

IDENTIFICATION OF DRUG TARGETS TOWARDS PROSTATE CANCER

by

Elif Esvap

B.S., Chemical Engineering, Boğaziçi University, 2016

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering
Boğaziçi University

2018

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude for my thesis supervisor Prof. Kutlu Ülgen and co-supervisor Assoc. Prof. Elif Özkırımlı Ölmez, who guided and motivated me all the time. I was very fortunate to work with them.

I would like to thank the members of my thesis committee, Prof. Berna Sarıyar Akbulut, Prof. Türkan Haliloğlu and Assoc. Prof. Burak Alakent, for their recommendations.

I am very lucky to have Dilara Göreke by my side from the first day of the freshman year to these last days. Aslı Şahin and Feride Eren were also essentials of my Boğaziçi years, and I would like to thank them for the fun times. I am also grateful to rest of the Qümene team for the adventures we had.

I would like to thank Selin Baç Bilgi and Coşkun Bilgi for opening their home to our starving group and making us feel always welcome. I am also grateful to Müge Kasım and Furkan Öztürk for not giving me a chance to binge-eat fries uncontrollably on Wednesday nights, as they finished each batch as soon as I turned my back.

I would like to express my special thanks to Begüm Yağcı as she was there at every step of this graduate journey. I possibly could have not handled these last couple of months without her. I also would like to express my appreciation to Merve Yüce, Özlem Özbek and the rest of the KB407 and KB440 members for the motivation and encouragement they gave. Although we did not share the same lab, I would like to thank Aysun Urhan, Feyza Öner, Aysegül Karakuş and İlke Kaykanat for their support.

Finally, I would like to thank to my family for their continuous support in any possible way. They made me who I am today. I love them all.

ABSTRACT

IDENTIFICATION OF DRUG TARGETS TOWARDS PROSTATE CANCER

Prostate cancer (Pca) is one of the leading cause of cancer deaths and second common cancer type in men. Identifying novel diagnostics and drug targets specific for Pca is critical since early detection and treatment are associated with high survival rates. As intracellular processes are interconnected, a systematic point of view is essential instead of analysis of individual components. Biological networks such as protein-protein interaction (PPI) networks have been proven effective representations of biological data. In this study, two types of networks were constructed and analyzed: Protein-protein interaction network (PPIN) was constructed by a selecting core set of Pca proteins and combining experimental interactions obtained from 4 different databases by core set proteins' mutual Gene Ontology (GO) terms. Co-Expression (Co-Exp) network was obtained by Weighted Correlation Network Analysis (WGCNA) using transcriptomics data from The Cancer Genome Atlas (TCGA). Both networks were divided into modules by Markov Clustering (MCL) and Topological Overlap Matrix (TOM), respectively. Gene Set Enrichment Analysis (GSEA) demonstrated that prostate tumors cover the eight hallmarks of cancer. Disease Ontology (DO) enrichments and disease identification of genes detected by the DIAMOnD algorithm revealed the link between Pca and other cancer types such as sarcoma, liver and bone cancer. Integration of topological features and differential expression of genes led to the identification of 272 potential drug targets. Further elimination of these targets was done by screening the Connectivity Map (C-Map) and 84 potential targets were identified, which are proteins taking part in hallmarks of cancer processes. 55 of these promising spots without any known drugs need further computational and experimental examination to be considered as new putative targets.

ÖZET

PROSTAT KANSERİNE YÖNELİK İLAÇ HEDEFLERİNİN TANIMLANMASI

Prostat kanseri (Pca), erkeklerde en yaygın ikinci kanser olmasının yanı sıra, kanser ölümlerinin en önde gelen sebeplerinden biridir. Erken teşhis ve tedavi yüksek sağkalım oranları ile ilişkili olduğundan, Pca'ya özgü yeni tanı ve ilaç hedeflerinin belirlenmesi önemlidir. Hücre içi süreçler birbirine bağlı olduğundan, bireysel bileşenlerin analizi yerine sistematik bir bakış açısı gereklidir. Protein-protein etkileşimi (PPI) ağları gibi biyolojik ağların biyolojik verilerin etkili temsilleri olduğu kanıtlanmıştır. Bu çalışmada, iki tip ağ oluşturulmuş ve analiz edilmiştir: Protein-protein etkileşim ağı (PPIN), 4 ayrı veri tabanından elde edilen deneysel etkileşimlerin seçilmiş çekirdek set Pca proteinlerinin ortak Gen Ontolojisi (GO) terimleriyle kesiştirilmesi ile elde edilmiştir. Eş Ekspresyon (Co-Exp) ağı, Kanser Genom Atlası'ndan (TCGA) transkriptomik veriler kullanılarak Ağırlıklı Korelasyon Ağı Analizi (WGCNA) ile oluşturulmuştur. Her iki ağ, sırasıyla Markov Kümeleme (MCL) ve Topolojik Örtüşme Matrisi (TOM) ile modüllere ayrılmıştır. Gen Set Zenginleştirme Analizi (GSEA), prostat tümörlerinin kanserin sekiz ayırt edici özelliğini kapsadığını göstermiştir. Hastalık Ontolojisi (DO) zenginleştirilmesi ve DIAMOnD algoritması tarafından tespit edilen genlerle ilişkili hastalıkların tanımlanması, Pca ile sarkom, karaciğer ve kemik kanseri gibi diğer kanser tipleri arasındaki bağlantıyı ortaya çıkarmıştır. Topolojik özelliklerin genlerin diferansiyel ekspresyonu ile birleşimi 272 potansiyel ilaç hedefinin belirlenmesine yol açmıştır. Bu hedeflerin daha az sayıya indirilebilmesi, Bağlantı Haritasının (C-Map) taranması ile gerçekleştirilmiştir ve kanser süreçlerinin ayırt edici özelliklerine sahip olan 84 potansiyel hedef belirlenmiştir. Bilinen herhangi bir ilaca sahip olmayan bu potansiyel hedeflerin 55'inin yeni varsayımsal hedef olarak kabul edilmesi için daha fazla hesaplamalı ve deneysel incelenmesi gerekmektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
ÖZET	v
LIST OF FIGURES	ix
LIST OF TABLES	xiii
LIST OF SYMBOLS	xvi
LIST OF ACRONYMS/ABBREVIATIONS	xvii
1. INTRODUCTION	1
2. THEORETICAL BACKGROUND	3
2.1. Prostate Cancer	3
2.2. Prostate Therapy	5
2.2.1. Tumor Receptors	5
2.2.1.1. GRPR Receptors	5
2.2.1.2. Steroid Receptors	6
2.2.1.3. Other Receptors	6
2.2.2. Drug target proteins specific for prostate cancer	7
2.3. Biological Networks	9
2.3.1. Metabolic Networks	10
2.3.2. Gene Regulatory Networks (GRNs)	11
2.3.3. Protein-Protein Interaction Networks	13
2.3.4. Genetic Interaction (GI) Networks	15
2.3.5. Signaling networks	16
2.4. Network Medicine	17
2.5. Sphingolipids and Cancer	19
3. METHODS	22
3.1. Databases	22
3.1.1. Protein-Protein Interaction Data	22
3.1.2. Gene Expression Data	23

3.1.3.	Genes/Proteins Data	24
3.2.	Network Construction	24
3.2.1.	PPI Network Construction	24
3.2.1.1.	Selecting Core Proteins	24
3.2.1.2.	Network Construction by GO Annotations	25
3.2.2.	Co-expression Network Construction	25
3.2.3.	Ligand-Centric Network Construction	28
3.3.	Clustering Methods	29
3.4.	Analysis Methods	31
3.4.1.	Topological Analysis	31
3.4.2.	Functional Enrichment Analysis	33
3.4.3.	Crosstalk Analysis	33
3.5.	Drug Target Identification	34
4.	RESULTS	35
4.1.	Construction of PPI Network	35
4.2.	Topological Analysis of Pca PPI Network	41
4.3.	Hub and Bottleneck Proteins	53
4.4.	Module Detection and Analysis	57
4.5.	Construction of Co-Expression Network	57
4.6.	Enrichment Analysis	64
4.6.1.	Gene Ontology Enrichment Analysis	64
4.6.1.1.	Prostate Cancer PPIN Proteins	64
4.6.1.2.	Prostate Cancer PPI Network Modules	66
4.6.1.3.	Prostate Cancer PPI Network Hubs and Bottlenecks	68
4.6.1.4.	Prostate Tumor Co-Expression Modules	70
4.6.2.	Pathway Enrichment Analysis	72
4.6.2.1.	Prostate Cancer PPIN Proteins	72
4.6.2.2.	Prostate Cancer PPI Network Modules	72
4.6.2.3.	Prostate Cancer PPI Network Hubs and Bottlenecks	73
4.6.2.4.	Prostate Tumor Co-Expression Modules	74
4.6.3.	Disease Ontology Enrichment Analysis	74

4.6.3.1.	Prostate Cancer PPIN Proteins	74
4.6.3.2.	Prostate Cancer PPI Network Modules	75
4.6.3.3.	Prostate Cancer Network Hubs and Bottlenecks	76
4.6.3.4.	Prostate Tumor Co-Expression Modules	76
4.6.4.	Gene Set Enrichment Analysis	77
4.6.4.1.	Hallmarks of Cancer	83
4.7.	Prognostic Genes	87
4.8.	DIAMOND Algorithm for Detection of Disease Modules	89
4.8.1.	Identification of Diseases	93
4.9.	Crosstalk Analysis	96
4.10.	Identifying Drug Targets	101
5.	CONCLUSIONS AND RECOMMENDATIONS	106
5.1.	Conclusions	106
5.2.	Recommendations	108
	REFERENCES	110
	APPENDIX A: TOPOLOGY OF CORE PROTEINS	135
	APPENDIX B: TOPOLOGY OF DISEASE PROTEINS	139
	APPENDIX C: PPIN CLUSTERS	142
	APPENDIX D: UNFAVORABLE PROGNOSTIC GENES	145
	APPENDIX E: FAVORABLE PROGNOSTIC GENES	151
	APPENDIX F: OMIM AND GWAS ENTRIES	153
	APPENDIX G: TOP 100 DIAMOND DISEASE GENES (alpha=2)	154
	APPENDIX H: TOP 100 DIAMOND DISEASE GENES (alpha=10)	155
	APPENDIX I: TARGETS FOR DRUG REPOSITIONING	156
	APPENDIX J: POTENTIAL DRUG TARGETS	157

LIST OF FIGURES

Figure 3.1.	Flowchart of Co-Expression Network Construction.	26
Figure 4.1.	Flowchart of PPI Network Construction.. . . .	35
Figure 4.2.	Prostate Cancer pathway (map05215) from KEGG.	36
Figure 4.3.	GO Biological Process Venn Diagram.	39
Figure 4.4.	GO Molecular Function Venn Diagram.	39
Figure 4.5.	GO Cellular Component Venn Diagram.	40
Figure 4.6.	PPI network of prostate cancer.	42
Figure 4.7.	Number of proteins and interactions in various networks [166]. . . .	42
Figure 4.8.	Characteristics of yeast PPI network [168].	43
Figure 4.9.	Characteristics of various networks [155].. . . .	44
Figure 4.10.	Degree vs number of nodes with fitted line.	45
Figure 4.11.	Degree distribution of nodes on logarithmic scale.	46
Figure 4.12.	Degree distributions for various species [170].	46
Figure 4.13.	Degree vs. degree distribution of PPI, core and disease proteins. . .	48
Figure 4.14.	Normalized edge betweenness vs. its distribution of PPI, core and disease proteins.. . . .	49

Figure 4.15. Eigenvector centrality vs. its distribution of PPI, core and disease proteins.	50
Figure 4.16. Average shortest path length vs. its distribution of PPI, core and disease proteins.	51
Figure 4.17. Eccentricity vs. its distribution of PPI, core and disease proteins. . .	52
Figure 4.18. Coreness vs. its distribution of PPI, core and disease proteins. . . .	52
Figure 4.19. Clustering of the Pca PPIN by MCL.	58
Figure 4.20. Clustering of tumor samples.	59
Figure 4.21. Soft threshold vs. model fit.	59
Figure 4.22. Relationship among the modules.	62
Figure 4.23. Visualized network of black module.	63
Figure 4.24. Visualized network of purple module.	63
Figure 4.25. Enriched GO Biological Process terms for PPI network proteins. . .	65
Figure 4.26. Enriched GO Molecular Function terms for PPI network proteins. .	65
Figure 4.27. Enriched GO Cellular Component terms for PPI network proteins. .	66
Figure 4.28. Enriched GO Biological Process terms for PPIN clusters.	67
Figure 4.29. Enriched GO Molecular Function terms for PPIN clusters.	67
Figure 4.30. Enriched GO Cellular Component terms for PPIN clusters.	68

Figure 4.31. Enriched GO Biological Process terms for hubs and bottlenecks. . .	68
Figure 4.32. Enriched GO Molecular Function terms for hubs and bottlenecks. .	69
Figure 4.33. Enriched GO Cellular Component terms for hubs and bottlenecks. .	69
Figure 4.34. Enriched GO Biological Process terms for first 11 Co-Expression modules.	70
Figure 4.35. Enriched GO Molecular Function terms for first 11 Co-Expression modules.	71
Figure 4.36. Enriched GO Cellular Component terms for first 11 Co-Expression modules.	71
Figure 4.37. KEGG pathway enrichment of Prostate Cancer PPIN proteins. . . .	72
Figure 4.38. KEGG pathway enrichment of Prostate Cancer PPIN modules. . . .	73
Figure 4.39. KEGG pathway enrichment of Prostate Cancer PPIN Hubs and Bottlenecks.	73
Figure 4.40. KEGG pathway enrichment of Prostate Tumor for first 11 Co- Expression modules.	74
Figure 4.41. DO Enrichment terms for Prostate Cancer PPIN proteins.	75
Figure 4.42. DO Enrichment terms for Prostate Cancer PPIN modules.	75
Figure 4.43. DO Enrichment terms for Prostate Cancer PPIN hubs bottlenecks.	76
Figure 4.44. DO Enrichment terms for Prostate Tumor Co-Expression modules..	76
Figure 4.45. First 6 Hallmarks of Cancer [194].	83

Figure 4.46. Emerging Hallmarks of Cancer [194].	84
Figure 4.47. Biological Process GO terms of prognostic genes shared with Pca PPIN.	89
Figure 4.48. GO Biological process enrichment for shared nine genes.	90
Figure 4.49. Seed and top 100 DIAMOnD genes for $\alpha=2$	92
Figure 4.50. Seed and top 100 DIAMOnD genes for $\alpha=10$	92
Figure 4.51. DO Enrichment comparison of genes for $\alpha=2$ and $\alpha=10$ identified by DIAMOnD algorithm.	96
Figure 4.52. Ligand-centric inhibition network for Sphingolipids.	98
Figure 4.53. Ligand-centric activation network for Sphingolipids.	98
Figure 4.54. Flowchart of drug target identification algorithm.	102
Figure 4.55. GO Biological Process terms associated with 272 drug targets. . . .	103
Figure 4.56. GO Biological Process terms associated with remaining 84 drug targets.	104

LIST OF TABLES

Table 2.1.	Receptors targeted for theranostic purposes treating prostate cancer and their ligands	8
Table 3.1.	Number of PPI from various databases	22
Table 3.2.	Number of samples and genes in GTEx and TCGA databases	23
Table 3.3.	Activity types and thresholds for inhibition and activation networks	29
Table 4.1.	Prostate Cancer pathway proteins in KEGG	36
Table 4.2.	KLK3 and its protein interaction partners from various databases .	37
Table 4.3.	Proteins having k-fold>2 (absolute).	38
Table 4.4.	Core protein set	40
Table 4.5.	Number of proteins and interactions between them for each repetition of the process	41
Table 4.6.	Topological properties of PPIN of prostate cancer	43
Table 4.7.	Definitions of graph theory concepts	43
Table 4.8.	Average topological properties of network, core set and disease proteins	47
Table 4.9.	Minimum/maximum topological properties of network, core set and disease proteins	47
Table 4.10.	Top 20 Degree and Betweenness Centrality results	53

Table 4.11.	Degree and Betweenness Centrality ranks of top proteins	54
Table 4.12.	Drugs and their indications for hub and bottleneck proteins	55
Table 4.13.	Modules and corresponding number of genes	61
Table 4.14.	Number of edges, interactions and genes in black module	62
Table 4.15.	Number of edges, interactions and genes in purple module	62
Table 4.16.	Description of Hallmark Gene Sets [193]	77
Table 4.17.	Number of genes in hallmark and Pca gene expression set (present study)	79
Table 4.18.	Significance of down- and upregulated gene sets in tumor.	81
Table 4.19.	Common Pca network proteins shared by prostate cancer prognostic genes/proteins	88
Table 4.20.	Nine of the core proteins that are shared with the seed gene list. . .	91
Table 4.21.	DisGeNET results for top 10 disease genes ($\alpha=2$)	93
Table 4.22.	DisGeNET results for top 25 disease genes ($\alpha=2$)	93
Table 4.23.	DisGeNET results for top 100 disease genes ($\alpha=2$)	94
Table 4.24.	DisGeNET results for top 10 disease genes ($\alpha=10$)	94
Table 4.25.	DisGeNET results for top 25 disease genes ($\alpha=10$)	94
Table 4.26.	DisGeNET results for top 100 disease genes ($\alpha=10$)	95
Table 4.27.	Total number of proteins, ligands and interactions for the ligand-centric inhibition network for Sphingolipids	97

Table 4.28.	Total number of proteins, ligands and interactions for the ligand-centric activation network for Sphingolipids	97
Table 4.29.	Mutual proteins of Pca PPIN with inhibition and activation networks 99	
Table A.1.	Topology of Core Proteins.	135
Table B.1.	Topology of Disease Proteins	139
Table C.1.	Number of proteins, interactions and core proteins in clusters. . . .	142
Table D.1.	Unfavorable prognostic genes for prostate cancer [206].	145
Table E.1.	Favorable prognostic genes for prostate cancer [206]	151
Table F.1.	Prostate cancer related genes form OMIM and GWAS	153
Table G.1.	Top 100 DIAMOnD disease genes for the network ($\alpha=2$)	154
Table H.1.	Top 100 DIAMOnD disease genes for the network ($\alpha=10$)	155
Table I.1.	Targets for Drug Repositioning.	156
Table J.1.	Potential drug targets	157

LIST OF SYMBOLS

a_{ij}	Symmetric adjacency matrix between genes i and j
C	Constant for power law
$FPKM_i$	Fragments per Kilobase Million for gene i
p_k	Probability of a node with exactly k links
s_{ij}	Co-expression similarity between genes i and j
TPM_i	Transcripts per Million for gene i
β	Soft threshold
γ	Degree exponent

LIST OF ACRONYMS/ABBREVIATIONS

AR	Androgen receptor
ASAH1	Acid ceramidase
B	Betweenness
BP	Biological Process
BIOGRID	Biological General Respiratory for Interaction Datasets
CC	Cellular Component
C-Map	Connectivity Map
Co-Exp	Co-Expression
DEG	Differentially Expressed Genes
DO	Disease Ontology
EC50	Half maximal effective concentration
FDR	False Discovery Rate
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
GTE _x	Genotype-Tissue Expression
GWAS	Genome Wide Association Studies
HPA	Human Protein Atlas
HPRD	Human Protein Reference Database
IC50	Half maximal inhibitory concentration
KEGG	Kyoto Encyclopedia of Genes and Genomes
KF	K-fold
KLK3	Kallikrein related peptidase 3
max	Maximum
MAPK	Mitogen-activated protein kinase
MCL	Markov Clustering
MeRAV	Metabolic Gene Rapid Visualizer
MF	Molecular Function
min	Minimum

MSigDB	Molecular Signature Database
NB	Normalized Betweenness
OMIM	Online Mendelian Inheritance in Man
Pca	Prostate cancer
PPI	Protein-protein interaction
PPIN	Protein-protein interaction network
PSA	Prostate specific antigen
SPHK1	Sphingosine kinase 1
SPHK2	Sphingosine kinase 2
TGCA	The Cancer Genome Atlas
TOM	Topological Overlap Matrix
WGCNA	Weighted Gene Co-Expression Network Analysis

1. INTRODUCTION

Prostate cancer (Pca) is one of the most common cancer types among men with high survival rates when it is treated in early stages. It is the most common cancer type excluding non-melanoma skin cancer and the third-leading cause of cancer deaths in men in the USA [1]. Only in 2018, 30,000 patients are predicted to die because of prostate cancer, which constitute 9% of all estimated deaths. Around 22,000 prostate cancer incidents were reported in Turkey between the years 2009 and 2013 becoming the second most common carcinoma in men. Although it is one of the significant cause of deaths, the scientific research in Turkey on prostate cancer is limited.

Since the biological systems like prostate cancer are complex with hundreds of cell types and regulatory and effector molecules, networks are valuable to investigate [2]. Networks are used to represent complex real-world systems, and network analysis is applied to many systems such as airport connectivity networks in transportation, electrical power grid networks in engineering and acquaintance networks in sociology [3]. Network analysis became a growing research area in computational and systems biology following the development of high throughput data collection techniques. Since the biological processes involve thousands of reactions and interactions, it is necessary to study the living cells as interconnected systems rather than individually by constructing networks [4]. These networks are at great value the understanding the biological systems as they provide a framework for modelling, visualizing and generating hypotheses about cellular events [3].

The aim of this work is to construct a PPIN of Pca along with a Co-Exp network with an ultimate goal of identifying drug targets. For the analysis of the constructed PPIN, topological and functional properties will be estimated. Then, highly connected proteins in these networks will be clustered using MCL algorithm. KEGG Pathway, Gene and Disease Ontology enrichments are performed in order to identify functional

modules. Further goal of this work is to identify proteins that participate in crosstalk between sphingolipid metabolism and Pca.

The following work consists of four main sections. In Theoretical Background, prostate cancer and its therapy options are explained, as well as biological networks and network medicine. In Methods section, the databases used in this study are briefly explained. The construction methods for three different types of networks and the analysis methods are detailly elaborated. The Results section includes the constructed networks and their enrichment analyses. In addition, topological analysis of the Pca PPIN is conducted and compared with the literature values, later to be used in drug target identification section. The last section is the conclusions and recommendations, where the end remarks are given.

2. THEORETICAL BACKGROUND

2.1. Prostate Cancer

Prostate cancer is the most common cancer type excluding non-melanoma skin cancer and the third-leading cause of cancer deaths in men in the USA [1]. The estimations by American Cancer Society suggest that there will be more than 150,000 new cases of prostate cancer diagnoses in the USA in 2018, which constitutes 19% of all cancer cases in this year. In 2012, prostate cancer is the second common cancer with over 1 million cases and fifth-leading cause of cancer deaths in men with 300,000 incidents worldwide [5]. Between the years 2009 and 2013, 21,797 prostate cancer incidents were reported in Turkey becoming the second most common carcinoma in men after lung cancer. The overall age standardized incidence rate of prostate cancer among all ages was reported as 36.4 per 100000 people in 2003 [6]. Between the years 2009 and 2017, cause of death of 29459 people in Turkey were reported to be malignant neoplasm of prostate [7].

Although prostate cancer is one of the most common malignancies in Turkey, only 325 articles on prostate cancer has been published in Turkey between 2006 and 2015, compared to 28714 articles worldwide making the scientific contribution from Turkey around 1% of the world. Although the number of articles is low, total number of articles between 2011 and 2015 increased by 70.8% with respect to years between 2006 and 2010 [8].

AR, PTEN and p53 protein network has a key role in the progression of many human cancers, which brings a new perspective to cancer therapeutics. Robinson *et al.* (2015) have been reported the alteration of expression levels (between 40-63%) of certain genes such as AR, ETS family, TP53 and PTEN with in castration-resistant prostate cancer (mCRPC), which is obtained from the clinical analysis of 150 affected individuals with metastatic disease [9,10].

To classify men according to low, intermediate, and high risk for tumor progression, numerous clinical features such as tumor stage, degree of tumor cell differentiation and the serum PSA of prostate cancer are used in clinical practice. A study was conducted by Singh *et al.* (2002) based on the expression-based models which help to distinguish patients according to risk for recurrence. Among remarkable factors for prostate cancer, serum PSA and measures of local invasion were found not related to critical gene expression signals [11].

Lotan *et al.* (2014) investigated the Gleason score, the most important variable in biopsies, to determine PTEN protein, a tumor suppressor, status in biopsies taken from 103 patients with prostate cancer. The expression level of PTEN was reported lost in prostate tumors with high Gleason grade and stage [12].

TP53 is another key tumor suppressor gene regulating apoptosis, senescence and several features important for cellular growth and death. TP53 mutations are found in several cancers and the activity of which is regulated by numerous biochemical mechanisms such as phosphorylation, acetylation and ubiquitination. Recently, TP53 has been reported to regulate microRNA (miR) gene expression. A group of miRs leads to development of oncogenesis (oncomiR), whereas other miRs obstruct tumor progression and dissemination (tumor suppressor miRs). Mutation of TP53 in carcinomas results in the growth and metastasizing of tumor cells [13].

Androgen receptor (AR) is a steroid-hormone activated transcription factor expressed throughout the progression of Pca. Majority of androgen-independent and hormone refractory Pca express AR and the growth of these types of cancer is nearly always dependent on the AR pathway [14]. Inhibition of androgen synthesis and modulation of AR signaling pathways is a method for treating prostate cancer [15].

2.2. Prostate Therapy

2.2.1. Tumor Receptors

Tumor receptors are one of the earliest targets for cancer therapy as they take part in carcinogenesis. The treatment of endocrine-related cancers such as prostate and breast cancer with receptor-target therapy is especially successful. Detecting the expression level of tumor receptors is the key for receptor-targeted therapy and traditionally, the expressions of receptors have been obtained by in vitro assay of biopsies. Recently, a method called tumor receptor imaging emerged for detecting expression profile of receptors instead of biopsy of tumor material, which can be heterogeneous. Advantages of tumor receptor imaging are the ability to characterize tumor biology, to identify therapeutical targets and to measure the receptor expression for the whole tumor rather than a part of it. Some of the receptors used in imaging are growth factor receptors (EGFR), somatostatin receptors (SSTRs) and steroid receptors such as AR and ER.

2.2.1.1. GRPR Receptors. Gastrin-releasing peptide receptor (GRPR) is a receptor of the protein GRP, which regulates several functions in gastrointestinal and central nervous systems such as release of gastrointestinal hormones and epithelial cell proliferation. GRPR is overexpressed in many malignancies including prostate, breast and lung cancer, whereas GRPR expression in normal cells are very low or absent. Thus, ligands binding to GRPR are potential diagnostic and therapeutic agents for tumors overexpressing GRPR. The amphibian homolog of GRP is bombesin (BBN) and its antagonist and agonist radiolabeled analogs are investigated for PET and SPECT imaging of prostate cancer [16]. One study investigated ^{99m}Tc -HABBN for visualization of Pca. Although Pca is visualized in pre-studies on mice, Pca is not detected in humans, probably because of ^{99m}Tc -HABBN's low metabolic stability [17]. Another study investigated two GRPR ligands, antagonist RM1 and agonist AMBA, with ^{18}F and ^{64}Cu labels. ^{18}F and ^{64}Cu -labeled AMBA probes showed low stability and rapid clearance from tumor. Both RM1 probes exhibit high in vitro stability and good in

vivo tumor imaging. Especially, ^{18}F -labeled RM1 showed increased tumor uptake and favorable pharmacokinetics [16]. Another study compared 4 agonists and 1 antagonists and decided that the antagonist bombesin Demobesin-1 labeled with ^{99m}Tc has the highest in vivo stability and tumor uptake. Agonists ^{111}In -PESIN and ^{111}In -AMBA have less but sufficient in vivo stabilities and their uptakes by tumor are high [18].

2.2.1.2. Steroid Receptors. Androgen receptor (AR) is a steroid-hormone activated transcription factor and the growth of Pca is dependent on the AR pathway. The inhibition of androgen synthesis and AR signaling could be a method for imaging or treating Pca. ^{18}F -FDHT is a promising agent for PET imaging of AR. Studies showed that many tumor sites that had not been identified before were revealed by this probe [19]. Although there is not any therapeutic agent targeting stromal AR, paracrine factors regulated by AR, such as transforming growth factors (TGFs), fibroblast growing factors (FGFs) and hepatocyte growing factors (HGFs) have ligands targeting them [20]. Estrogen receptor β (ER β) modulates AR signaling and loss of ER β results in upregulation of genes associated with poor patient survival in Pca. Studies on mice showed that the ER β agonist LY3201 upregulates ER β and inhibits AR expression by repressing AR coactivator RAR related orphan receptor c (RORc) and increasing AR corepressor dachshund family transcription factor 1 (DACH1). LY3201 also upregulates PTEN, NF κ B and TGF β and may be used for preventing the progression of Pca [21].

2.2.1.3. Other Receptors. Corticotropin-releasing factor receptor type 2a (CRFR2a) is a G protein-dependent receptor and overexpressed in many types of tumors such as prostate, pancreatic and breast. Des-acyl ghrelin (DAG) is a derivative of ghrelin, a neurotransmitter taking part in gastrointestinal tract. DAG binds to CRFR2a and the ^{177}Lu -labeled DAG may serve as a radiopharmaceutical against prostate cancer. It is a relatively stable analogue and a potential candidate for the treatment of tumor cells overexpressing CRFR2a [22]. Somatostatins are agonist ligands secreted by endocrine cells and six types of somatostatin receptors (SSTRs) have been identified. Most important action of somatostatins is to regulate the gastrointestinal system. Many endocrine

tumors, such as gastrointestinal and pancreatic tumors, express SSTRs. They are also expressed and localized in various tissue components of prostate cancer [23]. Radionuclide SSTR imaging and therapy have been important for detection and treatment of neuroendocrine tumors. ^{111}In -DTPA-pentetreotide is an agent approved by FDA for clinical SSTR imaging and therapy with up to 80% disease stabilization rate [19].

2.2.2. Drug target proteins specific for prostate cancer

Proprotein convertase subtilisin/kexin type 6 (PACE4) is a proprotein convertase, which processes the proproteins to active products. It has functions in secretory pathways and is overexpressed in 90-95% of the prostate tumors. Its role in cancer cell proliferation, tumor growth and angiogenesis make PACE4 a target in prostate cancer. High-affinity PACE4 inhibitor multi-leucine (ML) is investigated whether it can be used for diagnosis and treatment. ML is radiolabeled with ^{64}Cu for PET suitability. Although the stability of this probe in blood circulation is short-term, the prostate tumor uptake is high and visualization by PET is possible. Since the tumor uptake is specific, ML-based PACE4 targeting can also be used therapeutically, as ML shows anti-proliferative effects on prostate cancer cells [24].

Cell surface 78-kDa glucose-regulated protein (GRP78) is a endoplasmic reticulum chaperone protein involved in many processes such as folding and assembly of proteins, targeting misfolded proteins and translocating polypeptides across ER membrane [25]. GRP78 is highly expressed in cancer cell lines and solid tumors, including prostate cancer and correlated with poor survival. Aggressive variant prostate cancers (AVPC) is a group of tumors characterized by poor patient survival, and AVPC cells containing GRP78 on their surface can be targeted with a highly specific GRP78 ligand, SNTRVAP. Adeno-associated virus/phage (AAVP) particles displaying SNTRVAP provide feasible approach for AVPC imaging with PET and therapeutic purposes [26].

Prostate-specific membrane antigen (PSMA) is a cell membrane protein overexpressed in prostate cancer and has been investigated as a target for imaging and

therapeutic agents. Various PSMA ligands have been developed and labeled with several radionuclides such as ^{68}Ga , ^{18}F and ^{99m}Tc . ^{68}Ga -PSMA-HBED-CC has the ability to detect prostate cancer relapses and metastases by targeting extracellular domain of PSMA [27].

Chaperonin containing TCP-1 (CCT) is a macromolecular complex responsible for protein folding and overexpressed in cancer cells. Cytotoxic peptide CT20p is a selective toxic agent targeting CCT. It's only cytotoxic to cancer cells that highly express CCT and less toxic to normal cells. Folate conjugated hyperbranched nanoparticles are used as theranostic carriers with fluorescent dyes and therapeutic peptide CT20p via PSMA. Incubating PSMA-expressing Pca cells with Folate-HBPE(CT20p) results in cell death and reduces the tumor size [28].

Table 2.1. Receptors targeted for theranostic purposes treating prostate cancer and their ligands.

