibronectin
Laminin
Vitronectin
αV
β8
Laminin
α6
β4
Focal adhesion
ECM
Proteoglycan
HA
Collagen
Fibronectin
Laminin
CD44
Collagen
Fibronectin
THBS
Tenascin
Syndecan
Laminin
SV2
Glycoprotein
THBS
CD36
VWF
GPV
GPIbα
GPIbβ
GPIX
Collagen
GPVI
Laminin
Agrin
Perlecan
αDG
βDG
Ig-SF
THBS
CD47
HA
RHAMM
Data on KEGG graph
Rendered by Pathview


COMPLEMENT AND COAGULATION CASCADES
-1
0
1
Coagulation cascade
Complement cascade
Extrinsic pathway
Intrinsic pathway
Alternative pathway
Classical pathway
Lectin pathway
Tissue damage
Contact with damaged vessel
Microbe
Antigen-antibody complex
Carbohydrate
Kallikrein-kinin system
Inflammatory mediator regulation of TRP channels
Inflammation, vasodilation
Cell adhesion, migration, proliferation
Amplification loop
Plasmin
Anti-inflammatory responses,vasodilation, increased endothelial permeability, release of Weibel-Palade bodies
Platelet activation
Fibrinolytic system
Fibrin degradation products
Fibrinogen
Fibrin monomer
Cross-linked fibrin polymer
Fibrin clot
Membrane attack complex (C5b-9)
Cell lysis
Degranulation, chemotaxis
Phagocytosis
B cell receptor signaling pathway
Data on KEGG graph
Rendered by Pathview

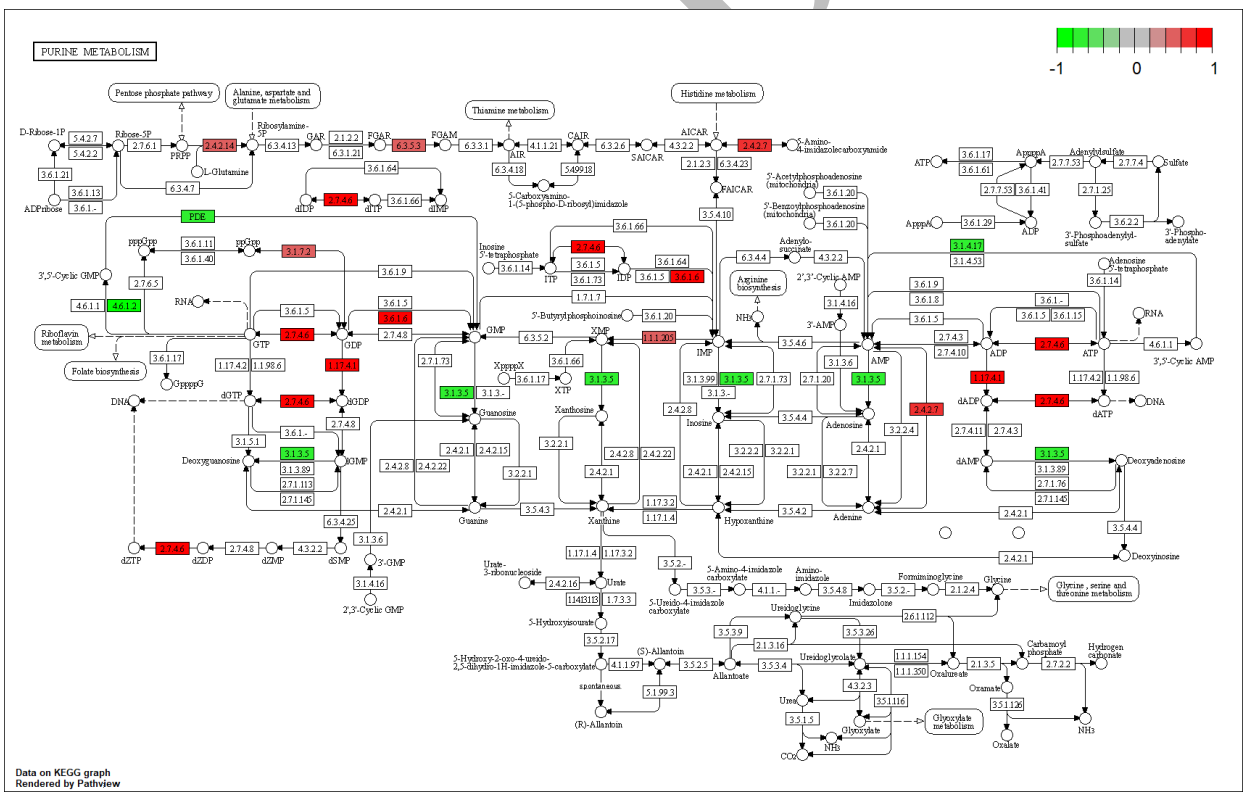

PURINE METABOLISM
-1
0
1
Pentose phosphate pathway
Alanine, aspartate and glutamate metabolism
Thiamine metabolism
Histidine metabolism
Riboflavin metabolism
Folate biosynthesis
Arginine biosynthesis
Glycine, serine and threonine metabolism
Glyoxylate metabolism
Data on KEGG graph
Rendered by Pathview

Figure 11. Visualization of pathways. Green boxes are downregulated genes and red boxes are upregulated genes.

PPI network construction and selection of hub genes

In order to find the hub genes, a PPI network of DEGs (supplementary file) with 411 nodes and 555 edges that was acquired from STRING was loaded into the Cytohubba plugin of Cytoscape 3.9. The 20 hub genes with the highest betweenness of connectivity were EGF, PRKCA, FLNA, CAV1, RGS9, RGS2, CD3EAP, RRM2, ITGA1, PPP1R12B, SDC2, MLC1, PRKG1, BIRC5, P4HB, FGFR2, POLR2H, VCL, PIK3R1 and RGS17.