Receptor	Ligand	Agonist/Antagonist	References
GRPR	^{99m}Tc -HABBN	Antagonist	Ananias <i>et al.</i>
GRPR	^{18}F -labeled RM1	Antagonist	Liu <i>et al.</i>
GRPR	^{99m}Tc -Demobesin1	Antagonist	Schroeder <i>et al.</i>
GRPR	^{111}In -PESIN	Agonist	Schroeder <i>et al.</i>
GRPR	^{111}In -AMBA	Agonist	Schroeder <i>et al.</i>
PACE4	^{64}Cu -labeled ML	Antagonist	Couture <i>et al.</i>
GRP78	AAVP displaying SNTRVAP	Antagonist	Ferrara <i>et al.</i>
PSMA	Folate-HBPE(CTP20)	Antagonist	Flores <i>et al.</i>
PSMA	^{68}Ga -PSMA-HBED-CC	Antagonist	<i>et al.</i>
AR	^{18}F -FDHT	Antagonist	<i>et al.</i>
ER	LY3201	Agonist	<i>et al.</i>
SSTR	^{111}In -DTPA-pentetreotide	Antagonist	<i>et al.</i>

N-(4-hydroxyphenyl) retinamide (4HPR) is a member of the retinoids, which are the natural or synthetic derivatives of vitamin A. 4HPR is found to be cytotoxic to many cancer types including prostate, breast, colorectal and leukemia. Reactive oxygen species (ROS) induce the apoptosis and it has been reported that 4HPR increase the ROS levels in several cancer cell lines. Ceramides, a family of sphingosine-based

lipid messengers, are also have roles in apoptosis and their levels are increased by the induction of 4HPR as well [29].

2.3. Biological Networks

Biological networks provide important and novel insights into biological function, complex cellular mechanism and evaluation that could not be determined by the characteristics of the individual parts, since intracellular processes involve thousands of reactions and interactions, and to understand them they should be studied as an interconnected system [3]. Interactions between biological molecules can be modeled as a network. A network, which is also called graph, consists of nodes and edges connecting them. In biological systems, nodes can be genes, proteins or metabolites, and edges can be physical, functional or chemical interactions between these molecules. The graphs can be directed or undirected, depending on the type of interactions. In directed networks, the source and target are distinguishable, and the information flow is one-sided. Since there are many types of biological molecules and interactions, various biological networks have been designed to model them. Generally, five types of biological networks are used. These are Metabolic Networks, Gene Regulatory Networks (GRNs), Protein-Protein Interaction (PPI) Networks, Genetic Interaction (GI) Networks and Signaling Networks. Metabolic networks were the first biological network type to be created and they still remain the most comprehensive one. The availability of high-throughput technologies in the last decade led to the development of other biological networks. The advances in experimental techniques led to the development of computational methods, which process the experimental data and construct the biological networks using the data. The networks mentioned above generally represent all possible interactions. At a particular time, only a subset of these interactions may be active. Networks should be integrated with experimental data to identify the conditionally activated regions of the networks. Additional pathways associated with a particular condition can be detected by integrating networks and high-throughput data [30].

2.3.1. Metabolic Networks

A metabolic network is the collection of metabolites, biochemical reactions, enzymes and genes in an organism [31]. It is a representation of cellular metabolism. Metabolites are small biomolecules that are intermediates or products of the metabolism. Biochemical reactions are needed in cells for the extracting energy or materials from the environment and using them to produce metabolites. Although some reactions are spontaneous, enzymes may catalyze them in order to accelerate the reaction speed [3]. The amount of metabolites and biochemical reactions varies from tissue to tissue and also from cell to cell. One of the aims of metabolic network construction is to clarify the links between genes, proteins and reactions. Most commonly used approaches for acquiring metabolomic data are NMR and MS [32]. MS-based techniques include chromatography-based MS techniques like Gas Chromatography (GC-MS) [33] and Liquid Chromatography (LC-MS) [34] and chromatography-free methods such as direct infusion [35], matrix-assisted laser desorption/ionization (MALDI) [36] and matrix-free laser desorption/ionization [37].

One study investigated potential drug targets for *Staphylococcus aureus* which is a multidrug-resistant, hospital- and community-acquired pathogen causing pneumonia, meningitis etc. Since it is expected that the mortality due to this pathogen will surpass that of HIV/AIDS in the United States, new drug targets should be discovered. The metabolic networks for different strains of *S. aureus* were constructed using genome annotation, functional-pathway analysis and comparative genomic approaches. While 90% of the metabolic enzymes and reactions are similar in the 13 strains, 70 single enzymes and 54 pairs of enzymes were found to be essential enzymes for growth. 44 of single enzymes and 10 of enzyme pairs were found to be common to all strains. 6 pairs of enzymes were found to be in amino sugar biosynthesis pathways which are involves in cell wall metabolism [38].

Another study reconstructed the genome-scale metabolic network of *Burkholderia cenocepacia*, an epidemic and antibiotic-resistant pathogen in patients with cystic

fibrosis (CF) or compromised immune system in order to find potential drug targets. Strain J2315 is found as the most infections strain of *B. cenocepacia* and the metabolic network reconstruction of this strain led to the identification of 45 essential enzymes such as ArnT, ArnB and ArnC which catalyze vital reactions for *B. cenocepacia* and may be used as potential therapeutic targets [39].

The genome-scale metabolic network of the gram-negative anaerobe *Porphyromonas gingivalis* responsible for adult periodontitis were reconstructed using integration of information from genome, online databases and literature screening and a model with 679 metabolic reactions were built. 8 putative targets involved in lipopolysaccharide biosynthesis pathway (LPS) were selected as potential therapeutic targets. LPS molecular types are known to affect immune system response and selectively blocking of specific LPS would be valuable to control the inflammatory consequences of LPS production [40].

2.3.2. Gene Regulatory Networks (GRNs)

Gene Regulatory Networks (GRNs) are focused to describe metabolic pathways and bioprocesses. Gene expression is an essential process, in which the information from a gene to synthesize either proteins or functional RNAs (rRNA, tRNA) depending on whether the gene is protein-coding or non-protein coding. The regulation of gene expression controls the differentiation of cell types by ensuring that the correct expression of a gene is conducted at the right time and location. Transcription factors regulate the transcription of genes in eukaryotes [3]. GRNs are useful tools to explain complex dependencies between key transcriptional factors (TFs) and their target genes. These networks consist of directed graphs with nodes representing genes and edges leading from regulators like TFs to their targets [41]. Generally, two experimental approaches are used for data collection: Yeast one-hybrid assay is a gene-centered approach, which identifies TFs regulating a gene [42], and Chromatin ImmunoPrecipitation (ChIP) combined with microarrays or sequencing techniques is a TF-centered one, which predicts potential targets of a TF [43]. Several different computational techniques are also

suggested for the construction of GRNs. Bayesian, Dynamic Bayesian, Boolean and Probabilistic Boolean are some examples of these computational techniques [44].

In a study, four healthy human subjects were administered bacterial endotoxin, which activates the innate immune responses, and the gene expressions in whole blood leukocytes were determined at various time intervals. Four additional subjects were also studied but without endotoxin for control. K-means analysis and principal component analysis (PCA) were used to describe the overall changes in gene expression profiles. A prototypical inflammatory cell containing 292 genes and their direct interactions was constructed using the Ingenuity Systems Inc. network analysis tool. The global representation of the human inflammatory response to endotoxin consists of 1556 genes and their interactions. New endotoxin-responsive modules were identified by evaluating the subnetworks. One example for these modules is the increased expression of components of the superoxide-producing phagocyte NADPH-oxidase system, which is an important enzyme for the host defense [45].

Another study by Seok *et al.* used the gene expression dataset of Calvona *et al.* and built a dynamic network of transcription events following endotoxin administration with 10 TFs, 99 target genes and 149 regulatory interactions. The activities of 10 TFs over time were predicted and the regulatory strength of TFs exerted on their target genes were determined using Network Component Analysis (NCA). It was found that when the regulatory strength matrix is clustered, the group of genes, which are co-regulated, can be determined. Incorporating the regulatory network information led to the identification of many additional interactions between TFs and their target genes. Also, it enabled to describe the activation of regulatory proteins and target genes over time [46].

The time-lagged expression correlation between TF genes and clusters of potential target genes were studied using a transcriptional network underlying TLR-stimulated murine macrophage activation. To determine the associations between TF genes and clusters of co-expressed target genes, microarray based expression profiling is used.

Time-lagged correlation (TLC) is a method used for temporal ordering of expression to determine whether the time lag between two correlated genes is consistent, and was used to identify the potential causal influences in the network. Using TLC, a network of 36 TF genes and 27 co-expressed clusters with statistically significant associations was identified. One finding of this study is that transcriptional regulator TGIF1 plays a role in macrophage activation [47].

2.3.3. Protein-Protein Interaction Networks

Protein-Protein Interactions are specific, direct or indirect physical contacts between proteins in a living organism [48]. PPI networks have great potential for identifying pathways and subnetworks of interest. The identification of essential gene networks and highly connected hub proteins may be used for discovering new drug targets. In PPI networks, the nodes represent the proteins whereas the edges represent the physical interactions between proteins. PPI networks are scale-free networks, meaning most proteins have only a few interactions while a few proteins have a large number of interactions. These proteins with lots of interactions are called hubs. Analysis of PPI networks showed that hub proteins are more likely to be essential than non-hub proteins [49]. PPIs are detected by in vitro, in vivo and in silico methods [50]. In in vitro methods, the procedure/test is performed outside a living organism while in in vivo techniques the procedure or the test/assay is conducted on the whole organism. Most commonly used in vitro methods are tandem affinity purification (TAP) [51], affinity chromatography [52], coimmunoprecipitation [53], protein fragment complementation (PCA) [54], protein arrays [55], X-ray crystallography [56], NMR spectroscopy [56]. In vivo methods are yeast two hybrid (Y2H, Y3H) [57] and synthetic lethality [58]. In silico methods are performed via computer simulation and there are sequence-based approaches [59], structure-based approaches [60], chromosome proximity, gene fusion, in silico 2 hybrid [61], mirror tree [62], phylogenetic tree [63] and gene expression-based approaches [64].

A dynamic PPI network of innate immune response was constructed by combining the gene expression data of lipopolysaccharide (LPS) stimulated dendritic cells with PPI data. Protein modules were identified in a differentially expressed network. Some of the identified modules were found to have a role in innate immunity, cell cycle and signaling etc. Also, for each protein complex, the enrichment analysis for biological processes and pathways were performed. Toll-like receptor (TLR) subnetwork was investigated, and it was found that S100A8 is downregulated in dendritic cells after LPS stimulation. The identified protein complexes such as TICAM1–TBK1–TRAF6–TLR3 have a crucial role in pathways involved in cancer as well as innate immunity [65].

Interleukin-10 (IL-10) is a well-known immune-modulatory cytokine with anti-inflammatory activity and its network is among the most important paths linking cancer and inflammation. A structural PPI network of IL-10, which also indicates the location of interaction sites and how the proteins interact, was constructed using Protein Interactions by Structural Matching Tool (PRISM). The effects of mutations on the interactions were analyzed and it is found that mutations that disrupt the interactions of IL-10 with its receptors IL-10RA and IL-10RB and A2M may enhance inflammation and modulate anti-tumor immunity. Mutations that disrupt the A2M–KLK13 interaction may promote cancer cell growth and metastasis. Also, another finding is that the blockage of A2M–APP interaction can lead to cancerous cellular proliferation due to free APP [66].

Another study constructed the normal and inflammatory PPI networks for type of influenza A virus (IAV) infections H5N1 and H1N1 which can trigger inflammatory diseases and cause respiratory tract infections. Using an optimization algorithm, the dynamical interactions among TLR2, IL-1, IL-10 and NFB between normal and inflammatory networks were identified. One result obtained by this study is that the protein complex TNFSF10–HDAC4–HDAC5 appeared in the normal and inflammatory networks at early stages of IAV infections and at the late stage in normal networks, thus this protein complex may be important for controlling IAV-induced inflammation [67].

2.3.4. Genetic Interaction (GI) Networks

Genetic interaction is an important concept for identifying the functional relationships between genes or their products. Genetic interactions are defined by the deviation of phenotype with double mutations from the neutral phenotype. In other words, if two independent mutations on two genes cause a difference in the neutral phenotype, there is a genetic interaction between these two genes [68]. Two types of genetic interactions are widely used: Synthetic-lethal interactions, in which two independent mutations cause a lethal effect on organism; and suppressor interactions, in which the lethality of one mutation is modified by the other mutation providing the cell's viability. Several methods have been developed for detecting genetic interactions. Synthetic genetic arrays (SGAs) [69], diploid synthetic lethality analysis by microarray (dSLAM) [70], epistatic miniarray profiles (E-MAP) [71] and RNA interference (RNAi) knockdowns [72] are some of the popular methods.

A genetic interaction network was constructed in order to analyze the autosomal dominant polycystic liver disease resulting from mutations in *PRKCSH* and *SEC63*. The results showed that these mutations also cause cyst formation in the kidney. *GII* or *Sec63p* were shown to be required in mice for the functional *PC1-PC2* complex which is a product for the polycystic kidney disease genes. It was also shown that the expression of *PC1*, a rate limiting component determining the severity of the cystic phenotype, can modify the level of cyst formation caused by the mutation in *Pkhd1* [73].

The cell wall protein *Ccw12p* is a mannoprotein which constructs the outer cover of the baker yeast's cell wall and the lack of this protein causes cell wall damage. The genetic interaction network identified 21 genes involved in cell wall integrity, chitin synthesis and cell polarity. Some of these genes such as *PFD1*, *WHI3* and *SRN2* have not been related to cell integrity before. It was suggested that *Ccw12p* has a specific role in reinforcing the expanding of cell wall [74].

2.3.5. Signaling networks

Signaling networks regulates the behavior of a cell by observing current conditions and making decisions about the appropriate use of resources [75]. They constantly monitor external stimuli like nutrient levels, growth factors and hormones, and internal stimuli like DNA damage and protein misfolding [3]. Receptors sense the information from both stimuli which is transmitted through various biochemical reactions and changes in transcriptome, proteome or metabolome may occur. Hence, signaling networks are the interface between stimuli and consequent phenotypic changes. Thus, reconstructing genome-scale signaling networks is crucial for understanding complex diseases.

Other than techniques like Yeast Two-Hybrid (Y2H) assays used for deciphering protein-protein interactions, recent experimental techniques have been developed for investigating the functional relationships in signaling networks [76]. Perturbation analysis that monitors the genome-wide changes in expression after disrupting specific components of a network [77], knockdown strategies like RNA interference (RNAi) [78], protein assays on the activity of proteins [79] and fluorescence imaging technologies generating data regarding protein localization and the dynamics of the signaling processes [80] are some of these techniques.

One study investigated the innate immune signaling networks regulating the production of type 1 interferon (IFN- α) by analyzing the protein complexes formed after microbial recognition. The dynamic changes in protein-protein interactions resulting from encounter with ligands were followed and analysis showed that 58 IFN- α -associated baits revealed connections with 260 proteins forming a framework of 401 protein interactions. Overexpression analysis identified 28 genes induced NF- κ B activity and five of them (TRIM32, MIB1, NEDD4L, TXLNA, SURF4) are positive regulators of NF- κ B. Mind bomb (MIB) proteins were shown to physically interact with TBK1, a kinase that phosphorylates IRF transcription factors controlling interferon production [81].

In another study, transcriptional profiling, genetic and small molecule perturbations and phosphoproteomics were combined to discover the signaling components involved in the Toll-like receptor (TLR) response of immune dendritic cells (DC). 35 signaling regulators, including 16 known regulators, were found to be involved in TLR signaling. Polo-like kinases (Plk) were discovered to have a role in host defense pathways inducing type 1 interferon networks. Especially Plk2 and Plk4 are important components of antiviral pathways and activate proteins such as Tnfrsf25 which was associated with autoimmune diseases although its role was not known [82].

Another study integrated the quantitative proteomics, genomics and small molecule perturbations to identify genes involved in the interferon-stimulatory DNA (ISD)-sensing pathway in which the immune system senses the viral DNA entering the mammalian cells and triggers type 1 interferon production. 809 candidate genes were silenced, the dsDNA responses were measured and resulting hits were connected with the known signaling network. ABCF1 was identified as a critical protein that associates with dsDNA and DNA-sensing components HMGB2 and IFI202. Another result is that CDC37 regulates the stability of the signaling molecule TBK1. Also, the inhibition of the CDC37-HSP90 interaction modulates the innate immune response to DNA and retroviral infection [83].

2.4. Network Medicine

The expanding number of acquired human specific molecular data led to an increase in knowledge about the relations between human metabolism. In the last years, human interactome is better understood as the development of high throughput techniques greatly decreased the money and time costs and analyses can be done in much shorter time. Today, hundreds of thousands of protein-protein interaction data can be acquired from several databases. To better analyze this big data, network biology is widely used. Network medicine is another concept benefitting from network theory. Its primary aim is to identify and treat diseases by looking at a larger perspective. Instead of identifying one abnormality that causes a defect, it aims to identify the molecular

pathways leading to that disease, since diseases are generally caused by a defect in a gene. A defect in this gene changes the activity of other interact with them, therefore a systematic method such as biological networks are vital for understanding the disease progression and proposing potential drug targets [84].

In disease networks, proteins encoded by disease genes have distinct features. At first, one may think hubs, proteins interacting with greater number of proteins, must be disease proteins. But it is shown that hub proteins are encoded by essential genes that are critical for survival and disease genes are not essential [85]. Essential genes tend to be expressed in many tissues, whereas disease genes are more tissue-specific.

Three different types of modules can be distinguished in disease networks: Topological modules, functional modules and disease modules. Topological modules show a dense subnetwork in which the proteins share more topological features with each other than other proteins outside the modules. In functional modules, proteins with similar or related functions are grouped together. The fact that the proteins involved in the same disease tend to interact with each other, leads to the generation of disease modules. Disease modules are likely to overlap with topological and functional modules, but it is not necessary [84].

Linkage, disease-module based and diffusion-based methods can be used for identification of new disease genes. In linkage methods, it is assumed that the interaction partners of disease proteins are more likely to be involved in the same disease. Disease-based methods assumes that proteins in the same topological, functional or disease module are most likely to be involved in the same disease. Diffusion-based methods use random-walk in which the one random walker moves to any neighbor of the disease protein with equal probability. At the end, proteins with visited the most by the random walkers are identified. Among these three methods, diffusion-based methods are found to be prior to disease-based and linkage-based methods, with that order.

2.5. Sphingolipids and Cancer

Sphingolipids are essential lipids defined by their 18 carbon-amino alcohol backbone and modification of this basic structure gives a vast family of sphingolipids that play significant roles in membrane biology and regulating cell function [86]. They have a rapid turnover and their levels are controlled by the balance between synthesis and degradation [87]. Sphingosine-1-phosphate (S1P), ceramide and ceramide-1-phosphate (C1P) are sphingolipids take part in inflammation and immune responses.

Sphingolipids play an important role in cell signaling, growth, apoptosis and proliferation. Recently, sphingolipids and their membrane receptors have been related to oestrogen-activated mediators. In breast cancer, sphingolipid signalling pathway is stimulated by the steroid hormone oestrogen. The analysis of samples taken from patients indicated significantly escalated levels of sphingosine kinase 1 (SPHK1) mRNA expression in breast carcinoma. The expression levels of SPHK1 and S1P receptors can be analysed in order to improve diagnosis and prognosis of breast cancer [88].

Ceramidases have been widely studied as mediators of sphingolipid metabolism and cancer. Beckham *et al.* (2014) demonstrated that acid ceramidase dysregulated in 60 percent of human prostate tumors. They reported that acid ceramidase overexpression enhanced prostate cancer cell resistance to chemotherapy by hindering accumulation of proapoptotic C16-ceramide following introduction of drugs such as etoposide, cisplatin, doxorubicin, and gemcitabine. Sphingosine kinases and S1P production could be targets to analyze the stimulation of ceramide and sphingosine levels in prostate cancer [89].

Ceramide can be produced by the activation of sphingomyelinase (SMase) by tumor necrosis factor alpha (TNF- α) which also activates the nuclear factor- κ B (NF- κ B). NF- κ B is a family of pro-inflammatory transcription factor encoding more than 150 genes including cytokines like IL-1 β , IL-6 and IL-8 and pro-inflammatory enzymes like COX-2 [90]. Ceramide itself can also activate inflammatory pathways via NF- κ B

transcription and up-regulate another family of transcription factors, c/EBP, which induces the gene expression of several inflammatory proteins.

Ceramide-1-phosphate (C1P) is a sphingolipid regulating cell migration and promoting cell growth and survival. C1P and ceramide kinase, the enzyme that catalyzes the synthesis of C1P, are involved in the inflammatory responses [91]. It can exert both pro- and anti-inflammatory functions. Some of the anti-inflammatory effects of C1P are blockade or counteraction of ceramide-induced inflammatory responses, whereas the pro-inflammatory effects are activation of NF- κ B transcription factor and upregulation of pro-inflammatory prostaglandin E2, PGE2 [92].

Apoptosis is a gene-directed program which activates specific metabolic pathways and results in cell shrinkage, membrane blebbing, chromatin condensation and nuclear fragmentation. This cellular process is usually not associated with an inflammatory response, while necrosis, a process which results in cells death because of membrane integrity, is usually pursued by inflammatory response. It has been reported that most cytotoxic anticancer agents lead apoptosis and induce cell death. Sphingolipid molecules can regulate the ability of malignant cells to grow and resist anticancer treatments. Ceramide, an intracellular proapoptotic signaling molecule, is produced by cancer cells in response to chemotherapeutics. The predominant pathway for the generation of ceramides against chemotherapy is the activation of SMase which hydrolyzes sphingomyelin and regulates apoptotic pathways. Ceramides and its metabolites have been reported as regulators of apoptosis and cell senescence. The introduction of water-soluble ceramides to cells in vitro causes apoptosis, whereas lower level of ceramide in the cells confers drug resistance. Thus, ceramides could be potential molecules to increase apoptotic potential of cancer cells and develop new strategies for cancer treatment [93,94].

S1P has diverse roles in innate and adaptive immunity. It regulates a variety of processes including immune cell trafficking, cell survival, cell differentiation and cytokine production etc. S1P and its receptor S1PR1 controls the trafficking and

migration of immune cells such as T and B lymphocytes, dendritic cells, macrophages and neutrophils [95]. S1P and SIPR1 also regulate allergic responses, lymphocyte differentiation and cytokine expression. S1P levels are found to be elevated in people with multiple sclerosis (MS), Crohn's disease and asthma. S1P has also been shown to exert anti-inflammatory roles by promoting the switch from pro-inflammatory M1 to the anti-inflammatory M2 macrophage [96]. In mast cells, S1P is involved in the activation and development of the inflammatory response. In other cells, S1P has a role in inflammatory process, particularly in response to $\text{TNF-}\alpha$.

3. METHODS

3.1. Databases

3.1.1. Protein-Protein Interaction Data

Protein-interaction data for human are acquired from several databases: the Biological General Respiratory for Interaction Datasets (BIOGRID) [97], IntAct molecular database (IntAct) Molecular Interaction database (MINT) [98], Human Protein Reference Database (HPRD) [99] and STRING [100].

BIOGRID is a database with over 410000 raw and 300000 non-redundant interactions for humans. IntAct database was developed by EMBL-EBI and over 720000 interactions are present where 51% of these interactions belong to Homo sapiens. HPRD includes the over 41000 protein-protein interactions for human. STRING is a database with experimental and predicted protein interactions from various sources and contains 9.6 million proteins from 2031 organisms. For this work, only experimental protein-interactions were used and assumed as undirected. Number of PPIs in each database is tabulated below. Totally, 350,895 unique protein-interactions between 18817 proteins were gathered.

Table 3.1. Number of PPI from various databases.

Database	Protein-Protein Interactions
BIOGRID	266089
IntAct & MINT	97042
STRING	166597
HPRD	39008
Total	350895

3.1.2. Gene Expression Data

The Metabolic gEne RApid Visualizer (MERAV) [101] is a database for gene expression of various normal tissues, primary tumors, cancer and non-cancer cell lines.

GTEX (Genotype-Tissue Expression) [102] and HPA (Human Protein Atlas) [103] contain the gene expression profiles for healthy tissues. The gene expressions are given in TPM (Transcripts per Million; for every 1,000,000 RNA molecules in the sample, x number came from this gene/transcript). TCGA (The Cancer Genome Atlas) [104] contains the data for 33 types of cancer including prostate cancer. Here, the gene expressions are given in FPKM (Fragments per Kilobase Million; rate of fragments per base multiplied by the number of fragments sequenced).

The Gene Expression Omnibus (GEO) [105] is a public database with high-throughput genomic datasets such as microarrays and chips.

The gene expression data from TCGA was converted from FPKM to TPM, since converting TPM to FPKM is more difficult. The formula for converting FPKM to TPM is [106]

$$TPM_i = \frac{FPKM_i}{\sum_j FPKM_j} \times 10^6 \quad (3.1)$$

Table 3.2. Number of samples and genes in GTEX and TCGA databases.

	Number of samples	Number of genes
GTEX	152	56201
TCGA	551	60483
Mutual	703	53196

3.1.3. Genes/Proteins Data

Online Mendelian Inheritance in Man (OMIM) [107] is a catalog of human genes and genetic phenotypes and focuses on phenotype-genotype relationships.

Genome Wide Association Studies (GWAS) [108] focus on identifying single nucleotide polymorphisms (SNP) that occur frequently in a disease. PheGenI database is used for obtaining GWAS data for prostate cancer [109].

KEGG (Kyoto Encyclopedia of Genes and Genomes) [110] is a bioinformatics database containing genomes, biological pathways, reactions and diseases.

InnateDB [111] is a database of innate immune response of humans and mice and includes genes, proteins, interactions and signaling pathways related to innate immune system.

Connectivity-Map (C-Map) is a collection containing over 1.5 million gene expression profiles representing over 5000 compounds and 3000 genetic reactants. The similarities between the expression data of human cells lines perturbed with compounds and reactants are found by a pattern matching algorithm. Diseases, drugs and genes can be linked together and functional connections can be made [112].

3.2. Network Construction

3.2.1. PPI Network Construction

3.2.1.1. Selecting Core Proteins. To obtain the core protein set, three different set of proteins were investigated: KEGG Prostate cancer pathway, neighbors of KLK3 protein and k-fold analysis between the gene expression of normal prostate tissue and prostate cancer cell lines. KEGG Prostate cancer pathway includes the genes/gene products and molecules associated with the selected disease and mapped as a pathway. Prostate

specific antigen (PSA) encoded by Kallikrein-related peptidase 3 (KLK3) is a protein produced almost only in prostate tissue and is used as a biomarker for prostate cancer, since PSA levels increase in the case of prostate cancer [113]. KLK3 related terms were obtained by screening the protein-protein interaction databases STRING, BIOGRID, IntAct, InnateDB and HPRD. As the last dataset, 10 samples of normal prostate tissue and 17 samples of prostate cancer cell line gene expression data of 17724 genes were acquired from MeRAV, and for each gene, k-fold was calculated. KF threshold was chosen as 2 and the up/downregulated genes were determined. Combining these three sets of proteins, 113 core proteins were selected as the core protein set for construction of Pca PPI network.

3.2.1.2. Network Construction by GO Annotations. The interaction partners of the core proteins given in Table 8 were investigated and those, that do not share any term from the mutual GO list, were eliminated, i.e. the new interacting partner should have at least one GO term from each aspect (MF, CC, BP) to be included into the network. This process was repeated for the next neighbors until no more protein can be added to the system (Table 4.5). At the end, all the interactions among all the proteins were checked to see whether they are included in the reconstructed network.

After 5th repetition, no more proteins or interactions could be added to the system and the final PPI network for the prostate cancer is obtained. After the self-loops were eliminated, the PPI network for the prostate cancer with all proteins includes 160666 interactions between 8709 proteins including 113 core proteins (Figure 4.6).

3.2.2. Co-expression Network Construction

Gene co-expression networks are constructed from transcriptome data and an effective way to identify the correlation among genes. By detecting genes functioning in the same manner, it is possible to identify if they take part in the same biological processes and play key roles in human diseases [114]. In co-expression networks, nodes show genes and they are connected by an undirected edge, they are significantly

correlated. One popular method to construct gene co-expression network is Weighted Gene Co-Expression Network Analysis (WGCNA). It also detects the gene modules and identifies hubs [115].

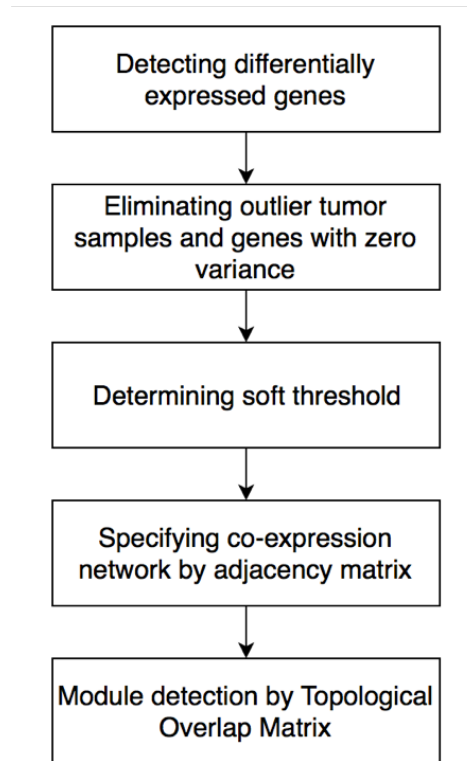


Figure 3.1. Flowchart of Co-Expression Network Construction.

A flowchart describing the algorithm of construction of Co-Expression networks are given in Figure 3.1. TCGA prostate cancer tumor gene expression data was used for the analysis by WGCNA. The dataset consists of expression data of over 50000 genes from 551 samples and constructing a network from such a large dataset is computationally expensive. Therefore, differentially expressed genes (DEGs) were detected with the help of the R software package limma. The package contains the capabilities of performing differential expression analysis and detecting co-expressed sets of genes by analyzing expression profiles [116]. Although co-expressed genes can be identified by limma, WGCNA has superior module detection and visualization skills.

Differentially expressed genes between normal prostate tissue and Pca tumor were determined by using the cut-off values $FDR < 0.05$ and $|\log_2 FC| > 1.5$ which corresponds to $FC = 3$. Next step was the construction of the co-expression network from DEGs. First, the genes and samples were checked and if they miss too many values, they were eliminated. Second, the samples were hierarchically clustered in order to remove the outlier samples. To continue constructing the network, the symmetric adjacency matrix a_{ij} , the connection strength between the nodes i and j , with values between 0 and 1 should be calculated. For a weighted network, adjacency is defined as

$$a_{ij} = s_{ij}^{\beta} \quad (3.2)$$

where s_{ij} is the co-expression similarity and β is the soft threshold. WGCNA gives different options for calculation of s_{ij} and for this analysis Pearson correlation was chosen as the similarity measure as it is given as default. Soft threshold assigns a weight to each connection and is required for the scale-free topology of the co-expression network. Linear regression model fitting index R^2 was used to detect how well a network satisfies scale-free network topology [117]. Although a R^2 value close to 1 means a good fit, it also leads to a network with few connections. Thus, the choice of soft-threshold is a significant parameter of network construction. The criterion determined by the developers of WGCNA is that the R^2 value should be at least 0.80 for a scale-free network. Soft threshold was chosen as 7 in this analysis.

After the co-expression network was specified by the adjacency matrix, clusters composed of densely interconnected genes were detected as modules. Interconnectedness of genes were determined by Topological Overlap Measure (TOM). While adjacency matrix describes the correlation between every pair of genes separately, TOM compares the relation of a pair of genes with all other genes. A high TOM between two genes means they connect to the same group of genes [118]. The type of the network was chosen as unsigned, meaning the absolute value of the Pearson correlations are taken into account for modules. The reason for this selection is that a network with only positive correlated members may lack important information. Minimum mod-

ule size was determined as 50 as the number of genes in the network large and small modules may give little information. Another parameter in module detection, merge-CutHeight, is the threshold for merging different modules and was selected as 0.10, so that the network is divided into a greater number of modules.

To visualize the networks as edges and nodes representing the correlated genes in Cytoscape, node and edge files were generated from the TOM by determining a threshold for node connections.

3.2.3. Ligand-Centric Network Construction

To obtain the crosstalk between sphingolipids and Pca, the proteins related to sphingolipids were acquired from UniProt database by PLITool, a tool using BioServices package of Python. Sphingolipid related proteins were acquired by searching the keywords “Sphingolipid”, “Glycosphingolipid”, “Sphingomyelin”, “GO:0006665” Sphingolipid Metabolic Process and “GO:0090520” Sphingolipid Mediated Signaling Pathway in Uniprot. Since the Sphingolipid Mediated Signaling Pathway has little number of entries, all the analysis was done by adding the Sphingolipid Metabolism and Signaling Pathway terms from the KEGG database. Then, the protein-ligand interactions were acquired from ChEMBL database and these interactions were grouped and eliminated according to their activity type and activity value. The activity types and corresponding thresholds are given in Table 3.3.

As the last step, ligand-centric inhibition and activation protein-protein interaction networks were constructed using the source code given by Öztürk *et al.* [119]. Although the edges represent physical interactions between two proteins in PPI networks, the proteins in ligand-centric networks are connected if they share at least one ligand.

Table 3.3. Activity types and thresholds for inhibition and activation networks.

Network Type	Activity Type	Threshold
Inhibition	Activity	<50%
	EC50	<1000 nM
	IC50	<1000 nM
	Inhibition	>50%
	Potency	<1000 nM
	Residual Activity	<50%
Activation	Activity	>101%
	Residual Activity	>101%

3.3. Clustering Methods

To analyze the large-scale PPI networks, some methods including network clustering algorithms have been developed throughout the years. Clustering became a popular method for analyzing the topological and functional properties of a network. The main aim of network clustering is to construct subnetworks from a larger network which shares certain clustering properties so that topological or functional modules are discovered. For this aim, different clustering algorithms with various characteristics have been introduced in the past years [120]. These algorithms were proposed by taking some features of the PPI networks such as the possibility of overlapping clusters, edge weights of interactions, annotation of attributes like Gene Ontology (GO) or the coverage of the network into account. Generally, the clustering algorithms can be divided into the following categories: Heuristic-based, Flow-based, Complete Enumeration, Random Walks and Message Passing, Graph-cut and Hierarchical Clustering, Multiple Clustering-based and Genomic Data-driven Clustering Algorithms.

Heuristic-based algorithms find dense network clusters by searching heuristically for potential cluster regions using an iterative seed and extend strategy. Heuristic-based algorithms are based on topological properties of networks to find functional modules and allow overlapping clusters. These methods are partial coverage methods;

i.e., they find the locally dense subgraphs which do not cover the network completely. Methods with higher coverage scores are advantageous than the rest of the algorithms. Some of the heuristic-based algorithms are MCODE [121], DPCLUS [122], IPCA [123], RNSC [124] and ClusterONE [125].

The second group of clustering algorithms are flow-based algorithms. In these algorithms, the clustering is achieved by a series of flow expansions and contractions to identify clusters. Although these algorithms cover the whole network, most of them cannot utilize the annotations such as GO. Markov Clustering (MCL) is the most used method among the flow-based algorithms. This approach was improved by several methods such as TRIBE-MCL [126], MLR-MCL [127] and SR-MCL [128]. Other studies conducted by Pereira-Leal *et al.* [129] and Cho *et al.* [130] also proposed algorithms using flow-based methods.

Complete enumeration algorithms try to compute all possible subnetworks above certain density. The problem with these types of algorithms is that they are computationally expensive methods and they can only be applied to small networks, although all relevant modules can be identified. Main algorithms using complete enumeration are super-paramagnetic clustering (SPC), Monte-Carlo optimization based technique (MC) [131], CFinder [132], Clique Percolation Method (CMC) [133] and DME [134].

The next method is random walks and message passing algorithms. In order to identify the clusters, they model the graph as a series of random walks. Nibble [135], Repeated Random Walk (RRW) [136] and Affinity Propagation (AP) [137] are typical examples of this kind of clustering algorithm. All three algorithms cover the whole network, but they do not utilize the GO just as flow-based algorithms. Moreover, only RRW enables overlapping clusters.

Hierarchical clustering algorithms are popular clustering methods, which view the hierarchy of a network in form of a tree. Full coverage and imitation of the hierarchical organization of the networks are the advantages of these methods, however the lack of

overlapping clusters and its sensitivity to noise favors other techniques. The algorithm by Chen *et al.* [138], MoNet [139], Tree-Snipping [140], VI-Cut [141], SCAN [142] and NeMo [143] use hierarchical clustering as their base method.

Multiple clustering-based algorithms have diverse approaches for clustering. They generate a set of clusters rather than a single cluster and then combine them. One technique developed by Asur *et al.* [144] and MOD-ILP [145] are full coverage, no overlapping clusters methods without annotation utilization. Non-negative matrix factorization-based (NMF) [146] has the same features as the previous two methods except NMF allows overlapping clusters.

The last category, genomic data-driven clustering algorithms integrate the genomic and PPI data to eliminate the noise in PPI networks. All of these methods have no utilization of GO annotations. While the algorithms by Segal *et al.* [147] and Maraziotis *et al.* [147] do not cover the full network, Lu *et al.* [148], Zheng *et al.* [149], CEZANNE [150] and Shi *et al.* [151] enable full coverage. Among these methods, only the technique of Shi *et al.* allows overlapping clusters.

Flow-based MCL is one of the widely used clustering algorithms for the analysis of different type of networks and clusterMaker application of Cytoscape was used as the main tool for constructing subnetworks in order to identify functional or topological modules.

3.4. Analysis Methods

3.4.1. Topological Analysis

Graph theory is a framework for the mathematical treatment of complex network and can be applied to variety of systems such as social networks, food-webs and biological networks [152]. The graphs can be directed, having a starting and end node or undirected, no directions are associated with the edges. Gene regulatory networks

and metabolic networks are generally directed, whereas protein-protein interaction networks are undirected.

Degree or connectivity of a node is defined as the number of connections or edges the node has to other nodes [153]. The graph density shows how dense a graph is according to the number of connections per node. A path is a pass through in which nodes are only visited once and the minimal length between two nodes are defined as the shortest path [154]. The average shortest path length, in other names characteristic path length, and the diameter of a network are defined as the average and maximum values of the shortest paths. Shortest paths play an important role in the transport and communication within a network.

Clustering coefficient is the measurement that shows the tendency of a graph to be divided into clusters. The local clustering coefficient measures the ratio of number of edges of between the neighbors of a specific node to the possible number of edges and the closer the local clustering coefficient is to 1, the more likely it is for the network to form clusters. Degree centrality shows that an important node is involved in a large number of interactions and equals to the degree of a node in an undirected graph. Nodes connected to many neighbors are called hubs and they have a high degree centrality. Removal of a hub has greater effect on the network connectivity than removal of a node with low degree centrality [152]. Closeness centrality is a measure which indicates important nodes that can communicate quickly with other nodes, so they are closer. It is a concept used for identifying the top central metabolites in large-scale metabolic networks. The idea behind the last of the commonly used topological measure, betweenness centrality, is that an important node lies on a higher number of paths between other nodes. The mean value of the betweenness centrality of essential proteins in a yeast protein-protein interaction network is 80% greater than non-essential proteins [152]. Degrees, clustering coefficients and centrality measures such as betweenness centrality are especially important factors which will be investigated and used for the analysis of the immune system PPI network.

3.4.2. Functional Enrichment Analysis

Functional Enrichment Analysis, or in other words Gene Set Enrichment Analysis, is a method focused on gene sets rather than single genes [155]. The method was firstly developed for discovering the metabolic pathways changed in diabetes, then it was modified to apply to larger datasets. Functional Enrichment Analysis makes it easier to interpret the results of the dataset by identifying statistically over-presented pathways and processes for a gene set rather than single genes. It groups the genes that share common biological function, chromosomal location or regulation based on prior biological knowledge such as biochemical pathways. In this proposed work, Functional Enrichment Analysis will be used for identifying the pathways and biological processes shared commonly by the members of the subnetworks. clusterProfiler is a widely used R package for enrichment analysis and visualization of various gene sets such as Gene Ontology, KEGG Pathway, Disease Ontology (DO) etc. [156].

3.4.3. Crosstalk Analysis

Crosstalk is the connectivity between (inter) or within (intra) biological groups like experimental gene set, a functional module or a pathway [157]. Especially, understanding the crosstalk between signaling networks is crucial for drug design against diseases caused by defects in signaling pathways [158]. A probabilistic estimate is essential for statistically meaningful conclusions about crosstalk enrichment. It is produced by comparing an observation to a reference, the former being the number of links in a network structure like a module and the latter being determined by the null model hypothesis expressing a distribution of the number of observations expected. A null model hypothesis is needed, since comparing a hub with many links to a regular node with small number of links does not makes sense, because same number of links may show an important biological pattern for regular nodes [157]. EnrichNet [159], CrossTalkZ [157], NEA [160] and BinoX [161] are some of the tools for crosstalk analysis.

3.5. Drug Target Identification

Drug discovery is a challenging process that is expensive and can last decades before clinical approval. It limits the testing of potential therapeutic agents [162]. Other than cheminformatics tools, network-based drug discovery can find possible drug targets. This approach collaborates genomics, transcriptomics, proteomics, metabolomics etc. which depends on the development of computational and system biology tools for data interpretation [163].

If inflammation is not controlled or resolved, several diseases such as asthma, multiple sclerosis and inflammatory bowel disease can be developed. Therefore, identification of drugs targeting prostate cancer proteins holds great importance. This proposed work aims to find possible drug targets for prostate cancer using network-based methods.

4. RESULTS

4.1. Construction of PPI Network

A flowchart showing the steps of PPI Network Construction is given in Figure 4.1.

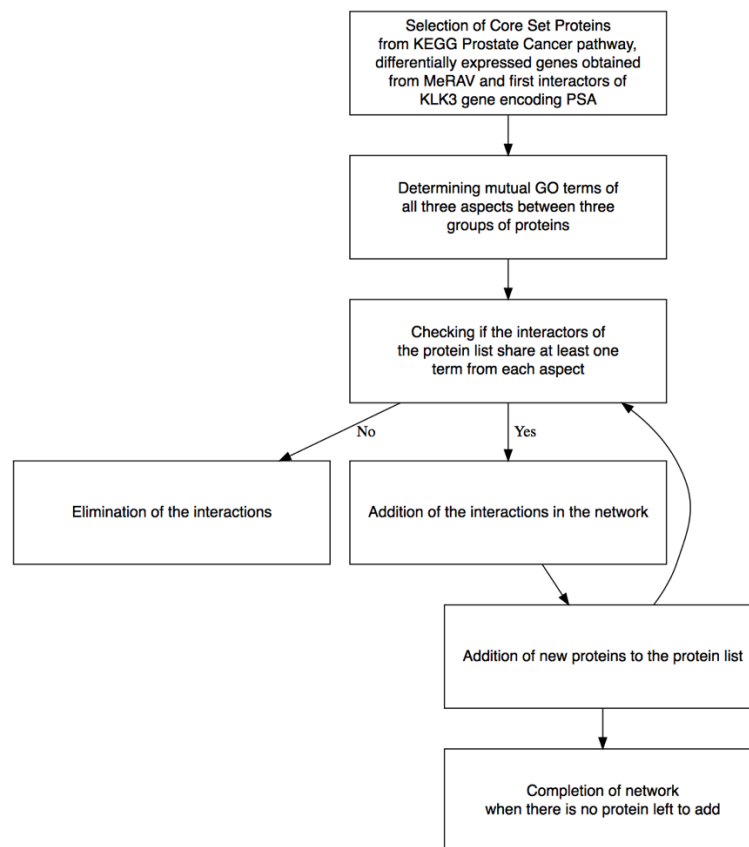


Figure 4.1. Flowchart of PPI Network Construction.

For the construction of Pca PPIN, three groups of proteins were accepted as the core set proteins. The first one is the members of Prostate Cancer pathway map given in KEGG database. The map is shown in Figure 4.2 and 86 proteins having roles in this pathway are taken as KEGG terms.

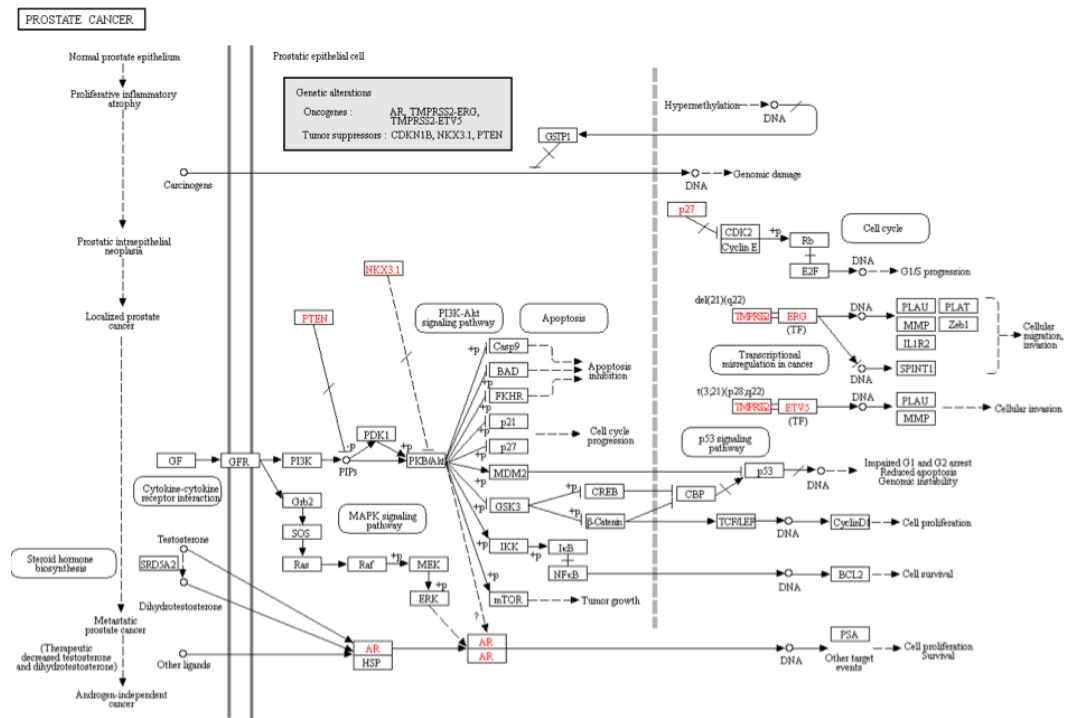


Figure 4.2. Prostate Cancer pathway (map05215) from KEGG.

Table 4.1. Prostate Cancer pathway proteins in KEGG.

AKT1	CDKN1B	EGFR	IKBK	NRAS	RAF1
AKT2	CHUK	EP300	INS	PDGFA	RB1
AKT3	CREB1	ERBB2	INSRR	PDGFB	RELA
AR	CREB3	FGFR1	KLK3	PDGFC	SOS1
ARAF	CREB3L1	FGFR2	KRAS	PDGFD	SOS2
ATF4	CREB3L2	FOXO1	LEF1	PDGFRA	SRD5A2
BAD	CREB3L3	GRB2	MAP2K2	PDGFRB	TCF7
BCL2	CREB3L4	GSK3B	MAPK1	PDPK1	TCF7L1
BRAF	CREB5	GSTP1	MAPK3	PIK3CA	TCF7L2
CASP9	CREBBP	HSP90AA1	MDM2	PIK3CB	TGFA
CCND1	CTNNB1	HSP90AB1	MTOR	PIK3CD	TP53
CCNE1	E2F1	HSP90B1	NFKB1	PIK3R1	
CCNE2	E2F2	IGF1	NFKBIA	PIK3R2	
CDK2	E2F3	IGF1R	NKX3-1	PIK3R3	
CDKN1A	EGF	IKBKB	NRAS	PTEN	

Elevated levels of prostate specific antigen (PSA) encoded by Kallikrein-related peptidase 3 (KLK3) can be an indicator of the prostate cancer, as well as benign prostatic hyperplasia. It is almost only produced in prostate tissue and normally present in blood in very low amount. Although an increase of PSA levels can be a sign for Pca, low levels of PSA in Pca patients and elevated PSA levels in non-Pca cases are possible, causing false negatives and positives. Interaction partners of KLK3 were gathered from various databases and selected as the second group of core set (Table 4.2).

Table 4.2. KLK3 and its protein interaction partners from various databases.

HPRD	BIOGRID	IntAct	InnateDB	STRING
A2M	A2M	A2M	SERPINA5	A2M
ALB	MIF4GD	ALB	AR	AR
EPOR	SERPINA1	AR	SERPINA1	FBN3
FN1	SERPINA5	FN1	ALB	FN1
IGFBP3	UBA52	IGHG1	A2M	IGKV3D-7
PLG		KNG1	FN1	MIF4GD
PTHLH		SERPINA3	SERPINA3	NGF
PZP				PDGFRB
SEMG1				PRG4
SEMG2				SERPINA1
SERPINA1				SERPINA3
SERPINA3				SERPINA5
SERPINA5				TENM1
SLPI				TENM2
				TENM3

For prostate tissue, k-fold values of 17724 genes were calculated with respect to normal and cancer cell line gene expression data from MeRAV. 14 genes are found to be upregulated 23 genes are downregulated. These 37 genes are taken as the KF dataset (Table 4.3).

Table 4.3. Proteins having k-fold>2 (absolute).

ACPP	COL15A1	NUF2
ACTG2	COL3A1	OGN
ANLN	DEPDC1B	OLFM4
AZGP1	DLGAP5	PBK
C11orf96	HLA-DPB1	RRM2
C8orf4	IGFBP7	SERPINB11
CCNB1	JCHAIN	SLC14A1
CDC20	KIF11	SMOC2
CDC42EP5	KIF23	SPARCL1
CDKN3	LUM	TGM4
CENPK	MGP	TOP2A
CEP55	MSMB	
CNN1	NEFH	

The next step is to find the Gene Ontology terms associated with each term. Mutual GO terms of aspects Molecular Function, Biological Process and Cellular Component were found between KEGG, KLK and KF terms. A gene was added to the core list if it is related to at least one GO term from each aspect. The resulting core set with 113 proteins is shown in Table 4.4.

A Venn diagram is a diagram showing the relationship between several sets. The Venn Diagrams between KEGG, KLK and KF terms for Molecular Function, Biological Process and Cellular Component were drawn in Matlab and shown in Figures 4.3-4.5. The sizes of the circles are proportional to the number shown inside. The following figures show that 663 biological process, 132 cellular component and 73 molecular function GO terms are mutual in all groups of proteins.

The interactors of core proteins were checked if they share at least one term from each aspect with the mutual GO list. If they do, they were included in the network and if they don't they were eliminated. Then, the interactors of previous step were investigated, and this process was repeated there is no protein left.

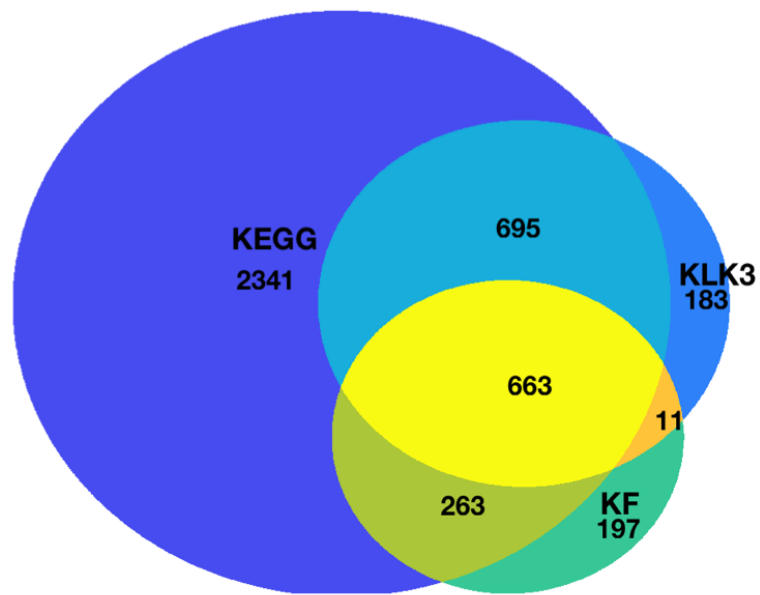


Figure 4.3. GO Biological Process Venn Diagram.

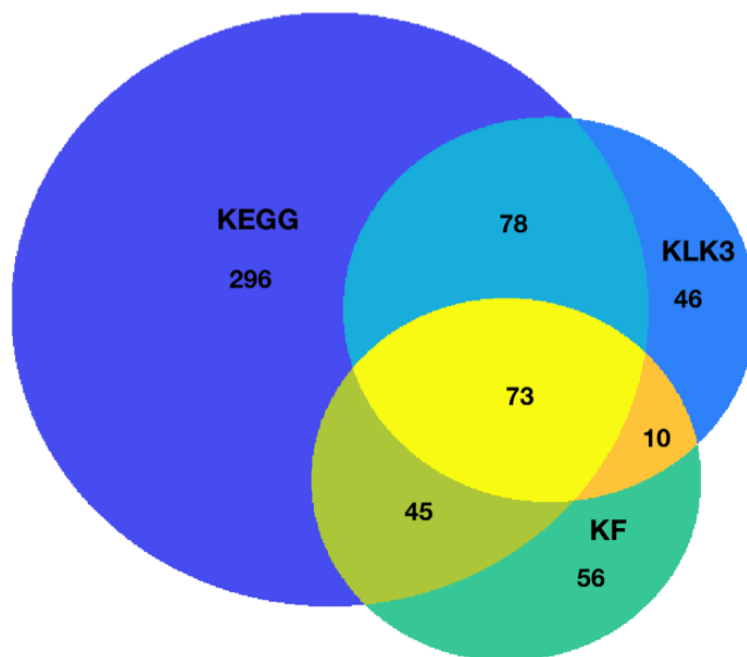


Figure 4.4. GO Molecular Function Venn Diagram.

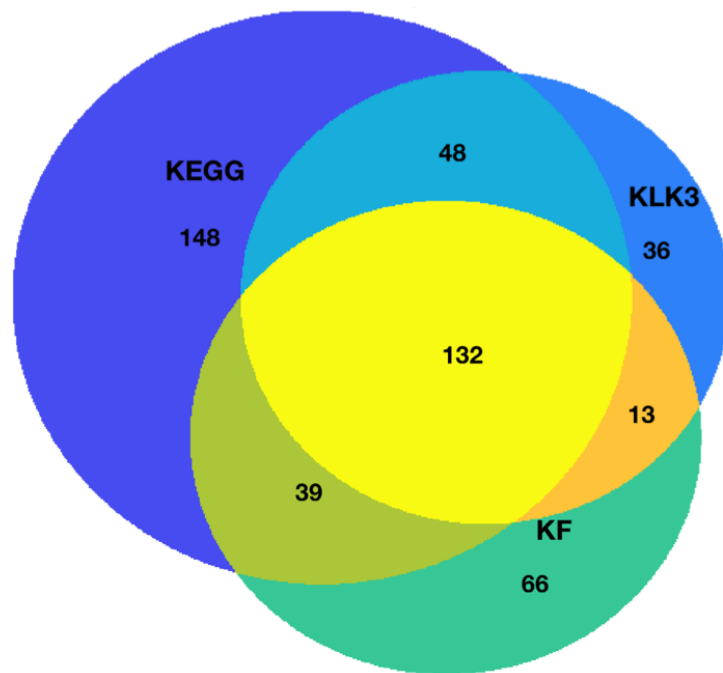


Figure 4.5. GO Cellular Component Venn Diagram.

Table 4.4. Core protein set.

A2M	CCNE2	DLGAP5	IGF1R	NFKB1	PIK3R2
ACPP	CDC20	E2F1	IGFBP7	NFKBIA	PTEN
ACTG2	CDC42EP5	E2F2	IKBKB	NKX3-1	RAF1
AKT1	CDK2	E2F3	IKBKG	NRAS	RB1
AKT2	CDKN1A	EGF	INS	NUF2	RELA
AKT3	CDKN1B	EGFR	JCHAIN	OGN	RRM2
ALB	CDKN3	EP300	KIF11	OLFM4	SERPINA1
AR	CEP55	ERBB2	KLK3	PBK	SERPINA3
ARAF	CHUK	FGFR1	KRAS	PDGFA	SERPINA5
ATF4	CNN1	FGFR2	LEF1	PDGFB	SOS1
AZGP1	COL3A1	FN1	LUM	PDGFC	SOS2
BAD	CREB1	FOXO1	MAP2K2	PDGFD	SPARCL1
BCL2	CREB3	GRB2	MAPK1	PDGFRA	TCF7
BRAF	CREB3L1	GSK3B	MAPK3	PDGFRB	TCF7L1
C8orf4	CREB3L3	GSTP1	MDM2	PDPK1	TCF7L2
CASP9	CREB5	HSP90AA1	MGP	PIK3CA	TGFA
CCNB1	CREBBP	HSP90AB1	MSMB	PIK3CB	TOP2A
CCND1	CTNNB1	HSP90B1	MTOR	PIK3CD	TP53
CCNE1	DEPDC1B	IGF1	NEFH	PIK3R1	

Table 4.5. Number of proteins and interactions between them for each repetition of the process.

	Number of Proteins	Number of Interactions
Core Proteins	113	595
Core Proteins and their 1. neighbors	4751	14877
Core Proteins and their 1. and 2. neighbors	8594	155747
Core Proteins and their 1., 2. and 3. neighbors	8707	163487
Core Proteins and their 1., 2., 3. and 4. neighbors	8709	163507
Core Proteins and their 1., 2., 3., 4. and 5. neighbors	8709	163508

After fifth repetition, no more proteins or interactions could be added to the system and the final PPI network for the prostate cancer was obtained. After the self-loops were eliminated, the PPI network for the prostate cancer with all proteins includes 160666 interactions between 8709 proteins including 113 core proteins (Figure 4.6).

To check whether the constructed network is a valid one, literature was screened. One study by Kar *et al.* constructed PIN and cPIN networks, which are human protein-protein interaction and cancer related subnetwork of human PPI, respectively. The number of proteins in cPIN and this work's PPI network are nearly the same, while the number of interactions is greater in the latter one (Figure 4.7). The compared gathered PPI data from DIP [164] and MIPS [165] databases, whereas four different databases were used in this study. This may be the cause the number of interactions difference between two networks.

4.2. Topological Analysis of Pca PPI Network

For considerably large protein networks, as in Figure 4.6, identifying and interpreting the structural properties of networks are at high importance. Graph theory, a mathematical tool, is introduced to biological networks for analysis of topology. The

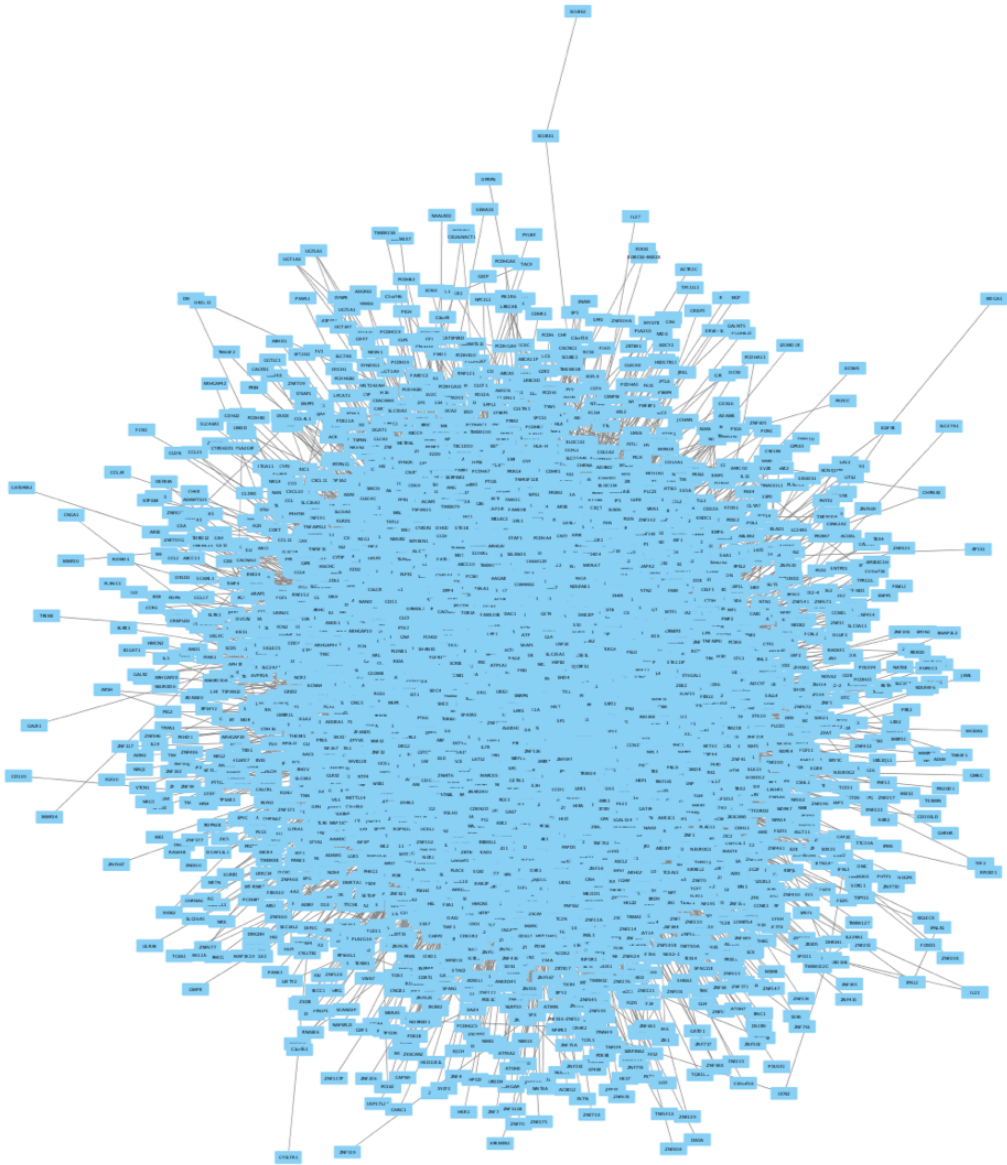


Figure 4.6. PPI network of prostate cancer.

Network name	Protein	Interaction	Known complex in PDB
PIN	13584	85083	206
cPIN	8990	27413	149

Figure 4.7. Number of proteins and interactions in various networks [166].

topological properties of Pca PPIN were estimated by graph theoretical approach and given in Table 4.6 and descriptions of some graph theory concepts are in Table 4.7.

Table 4.6. Topological properties of PPIN of prostate cancer.

Clustering Coefficient	0.124	Network Density	0.004
Network Diameter	8	Characteristic Path Length	2.885
Average Number of Neighbors	36.897	Network Heterogeneity	1.622

Table 4.7. Definitions of graph theory concepts.

Degree (Connectivity)	Number of connections or edges of a node.
Density	A measure showing how dense a network is populated with edges.
Characteristic Path Length	Average shortest path length, in other words the average of the minimal length between two nodes in which nodes are only visited once.
Network Diameter	Maximum value of the shortest path between two nodes
Network Heterogeneity	Tendency of a network to contain hub nodes
Clustering Coefficient	Tendency of a graph to be divided into clusters

For biological networks, the average path lengths and diameters are found to be relatively small although their sizes are generally quite large. This property of metabolic, genetic and protein interaction networks is called small world property [167]. The characteristic path lengths indicate the speed of communication flow in the networks, and a small value shows that only a small number of intermediate proteins are necessary in order to influence another [154]. Characteristic path lengths and diameters of a yeast PPI network are exemplified in Figure 4.8. The values for the essential proteins in the network are consistent with this work's findings.

	Average degree (K)	Clustering coefficient (C)	Characteristic path length (L)	Diameter (D)
Essential	18.7	0.182	3.84	10
Synthetic lethal	9.2	0.083	4.24	10
Non-essential	7.4	0.095	4.49	11
<i>P</i> -value	$<10^{-12}$	$<10^{-12}$	$<10^{-12}$	—

Figure 4.8. Characteristics of yeast PPI network [168].

The statistics of the prostate cancer PPIN shown in Table 4.6 were compared with the literature values presented in Figure 4.9. The clustering coefficient and characteristic path lengths were found to be coherent with the literature as in previous comparison. A low clustering coefficient such as 0.124 indicates that the network is not likely to cluster together.