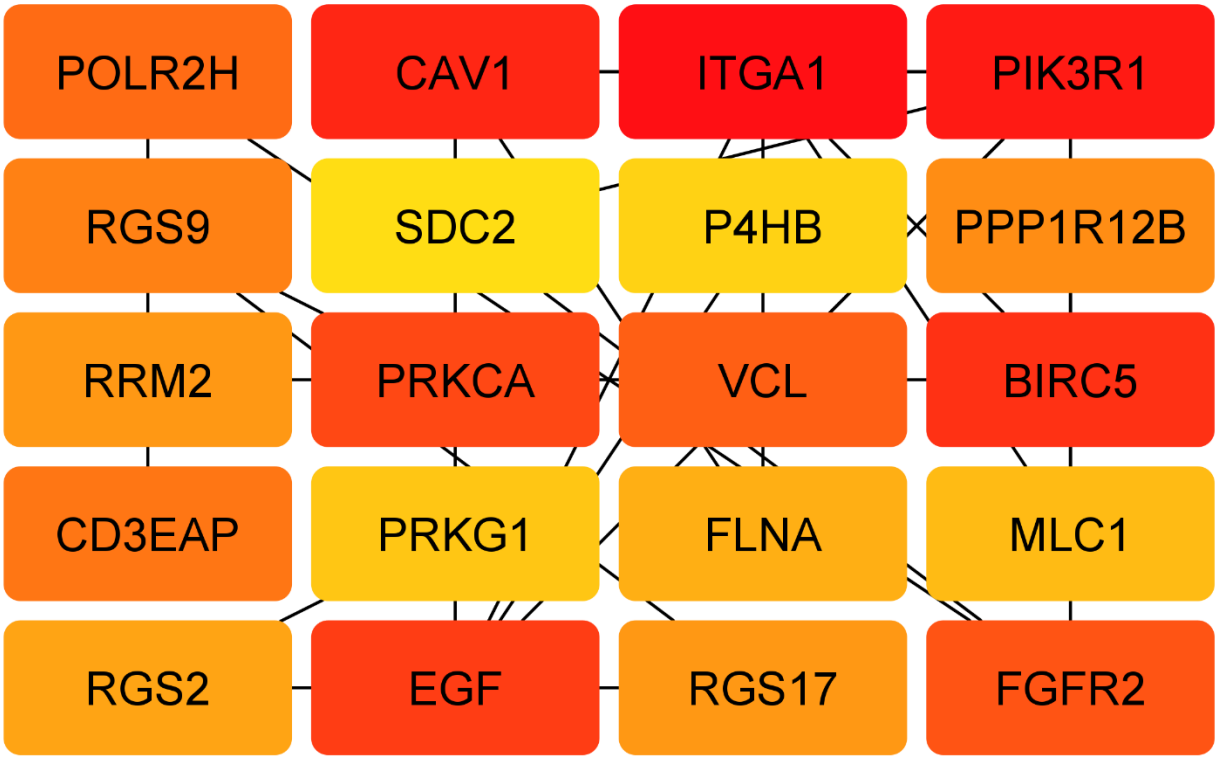


Figure 12. 20 hub genes with the highest betweenness of connectivity.

ceRNA network construction in prostate cancer

LncRNAs can act as an endogenous "sponge" to regulate the expression of mRNA by adsorbing miRNA, according to the ceRNA theory (23). The lncRNA-miRNA-mRNA ceRNA network was built using upregulated or down-regulated miRNAs, as well as lncRNAs or mRNAs (24). DElncRNAs and DEmiRNAs networks did not interact in our research. We utilized the miR2Disease database as a result. In miR2disease, we chose miRNAs that changed in prostate tumor tissues compared to adjacent normal tissue samples, both up- and down-regulating. We discovered 14 PC-related miRNAs using the miR2Disease database. The relationship between lncRNAs and miRNAs was then evaluated using miRcode. This step showed that 13 of 14 PC-specific miRNAs may target to the 13 of 39 DElncRNAs (Table 5). Then we used miRDB, miRTarBase, TargetScan and miRWalk to predict targeted mRNAs by these 13 miRNAs to discover the relationship between miRNAs and mRNAs. We found that 10 miRNAs might target 24 out of the 1312 mRNAs (Table 6). If miRNA-targeted mRNAs were not found in DEmRNAs, they were eliminated. Using this information (Table 5 and 6), we used Cytoscape 3.9 to construct the lncRNA-miRNA-mRNA ceRNA network. Once the targeted DEmRNAs and DElncRNAs' expression patterns were reversed, DElncRNAs, targeted DEmRNAs, and the interacting miRNAs were all eliminated from the ceRNA network. A total of 8 lncRNAs, 4 mRNAs, and 3 miRNAs were in-cluded in the ceRNA network (Figure 13). At last, we computed nodes degrees and displayed the top 7 nodes with the highest degree in the network using cytohubba app (Figure 14). We found ENO1-AS1, hsa-miR-182, hsa-miR-125b-5p, hsa-miR-145, MEG3, FOXF2 and MYO6 as 7 hub genes in ceRNA network.