Basic characteristics of a number of information/communication, biological and social networks from the real-world

Network	N	$\langle k \rangle$	L	C	γ	ν	Ref.
AS2001	11,174	4.19	3.62	0.24	2.38	< 0	[8,25,80]
Routers	228, 263	2.80	9.5	0.03	2.18	> 0	[8,25,80]
Gnutella	709	3.6	4.3	0.014	2.19	< 0	[100]
WWW	$\sim 2 \times 10^8$	7.5	16	0.11	2.1/2.7	Unknown	[101]
Protein	2,115	6.80	2.12	0.07	2.4	< 0	[86]
Metabolic	778	3.2	7.40	0.7	2.2/2.1	< 0	[85]
Math1999	57, 516	5.00	8.46	0.15	2.47	> 0	[81,82]
Actors	225,226	61	3.65	0.79	2.3	> 0	[28,83]
e-mail	59,812	2.88	4.95	0.03	1.5/2.0	Unknown	[84]

The quantities measured are: number of vertices N , characteristic path length L , clustering coefficient C , average degree $\langle k \rangle$, exponent of the degree distribution γ , and type of correlations. All networks, except the WWW, metabolic (*Escherichia coli*) and e-mail networks, are undirected. The two values of γ represent, respectively, the in/out-degree exponents when the network is directed.

Figure 4.9. Characteristics of various networks [155].

Biological networks tend to be scale-free networks, that is the node degree distribution follows the power law as in the case for many networks such as the Internet, the language or the web of Hollywood actors. Power law for scale free networks is given as

$$p_k = Ck^{-\gamma} \quad (4.1)$$

where p_k is the probability of a node with exactly k links and γ is the degree exponent [169]. The following power law was fitted to the degree of nodes vs number of nodes with an R^2 value of 0.876.

$$p_k = 7059.5k^{-1.46}$$

Since the R^2 value can be considered high, power law is a good fit showing that the constructed prostate cancer PPIN is a scale-free network as expected. The exponent of degree distribution was found 1.46, which is less than nearly all network types given in Figure 4.9.

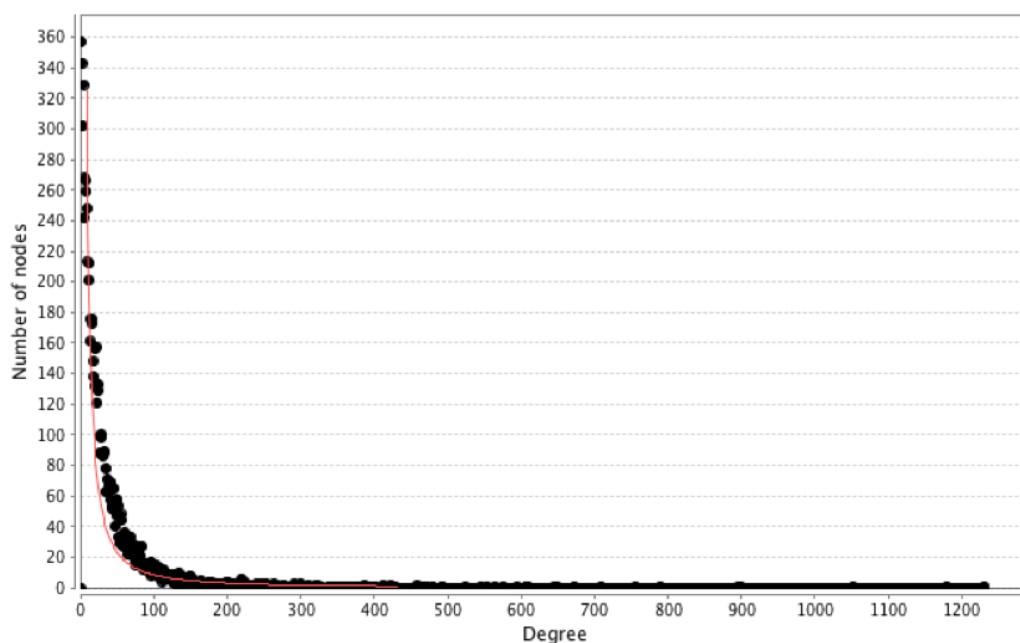


Figure 4.10. Degree vs number of nodes with fitted line.

In Figure 4.12, degree distributions of various species such as yeast, worm and fly are shown. Although the vertical axis is number of nodes in Figure 4.11 and degree distributions in Figure 4.11, since the degree distribution is proportional to number of nodes, they all show similar trends.

PPI networks provide a systematic point of view and they may uncover unidentified drug targets as the main targets of drugs are proteins of different kinds. Especially, the topological features can be invaluable for prediction of potential drug targets since their topological features can be different from normal proteins.

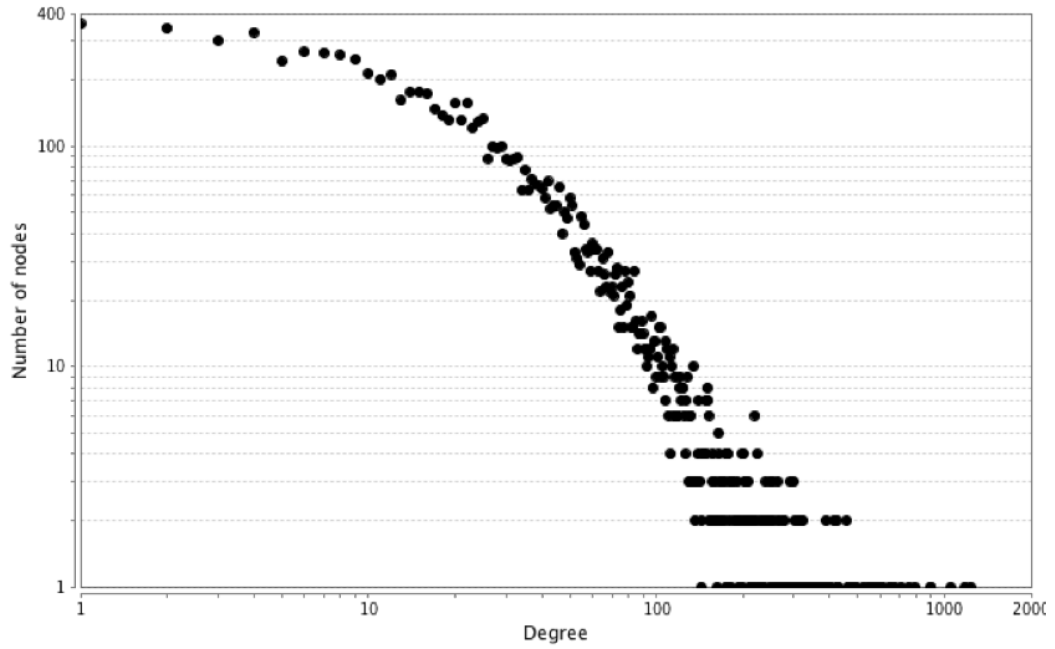
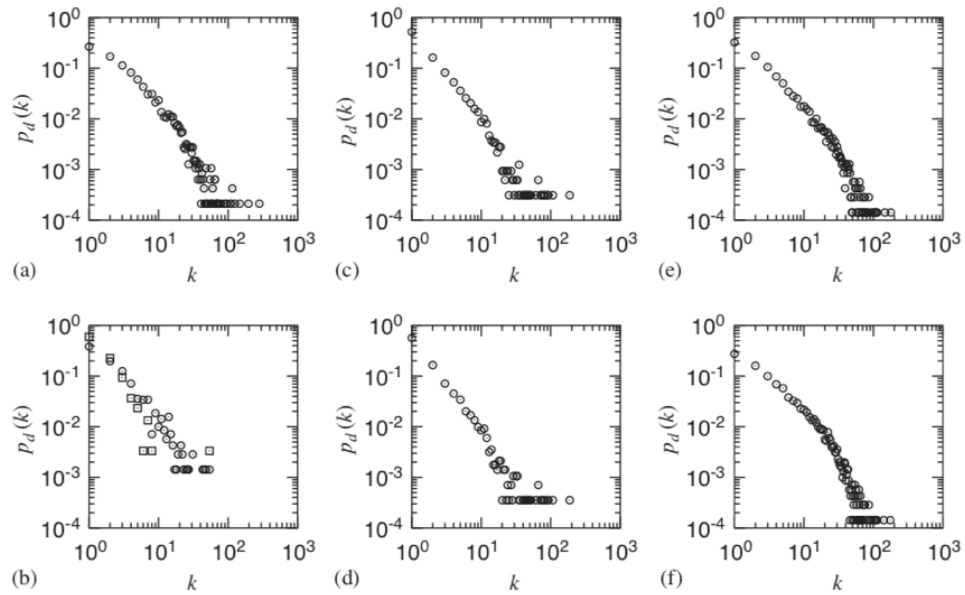


Figure 4.11. Degree distribution of nodes on logarithmic scale.



The degree distributions $p_d(k)$ for (a) the yeast, (b) the prokaryotes *Helicobacter pylori* ($\circ$) and *Escherichia coli* ($\square$), (c) the worm (Worm-All), (d) the Y2H subset of the worm dataset (Worm-Y2H), (e) the fly, and (f) the Fly + Interolog dataset.

Figure 4.12. Degree distributions for various species [170].

Table 4.8. Average topological properties of network, core set and disease proteins.

	Network Average	Core Set Protein Average	Disease (Prostate) Proteins Average
Degree	36.9	135.6	63.45
Betweenness Centrality	8209	70489	16970
Normalized Betweenness Centrality	0.003	0.028	0.007
Eigenvector Centrality	0.039	0.134	0.068
Average Shortest Path Length	2.89	2.61	2.74
Eccentricity	5.5	5.1	5.29
Coreness	18.9	30.8	25.6

Table 4.9. Minimum/maximum topological properties of network, core set and disease proteins.

	Network Min-Max	Core Set Protein Min-Max	Disease Proteins Min-Max
Degree	1 - 1231	1 - 888	1 - 493
Betweenness Centrality	0 - 2507453	0 - 1035840	0 - 351284
Normalized Betweenness Centrality	0 - 1	0 - 0.41	0 - 0.14
Eigenvector Centrality	$5.9 \times 10^{-5} - 1$	$2.5 \times 10^{-5} - 0.83$	$6.4 \times 10^{-5} - 0.43$
Average Shortest Path Length	2.0 - 5.4	2.0 - 4.1	2.2 - 4.5
Eccentricity	4 - 8	4 - 6	5 - 7
Coreness	1 - 45	1 - 45	1 - 45

Drug targets may be intermediaries through which the drug affects the proteins in the PPI [171]. Degree and betweenness centrality of proteins are two characteristics of intermediary function. Figure 4.13 shows the degree distributions of network, core and disease proteins on normal and logarithmic scale. The distributions follow power law for all groups of proteins. As it can be seen in Figure 4.13, Tables 4.8 and 4.9, although the highest degree proteins are network proteins, the degree average of disease proteins are two-folds the network average degree. This result implies that drug targets are not hub proteins, they rather take part in specific functions by interacting with a certain number of proteins. Previous findings also confirm that disease proteins have higher degrees than other proteins [172].

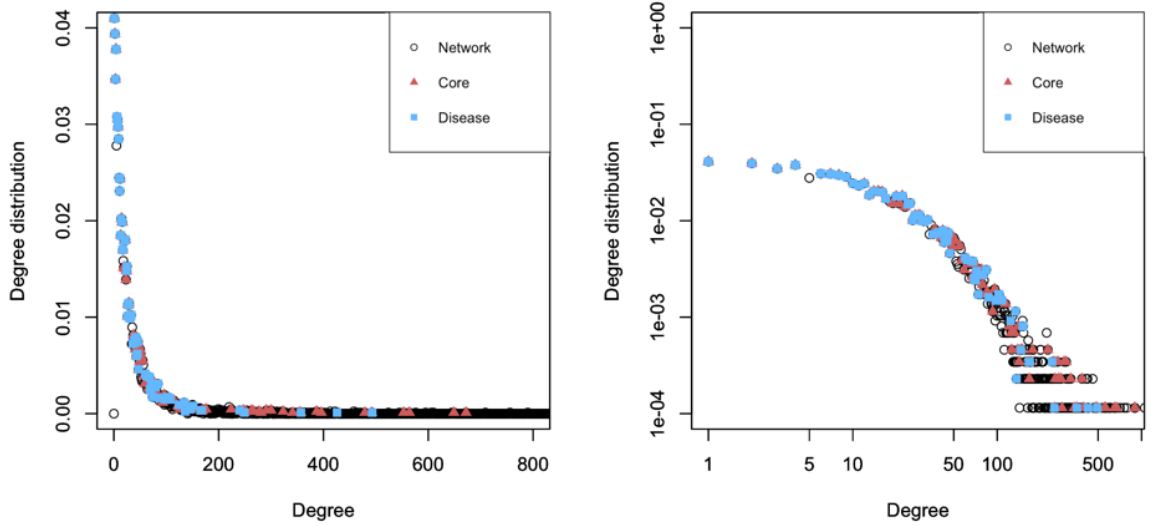


Figure 4.13. Degree vs. degree distribution of PPI, core and disease proteins.

Betweenness centrality is another feature representing the intermediary function of proteins. Normalized between can be also used for comparison since the values can be large. It is calculated as follows where $NB(v)$ and $B(v)$ is the normalized betweenness of node v , respectively.

$$NB(v) = \frac{B(v) - \min(B)}{\max(B) - \min(B)} \quad (4.2)$$

Distributions of normalized betweenness for protein sets (Figure 4.14) also similar to each other. Again, maximum normalized betweenness values belong to network proteins but average of disease proteins is higher. This is a sign that the drug targets are probably not bottlenecks.

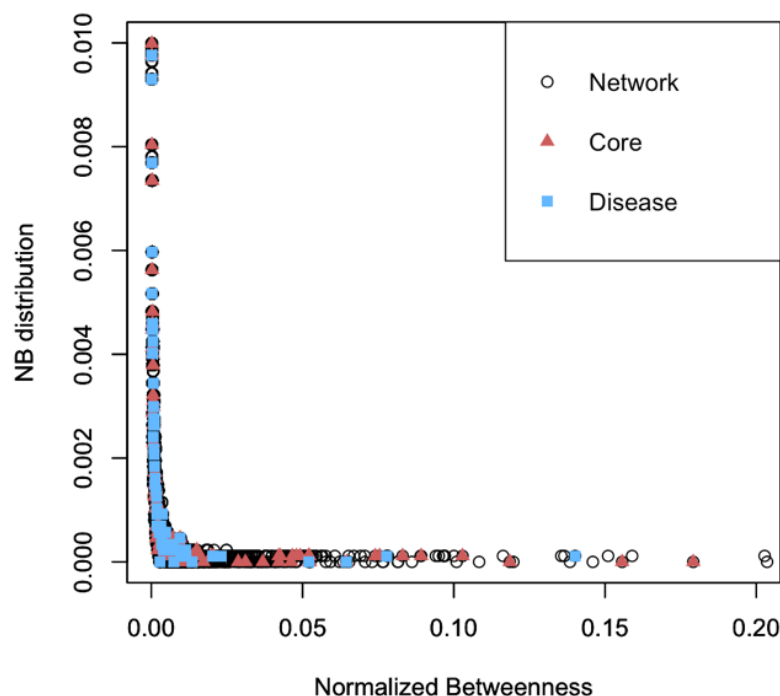


Figure 4.14. Normalized edge betweenness vs. its distribution of PPI, core and disease proteins.

Eigenvector centrality measures the importance of a node by assessing its connections with high degree. The eigenvector distributions (Figure 4.15) follow the same trend as degree and normalized betweenness: distributions of network, core and disease proteins are similar, highest eigenvector centrality is seen in network proteins and average of disease proteins are greater than network proteins. The last point indicates that although drug targets are neither hub nor bottleneck proteins, they interact with higher degree proteins.

Drug targets may also be the sources which receive one stimulus and convert to another. Being adjacent to other proteins may be convenient for transmitting the signal from one protein to another. So, average distance and eccentricity of a protein may be significant.

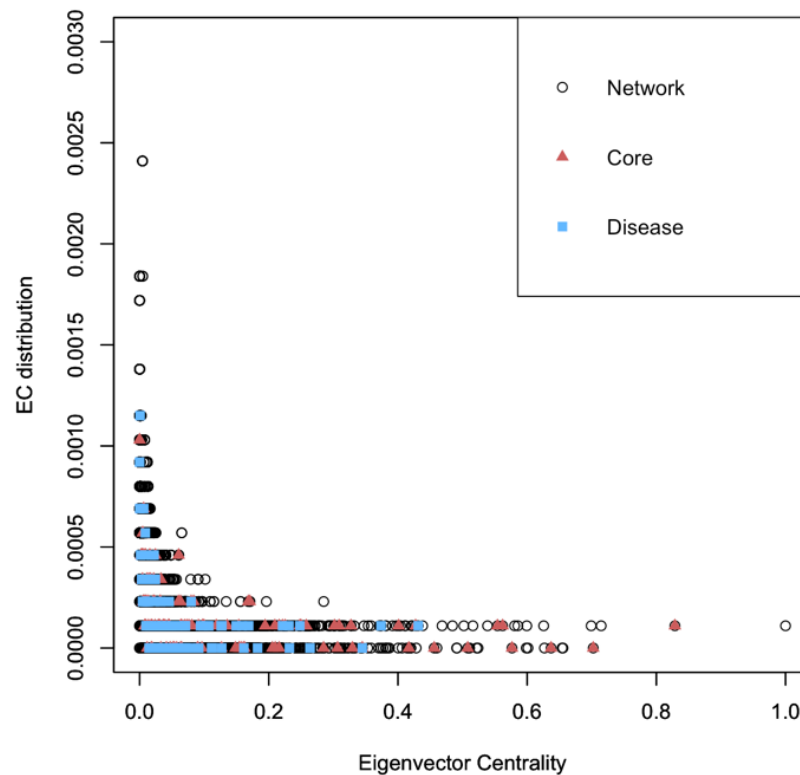


Figure 4.15. Eigenvector centrality vs. its distribution of PPI, core and disease proteins.

Protein with a smaller average distance indicates that it is more convenient for that protein to communicate with other proteins. Mean of average distances of disease proteins are less than network proteins (Figure 4.16 and Table 4.8), suggesting special proteins have lower average distances. Other studies also identified that cancer related proteins have lower average distances with respect to other proteins [167].

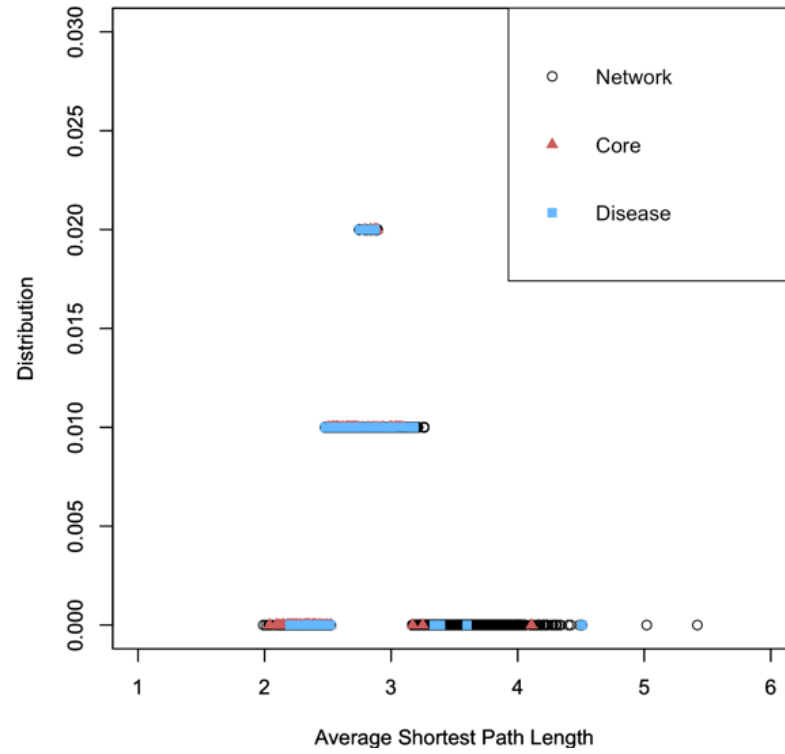


Figure 4.16. Average shortest path length vs. its distribution of PPI, core and disease proteins.

Eccentricity is the measure of the distance between a protein and another protein farthest from it. As in average distance, a low value indicates that the protein is easily reachable for other proteins and it can influence more proteins. The average of eccentricity for disease proteins is lower than network proteins, but there is no significant difference (Figure 4.17).

Coreness is another parameter that detects the significance of a protein by defining the structure of the PPI network. A k -core is the subgraph of the network with all edges having at least k degree. Coreness of a protein is the maximum reached k -core, so that it does not present in the $(k+1)$ th core. No significant differences are detected between network and disease proteins (Figure 4.18).

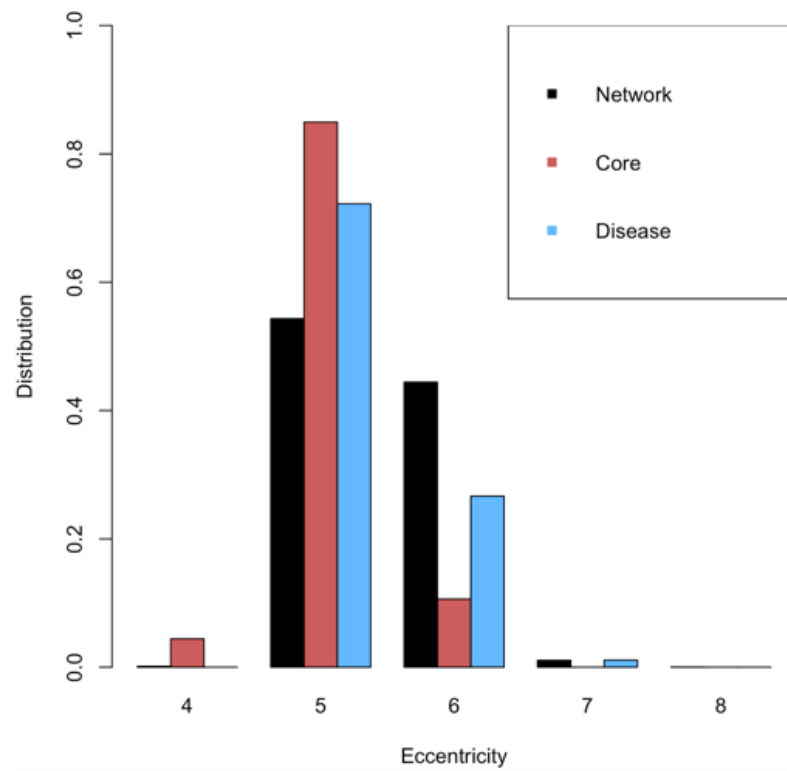


Figure 4.17. Eccentricity vs. its distribution of PPI, core and disease proteins.

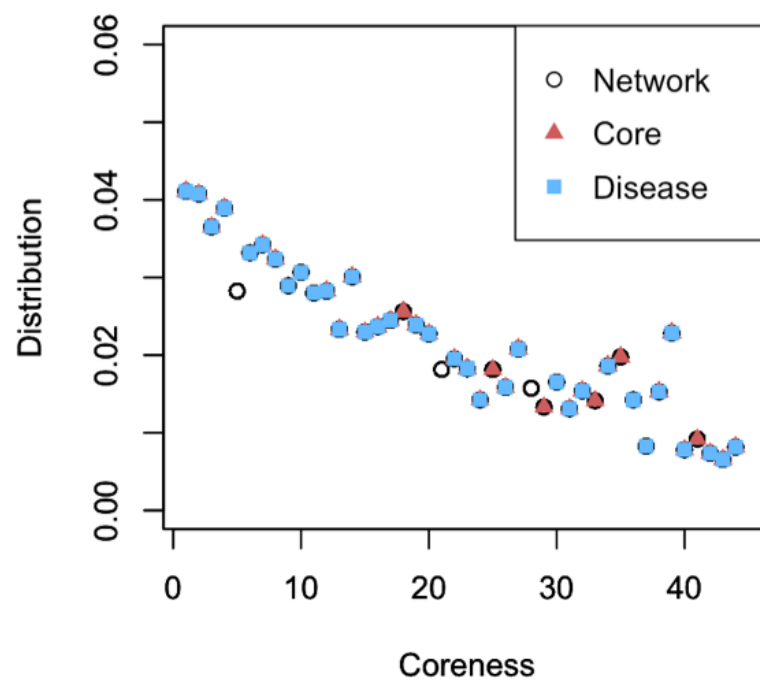


Figure 4.18. Coreness vs. its distribution of PPI, core and disease proteins.

4.3. Hub and Bottleneck Proteins

Hubs and bottlenecks are also important points in biological networks. Nodes with higher degrees, meaning they connect to many neighbors, are called hubs whereas bottlenecks, proteins with the high betweenness centrality, lie on many shortest paths between other nodes and control the information flow in the network. Hubs and bottlenecks are essential proteins in a network, removing them has a great effect on the network connectivity [173]. In Table 4.10, hub and bottleneck proteins in the prostate cancer PPIN are tabulated. Bottleneck proteins tend to be hub proteins as shown in Table 4.10 and Table 4.11.

Table 4.10. Top 20 Degree and Betweenness Centrality results.

Rank	name	Degree	Rank	name	Betweenness Centrality
1	NTRK1	1231	1	APP	0.0661
2	APP	1180	2	ELAVL1	0.0492
3	ELAVL1	1052	3	NTRK1	0.0385
4	EGFR	899	4	EGFR	0.0273
5	TP53	896	5	TP53	0.0252
6	XPO1	790	6	XPO1	0.0193
7	UBC	756	7	UBC	0.0187
8	CUL3	709	8	GRB2	0.0172
9	HSP90AA1	672	9	MOV10	0.0150
10	GRB2	649	10	HSP90AA1	0.0145
11	MYC	645	11	CUL3	0.0135
12	ESR1	610	12	MYC	0.0134
13	MCM2	606	13	FN1	0.0119
14	JUN	595	14	MEOX2	0.0105
15	MOV10	576	15	HSP90AB1	0.0103
16	HSP90AB1	564	16	MCM2	0.0100
17	FN1	553	17	JUN	0.0097
18	COPS5	545	18	CDH1	0.0093
19	YWHAZ	524	19	COPS5	0.0092
20	CDH1	493	20	ESR1	0.0090

Table 4.11. Degree and Betweenness Centrality ranks of top proteins.

	Degree Rank	Betweenness Centrality Rank
NTRK1	1	3
APP	2	1
ELAVL1	3	2
EGFR	4	4
TP53	5	5
XPO1	6	6
UBC	7	7
CUL3	8	11
HSP90AA1	9	10
GRB2	10	8
MYC	11	12
ESR1	12	20
MCM2	13	16
JUN	14	17
MOV10	15	9
HSP90AB1	16	15
FN1	17	13
COPS5	18	19
YWHAZ	19	-
CDH1	20	18
MEOX2	-	14

DrugBank is an online database containing drugs, chemical and pharmaceutical data and drug targets. Hub and bottleneck proteins were searched in Drugbank as targets and drugs associated with each protein are found. Only approved and investigational drugs with pharmaceutical action were taken into account and tabulated in Table 4.3. Four out of five top proteins are related to cancer, and two of them are used as a target for various cancers such as colorectal and lung cancer. Investigational drugs are marked as (I).

Rest of the analysis regarding hubs and bottlenecks are enrichment analysis given in Section 4.6.

Table 4.12. Drugs and their indications for hub and bottleneck proteins.

Protein	Protein Name	Drug	Indications
NTRK1	High affinity nerve growth factor	Regorafenib	Gastrointestinal Stromal Tumor, Metastatic Colorectal Cancer
APP	Amyloid beta A4 protein	Florbetaben, Flutemetamol, Florbetapir	Alzheimer's Disease, Cognitive Decline
ELAVL1	ELAV like protein		
EGFR	Epidermal growth factor receptor	Cetuximab	Metastatic Colorectal Cancer, Cell Carcinoma of Head and Neck
		Getifinib, Necitumumab, Afatinib, Osimertinib, Olmutinib, Brigatinib	Non-small Cell Lung Cancer
		Erlotinib	Non-small Cell Lung Cancer, Pancreatic Cancer
		Lapatinib, Neratinib	Breast Cancer

Table 4.12. Drugs and their indications for hub and bottleneck proteins (cont.)

Protein	Protein Name	Drug	Indications
		Panitumumab	Metastatic Colorectal Cancer
TP53	Cellular tumor antigen p53		
XPO1	Exportin-1		
UBC	Polyubiquitin-C		
CUL3	Cullin 3		
HSP90AA1	Heat shock protein HSP 90 alpha	Alvespimycin (I)	Acute Myelocytic Leukemia
GRB2	Growth factor receptor-bound protein 2		
MYC	Myc proto- oncogene protein		
ESR1	Estrogen receptor	>20 drugs	
MCM2	DNA replication licensing factor		
JUN	Transcription factor AP-1	Arsenic trioxide	Acute Promyelocytic Leukemia
MOV10	Putative helicase		
HSP90AB1	Heat shock protein HSP 90 beta		
FN1	Fibronectin	Ocriplasmin	Symptomatic Vitreomacular Adhesion
COPS5	COP9 signalosome complex subunit 5		
YWHAZ	14-3-3 protein zeta/delta		
CDH1	Cadherin 1		
MEOX2	Homebox protein MOX-2		

4.4. Module Detection and Analysis

Analyzing large-scale PPI networks requires network clustering, a method which constructs subnetworks from a larger network. These subnetworks share certain properties that can be used to identify topological or functional modules. Markov-Clustering Algorithm (MCL) is a flow-based clustering algorithm, which clusters the network by a series of flow expansions. The inflation parameter, the factor influencing the granularity of the clusters, was chosen as 2.5 and the general picture after clustering is given in Figure 4.19.

To have a better look, the clusters consisting of at least 35 proteins are tabulated below in Table C.1. Core proteins were also checked to see if they clustered together in one module. 4 proteins are included in cluster 4.

4.5. Construction of Co-Expression Network

TCGA prostate cancer tumor gene expression data consists of expression data of over 50000 genes from 551 samples and detecting differentially expressed genes (DEGs) is a preprocessing step in order to construct the co-expression network. The threshold for DEGs were determined as $FDR < 0.05$ and $FC > 3$.

16962 genes were found to be differentially expressed with 8654 up- and 8308 downregulated genes. To check whether choosing FC cut-off as 3 was rational, core proteins used for constructing the PPI network were screened. 91 genes out of 113 are found to be differentially expressed as 50 and 41 of them are up- and downregulated, respectively. Then, 136 genes were eliminated from the dataset as they miss too many values or have zero variance. 551 tumor samples with remaining 16824 genes were hierarchically clustered to detect obvious outliers.