Table 5. The MiRcode database demonstrated interactions between 13 DElncRNAs and 13 DEmiRNAs.

lncRNA	**miRNA**

PCA3, ERVH48-1, ADAMTS9-AS2, RBMS3-AS3, MEG3	miR-96
PCA3, ERVH48-1, ADAMTS9-AS2, BOLA3-AS1, RBMS3-AS3, MEG3	miR-182
PCA3, CRNDE, ADAMTS9-AS2, MEG3	miR-221
PCA3, CRNDE, ADAMTS9-AS2, MEG3	miR-222
MCF2L-AS1, CRNDE, ADAMTS9-AS2, HCG11, MEG3	miR-205
ENO1-AS1, CRNDE, ERVH48-1, MAGI2-AS3, ADAMTS9-AS2, MEG3	miR-145
CRNDE, SNHG9, MAGI2-AS3, ADAMTS9-AS2, HCG11, MEG3	miR-31
CRNDE, ERVH48-1, MAGI2-AS3, ADAMTS9-AS2, HCG11, MEG3	miR-181b
CRNDE, ADAMTS9-AS2	miR-183
ERVH48-1, SNHG9, ADAMTS9-AS2, BOLA3-AS1, MEG3	miR-184
POU6F2-AS2, ADAMTS9-AS2,	miR-375
POU6F2-AS2, MEG3	miR-125b-5p
MAGI2-AS3, MEG3	miR-16

Table 6. miRWalk, miRDB and TargetScan databases revealed interactions between 10 DEmiRNAs and 24 DEmRNAs.

miRNA	**mRNA**
miR-96	TP53INP1, NIPA1
miR-182	FOXF2
miR-221	TRPS1, KIT
miR-222	STOX2, TRPS1
miR-205	LRRK2
miR-145	ADD3, TGFB2, MYO6
miR-31	SPRED1
miR-181b	KLHL15, PLPP3, PLAG1
miR-16	RAB9B, GALNT7, PSAT1, TGFBR3, PDLIM5, SLC9A6
miR-125b-5p	MFSD9, STOX2, HK2

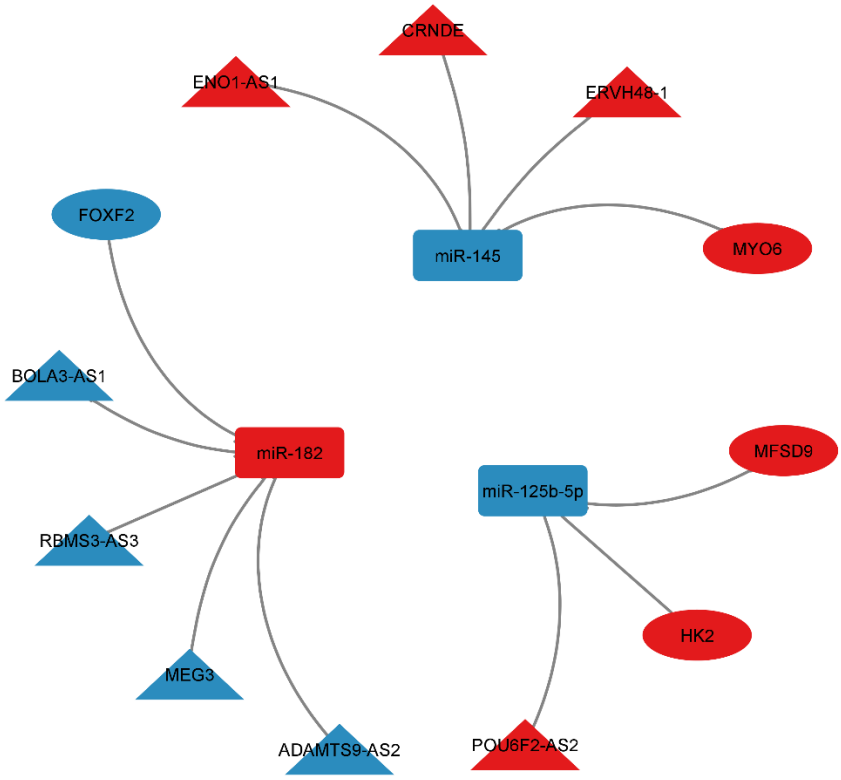


Figure 13. CeRNA network in prostate cancer. Red nodes mean a strong expression level, while blue nodes signify a low level of expression. Ellipses show protein-coding genes; rectangles show miRNAs; Triangles show lncRNAs; gray edges designate lncRNA-miRNA-mRNA interaction.

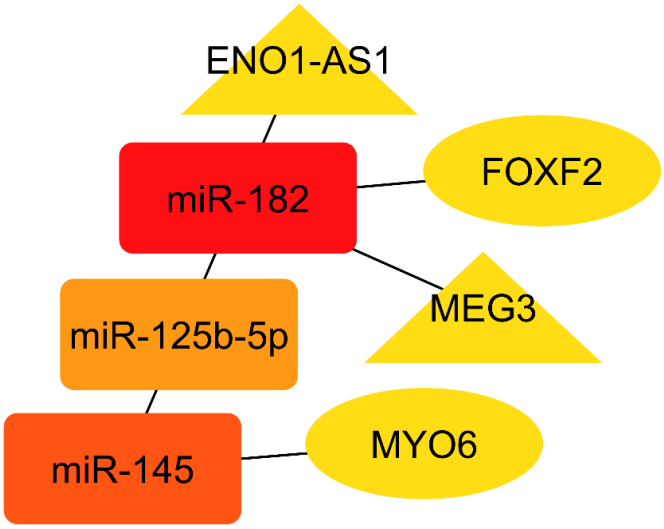


Figure 14. Top 7 genes with highest degree in ceRNA network.

We performed gene ontology enrichment analysis of the target genes in the PPI and ceRNA networks. The final result of this analysis is shown in the form of bar plot and dot plot in Figure 15.

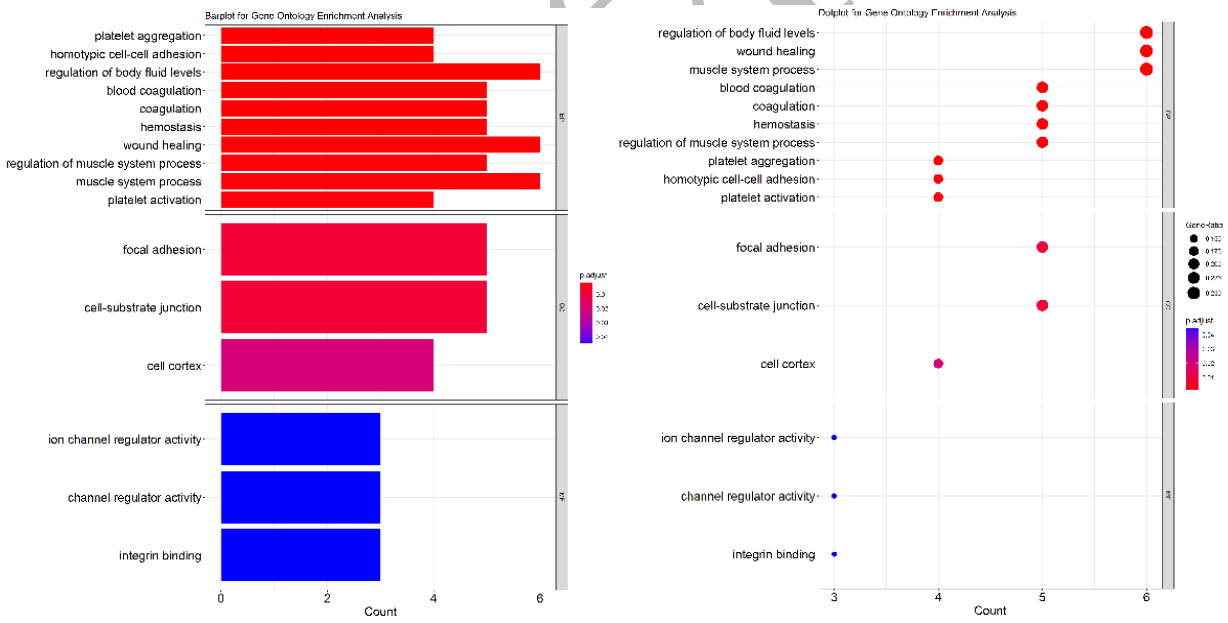


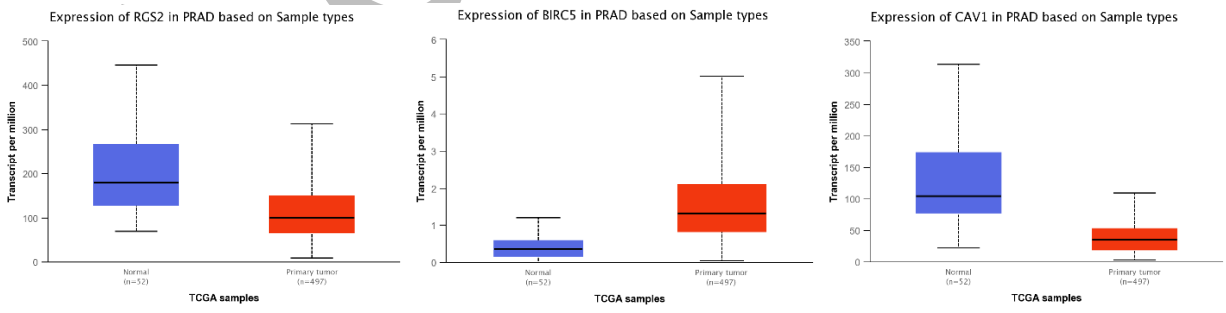
Figure 15. Gene ontology enrichment analysis of the target genes in the networks. Barplot and dotplot indicate top functional terms related to hub genes.

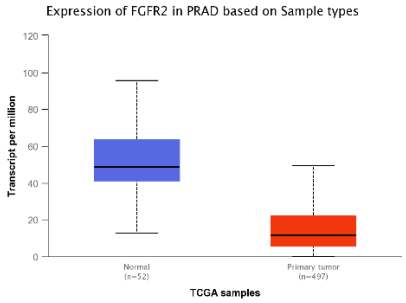
Confirmation of hub genes via expression value

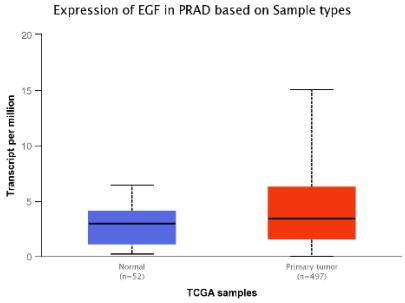
The expression value of hub genes was evaluated using the ualcan database. As a result, all hub genes in PPI network and MIR182 in ceRNA network indicated good statistical significance (Figure 16 and Table 7).

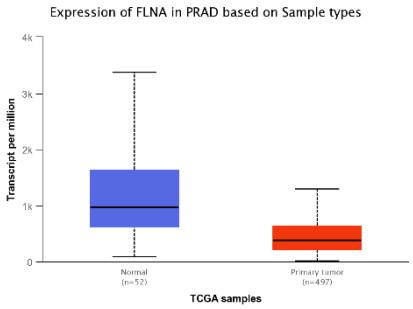
Table 7. Statistical significance of hub genes based on sample types in prostate cancer.