As can be seen in Figure 4.20, a large portion of the samples are clustered together. By choosing the height cut-off value as 48000 shown with the red line, 22 samples were

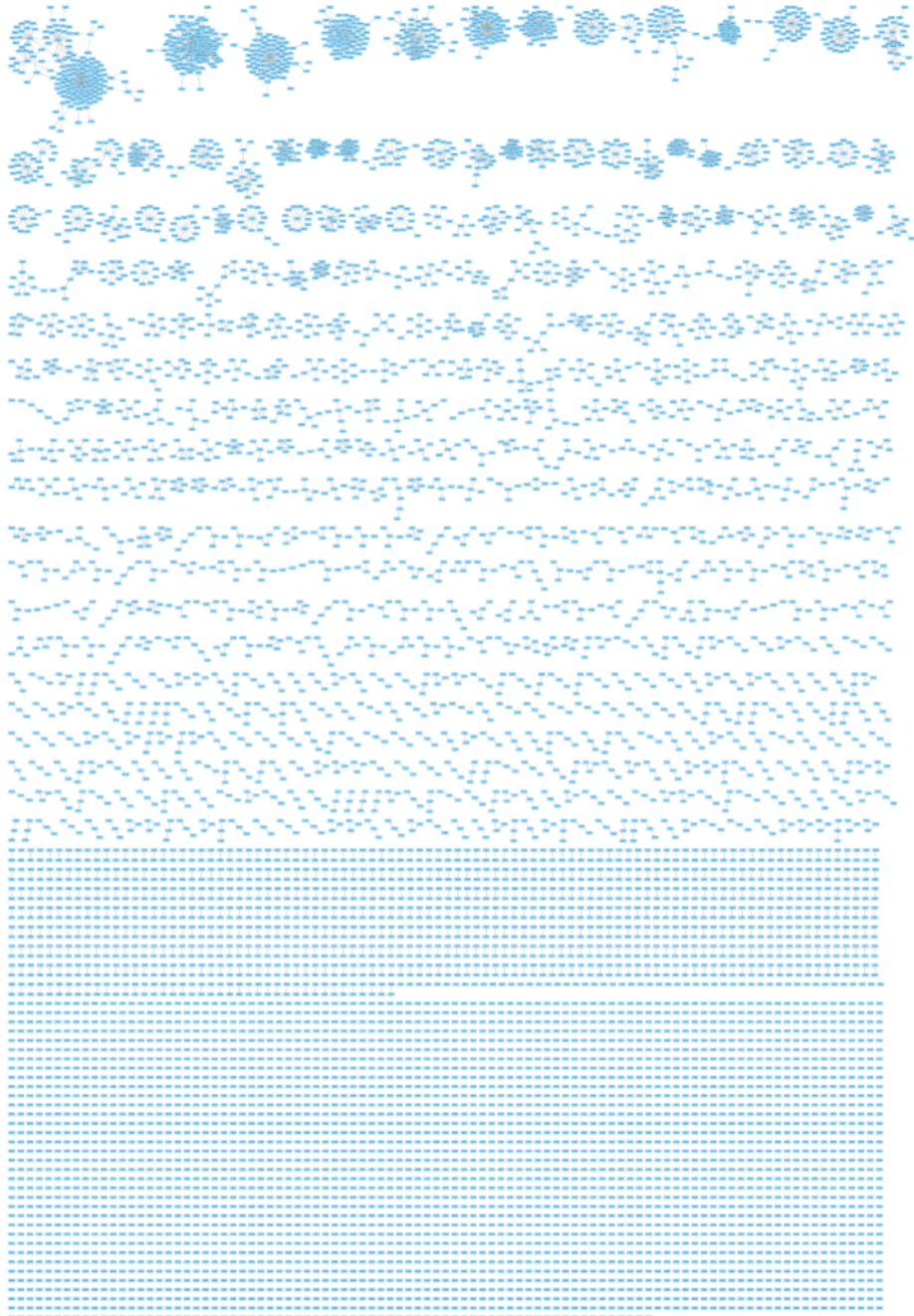


Figure 4.19. Clustering of the Pca PPIN by MCL.

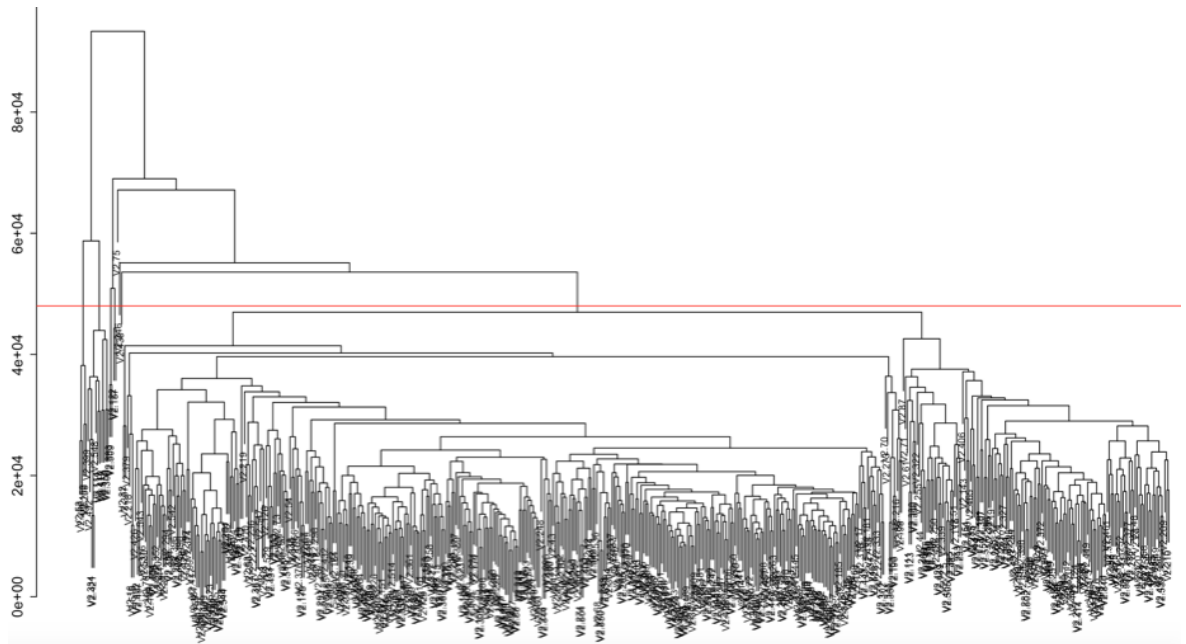


Figure 4.20. Clustering of tumor samples.

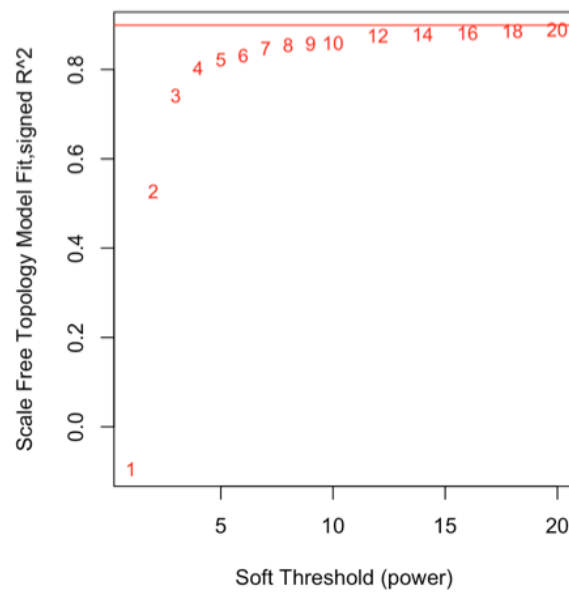


Figure 4.21. Soft threshold vs. model fit.

removed from the data. Figure 4.21 shows the soft threshold vs R^2 for the dataset of this analysis. Soft threshold was chosen as 7, since it is the lowest value where the curve flattens and for which the scale-free topology holds.

While the co-expressions of genes were specified by the adjacency matrix, modules were identified using TOM, which describes the relation of a pair of genes with every other genes. As in adjacency matrix, a high value between two genes indicates that they are connected to the same group of genes. While computing TOM, absolute Pearson correlations were used in order to capture not only the positive correlations, but also the negative ones. This way up- and downregulated genes could be clustered together. Table 4.13 shows the modules constructed with the specifications given in Methods section and the up/downregulated genes in each module.

The following figure shows the eigengene dendrogram of modules. Eigengenes are representations of expression profiles of genes inside modules. Modules on the same branch of this hierarchical clustered dendrogram are called meta-modules. For example, the correlations between green and magenta modules are higher than the one with module skyblue. The genes that cannot be clustered together since their correlations do not comply with any of the modules are put into the grey module. The expectation was that the grey module would be at the outer site of the dendrogram. Surprisingly, the correlations between grey and other large modules such as black and turquoise are higher than other modules such as yellow (Figure 4.22).

Next step is to visualize the modules as nodes and edges. A threshold of 0.03 was given for TOM, so only the genes in the same module having a TOM value greater than 0.03 were connected by an edge. Connecting all genes inside a module with each other would cause $\frac{n^2}{2!}$ edges and it is visually chaotic and also computationally expensive. To include as many genes and also have a simpler graph, several thresholds were tested, and best value is found as 0.03. Black and purple modules are shown in Figure 4.23 and 4.24, respectively.

Table 4.13. Modules and corresponding number of genes.

Module	No of genes	Upregulated Genes	Downregulated Genes
Turquoise	3692	2631 (71.3%)	1061 (28.7%)
Grey	3001	1957 (65.2%)	1044 (34.8%)
Blue	1345	917 (68.2%)	428 (31.8%)
Brown	1148	646 (56.3%)	502 (43.7%)
Yellow	1095	93 (8.5%)	1002 (91.5%)
Green	920	100 (10.9%)	820 (89.1%)
Red	639	163 (25.5%)	476 (74.5%)
Black	557	430 (77.2%)	127 (22.8%)
Pink	497	443 (89.1%)	54 (10.9%)
Magenta	462	81 (17.5%)	381 (82.5%)
Purple	436	137 (31.4%)	299 (68.6%)
Greenyellow	427	27 (6.3%)	400 (93.7%)
Tan	289	30 (10.4%)	259 (89.6%)
Salmon	270	256 (94.8%)	14 (5.2%)
Cyan	263	206 (78.3%)	57 (21.7%)
Midnightblue	252	43 (17.1%)	209 (82.9%)
Lightcyan	182	70 (38.5%)	112 (61.5%)
Grey60	178	56 (31.5%)	122 (68.5%)
Lightgreen	160	19 (11.9%)	141 (88.1%)
Lightyellow	128	108 (84.4%)	20 (15.6%)
Royalblue	126	56 (44.4%)	70 (55.6%)
Darkred	122	3 (2.5%)	119 (97.5%)
Darkgreen	121	10 (8.3%)	111 (91.7%)
Darkturquoise	87	9 (10.3%)	78 (89.7%)
Darkgrey	86	6 (7%)	80 (93%)
Orange	78	70 (89.7%)	8 (10.3%)
Darkorange	72	0 (0%)	72 (100%)
White	67	37 (55.2%)	30 (44.8%)
Skyblue	65	10 (15.4%)	55 (84.6%)
Saddlebrown	59	40 (67.8%)	19 (32.2%)

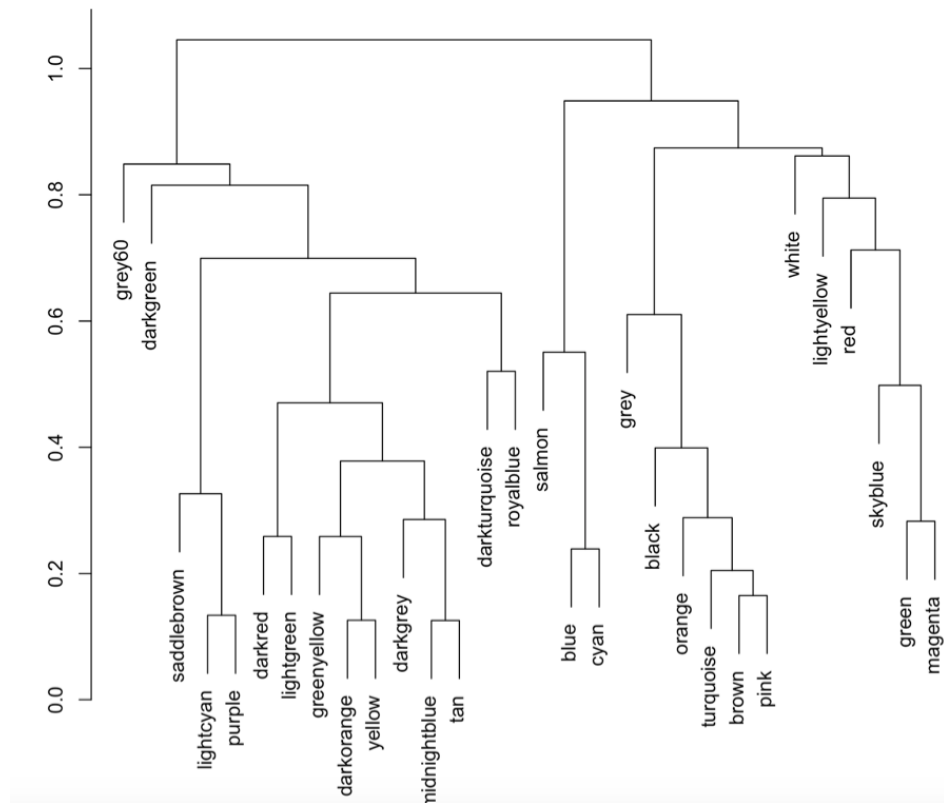


Figure 4.22. Relationship among the modules.

Table 4.14. Number of edges, interactions and genes in black module.

	Module Black
No of edges	411
No of interactions	15640
No of genes in module	557

Table 4.15. Number of edges, interactions and genes in purple module.

	Module Purple
No of edges	404
No of interactions	44527
No of genes in module	436

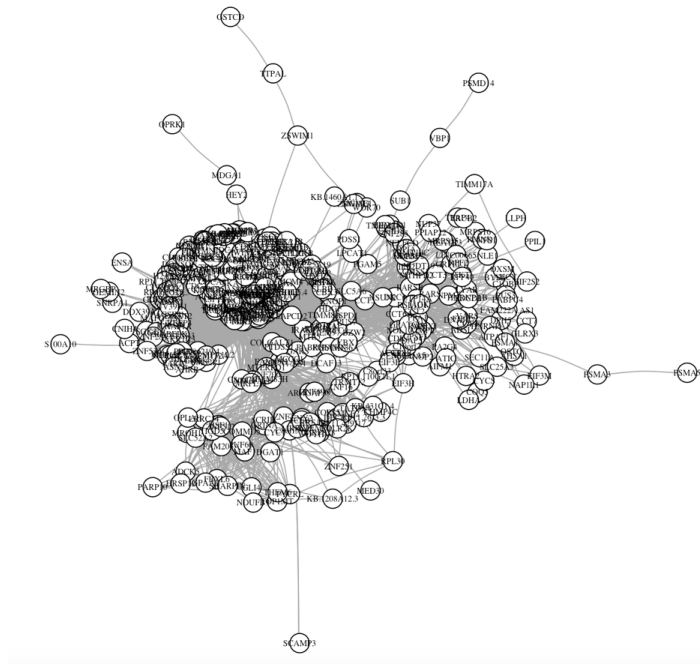


Figure 4.23. Visualized network of black module.

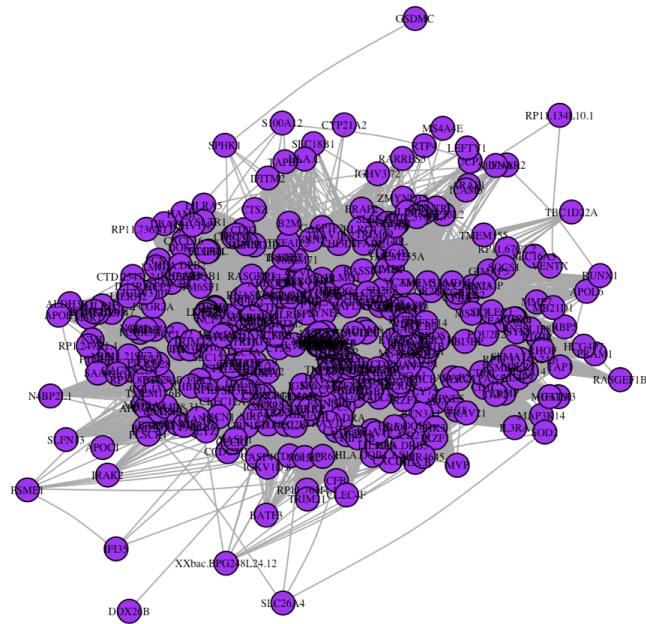


Figure 4.24. Visualized network of purple module.

4.6. Enrichment Analysis

The Functional Enrichment Analyses of several data sets were done by using R package clusterProfiler [156]. Hypergeometric distribution is the calculation method for the enrichment tests for Gene Ontology and KEGG terms. These tests were corrected by Benjamini Hochberg False Discovery Rate (FDR) correction. For all enrichment analysis, p-value and FDR-value thresholds were chosen as 0.01 and 0.05, respectively.

One advantage of clusterProfiler is that it gives the opportunity to calculate the enriched terms for each cluster and to visualize them together. This way, clusters can be compared more easily. In the following sections, the graphs show the clusters vs most enriched GO and KEGG terms. The sizes of the points represent the gene ratio, the ratio of number of genes for that enrichment term over number of genes in the investigated gene list. Also, the enrichment p-values are also represented by colors: red indicates higher enrichment, whereas blue indicates lower.

4.6.1. Gene Ontology Enrichment Analysis

In this section, the enrichment analysis of all three aspects of Gene Ontology were conducted for different data sets.

4.6.1.1. Prostate Cancer PPIN Proteins. Prostate cancer PPIN proteins were found to be highly enriched in GO Biological Process terms such as regulation of serine/threonine kinase activities, response to oxygen levels and regulation of mitotic cell cycle phase transition and apoptotic signaling pathway (Figure 4.25). Serine/threonine kinase activity is a hallmark of a great fraction of cancers and also proliferative diseases [174]. They regulate different pathways in normal and cancer cells. They take part in the development of castration-resistant prostate cancer and may cause bone metastatic disease [175]. Hypoxia, critically reduced oxygen level in tissues, is one of the reasons for resistance to therapy such as chemotherapy and radiation. Low oxygen levels are associated with tumor aggressiveness and metastasis [176]. AR is a required

protein for prostate tumor development as mentioned before and regulates the cell cycle progression. Also, cell cycle pathway has a critical role in modulating androgen response [177].

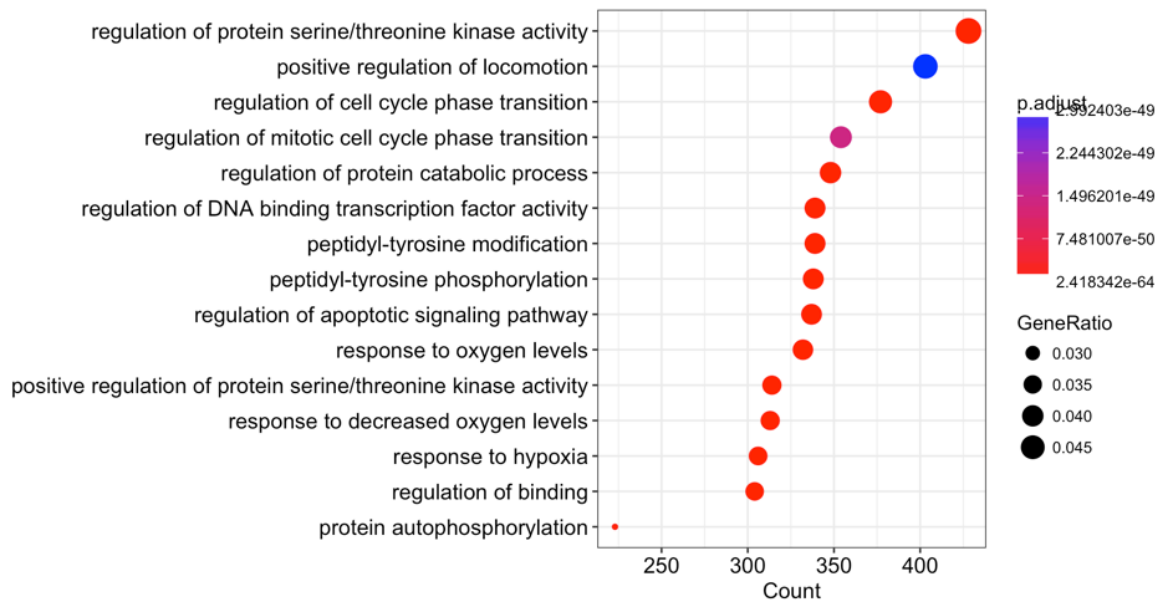


Figure 4.25. Enriched GO Biological Process terms for PPI network proteins.

GO Molecular functions of PPIN proteins involve similar terms as Biological Process. Serine/threonine kinase activities and terms related to transcription such as core promoter proximal region DNA binding are significantly enriched. Transcription is an important process in mitotic cell cycle and it makes sense that many terms regarding promoter binding and coactivator activity are present in Figure 4.26.

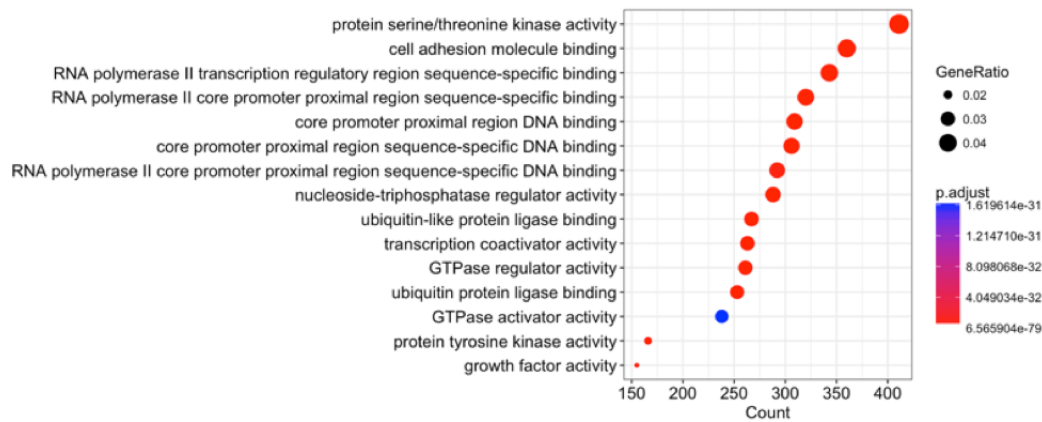


Figure 4.26. Enriched GO Molecular Function terms for PPI network proteins.

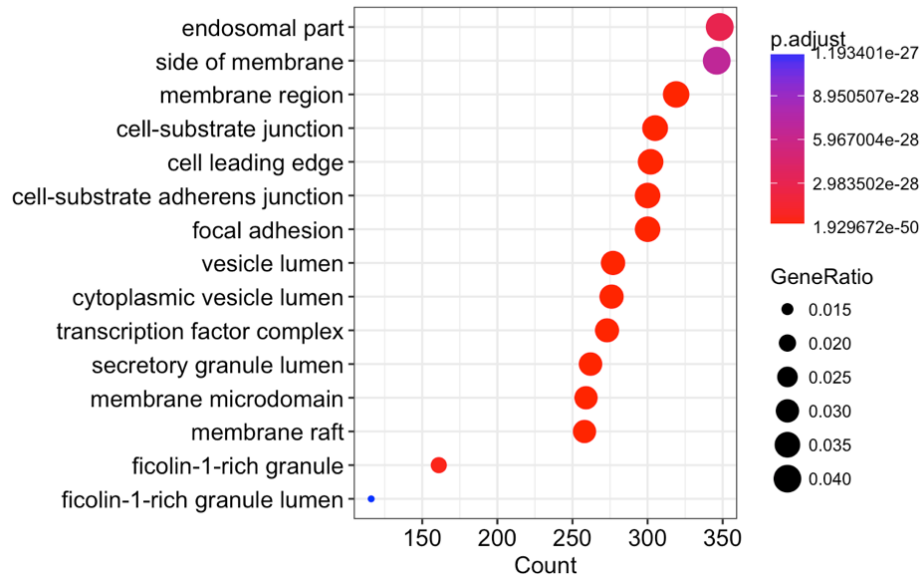


Figure 4.27. Enriched GO Cellular Component terms for PPI network proteins.

4.6.1.2. Prostate Cancer PPI Network Modules. Cluster 2 takes part in RNA related processes such as ribonucleoprotein complex biogenesis (Figure 4.28). In literature, it was reported that several ribonucleoproteins are associated with worse prognosis in many cancer types [178,179]. For instance, heterogenous nuclear ribonucleoprotein L (HNRNPL) is required for prostate cancer growth and genes regulated by HNRNPL are relevant to prostate cancer progression [180]. Among the members from HNRNP family, 3 out of 18 proteins are present in cluster 2. The processes enriched in Cluster 4 and 5 are signaling pathways involved in immune response. Immune system is a mechanism controlling several diseases such as cancer. It is reasonable for a cancer network to contain immune response-regulating terms. Deacetylation and phosphorylation of metabolites are enriched significantly in clusters 7 and 10, respectively. Deacetylation, especially histone deacetylation, is the major driving force of epigenetic gene regulation and takes part in prostate cancer progression [181]. On the other hand, phosphorylation process is a regulator of cell growth and apoptosis, which are important characteristics of cancer and will be mentioned later in Section 4.6.4.1 [182].

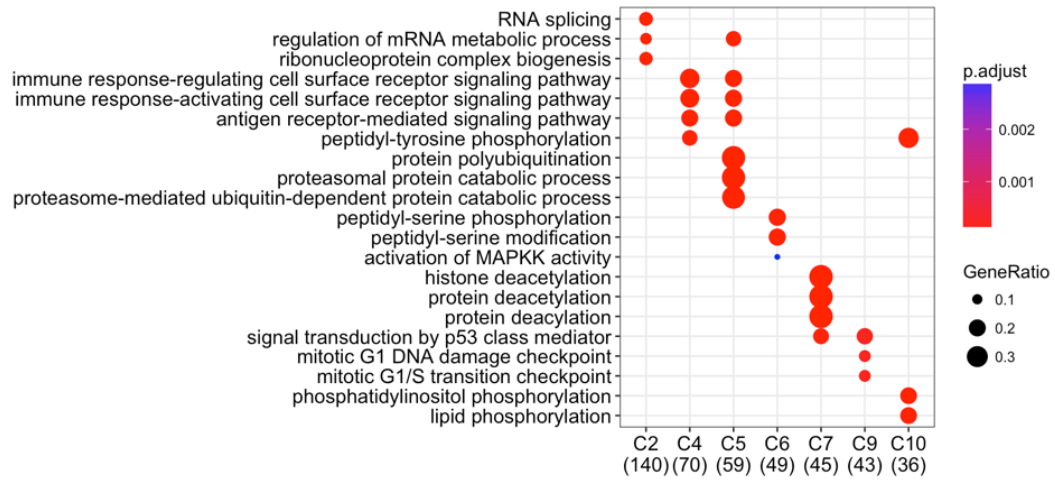


Figure 4.28. Enriched GO Biological Process terms for PPIN clusters.

Cluster 6, 10 and cluster 4, the cluster containing the most number of core proteins, are mostly enriched in kinase activities (Figure 4.29). As mentioned before, phosphorylation is a significant process in cancer progression, and kinases and phosphatases control this process. Also, they have roles in immune responses as they control the signaling mechanisms [183]. Deacetylase terms are again seen in cluster 7.

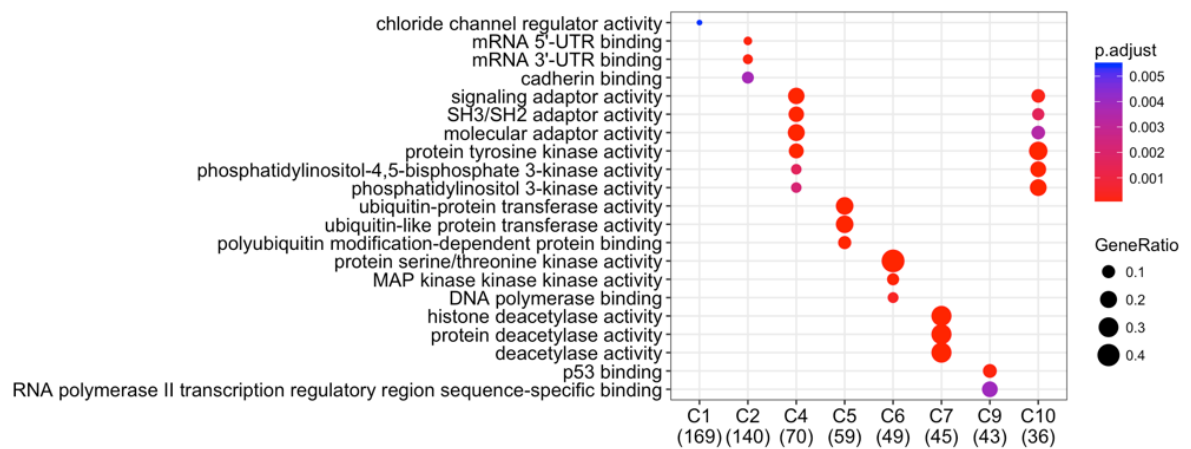


Figure 4.29. Enriched GO Molecular Function terms for PPIN clusters.

In parallel to biological process and molecular function, cellular components of cluster 7 are enriched in histone acetylase complex. Since cluster 4 is enriched in signaling processes, enrichment of plasma membrane proteins is reasonable (Figure 4.30).

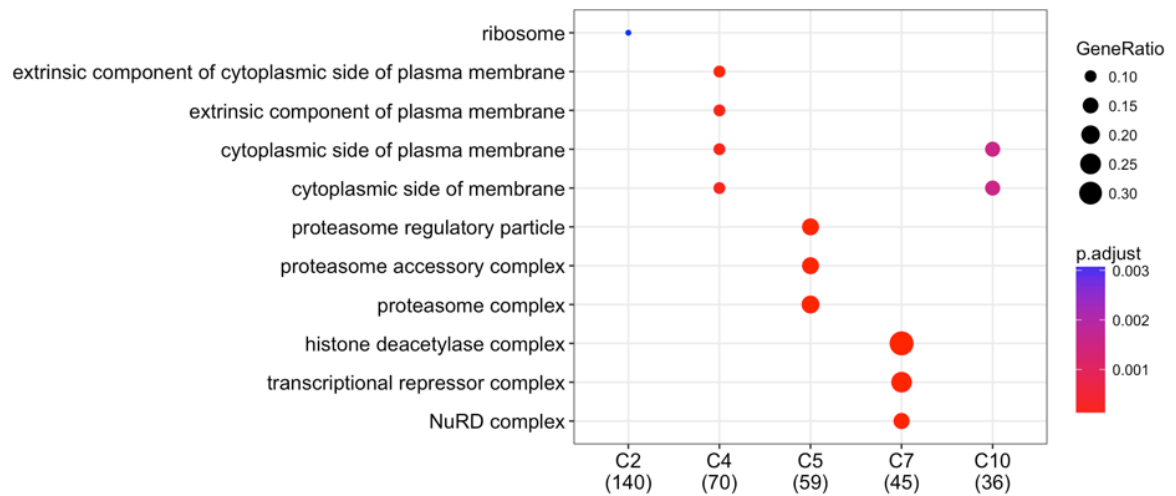


Figure 4.30. Enriched GO Cellular Component terms for PPIN clusters.

4.6.1.3. Prostate Cancer PPI Network Hubs and Bottlenecks. Gene Ontology enrichment analysis was performed for top 100 hub and bottleneck proteins. Since two groups share 70% of their proteins, the enrichments are alike with small significance differences (Figure 4.31-4.33). Bottleneck proteins have negligibly higher significance in regulation of apoptotic signaling pathway. Apoptosis is a process of self-destruction and cancer cells try to escape from this type of cell death. It is one of the hallmarks of cancer explained detailly in Section 4.6.4.1. Another enriched term is the regulation of DNA binding transcription factor activity. Transcription factor up- or downregulate the transcription of genes and can be drug targets as they take part in oncogenic pathways such as prostate cancer pathway [184].

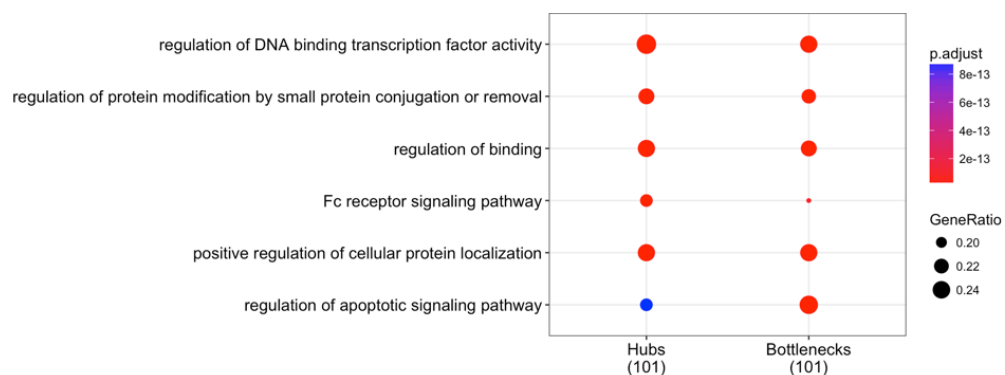


Figure 4.31. Enriched GO Biological Process terms for hubs and bottlenecks.

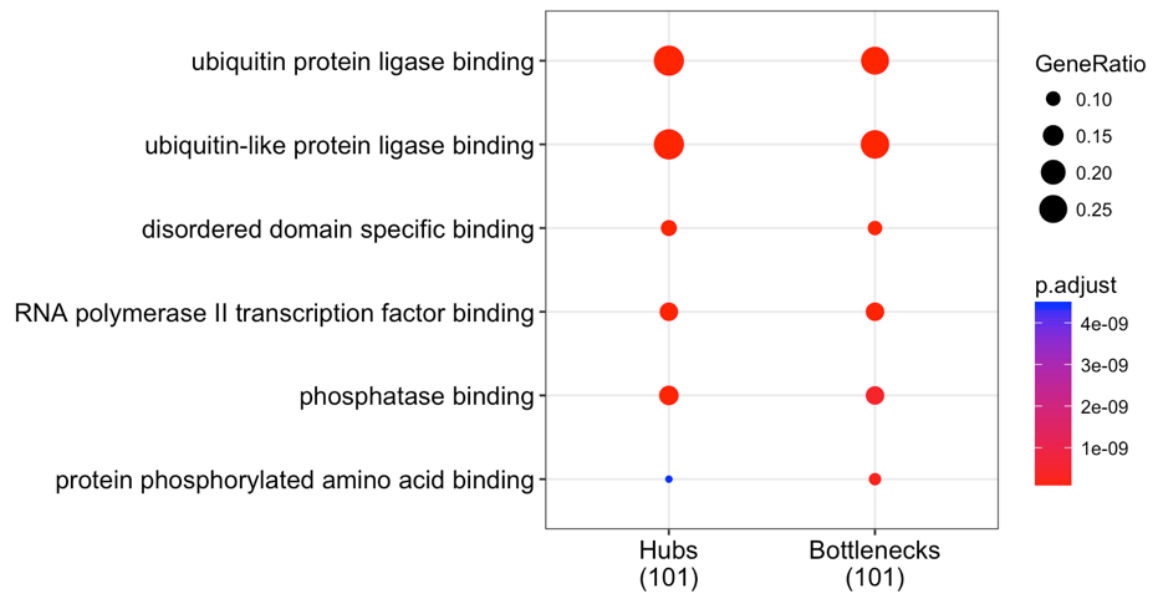


Figure 4.32. Enriched GO Molecular Function terms for hubs and bottlenecks.

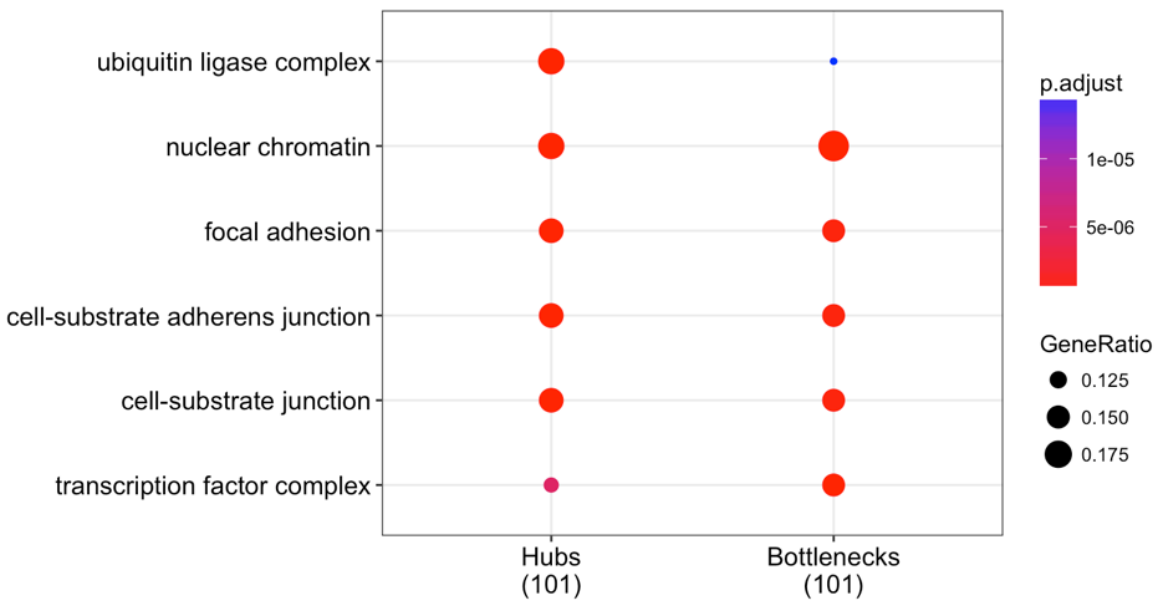


Figure 4.33. Enriched GO Cellular Component terms for hubs and bottlenecks.

4.6.1.4. Prostate Tumor Co-Expression Modules. The threshold for module size was chosen as 400 and for the passing 11 modules enriched terms were found (Figure 4.34-4.36). Yellow and greenyellow modules are closely related with each other (Figure 4.22) and therefore similar enrichment terms were obtained. Both modules are enriched in muscle system and development.

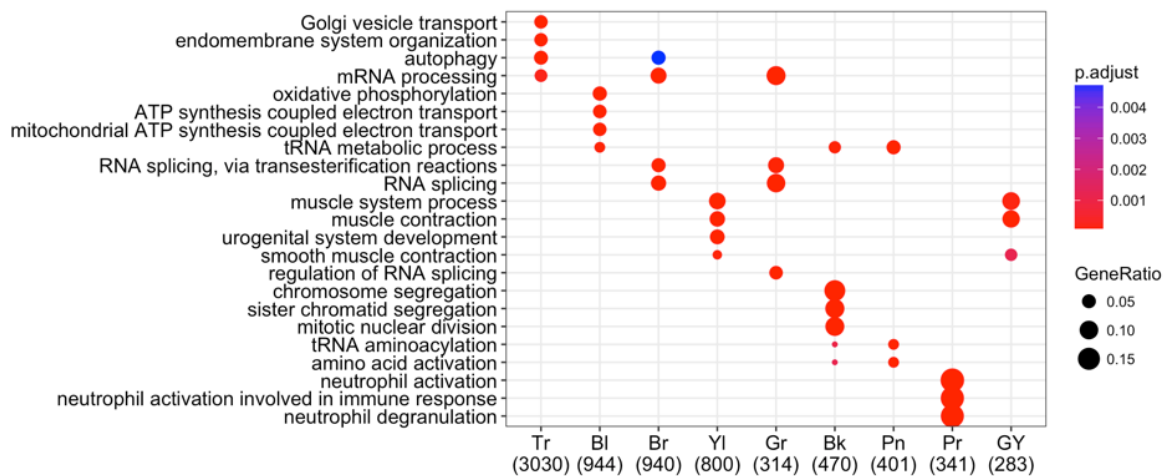


Figure 4.34. Enriched GO Biological Process terms for first 11 Co-Expression modules.

At first glance, prostate cancer and muscle may seem irrelevant. Androgens, consequently AR, have roles in male sexual differentiation and also in modulating physiological responses of muscles. Specifically, androgen-AR signaling pathway takes a part in skeletal muscle development [185]. Purple module is significantly enriched and has many members in neutrophil related biological processes. Neutrophils are oncogenic molecules participating in tumor initiation, growth, proliferation and metastasis. They play a significant role in inflammation, which is a hallmark of cancer [186].

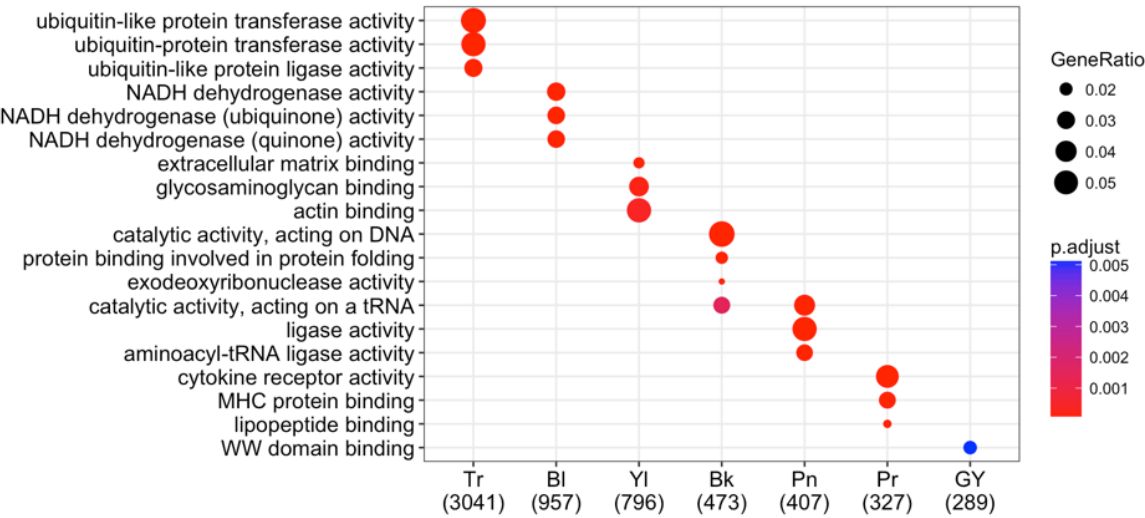


Figure 4.35. Enriched GO Molecular Function terms for first 11 Co-Expression modules.

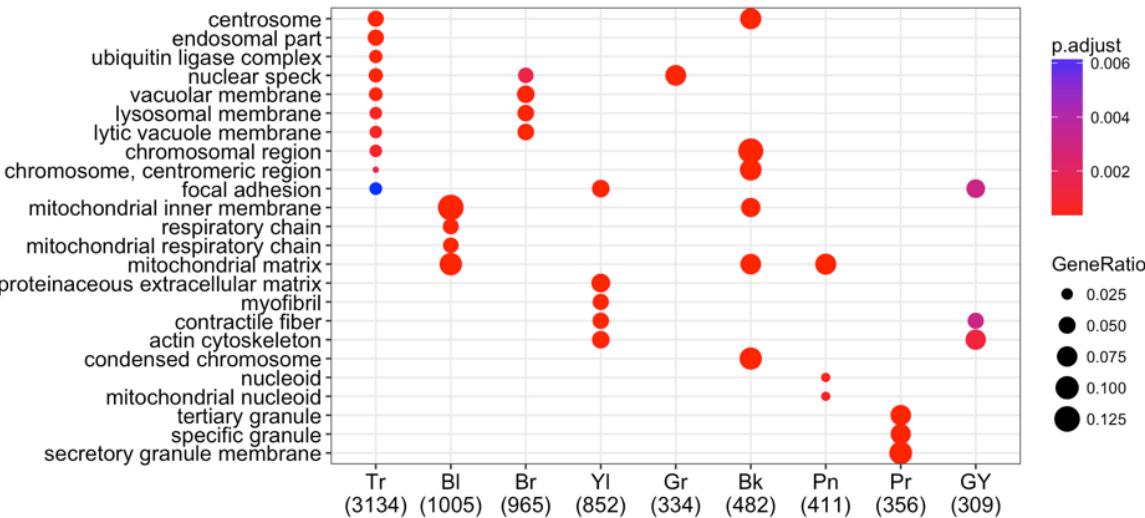


Figure 4.36. Enriched GO Cellular Component terms for first 11 Co-Expression modules.

4.6.2. Pathway Enrichment Analysis

4.6.2.1. Prostate Cancer PPIN Proteins. Several signaling pathways were found to be enriched among PPIN proteins. PI3K-Akt, Ras, MAPK, T cell receptor and ErbB signaling pathways are present in Figure 4.37. Ras and MAPK activations were reported as increasing with increasing tumor grade and stage. Increased AR activity may be also caused by Ras signaling [187]. There is also a crosstalk between AR and ERbB signaling pathways [188].

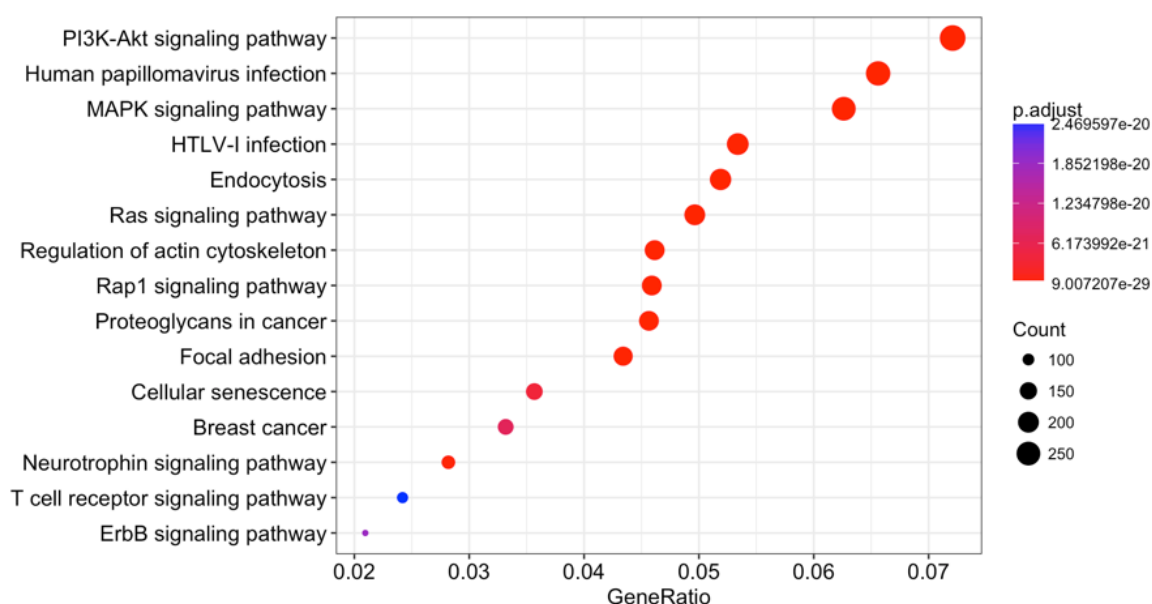


Figure 4.37. KEGG pathway enrichment of Prostate Cancer PPIN proteins.

4.6.2.2. Prostate Cancer PPI Network Modules. Cluster 4 and 10 are enriched in many signaling pathways such as Ras and ErbB. In addition to these two pathways, cluster 4 is enriched also in Neutrophin and T cell receptor pathways, whereas cluster 10 is enriched in MAPK signaling pathway. The pathway enrichments are nearly coherent. Ras, MAPK, ErbB and T cell receptor signaling pathways are also present in Pca PPIN proteins' enrichments. As mentioned above, AR protein is associated many of these pathways (Figure 4.38).

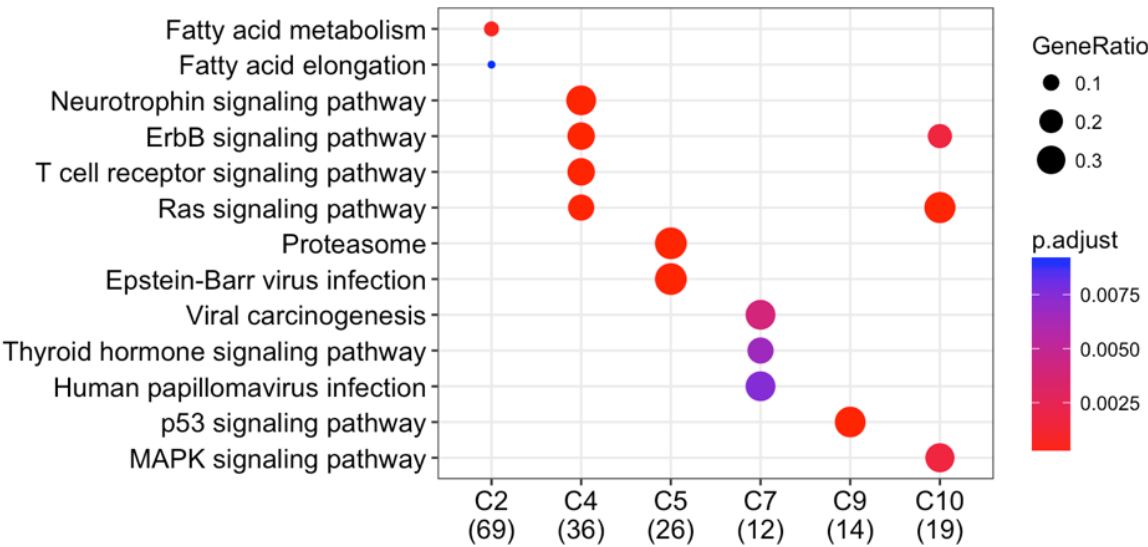


Figure 4.38. KEGG pathway enrichment of Prostate Cancer PPIN modules.

4.6.2.3. Prostate Cancer PPI Network Hubs and Bottlenecks. Although the number of proteins in the pathway is low, prostate cancer was found to be significantly enriched in both PPI hubs and bottlenecks (Figure 4.39). As discussed before, cell cycle is a process affected by AR protein, consequently by prostate cancer. Enrichments in cell cycle pathway is reasonable. Unexpectedly, viral processes such as viral carcinogenesis and Hepatitis B were highly enriched.

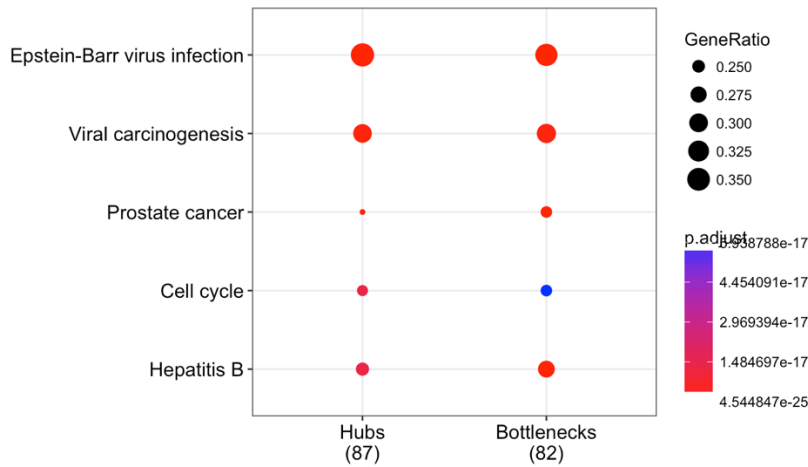


Figure 4.39. KEGG pathway enrichment of Prostate Cancer PPIN Hubs and Bottlenecks.

4.6.2.4. Prostate Tumor Co-Expression Modules. Autophagy, enriched in turquoise module, manages the cellular homeostasis by facilitating nutrient utilization and toxin removal. Autophagy can both suppress tumor development and enable cancer cell survival during therapy [189]. Cell cycle is again found as an enrichment term. The enrichments among different modules nearly never overlap, suggesting each module regulates different processes and exist in different pathways. Gene expression similarities and dissimilarities clearly distinguish the pathways and metabolisms (Figure 4.40).

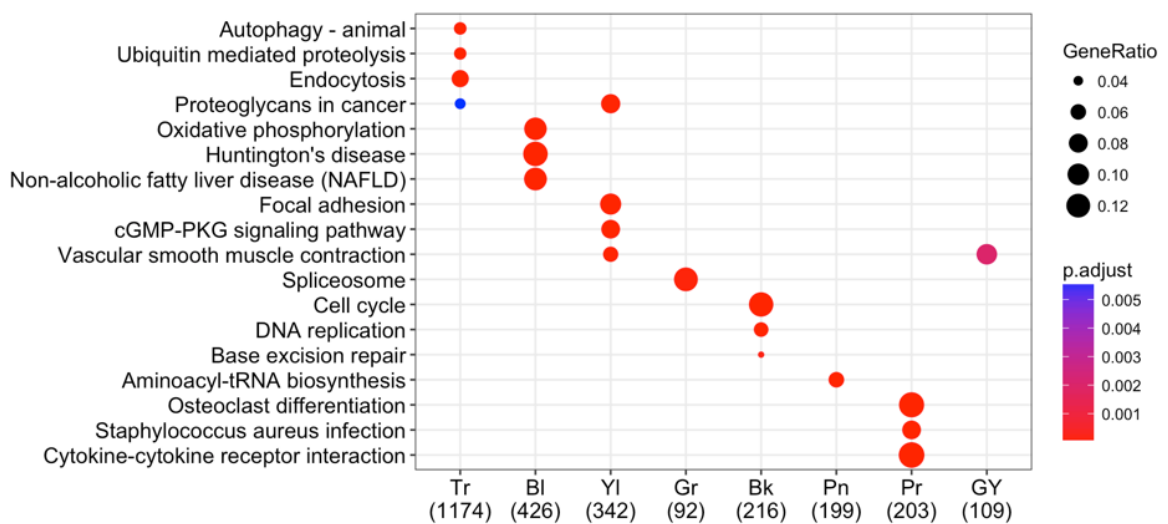


Figure 4.40. KEGG pathway enrichment of Prostate Tumor for first 11 Co-Expression modules.

4.6.3. Disease Ontology Enrichment Analysis

Disease Ontology (DO) is an ontology database which aims to construct a standardized ontology for human diseases by connecting clinical, genetic data and symptoms [190]. Using this ontology, enrichment of disease for given data sets can be acquired.

4.6.3.1. Prostate Cancer PPIN Proteins. Top DO enrichment results for the Pca PPIN proteins are given in Figure 4.41. 10 out of 15 enrichments are cancer related diseases such as bone marrow, lung and connective tissue cancer. Among 8709 proteins present

in the network, above mentioned cancers have around 300 proteins associated with them. The relation between Pca and other diseases will be discussed in Section 4.8.1.

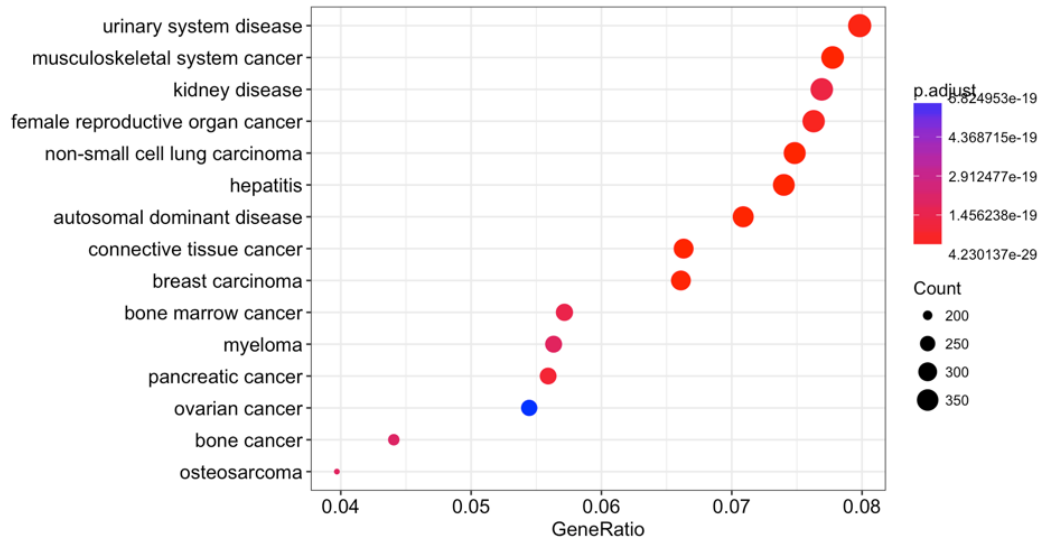


Figure 4.41. DO Enrichment terms for Prostate Cancer PPIN proteins.

4.6.3.2. Prostate Cancer PPI Network Modules. Only two clusters are significantly enriched in DO terms (Figure 4.42). Cluster 10 includes the synovium cancer and sarcoma terms, which are a type of carcinoma found in joint capsules such as in arms and legs. Although it is very rare, synovial sarcoma can be seen in prostate tissue [191]. Hypothyroidism is a condition in which the thyroid gland slows the production of thyroid, resulting in significant changes in energy and metabolism. Studies investigated the effects of thyroid disease on prostate cancer, and found that thyroid disease is associated with increased risk in prostate cancer [192].

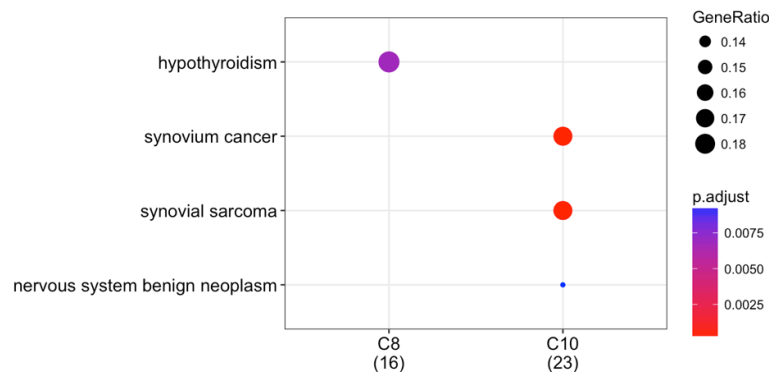


Figure 4.42. DO Enrichment terms for Prostate Cancer PPIN modules.

4.6.3.3. Prostate Cancer Network Hubs and Bottlenecks. Connective tissue and bone cancer are again present in the enrichment results for Pca PPI network hubs and bottleneck as they are in Pca PPI network (Figure 4.43). The disease identification Section 4.8.1 will further discuss the relations of these diseases with prostate cancer.

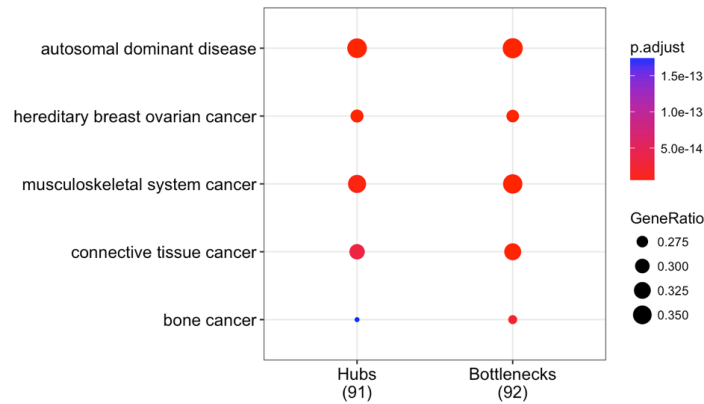


Figure 4.43. DO Enrichment terms for Prostate Cancer PPIN hubs bottlenecks.

4.6.3.4. Prostate Tumor Co-Expression Modules. Among 11 modules, the enriched terms are only for yellow and purple modules, and also their adjusted p-values are small in comparison to other enrichment results (Figure 4.44). Same case can be seen in DO enrichments of Pca PPIN modules. For only two modules enrichments are acquired and although the gene ratios are high, p-values are even lower than co-expression modules' p-values (Figure 4.42). One can argue that, even the gene expressions within a module are similar, they do not carry the characteristics of a specific disease.

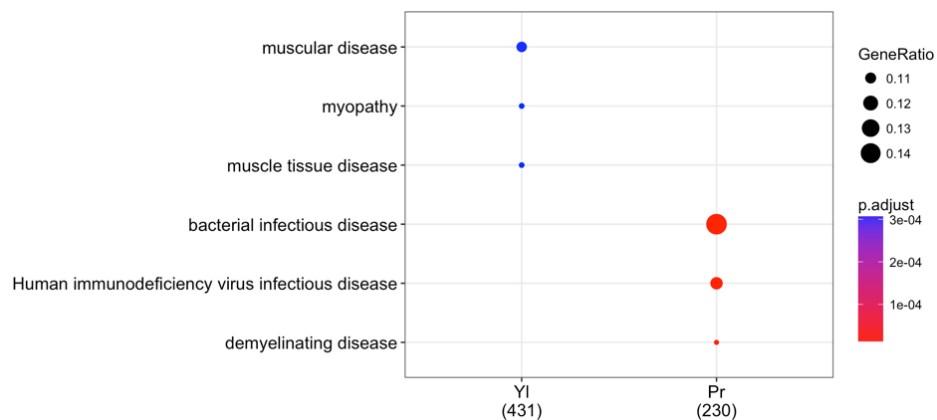


Figure 4.44. DO Enrichment terms for Prostate Tumor Co-Expression modules.

4.6.4. Gene Set Enrichment Analysis

Analysis of gene expression data may be challenging since a small number of statistically significant genes gives nearly no results and a large number of significant genes makes it difficult to decide the most important genes. Gene Set Enrichment Analysis (GSEA) is a method that finds the statistically significant difference of phenotypes for a given gene set. Gene expression data with its corresponding phenotypes are analyzed if they are enriched for the selected gene sets. These gene sets may be Gene Ontology terms, KEGG pathways or hallmark gene sets curated by Molecular Signatures Database (MSigDB).

Overlapping genes between previously defined gene sets and heterogeneity of these sets urged the necessity of defining hallmark genes, which represent well-defined biological processes and assist the GSEA analysis. Each hallmark is given under a process category as seen in Table 4.16.

Table 4.16. Description of Hallmark Gene Sets [193].

Hallmark Name	Process Category	Description
Apical Junction	cellular component	apical junction complex consisting of adherens and tight junctions
Apical Surface	cellular component	membrane proteins in the apical domain
Peroxisome	cellular component	peroxisomes
Adipogenesis	development	adipocyte development
Angiogenesis	development	blood vessel formation
Epithelial Mesenchymal Transition	development	epithelial mesenchymal transition
Myogenesis	development	muscle differentiation
Spermatogenesis	development	sperm development and male fertility
Pancreas Beta Cell	development	genes specific to pancreatic beta cells
DNA Repair	DNA damage	DNA repair
UV Response Down	DNA damage	UV response: downregulated genes
UV Response Up	DNA damage	UV response: upregulated genes
Allograft Rejection	immune	allograft rejection

Table 4.16. Description of Hallmark Gene Sets [193] (cont.)

Hallmark Name	Process Category	Description
Coagulation	immune	blood coagulation cascade
Complement	immune	complement cascade
Interferon Alpha Response	immune	interferon alpha response
Interferon Gamma Response	immune	interferon gamma response
IL6-JAK-STAT3 Signaling	immune	IL6 STAT3 signaling during acute phase response
Inflammatory Response	immune	inflammation
Bile Acid Metabolism	metabolic	biosynthesis of bile acids
Cholesterol Homeostasis	metabolic	cholesterol homeostasis
Fatty Acid Metabolism	metabolic	fatty acid metabolism
Glycolysis	metabolic	glycolysis and gluconeogenesis
Heme Metabolism	metabolic	heme metabolism and erythroid lineage
Oxidative Phosphorylation	metabolic	oxidative phosphorylation and citric acid cycle
Xenobiotic Metabolism	metabolic	metabolism of xenobiotics
Apoptosis	pathway	programmed cell death; caspase pathway
Hypoxia	pathway	response to hypoxia; HIF1A targets
Protein Secretion	pathway	protein secretion
Unfolded Protein Response	pathway	unfolded protein response; ER stress
Reactive Oxygen Species Pathway	pathway	reactive oxygen species pathway
E2F Targets	proliferation	cell cycle progression: E2F targets
G2M Checkpoint	proliferation	cell cycle progression: G2/M checkpoint
MYC Targets V1	proliferation	MYC targets, variant 1
MYC Targets V2	proliferation	MYC targets, variant 2
P53 Pathway	proliferation	p53 pathway
Mitotic Spindle	proliferation	cell cycle progression: mitotic spindle assembly
Androgen Response	signaling	androgen response
Estrogen Response Early	signaling	early estrogen response

Table 4.16. Description of Hallmark Gene Sets [193] (cont.)

Hallmark Name	Process Category	Description
Estrogen Response Late	signaling	late estrogen response
IL2-JAK-STAT5 Signaling	signaling	IL2 STAT5 signaling
KRAS Signaling Up	signaling	KRAS signaling, upregulated genes
KRAS Signaling Dn	signaling	KRAS signaling, downregulated genes
MTORC1 Signaling	signaling	mTORC1 signaling
NOTCH Signaling	signaling	Notch signaling
PI3K-AKT-MTOR Signaling	signaling	PI3K signaling via AKT to mTORC1
Hedgehog Signaling	signaling	Hedgehog signaling
TGF Beta Signaling	signaling	TGF beta signaling
TNFA Signaling Via NFKB	signaling	TNFA signaling via NFB
WNT Beta Catenin Signaling	signaling	canonical beta catenin pathway

Table 4.17. Number of genes in hallmark and Pca gene expression set.

Hallmark Name	Number of Genes in Hallmark	Number of Genes Present in the Hallmark Gene Set (present study)
TNFA Signaling via NFKB	200	196
Hypoxia	200	197
Cholesterol Homeostasis	74	73
Mitotic Spindle	200	198
WNT Beta Catenin Signaling	42	42
TGF Beta Signaling	54	54
IL6-JAK-STAT3 Signaling	87	87
DNA Repair	150	143
G2M Checkpoint	200	198
Apoptosis	161	159
NOTCH Signaling	32	32
Adipogenesis	200	195
Estrogen Response Early	200	199
Estrogen Response Late	200	199
Androgen Response	101	99
Myogenesis	200	198

Table 4.17. Number of genes in hallmark and Pca gene expression set (cont.)