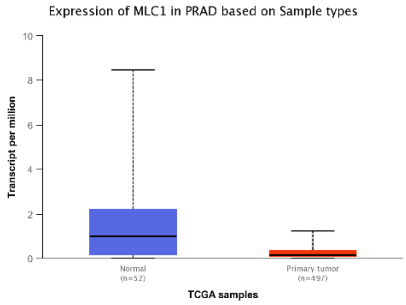
Hub genes	Statistical significance of expression value
POLR2H	<1E-12
CAV1	2.23E-08
ITGA1	4.22E-05
PIK3R1	3.79E-10
RGS9	5.79E-05
SDC2	1.25E-04
P4HB	<1E-12
PPP1R12B	1.13E-06
RRM2	2.32E-07
PRKCA	2.07E-06
VCL	7.55E-07
BIRC5	7.23E-11
CD3EAP	4.68E-08
PRKG1	4.22E-06
FLNA	3.27E-07
MLC1	1.03E-04
RGS2	3.40E-04
EGF	4.02E-13
RGS17	9.88E-05
FGFR2	<1E-12
MIR182	<1E-12

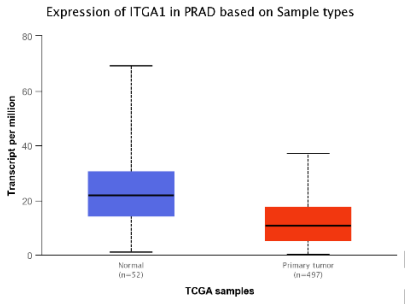


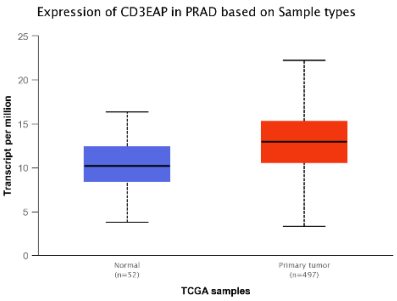

Expression of FGFR2 in PRAD based on Sample types
Transcript per million
120
100
80
60
40
20
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples

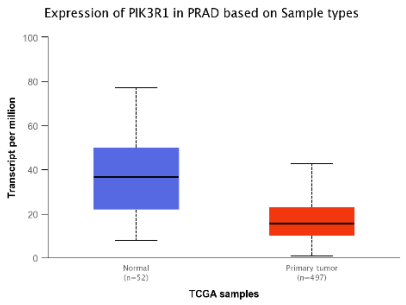

Expression of EGF in PRAD based on Sample types
Transcript per million
20
15
10
5
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples

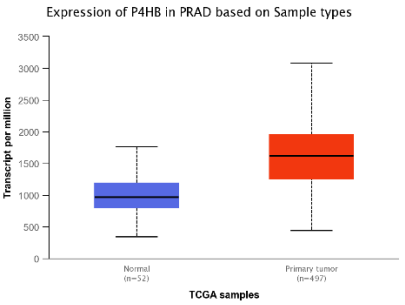

Expression of FLNA in PRAD based on Sample types
Transcript per million
4k
3k
2k
1k
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples

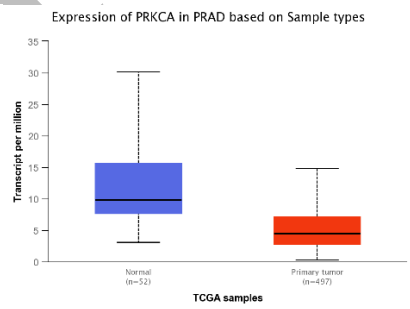

Expression of MLC1 in PRAD based on Sample types
Transcript per million
10
8
6
4
2
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples

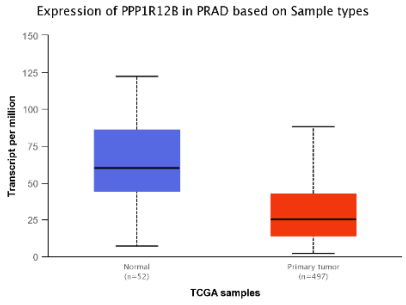

Expression of ITGA1 in PRAD based on Sample types
Transcript per million
80
60
40
20
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples

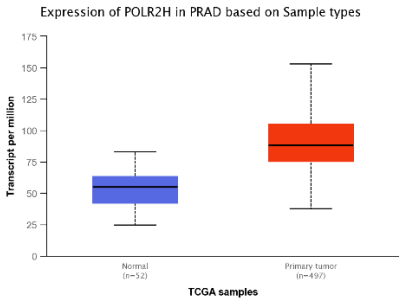

Expression of CD3EAP in PRAD based on Sample types
Transcript per million
25
20
15
10
5
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples

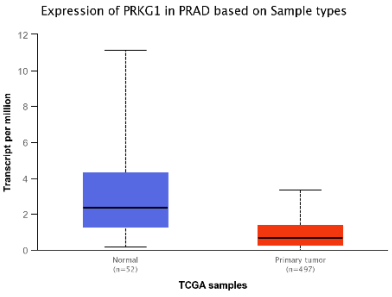

Expression of PIK3R1 in PRAD based on Sample types
Transcript per million
100
80
60
40
20
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples


Expression of P4HB in PRAD based on Sample types
Transcript per million
3500
3000
2500
2000
1500
1000
500
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples


Expression of PRKCA in PRAD based on Sample types
Transcript per million
35
30
25
20
15
10
5
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples


Expression of PPP1R12B in PRAD based on Sample types
Transcript per million
150
125
100
75
50
25
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples


Expression of POLR2H in PRAD based on Sample types
Transcript per million
175
150
125
100
75
50
25
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples


Expression of PRKG1 in PRAD based on Sample types
Transcript per million
12
10
8
6
4
2
0
Normal (n=52)
Primary tumor (n=497)
TCGA samples

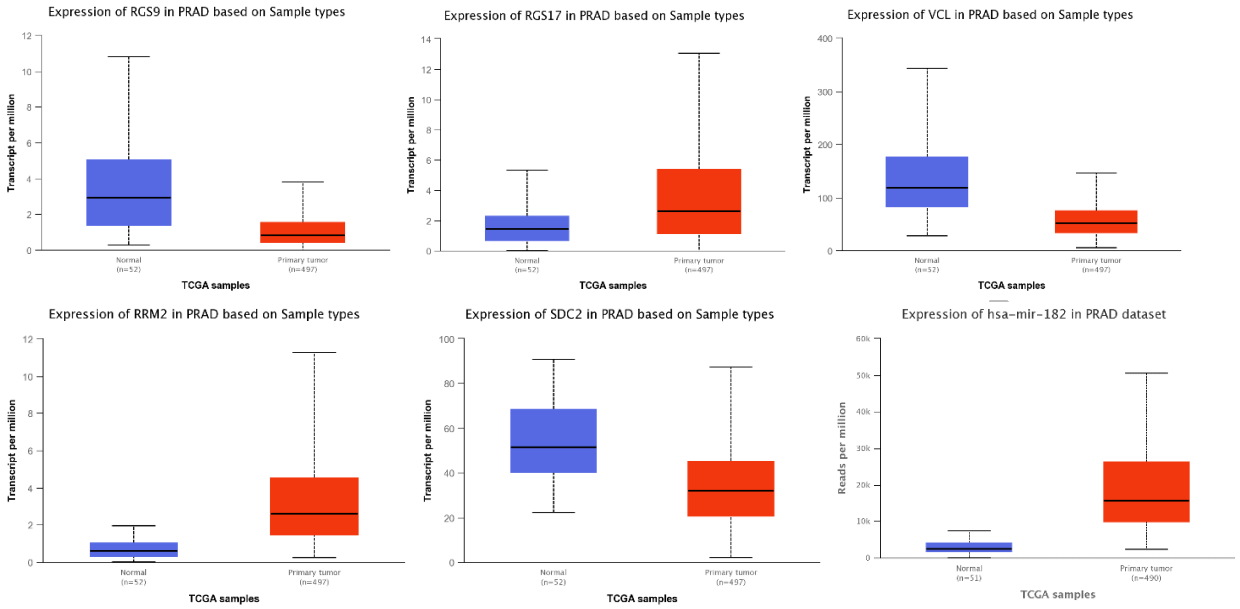
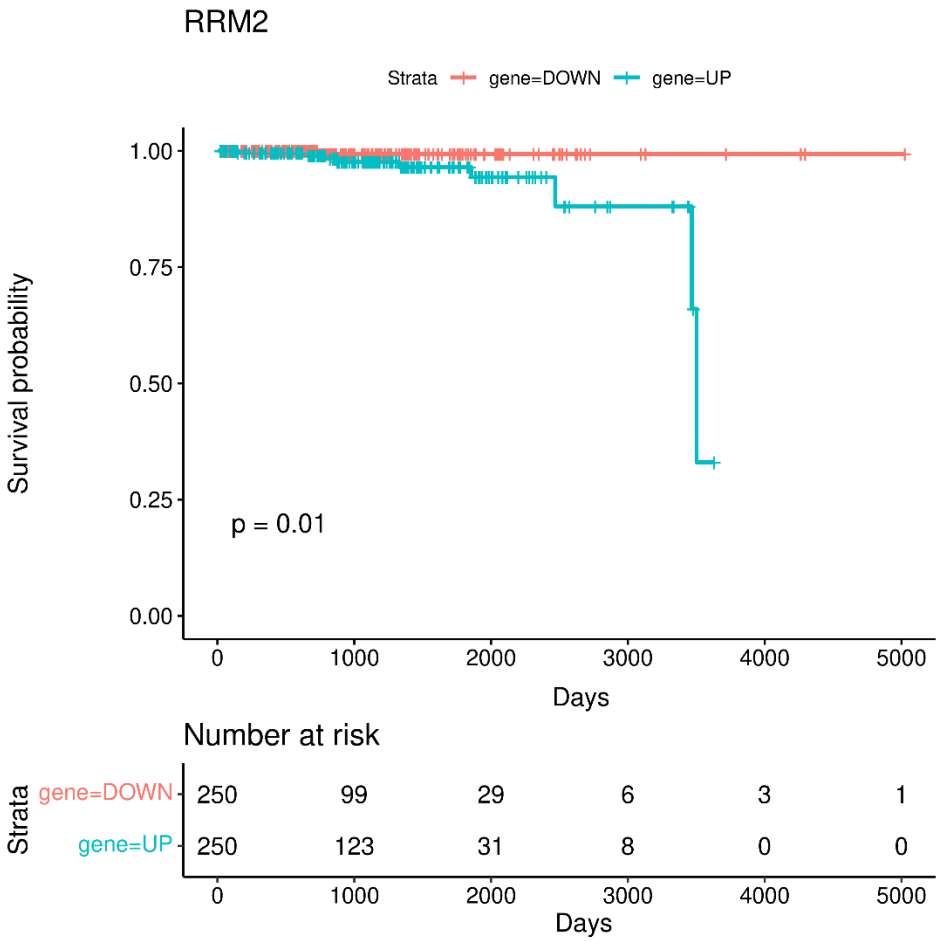


Figure 16. Box plots of gene expression of hub genes in prostate tumor and healthy samples based on TCGA. Red and green boxes show gene expression of hub genes in prostate tumor and healthy samples, respectively.

Survival analysis

For survival analysis, we downloaded and analyzed transcriptome profiling of prostate cancer samples (TCGA-PRAD) using TCGAbiolinks (version 2.24.3) (25) and edgeR (version 3.38.4) (26) packages. Survival was analyzed based on Kaplan-Meier curves using survival package in R. We performed survival analysis based on hub genes in PPI and ceRNA networks. The difference was statistically significant with log-rank $P < 0.05$. RRM2 and MYO6 showed a strong correlation with a reduced overall survival time in individuals with prostate cancer (Figure 17).