Hallmark Name	Number of Genes in Hallmark	Number of Genes Present in the Hallmark Gene Set (present study)
Protein Secretion	96	95
Interferon Alpha Response	97	93
Interferon Gamma Response	200	196
Apical Junction	200	196
Apical Surface	44	44
Hedgehog Signaling	36	36
Complement	200	193
Unfolded Protein Response	113	111
PI3K-AKT-MTOR Signaling	105	104
MTORC1 Signaling	200	198
E2F Targets	200	199
MYC Targets v1	200	198
MYC Targets v2	58	58
Epithelial Mesenchymal Transition	200	198
Inflammatory Response	200	197
Xenobiotic Metabolism	200	200
Fatty Acid Metabolism	158	157
Oxidative Phosphorylation	200	199
Glycolysis	200	198
Reactive Oxygen Species Pathway	49	47
P53 Pathway	200	199
UV Response Up	158	156
UV Response Down	144	144
Angiogenesis	36	36
Heme Metabolism	200	197
Coagulation	138	134
IL2-STAT5 Signaling	200	198
Bile Acid Metabolism	112	112
Peroxisome	104	103
Allograft Rejection	200	198
Spermatogenesis	135	133
KRAS Signaling Up	200	196
KRAS Signaling Down	200	193
Pancreas Beta Cells	40	40

Normal prostate tissue and prostate cancer tumor gene expression data from GTEx and TCGA were analyzed using GSEA for hallmark genes and the results are given in Tables 4.17 and 4.18. Out of 51875 genes, 21456 (41.4%) genes are markers for Normal phenotype and 30419 (58.6%) of them are for phenotype Tumor. Nearly all genes in hallmark gene sets are present in the investigated dataset.

20 gene sets were found to be downregulated and 30 gene sets are upregulated for phenotype Pca tumor. From 20 downregulated gene sets only two gene sets, namely Epithelial Mesenchymal Transition, Myogenesis, are significantly enriched with a p-value $<1\%$ and Apical Junction is added to that list if p-value threshold is $<5\%$.

On the other hand, 18 upregulated gene sets for Pca tumor are significantly enriched with a p-value $<1\%$. These gene sets are Androgen Response, Protein Secretion, MTORC1 Signaling, MYC Targets v1, Fatty Acid Metabolism, Oxidative Phosphorylation, Peroxisome, Adipogenesis, Cholesterol Homeostasis, G2M Checkpoint, E2F Targets, Unfolded Protein Response, MYC Targets v2, Glycolysis, Bile Acid Metabolism, Estrogen Response Late, UV Response Up and Estrogen Response Early. If the threshold is increased to 5%, PI3K-AKT-MTOR Signaling, Apoptosis and Xenobiotic Metabolism gene sets are also enriched. The results of the GSEA analysis are discussed in Section 4.6.4.1.

Table 4.18. Significance of down- and upregulated gene sets in tumor.

Hallmark Name	Process Category	Down- or Upregulated in Phenotype Tumor	Significance (p-value)
Apical Junction	cellular component	Downregulated	5%
Apical Surface	cellular component	Downregulated	
Peroxisome	cellular component	Upregulated	1%
Adipogenesis	development	Upregulated	1%
Angiogenesis	development	Downregulated	
Epithelial Mesenchymal Transition	development	Downregulated	1%
Myogenesis	development	Downregulated	1%
Spermatogenesis	development	Upregulated	
Pancreas Beta Cell	development	Upregulated	

Table 4.18. Significance of down- and upregulated gene sets in tumor (cont.)

Hallmark Name	Process Category	Down- or Upregulated in Phenotype Tumor	Significance (p-value)
DNA Repair	DNA damage	Upregulated	5%
UV Response Down	DNA damage	Downregulated	
UV Response Up	DNA damage	Upregulated	1%
Allograft Rejection	immune	Upregulated	
Coagulation	immune	Downregulated	
Complement	immune	Downregulated	
Interferon Alpha Response	immune	Downregulated	
Interferon Gamma Response	immune	Downregulated	
IL6-JAK-STAT3 Signaling	immune	Downregulated	
Inflammatory Response	immune	Downregulated	
Bile Acid Metabolism	metabolic	Upregulated	1%
Cholesterol Homeostasis	metabolic	Upregulated	1%
Fatty Acid Metabolism	metabolic	Upregulated	1%
Glycolysis	metabolic	Upregulated	1%
Heme Metabolism	metabolic	Upregulated	1%
Oxidative Phosphorylation	metabolic	Upregulated	1%
Xenobiotic Metabolism	metabolic	Upregulated	5%
Apoptosis	pathway	Upregulated	5%
Hypoxia	pathway	Downregulated	
Protein Secretion	pathway	Upregulated	1%
Unfolded Protein Response	pathway	Upregulated	1%
Reactive Oxygen Species Pathway	pathway	Upregulated	
E2F Targets	proliferation	Upregulated	1%
G2M Checkpoint	proliferation	Upregulated	1%
MYC Targets V1	proliferation	Upregulated	1%
MYC Targets V2	proliferation	Upregulated	1%
P53 Pathway	proliferation	Downregulated	
Mitotic Spindle	proliferation	Downregulated	
Androgen Response	signaling	Upregulated	1%
Estrogen Response Early	signaling	Upregulated	1%
Estrogen Response Late	signaling	Upregulated	1%
IL2-JAK-STAT5 Signaling	signaling	Downregulated	
KRAS Signaling Up	signaling	Upregulated	
KRAS Signaling Dn	signaling	Downregulated	
MTORC1 Signaling	signaling	Upregulated	1%

Table 4.18. Significance of down- and upregulated gene sets in tumor (cont.)

Hallmark Name	Process Category	Down- or Upregulated in Phenotype Tumor	Significance (p-value)
NOTCH Signaling	signaling	Upregulated	
PI3K-AKT-MTOR Signaling	signaling	Upregulated	5%
Hedgehog Signaling	signaling	Downregulated	
TGF Beta Signaling	signaling	Upregulated	
TNFA Signaling Via NFkB	signaling	Downregulated	
WNT Beta Catenin Signaling	signaling	Downregulated	

4.6.4.1. Hallmarks of Cancer. Normal cells undergo several changes and attain new features as they transform to an uncontrollable neoplastic state. These new features are crucial for the development of neoplastic cells. Hanahan *et al.* proposed that cancer has eight common hallmarks (Figure 4.45 and 4.46). These are sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and two emerging hallmarks reprogramming of energy metabolism and evading immune destruction [194].

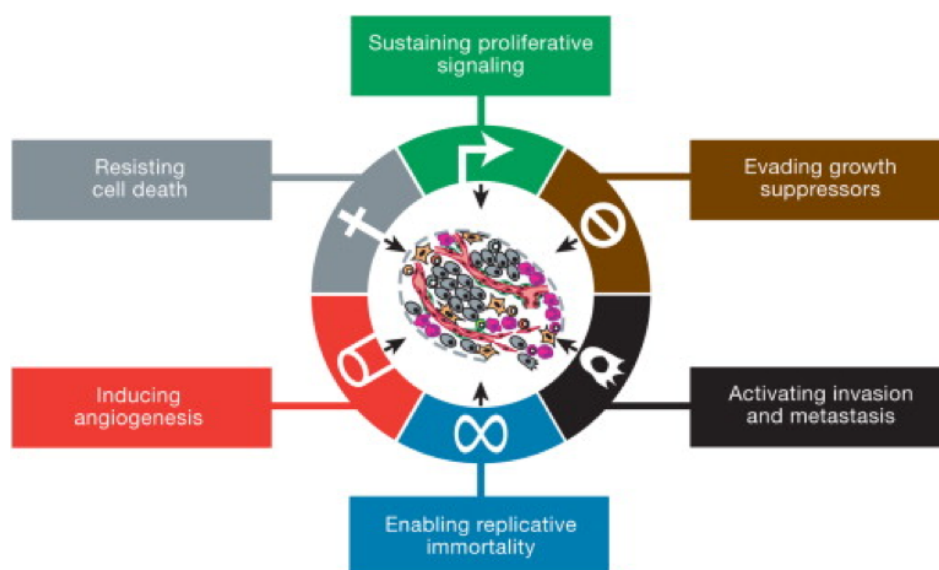


Figure 4.45. First 6 Hallmarks of Cancer [194].

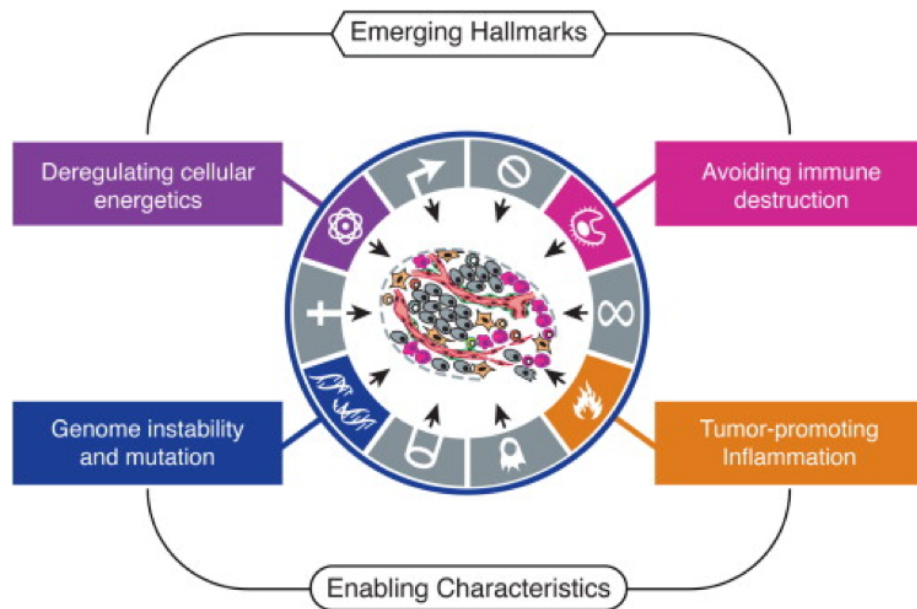


Figure 4.46. Emerging Hallmarks of Cancer [194].

Sustaining proliferative signaling is the most essential feature of the neoplastic cells. As normal cells control the progression of the cell cycle by controlling the growth factors that bind to receptors containing tyrosine kinase domains, cancer cells deregulate the growth-promoting signals by either producing the growth factor ligands, by signaling nearby normal cells to supply growth factors or by elevating the receptor proteins at cell surface [195]. As can be seen in GO Molecular Function Enrichment of clusters 4 and 10 shown in 4.29, tyrosine kinase activity is significantly enriched. In addition to tyrosine kinase, regulation of Ras protein signal transduction is also enriched in cluster 4, which is one of the signaling pathways activated by the tumor cells for growth factor independence (Figure 4.38). Somatic mutations result in activation of the signaling pathways of mitogen-activated protein (MAP)-kinase and phosphoinositide 3-kinase (PI3-kinase) [196, 197]. Both of these pathways are found to be enriched in KEGG pathways of PPIN as shown in Figure 4.37. On the other hand, the GSEA analysis in Section 4.6.4 shows that four out of six hallmarks under the process category proliferation are significantly upregulated in prostate cancer tumors with a p-value $<1\%$. Also, PI3K-AKT-MTOR signaling is upregulated with a p-value $<5\%$. This signaling mechanism is quite interesting, since the inhibition of MTOR is expected to have anti-proliferative effects [198], but in some cancer cells inhibition of MTOR causes activation of PI3K.

While activating the growth-stimulatory signals, tumor cells should also block the negative feedback mechanisms involved in proliferation. Tumor suppressors retinoblastoma (RB) and tumor protein 53 (P53 or TP53 proteins) control the mechanisms of proliferation and apoptosis and known as the two typical tumor suppressors, one transducing the signals from outside the cell and the other from inside the cell [199]. Although it is not statistically significant, the GSEA analysis shows that P53 pathway is downregulated in prostate cancer as expected. Another suppressor, TGF- β is actually anti-proliferative, but in late stage tumors it is found to be tumor-promoting. The signaling of TGF- β is also upregulated in prostate cancer, but not significantly. One reason may be that, TCGA database does not include the stages of the prostate cancer and the dataset consists of all stages of PCa. A GSEA analysis between early and late stages of Pca may give significantly upregulated TGF- β signaling in late-stage tumors.

Apoptosis, a process of programmed cell death, is a homeostatic mechanism to maintain aging and population of cells. It is also a defense mechanism against diseases such as cancer. Tumor cells try to limit apoptosis and a common method is to cause loss of function of TP53 as tumor suppressor [200]. As mentioned before, its signaling pathway is downregulated in tumors, proving this fact. The signaling pathway of PI3K-AKT-MTOR, an upregulated pathway in prostate cancer as mentioned above, blocks apoptosis and autophagy, helps the tumor to proliferate. However, GSEA results show that the apoptosis is upregulated in prostate cancer significantly (p-value <5%) on the contrary of the expectations.

The cells have only a limited number of cell growth and division cycles and at the end they either enter senescence, a non-proliferative but viable state, or encounter cell death. The length of telomeric DNA is a property that decides the number of cycles [201]. Telomerase is expressed at significant levels in immortal cells, such as cancer cells. DNA damage repair is also a part of the immortality of cancer cells [202] and it is significantly upregulated in prostate cancer tumors (p-value <5%).

Vascular networks are important for cell survival since they provide oxygen and other nutrients to cell and take away wastes and carbon dioxide. Development of these networks by angiogenesis is especially important for these needs of tumor cells as they increase in vast amounts. Angiogenesis is induced in very early stages of tumorigenesis and expected to be upregulated in cancers [203]. However, GSEA results show the opposite. The angiogenesis process is downregulated in prostate cancer, although it is not significant.

Uncontrolled cell proliferation also calls for adjustment in energy metabolism in order to sustain the cell growth and division, and in case of angiogenesis it cannot provide enough sources. Cancer cells can reprogram glucose metabolism. In the presence of oxygen, the cells process the glucose first to pyruvate via glycolysis and then to carbon dioxide in mitochondria. But under anaerobic conditions, only glycolysis process is present. The cancer cells can change their energy production method to glycolysis although oxygen is present [204]. The process category ‘metabolic’ stated by MSigDB consists of Bile Acid, Fatty Acid, Heme and Xenobiotic Metabolisms, Cholesterol Homeostasis, Oxidative Phosphorylation and Glycolysis. All of these processes are significantly upregulated in prostate cancer. (p-value <1% except Xenobiotic Metabolism) These results are in agreement with the given hallmark of cancer.

The immune system plays an important role preventing the formation and progression of neoplasms. Most cancer cells are eliminated by the constantly monitoring immune system, which is acting as a barrier. The remaining tumors are the ones to avoid detection or limit the immunological responses. Immunosurveillance hypothesis suggests that, natural killer (NK) cells and cytotoxic T lymphocytes (CTCs) take part in tumor destruction [205]. Cancer cells may prevent these cells from functioning by producing more immunosuppressive factors such as tumor growth factors (TGFs) [194] Although the significance level is low, TGF- β signaling is upregulated in prostate cancer verifying the cells increase the immunosuppressive functions in order to escape from immune responses. On the other hand, inflammatory and interferon responses are downregulated in Pca, as expected.

4.7. Prognostic Genes

Prognostic genes are biological markers that can be used to predict the progress of a disease. If higher expression level of a gene shows a significantly low survival rate for a specific disease, that gene is called unfavorable prognostic gene. For favorable prognostic genes, higher expression level is correlated with high survival rate.

A study by Human Protein Atlas (HPA) identified 134 unfavorable and 25 favorable prognostic genes [206] for prostate cancer by analyzing the transcriptomics data present in The Cancer Genome Atlas (TCGA). Table D.1 and E.1 show the unfavorable and favorable prognostic genes and the cancers they are associated with. U stands for unfavorable and F stands for favorable.

Among the unfavorable prognostic genes for prostate cancer in Table D.1, 50 out of 136 genes are found to be unfavorable prognostic genes for liver cancer. Similarly, 43 genes are found to be unfavorable prognostic genes for renal (kidney) cancer, whereas 15 are found to be favorable. Moreover, 19 genes are found to be favorable genes for urothelial cancer.

The favorable prognostic genes for prostate cancer given in Table E.1 share four unfavorable prognostic genes for renal (kidney) cancer and three favorable prognostic genes for renal cancer. Two genes are found to be unfavorable prognostic genes for pancreatic cancer and one gene is found to be favorable. Two genes are found as unfavorable prognostic genes for liver cancer.

Prostate cancer shares many unfavorable prognostic genes with liver and renal cancer, suggesting a patient with one of these diseases are under the risk of developing another one.

Common proteins of core protein set given (Table 4.4) with prognostic genes are all unfavorable prognostic genes. 15 out of 25 favorable and 74 out of 134 unfavorable prognostic genes are present in the Pca PPI network. GO Biological Process terms of these 89 genes were conducted and nearly all of the proteins take part in cellular processes. Cell proliferation, reproduction and growth are hallmarks processes of cancer and it is expected that prognostic genes carry out the missions essential for the survival of cancer cells (Figure 4.47).

Table 4.19. Common Pca network proteins shared by prostate cancer prognostic genes/proteins.

Core Protein Set (2)	AKT2, TP53
Pca PPI proteins (89/159)¹ (56%)	Favorable: AOC1, ARL2, COMT, CRACR2A, CSRP2, EPHX1, FOS, HBEGF, MAP6, MYB, NTNG2, RNF122, SESN1, TNS2, TXN2
	Unfavorable: ABCC5, AKT2, ARMC8, ARRB2, ASB9, ATG3, BMP8B, BRPF1, C3orf38, CCNT2, CCT2, CDK5RAP1, CEP70, CLK3, COPS3, DCUN1D5, DPP4, DRG2, DVL3, EID2, EXOG, FBXO45, FGD6, GNL3, HAUS1, HBEGF, HKR1, HNRNPA2B1, HOOK2, MBD1, MDM4, MINK1, MTF2, MXI1, NCBP1, NCBP2, ODF2, PAK2, PAK4, PNN, POLG, POLG2, RAD18, RAN, RBM25, RIOK3, RNF8, RUBCN, RUSC1, SETD6, SH2D3A, SHTN1, SIN3B, SRC, SRSF10, SYNJ1, TDP2, TIPIN, TOB1, TP53, USP53, ZNF195, ZNF197, ZNF248, ZNF274, ZNF330, ZNF335, ZNF419, ZNF440, ZNF503, ZNF550, ZNF668, ZNF830, ZRANB2, ZWILCH

Expression levels of prognostic genes are also investigated. 11 (44%) favorable genes were found to be upregulated, whereas 9 (36%) of them are downregulated. For unfavorable genes, these numbers are 45 (33%) and 69 (51%), respectively. TCGA reported only 2% of the cases given in the database is dead. Downregulation of nearly half of the prognostic genes in patients seems to be in accordance with TCGA reports, although higher number of overexpressed favorable genes were expected.

¹Total prognostic genes (favorable and unfavorable)

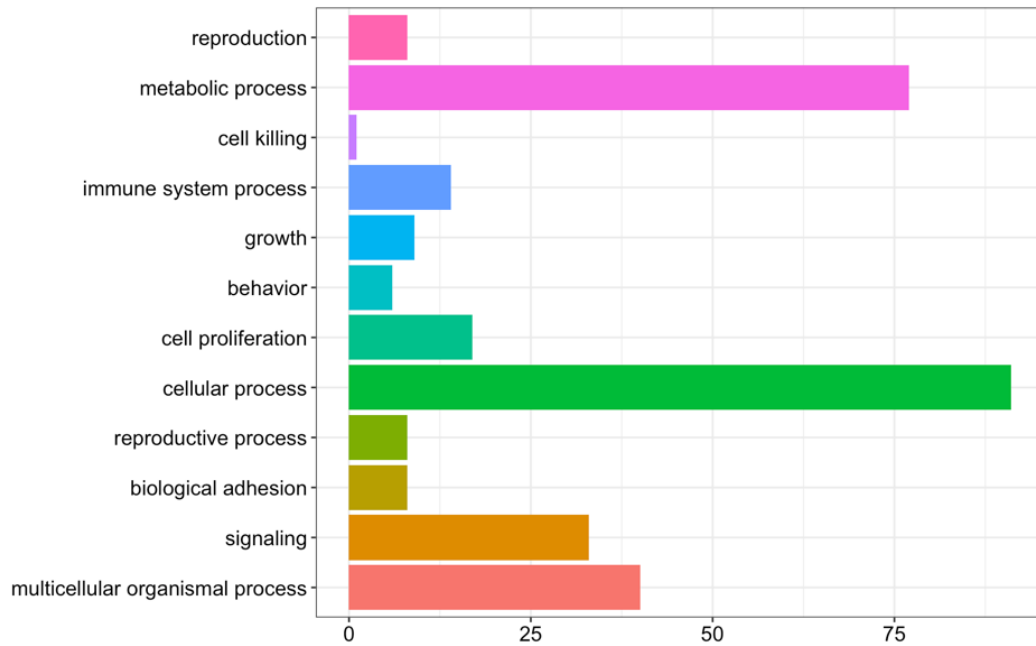


Figure 4.47. Biological Process GO terms of prognostic genes shared with Pca PPIN.

4.8. DIAMOnD Algorithm for Detection of Disease Modules

Proteins associates with a certain disease are likely to interact with each other, work in similar pathways and cluster together in the same neighborhood in a PPI network. These neighborhoods are called disease modules. Although there are many methods such as using connectivity degree to identify disease modules, DIAMOnD algorithm [207] uses connectivity significance for identification. This method evaluates the significance of connections with the seed genes, genes related with a certain disease, rather than the number of connections.

Seed genes for prostate cancer were acquired from Online Mendelian Inheritance in Man (OMIM) [107] and Genome Wide Association Studies (GWAS) [108]. In OMIM, “Prostate Cancer” term was searched and in addition to the genes in Phenotype-Gene Relationship table, the text was manually screened. The prostate cancer gene associations were obtained from PheGenI database by choosing a threshold value of $p\text{-value} \leq 5 \times 10^{-8}$ as stated in the article [109,207]. 68 and 138 genes were found to be associated with prostate cancer in OMIM and GWAS, respectively (Table F.1).

10 of these genes, namely AR, BRCA2, EHBP1, HNF1B, JAZF1, KLK3, LMTK2, MSMB, PRNCR1 and TERT, were found to be present in both databases. GO enrichment analysis gives no statistically significant results. Totally, 196 genes associated with prostate cancer were selected as the seed genes for the DIAMOnD algorithm.

When these seed genes were compared with the core protein list, it was found that coding genes of nine of the core proteins are shared with seed gene list. These nine genes are AR, AZGP1, CDKN1B, CHUK, FGFR2, KLK3, MSMB, NKX3-1 and PTEN (Table 4.20). GO enrichment for these 9 genes were investigated. For molecular function and cellular component, none of the GO terms are enriched. For biological process, enriched processes are related to prostate and its gland, as shown in Figure 4.48.

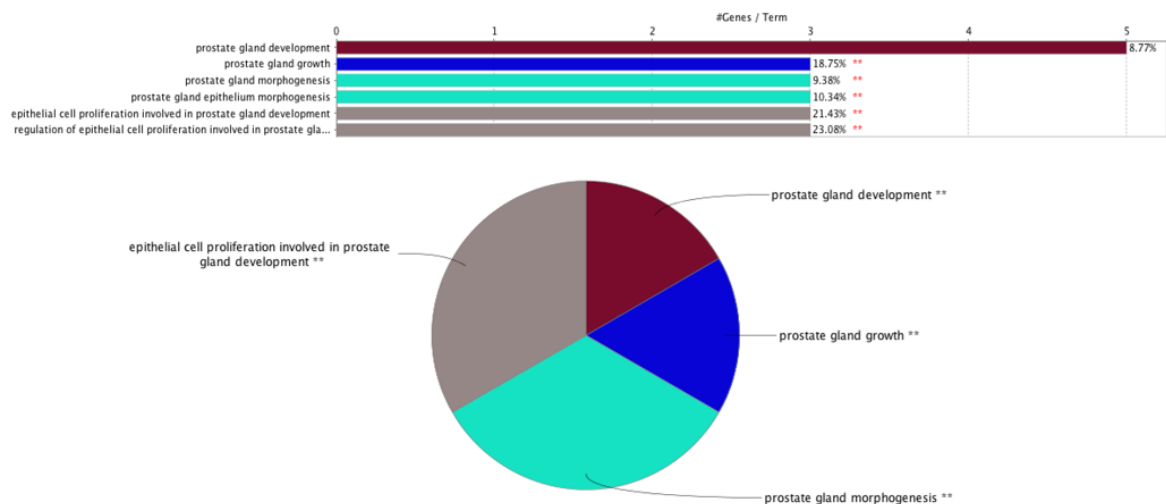


Figure 4.48. GO Biological process enrichment for shared nine genes.

This DIAMOnD algorithm is extending by an alpha value, which gives a higher weight to seed proteins so that direct neighbors of seed proteins are more likely to be identified. Although the optimal alpha value depends on the disease type, alpha=10 gives the best results for many diseases. The results of the algorithm for alpha values 2 and 10 are given in Tables G.1 and H.1. The subnetworks of PPIN, consisting of only OMIM & GWAS seed genes and top 100 proteins detected by DIAMOnD algorithm with alpha values 2 and 10 are given in Figures 4.49 and 4.50.

Table 4.20. Nine of the core proteins that are shared with the seed gene list.

Gene Name	Gene Description	Diseases Associated with Gene
AR	Androgen receptor	Androgen insensitivity, Androgen insensitivity, partial, with or without breast cancer, Hypospadias 1, Spinal and bulbar muscular atrophy of Kennedy, susceptibility to Prostate Cancer
AZGP1	Zinc-alpha-2-glycoprotein	
CDKN1B	Cyclin-dependent kinase inhibitor 1B	Multiple endocrine neoplasia
CHUK	Inhibitor of nuclear factor kappa-B kinase subunit alpha	Cocoon syndrome
FGFR2	Fibroblast growth factor receptor 2	Antley-Bixler syndrome without genital anomalies or disordered steroidogenesis, Apert syndrome, Beare-Stevenson cutis gyrata syndrome, Bent bone dysplasia syndrome, Craniofacial-skeletal-dermatologic dysplasia, Craniosynostosis, Crouzon syndrome, Gastric cancer, Jackson-Weiss syndrome, LADD syndrome, Pfeiffer syndrome, Saethre-Chotzen syndrome, Scaphocephaly and Axenfeld-Rieger anomaly, Scaphocephaly, maxillary retrusion, and mental retardation
KLK3	Prostate specific antigen	
MSMB	Beta-microseminoprotein	Prostate cancer
NKX3-1	Homebox protein	
PTEN	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase	Bannayan-Riley-Ruvalcaba syndrome, Cowden syndrome 1, Endometrial carcinoma, Lhermitte-Duclos syndrome, Macrocephaly/autism syndrome, Malignant melanoma, PTEN hamartoma tumor syndrome, Head and neck Squamous cell carcinoma, VATER association with macrocephaly and ventriculomegaly, Glioma susceptibility 2, Meningioma, Prostate cancer

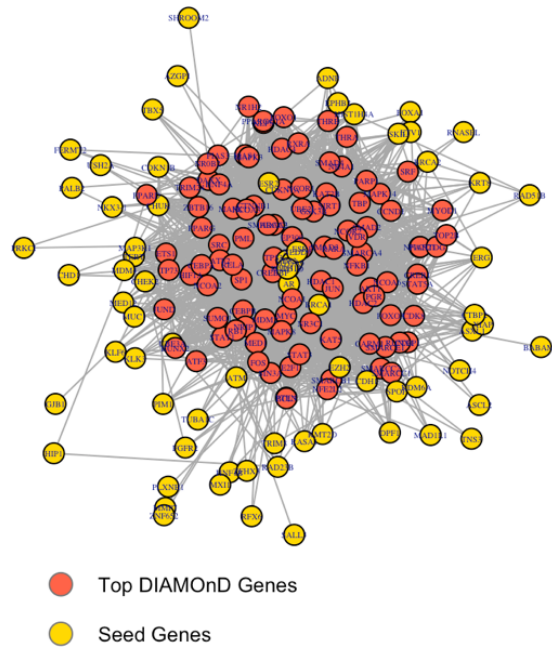


Figure 4.49. Seed and top 100 DIAMOnD genes for $\alpha=2$.

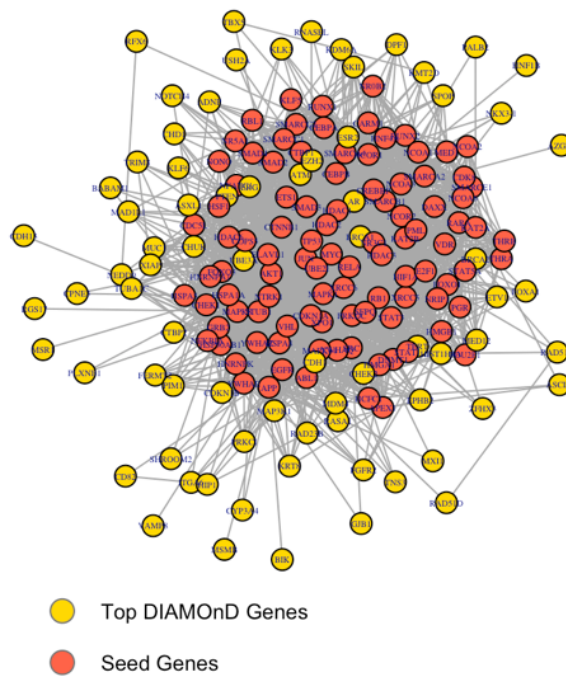


Figure 4.50. Seed and top 100 DIAMOnD genes for $\alpha=10$.

4.8.1. Identification of Diseases

Next, the diseases associated with the genes in Tables G.1 and H.1 were identified using DisGeNET, a platform for the human disease-associated genes. The analysis was done for the top 10, 25 and 100 genes for alpha values 2 and 10 (Tables 4.21-4.26). The score is computed by taking the number of type of sources and publications supporting the association.

Table 4.21. DisGeNET results for top 10 disease genes (alpha=2).

Disease	Entry Name	Score	PMIDs
C0035934	Rubinstein-Taybi Syndrome	0.709	50
C2239176	Liver carcinoma	0.672	599
C1261473	Sarcoma	0.641	173
C0206686	Adrenocortical carcinoma	0.621	50
C0205770	Choroid Plexus Papilloma	0.606	5
C0029463	Osteosarcoma	0.549	165
C0037286	Skin Neoplasms	0.543	61
C0085390	Li-Fraumeni Syndrome	0.529	315
C0038356	Stomach Neoplasms	0.526	56
C0007137	Squamous cell carcinoma	0.52	694

Table 4.22. DisGeNET results for top 25 disease genes (alpha=2).

Disease	Entry Name	Score	PMIDs
C0035934	Rubinstein-Taybi Syndrome	0.709	50
C0023487	Acute Promyelocytic Leukemia	0.707	423
C2239176	Liver carcinoma	0.672	599
C1261473	Sarcoma	0.641	173
C0206686	Adrenocortical carcinoma	0.621	50
C0025149	Medulloblastoma	0.607	21
C0205770	Choroid Plexus Papilloma	0.606	5
C0206711	Pilomatrixoma	0.602	13
C0029463	Osteosarcoma	0.549	165
C0037286	Skin Neoplasms	0.543	61

Table 4.23. DisGeNET results for top 100 disease genes (alpha=2).

Disease	Entry Name	Score	PMIDs
C0023467	Leukemia, Myelocytic, Acute	0.796	169
C0006413	Burkitt Lymphoma	0.765	234
C0008928	Cleidocranial Dysplasia	0.731	99
C0011860	Diabetes Mellitus, Non-Insulin-Dependent	0.729	157
C0028754	Obesity	0.727	224
C0035934	Rubinstein-Taybi Syndrome	0.709	50
C0345893	Juvenile polyposis syndrome	0.708	70
C0023487	Acute Promyelocytic Leukemia	0.707	423
C0035335	Retinoblastoma	0.706	157
C0342646	Vitamin D-Dependent Rickets, Type 2A	0.68	10

Table 4.24. DisGeNET results for top 10 disease genes (alpha=10).