RRM2
Strata
gene=DOWN
gene=UP
1.00
0.75
0.50
0.25
0.00
Survival probability
p = 0.01
0
1000
2000
3000
4000
5000
Days
Number at risk
Strata
gene=DOWN
gene=UP
250
99
29
6
3
1
250
123
31
8
0
0
0
1000
2000
3000
4000
5000
Days

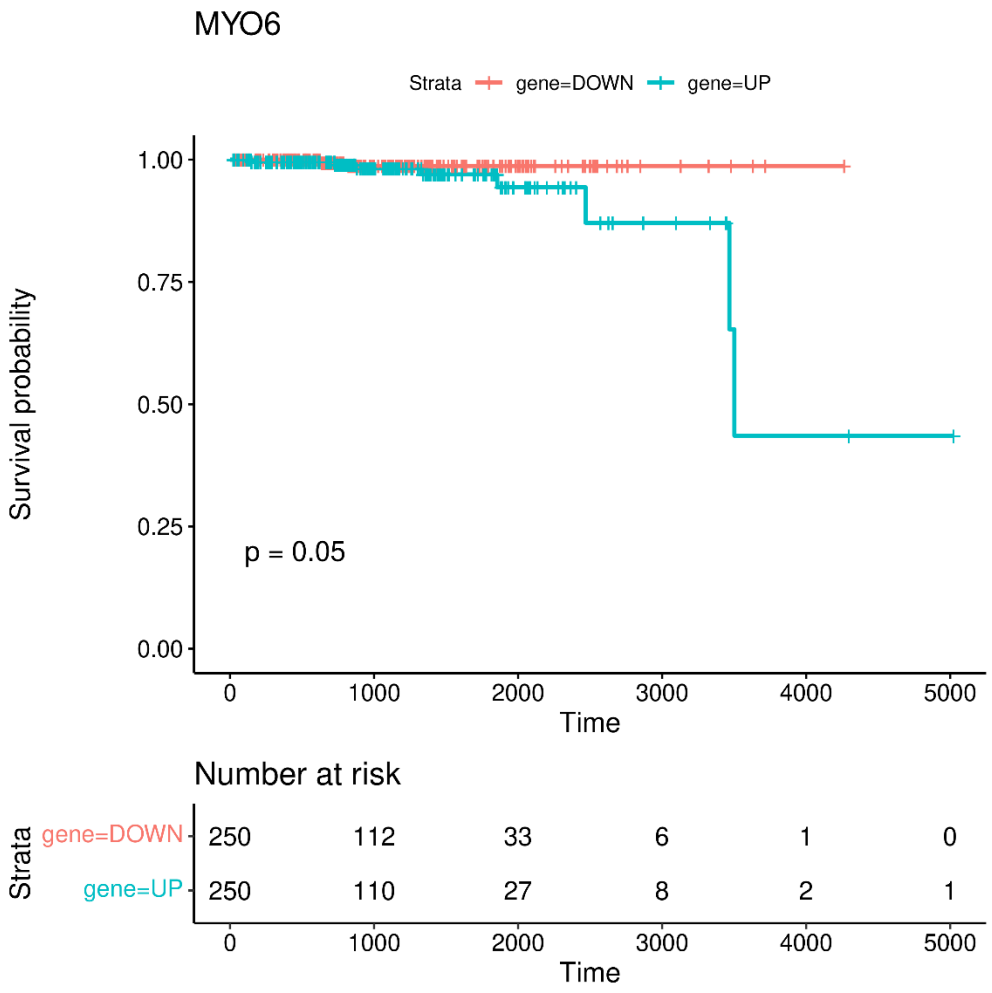


Figure 17. RRM2 and MYO6 Kaplan-Meier survival curves is related to patients with prostate cancer's overall survival.

Discussion

ceRNA networks have been found to participate in the pathoetiology of several cancers, including prostate cancer. Unraveling the interactions between constituents of these networks can facilitate identification of the most relevant cancer-specific pathogenic events. The current study aimed at construction of ceRNA network in prostate cancer.

We obtained 1312 DEmRNAs, including 778 downregulated DEmRNAs (such as CXCL13 and BMP5) and 584 upregulated DEmRNAs (such as OR51E2 and LUZP2). BMP5 has been previously shown to be an important regulator of basal prostate stem cell homeostasis being involved in the initiation of prostate cancer (27). In addition, CXCL13 is an androgen-responsive gene participating in the androgen-regulated migration and invasion of prostate cancer cells (28). On the other hand, OR51E2 has been shown to inhibit proliferation and induce prostate cancer cell death (29). LUZP2 has been previously reported to be over-expressed in hormone-naive prostate cancer but its expression has been decreased in the course of evolution of hormone-naive prostate cancer to castration-resistant ones (30).

Moreover, we found 39 DElncRNAs, including 10 down-regulated DElncRNAs (such as UBXN10-AS1 and FENDRR) and 29 up-regulated DElncRNAs (such as PCA3 and LINC00992) and 10 DEmiRNAs, including 2 down-regulated DEmiRNAs (such as MIR675 and MIR1908) and 8 up-regulated DEmiRNAs (such as MIR6773 and MIR4683). PCA3 has been shown to regulate important pathways and targets and contribute in the development of prostate cancer (31).

We constructed the ceRNA network between these transcripts. A total of 13 lncRNAs, 24 mRNAs, and 13 miRNAs were included in the ceRNA network. Dysregulated pathways included focal adhesion, protein digestion and absorption, vascular smooth muscle contraction, ECM-receptor interaction, complement and coagulation cascades and purine metabolism. Dysregulation of focal adhesion is an important step in tumorigenesis leading to metastasis. In line with this, the smooth muscle contraction with myosin is known to regulate the redistribution of actin-controlled factors during call migration. Also, extracellular matrix (ECM) and ECM-receptor plays an important role in the cell-cell contact. Changes and plasticity of ECM is suggested to control progression and invasion potential of prostate cancer cells (Luthold et al., 2022). Moreover, oncogenic activity of

the complement cascade has been suggested to play a role in facilitating cancer cell proliferation and dysregulation of mitogenic pathways (Rutkowski et al., 2010). The biosynthesis of purines, as a basic component of nucleic acids, is linked to prostate cancer progression by providing the increased need accompanied with increased growth rate and proliferation of cancer cells (Khalafalla et al., 2022, BBA review). The purine metabolism by purinosome is a multi-enzyme complex located around mitochondria and microtubules. Purinosome has been emphasized for its therapeutic potential in cancers (Yin et al., 2018; Frontiers Immunology). These analyses point towards a novel identified ceRNA network of metastatic potential in prostate cancer.

Thus, several important cancer-related pathways linked to each other are modulated by the identified ceRNA networks in the current study.

We also evaluated the significance of these RNAs in the determination of survival of patients with prostate cancer. Among the dysregulated genes, RRM2 showed a strong correlation with a reduced overall survival time in individuals with prostate cancer. RRM2 codes one of two subunits of ribonucleotide reductase. This enzyme facilitates conversion of ribonucleotides to deoxyribonucleotides. Expression of the encoded protein by this gene is controlled during the progression of cell-cycle. This protein is up-regulated in several cancers and is involved in the gemcitabine metabolism. Thus, RRM2 has been suggested as a marker for chemotherapy response and prognosis. It's up-regulation can facilitate DNA damage repair and affect activity of signaling cascades(32). Future studies are needed to find the underlying mechanism of participation of this gene in the course of prostate cancer.

This study provides novel candidates for design of specific treatment modalities for prostate cancer and broadens the current insight about the role of non-coding RNAs in the pathogenesis of prostate cancer.

Declarations

Ethics approval and consent to Participant

Not applicable.

Consent of publication

Not applicable

Availability of Data and Materials

The analyzed data sets generated during the study are available from the corresponding author on reasonable request.

Competing Interest

The authors declare they have no conflict of interest

Funding

Not applicable.

Authors' contributions

MT and SGF wrote the draft and revised it. MT and AB designed and supervised the study. AS and MT collected the data and designed the figures and tables. MT and AS performed the bioinformatic analysis. All the authors read the submitted version and approved it.

Acknowledgement

Not applicable.

References

1. Zhang X, Wang W, Zhu W, Dong J, Cheng Y, Yin Z, et al. Mechanisms and Functions of Long Non-Coding RNAs at Multiple Regulatory Levels. International journal of molecular sciences. 2019;20(22).

2. Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. Nature Reviews Genetics. 2016;17(5):272-83.

3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2021;71(3):209-49.

4. Li F, Li H, Hou Y. Identification and analysis of survival-associated ceRNA triplets in prostate adenocarcinoma. Oncology letters. 2019;18(4):4040-7.

5. Guo Z, Han L, Fu Y, Wu Z, Ma Y, Li Y, et al. Systematic Evaluation of the Diagnostic and Prognostic Significance of Competitive Endogenous RNA Networks in Prostate Cancer. Frontiers in genetics. 2020;11:785.

6. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882-3.

7. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

8. Guo K, Jin F. NFAT5 promotes proliferation and migration of lung adenocarcinoma cells in part through regulating AQP5 expression. Biochemical and Biophysical Research Communications. 2015;465(3):644-9.

9. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2(3):100141.

10. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.

11. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447-52.

12. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.

13. Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Systems Biology. 2014;8(4):S11.

14. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. 2009;37(Database issue):D98-104.

15. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res. 2020;48(D1):D127-d31.

16. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. Nucleic Acids Res. 2020;48(D1):D148-d54.

17. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015;4.

18. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. PLoS One. 2018;13(10):e0206239.

19. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. Neoplasia. 2017;19(8):649-58.

20. Wu J, Liu T, Rios Z, Mei Q, Lin X, Cao S. Heat shock proteins and cancer. Trends in pharmacological sciences. 2017;38(3):226-56.

21. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics. 2013;29(14):1830-1.

22. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics. 2009;10(1):161.

23. Guo LL, Song CH, Wang P, Dai LP, Zhang JY, Wang KJ. Competing endogenous RNA networks and gastric cancer. World J Gastroenterol. 2015;21(41):11680-7.

24. Xu Y, Chen J, Yang Z, Xu L. Identification of RNA Expression Profiles in Thyroid Cancer to Construct a Competing Endogenous RNA (ceRNA) Network of mRNAs, Long Noncoding RNAs (lncRNAs), and microRNAs (miRNAs). Med Sci Monit. 2019;25:1140-54.

25. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. 2016;44(8):e71.

26. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.

27. Tremblay M, Viala S, Shafer ME, Graham-Paquin AL, Liu C, Bouchard M. Regulation of stem/progenitor cell maintenance by BMP5 in prostate homeostasis and cancer initiation. Elife. 2020;9.

28. Fan L, Zhu Q, Liu L, Zhu C, Huang H, Lu S, et al. CXCL13 is androgen-responsive and involved in androgen induced prostate cancer cell migration and invasion. Oncotarget. 2017;8(32):53244-61.

29. Pronin A, Slepak V. Ectopically expressed olfactory receptors OR51E1 and OR51E2 suppress proliferation and promote cell death in a prostate cancer cell line. The Journal of biological chemistry. 2021;296:100475.

30. Zhao J, Zhao Y, Wang L, Zhang J, Karnes RJ, Kohli M, et al. Alterations of androgen receptor-regulated enhancer RNAs (eRNAs) contribute to enzalutamide resistance in castration-resistant prostate cancer. Oncotarget. 2016;7(25):38551.

31. Lemos AE, Ferreira LB, Batoreu NM, de Freitas PP, Bonamino MH, Gimba ER. PCA3 long noncoding RNA modulates the expression of key cancer-related genes in LNCaP prostate cancer cells. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2016;37(8):11339-48.

32. Zhan Y, Jiang L, Jin X, Ying S, Wu Z, Wang L, et al. Inhibiting RRM2 to enhance the anticancer activity of chemotherapy. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 2021;133:110996.