Disease	Entry Name	Score	PMIDs
C0002395	Alzheimer's Disease	0.886	1797
C0035934	Rubinstein-Taybi Syndrome	0.709	50
C2239176	Liver carcinoma	0.672	599
C1261473	Sarcoma	0.641	173
C0206686	Adrenocortical carcinoma	0.621	50
C0205770	Choroid Plexus Papilloma	0.606	5
C0029463	Osteosarcoma	0.549	165
C0037286	Skin Neoplasms	0.543	61
C0085390	Li-Fraumeni Syndrome	0.529	315
C0038356	Stomach Neoplasms	0.526	56

Table 4.25. DisGeNET results for top 25 disease genes (alpha=10).

Disease	Entry Name	Score	PMIDs
C0002395	Alzheimer's Disease	0.886	1797
C0006413	Burkitt Lymphoma	0.765	234
C0035934	Rubinstein-Taybi Syndrome	0.709	50
C0342646	Vitamin D-Dependent Rickets, Type 2A	0.68	10
C2239176	Liver carcinoma	0.672	599
C1261473	Sarcoma	0.641	173
C0206686	Adrenocortical carcinoma	0.621	50
C0025149	Medulloblastoma	0.607	21
C0205770	Choroid Plexus Papilloma	0.606	5
C0206711	Pilomatrixoma	0.602	13

Table 4.26. DisGeNET results for top 100 disease genes ($\alpha=10$).

Disease	Entry Name	Score	PMIDs
C0002395	Alzheimer's Disease	0.886	1797
C0023467	Leukemia, Myelocytic, Acute	0.796	169
C0019562	Von Hippel-Lindau Syndrome	0.782	213
C0006413	Burkitt Lymphoma	0.765	234
C0008928	Cleidocranial Dysplasia	0.731	99
C0011860	Diabetes Mellitus, Non-Insulin-Dependent	0.729	157
C0035934	Rubinstein-Taybi Syndrome	0.709	50
C0345893	Juvenile polyposis syndrome	0.708	70
C0023487	Acute Promyelocytic Leukemia	0.707	423
C0035335	Retinoblastoma	0.706	157

For prostate cancer network, many cancer types were found to be related to the disease genes detected by the DIAMOnD algorithm. Liver carcinoma, sarcoma (cancer growing on connective tissue, such as cancellous bone) and adrenocortical carcinoma (cancer originating in cortex of adrenal gland) are nearly always present in the top 10 diseases whether the α value is 2 or 10. By the computational methods used in the present study, it appears that a patient with prostate carcinoma may develop liver carcinoma, sarcoma or adrenocortical carcinoma. Disease Ontology enrichments for top 100 genes identified by DIAMOnD algorithm (Figure 4.51) are in accordance to DisGeNET results. Other than prostate cancer, connective tissue and bone cancer are found to be related with these genes. Surprisingly, ovarian cancer is also identified as an enriched disease. As ovarian and prostate cancers are specific to females and males, respectively, the association of these two diseases were not predicted.

One study investigating the prostate cancer metastasis identified that 35% of the autopsies performed on prostate cancer patients had metastasis on other parts of their body. Most common type was found as bone (90%), followed by lung (46%), liver (25%), pleura (21%) and adrenals (13%) [208]. The diseases found by DIAMOnD algorithm complies with the study. Prostate cancer patients with liver and bone metastases have a low survival rate [209]. Additionally, as mentioned in Section 4.7, liver and prostate cancer share many of their favorable and unfavorable prognostic genes. From all of this information one can conclude that prostate cancer has a strong relation

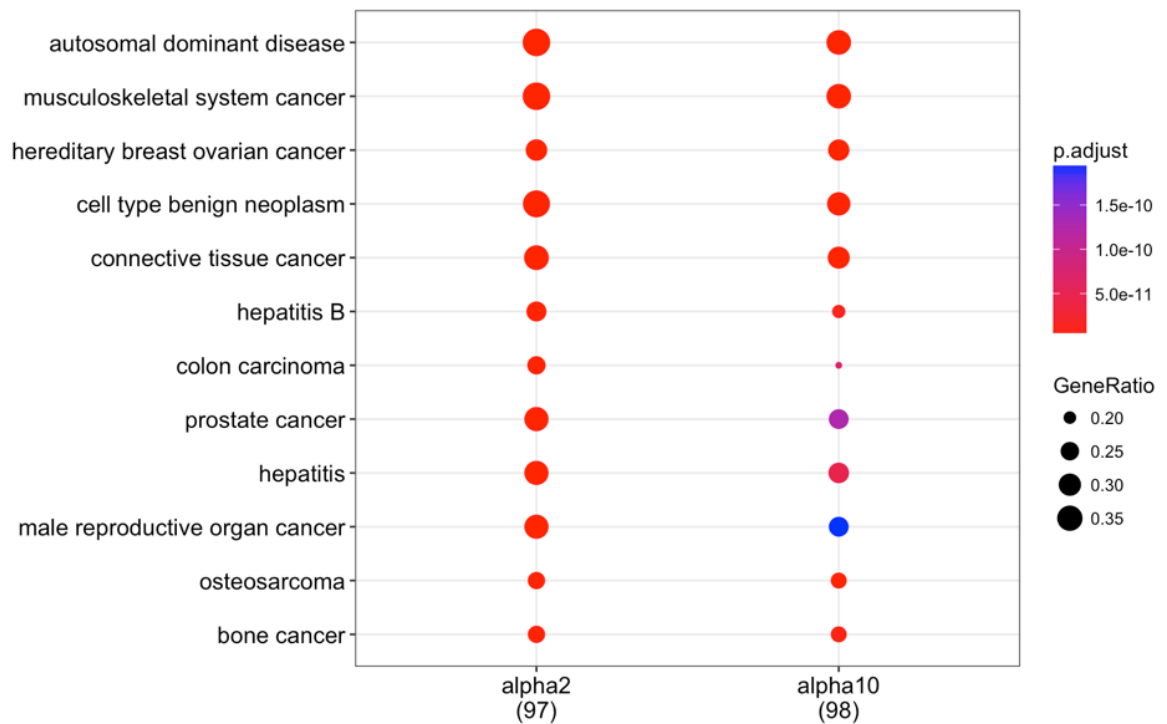


Figure 4.51. DO Enrichment comparison of genes for $\alpha=2$ and $\alpha=10$ identified by DIAMOnD algorithm.

with liver carcinoma and sarcoma. Another point where sarcoma term is encountered is the Disease Ontology enrichment analysis for cluster 10 (Figure 4.42). Prostate synovial sarcoma is a rare type of cancer and occurs in relatively younger patients [191]. Also DO enrichment analysis for Pca PPIN in Section 4.6.3.1 indicates that proteins somewhat related to prostate cancer play also roles in bone, connective tissue and lung cancer. Surprisingly, liver carcinoma was not one of the enriched diseases, since both DisGeNET and prognostic genes results shows a relation between liver and prostate cancer.

4.9. Crosstalk Analysis

Sphingolipids are important proteins in the mechanism of cancer and they regulate many processes taking part in cancer such as proliferation and cell growth. In order to identify the crosstalk between sphingolipid metabolism and Pca, inhibition and activation ligand centric networks are constructed. Ligand-centric networks are another

type of biologic networks in which the nodes (proteins) are connected by an edge if they bind to the same ligands. After the proteins in the sphingolipid metabolism were gathered, their ligands were obtained from ChEMBL database. If the ligands decrease the activity or concentration of the given protein, they were included in the inhibition network, and if they increase the activity, they were included in the activation network. Table 4.27 and 4.28 shows the numerical properties of inhibition and activation networks (Figure 4.52 and 4.53), respectively.

The inhibition and activation sphingolipid networks consist of 89 and 40 proteins, respectively. Gene expression data for genes encoding these proteins was screened and for all 129 genes, expression data is present. 35% and 40% of the the genes encoding the inhibition network proteins are up- and downregulated, respectively, and the rest of the genes are not significantly expressed. For activation network, these rates are 28% and 58%. Also, these two networks were compared with the Pca PPIN. 82% and 98% of the proteins in inhibition and activation networks are present in the PPIN. Most of the proteins that cannot be found in PPIN are not significantly expressed. (Table 4.29)

Table 4.27. Total number of proteins, ligands and interactions for the ligand-centric inhibition network for Sphingolipids.

No of Proteins	No of Ligands	No of Interactions
89	5707	911

Table 4.28. Total number of proteins, ligands and interactions for the ligand-centric activation network for Sphingolipids.

No of Proteins	No of Ligands	No of Interactions
40	289	509

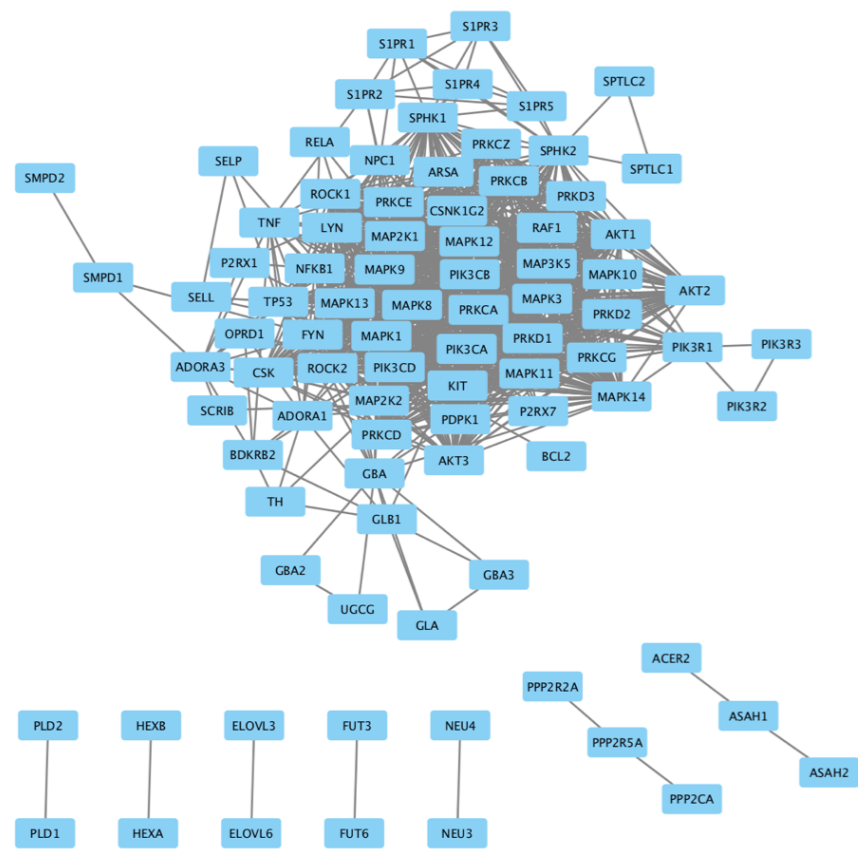


Figure 4.52. Ligand-centric inhibition network for Sphingolipids.

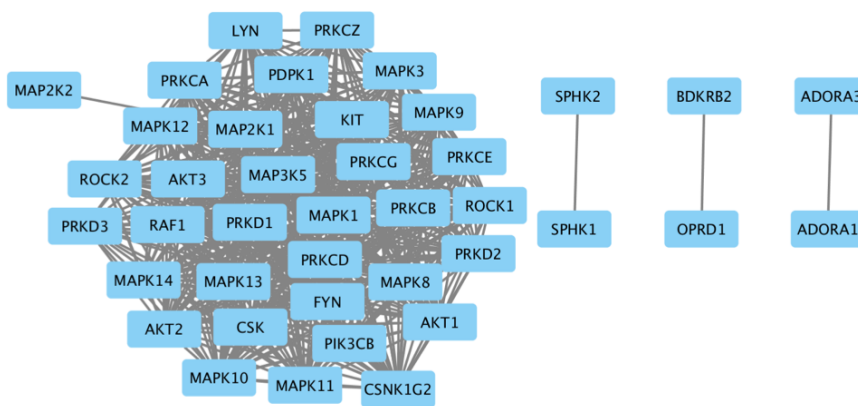


Figure 4.53. Ligand-centric activation network for Sphingolipids.

Table 4.29. Mutual proteins of Pca PPIN with inhibition and activation networks.

Inhibition Network/ PPIN Proteins (73/89) ²	Upregulated: ACER2* ³ , AKT3, ARSA, ASAH1, CSK, ELOVL6, GBA, GLA, GLB1, HEXB, MAP2K1, MAP3K5, MAPK1, MAPK13, MAPK9, NEU3*, NFKB1, PIK3CB, PPP2CA, PPP2R5A, PRKCD, PRKCZ, PRKD1, SCRIB, SELL, SMPD1, SMPD2, SPTLC1*, SPTLC2*, TP53, UGCG* (31)
	Downregulated: ADORA1, AKT1, AKT2, BDKRB2, CSNK1G2, FYN, HEXA, KIT, MAP2K2, MAPK10, MAPK11, MAPK12, MAPK14, MAPK3, NPC1, P2RX1, PDPK1, PIK3CA, PIK3CD, PIK3R1, PIK3R2, PLD1, PLD2, PRKCA, PRKCB, PRKD2, PRKD3, RAF1, RELA, ROCK2, S1PR1, S1PR3, SELL, SELP, SPHK1, SPHK2 (36)
	Not Significantly Expressed: ADORA3*, ASAH2*, BCL2, ELOVL3*, FUT3*, FUT6*, GBA2*, GBA3*, LYN, MAPK8, NEU4*, OPRD1, P2RX7, PIK3R3*, PPP2R2A, PRKCE, PRKCG, ROCK1, S1PR2, S1PR4, S1PR5*, TH, TNF (23)
Activation Network/ PPIN Proteins (39/40)	Upregulated: AKT3, CSK, MAP2K1, MAP3K5, MAPK1, MAPK13, MAPK9, PIK3CB, PRKCD, PRKCZ, PRKD1 (11)
	Downregulated: ADORA1, ADORA3*, AKT1, AKT2, BDKRB2, CSNK1G2, FYN, KIT, MAP2K2, MAPK10, MAPK11, MAPK12, MAPK14, MAPK3, PDPK1, PRKCA, PRKCB, PRKD2, PRKD3, RAF1, ROCK2, SPHK1, SPHK2 (23)
	Not Significantly Expressed: LYN, MAPK8, PRKCE, PRKCG, ROCK1 (5)

Sphingolipid signaling controls various molecular functions related to cancer development and progression. Since many of these functions causes cell death, cancer cells try to disrupt them by dysregulating the sphingolipid metabolism. Most important sphingolipid proteins associated with prostate cancer are acid ceramidase (ASAH1) and sphingosine kinase-1 (SPHK1). Ceramides, metabolized by ceramidases to generate sphingosines, are central molecules in sphingolipid pathways and take part in

²Total number of proteins in inhibition network is 89.

³*Proteins not included in the PPIN.

processes such as cell senescence and apoptosis. As cancer cells should avoid them for survival, acid ceramidase levels are high in prostate cancer cells, as converting ceramidase to sphingosine would decrease apoptosis [210]. In addition to upregulation of ASAH1 identified in this work, other studies remarked that as Gleason grade, a grading system for aggressiveness of prostate cancer, increases ASAH1 levels are also increased, suggesting tumor severity and ASAH1 expression are somehow linked [211].

Sphingosine-1-phosphate, a metabolite of sphingosine converted by SPHK1, is a promoter of cell proliferation and angiogenesis, which are two hallmarks of cancer. SPHK1, detected in both sphingolipid inhibition and activation networks, is an oncogene as acts as an anti-apoptotic and anti-inflammatory compound. SPHK1 mRNA levels are found higher in many cancer cells in comparison to normal cells, although differentially expressed genes analysis conducted in this work states SPHK1 is down-regulated in prostate cancer [212]. This result is in conflict with literature findings, as they state the expression of SPHK1 is increased in prostate cancer and high SPHK1 levels are associated with poor survival rates in both prostate cancer and breast cancer [213,214]. On the other hand, little is known about the homolog SPHK2 and its role in carcinogenesis. There are some studies claiming SPHK2 is apoptotic and some are suggesting it is anti-apoptotic [215,216]. As SPHK2 also converts sphingosine to sphingosine-1-phosphate, it promotes the cell proliferation and angiogenesis indirectly. The inhibition of SPHK2 leads to decrease in proliferation and castration-resistant prostate cancer cell viability [217]. Protein-kinase C (PKC) family members are key regulators of cell survival and overexpressed in prostate cancer, also confirmed by the DEG analysis and they are present in both inhibition and activation networks [218]. Activation of PKC results in inhibition of ceramide-mediated apoptosis and PKC's are directly linked with cancer. AKT3, a protein in PI3K-AKT signaling pathway, is overexpressed in prostate cancer cells lines and reported to promote the proliferation of cancer cells. The protein level in prostate tumors was found two times higher than the neighboring normal prostate cells, suggesting it may be a potential therapeutic target [219].

4.10. Identifying Drug Targets

Drug targets are molecules, generally proteins, that are critically involved in a disease process. Drugs either inhibit or activate the target for therapeutic effect by physically interacting with target or activate some signaling cascades. An important characteristic of a good drug targets is its specificity to the related process. Recent years, computational methods have been developed to identify new potential drug targets. Improved high-throughput technologies led to construction of biological models using networks such as PPI networks and gene regulatory networks [220].

Drug target identification by protein-protein interaction networks is a common method. PPI networks may help revealing important pathways and functions. Topological properties of networks such as degree and centrality may spot possible targets. Although disease proteins are not the ones with the highest degree and betweenness centrality, they are highly connected and not essential proteins like hubs and bottle-necks.

A flowchart describing the drug target identification algorithm is given in Figure 4.54. Topological characteristic of disease proteins related with prostate cancer obtained from OMIM and GWAS was used as elimination criteria of drug targets. In Section 4.2, topological feature differences between disease and other network proteins were compared. Degree, betweenness, normalized betweenness and eigenvector centrality were found to be higher, average shortest path lengths are found lower in disease proteins than other proteins (Table 4.8). However, the distributions of eccentricity and coreness did not show much difference (Figure 4.17 and 4.18). Therefore, they were not selected as elimination criteria. For other 5 feature the following threshold was determined. The proteins were selected if their feature is between the 40% and 60% percentile of the corresponding feature. For instance, the 40% and 60% quartiles for average short path lengths are calculated as 2.63 and 2.79, respectively. If the average short path length of a protein is not between these values, it is eliminated. 336 proteins held these criteria for all five topological characteristics. Then, their expression profiles

in prostate tumor were checked and among the genes encoding these 336 proteins, 152 are up- and 120 are downregulated. This reduces the number of possible drug targets to 272.

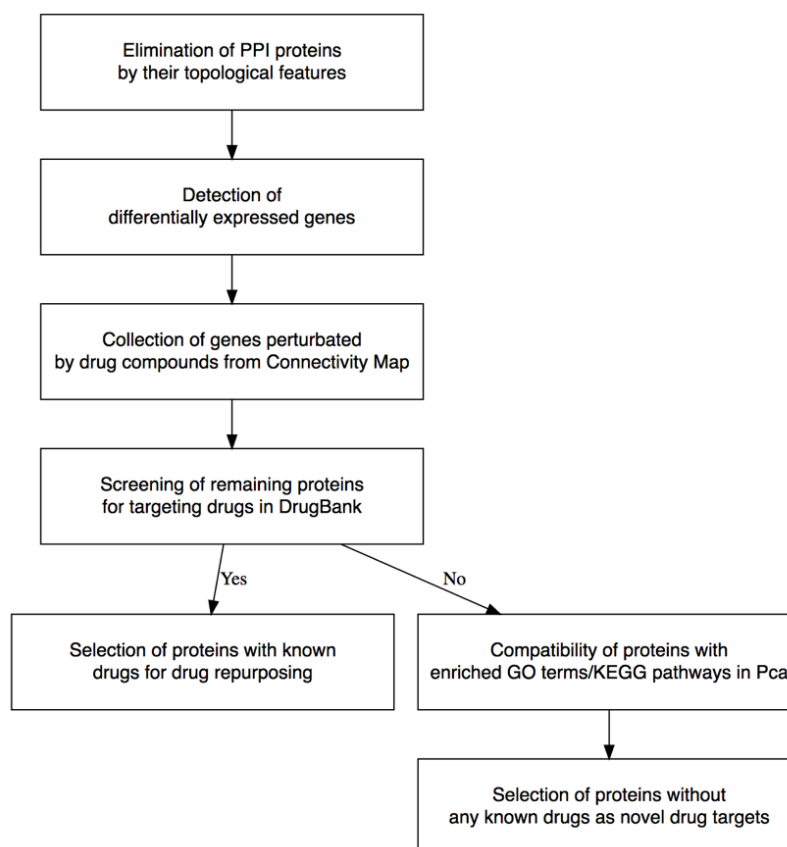


Figure 4.54. Flowchart of drug target identification algorithm.

Connectivity Map is a database which contains gene expression signatures from exposure of several cell types with perturbagens such as compounds and gene over-expressions. The cellular responses to these perturbations are used to identify the relationships between diseases, genes and drugs [112].

It is not known if the proteins show pro- or antitumor effects, so both negatively and positively correlated gene expression patterns associated with cells perturbed with drug are taken into account. The score threshold was chosen as 95%, meaning a gene is expected to be affected by at least one compound with a minimum score of 95%. The drug targets not matching these criteria were eliminated. At the end, 84 out

of 272 genes were found to be affected by compounds in cancer cell lines and taken as final potential drug targets.

To have a better idea about the characteristics of the above mentioned 272 potential drug targets, GO Biological Process Terms related to these proteins were found (Figure 4.55). Apart from general process terms such as metabolic and cellular process, cancer-related processes were found in remarkable amounts. Cell growth, proliferation, reproduction, immune system process and signaling are all associated with processes take part in hallmarks of cancer.

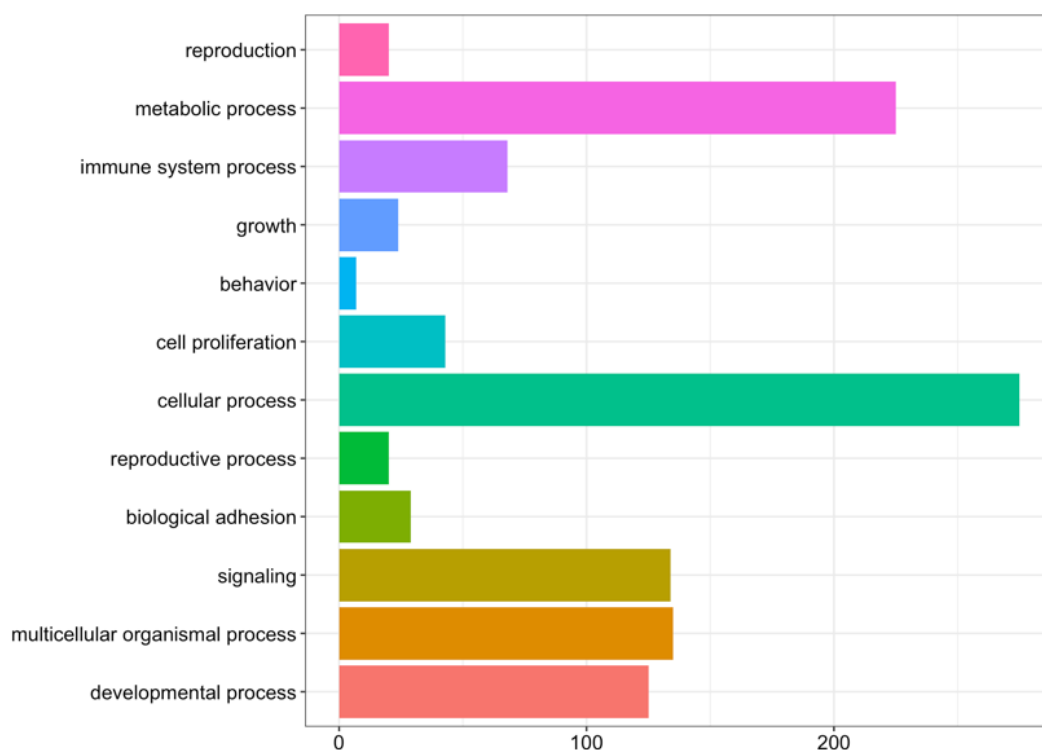


Figure 4.55. GO Biological Process terms associated with 272 drug targets.

GO Biological Process terms of the remaining drug targets (Figure 4.56) show a similar distribution as GO analysis of 272 drug targets (Figure 4.55), suggesting remaining drug targets are a presentation of the firstly identified targets.

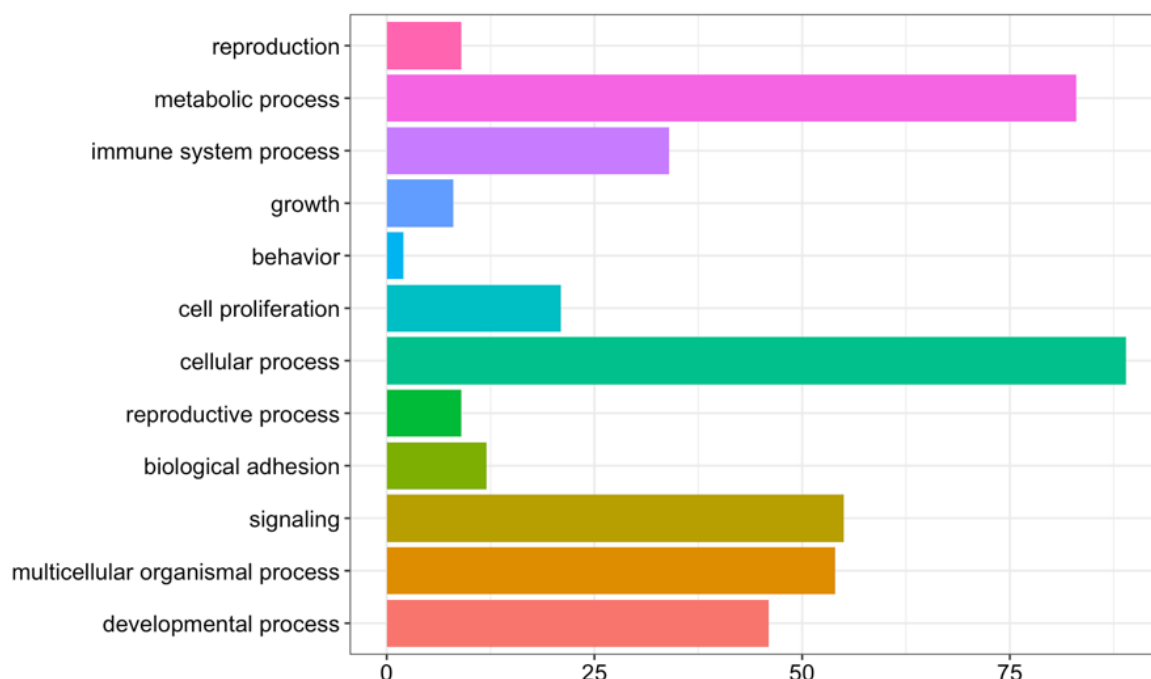


Figure 4.56. GO Biological Process terms associated with remaining 84 drug targets.

Remaining 84 proteins were screened in DrugBank, a database containing drugs and drug targets [221]. 29 of these proteins were found to be interacted with at least one drug (Table I.1). For six proteins, the targeting drugs have anti-tumor activities, they are for the treatment of various neoplasms such as leukemia, lung cancer and sarcoma. Drugs targeting these 29 proteins may be used for drug repurposing, the application of a known drug to a new disease. 55 proteins without any drug entries can here be proposed as new drug targets (Table J.1). GO Biological Process enrichments of these 55 proteins were regulation of MAPK signaling pathway, immune system process, regulation of apoptotic process and regulation of multicellular development. Among these enrichments, first 3 terms were also present in PPIN proteins and PPI hubs bottlenecks. As avoiding immune destruction and resisting cell death are common hallmarks of all cancer types, the enrichments seem reasonable.

One study by Özen *et al.* predicted CTNNB1P1, ITGB2, SOX9, KLF4, OCT4 and SOX2 as putative targets for prostate cancer computationally [222]. None of these targets were among the 55 proteins identified above. In addition, the experimentally

identified receptors mentioned in Section 2.2 and shown in Table 2.1 were not present in this work’s dataset.

Literature survey about disease drug targets were conducted and 39 of them were found to be reported as biomarkers or therapeutic targets for various types of cancer. 15 proteins were detected as candidate drug targets for several cancers such as leukemia and lung cancer. 16 of these targets were reported as potential targets for prostate cancer. Two out of 55 were stated as biomarkers for Pca, whereas five were biomarkers for other types of cancers (Table J.1).

ChEMBL database [223] was screened for ligand-target interactions. 24 targets were found to have at least one bioactivity entry. Among them, nine targets, which were reported as therapeutic target for Pca, have ligands binding them.

Thus, the ones proposed in this work need further experimental and computational investigations for the validation of these promising novel targets.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

Prostate cancer, the development of abnormal cells in prostate tissue, is one of the most common cancers worldwide. The risk of developing Pca is increased with age, family history and diet. Nowadays, several methods such as biopsy, PSA blood screening and MRI scans are used to detect Pca. For high survival rates, early diagnosis and therapy are crucial. To acquire a better understanding about the mechanisms that lead to Pca, network theory was employed. Biological networks can contribute new information about the mechanisms behind biological processes as they provide a systematic view on pathways.

The structure of any biological network can be predicted with the well-understood graph-theoretical analysis. In this work, protein-protein interaction (PPI) and Co-Expression (Co-Exp) networks were reconstructed using the available data in the literature. To construct a PPI network, a core set of proteins were acquired using three different methods. The mutual Gene Ontology (GO) terms of all three aspects were listed, and interaction partners of core set proteins were checked if they share the same GO terms. Addition of interaction partners were repeated until there is no interaction partner left. The resulting PPI network consisted of 8709 proteins and over 160,000 interactions. Graph theoretical analysis showed that the network has scale-free topology, which is a feature of biological networks. The distributions of the topological indices are in accordance with the literature values. Co-Exp network was constructed using the Weighted Correlation Network Analysis (WGCNA) package on transcriptomics data from The Cancer Genome Atlas (TCGA). Firstly, the differentially expressed genes between prostate tumor cells from TCGA and normal prostate tissue cells obtained from GTEx database. Around 15000 genes were determined as DEGs. These genes were used for network construction and modules were detected, in which the genes show similar expression patterns.

Gene Ontology (GO) enrichments for PPIN proteins were regulation of apoptotic signaling pathway and regulation of serine/threonine kinase activities, whereas immune system regulating signaling pathway terms were enriched in PPIN modules. As in PPIN proteins, regulation of apoptotic signaling pathway term was also present in enrichment analysis of PPI hubs and bottlenecks. Some of the Co-Expression modules were enriched in processes related to muscle contraction and neutrophil activation.

In KEGG Pathway enrichment analysis, PPIN proteins and modules were found to be enriched in kinase signaling networks such as MAPK, Ras and ErbB. However, in PPI hubs and bottlenecks the enriched terms were related to viral diseases.

Gene Set Enrichment Analysis (GSEA) conducted on normal and tumor prostate cells using Hallmark Gene Sets from MSigDB. Tumor cells were found to be enriched in processes associated with hallmarks of cancer such as apoptosis, angiogenesis and proliferation.

DIAMOnD is an algorithm for detecting disease modules around specified disease genes. Prostate cancer disease genes were obtained from GWAS and OMIM, and top 100 disease proteins were detected. DisGeNET was used for disease identification related with disease proteins. Also, Disease Ontology (DO) enrichments were conducted on PPIN proteins, PPIN modules, Co-Exp modules and hubs bottlenecks of the PPIN. All of the analysis showed that there is a strong relation between Pca and sarcoma, liver and bone cancer. Literature screening was also in agreement with this study's findings.

Another type of networks, ligand-centric networks were also constructed for the crosstalk analysis between sphingolipids and Pca. Important sphingolipid metabolites such as ASAH1 and SPHK1 were detected in ligand-centric and PPI networks. High expressions of both proteins are associated with poor survival.

As the last step, drug potential drug targets were identified. Using 40 and 60% percentiles of the topological indices of disease genes found in graph theoretical analysis, PPIN proteins were eliminated if they are outside the bounds. Then, the expression levels of genes encoding these proteins were checked for differential expression and 272 targets remained. To decrease the number of the targets further, the proteins were screened in the Connectivity Map (C-Map). This tool uses a pattern matching algorithm to find the similarities in gene expressions, thus linking diseases with compounds like drugs. The remaining 84 proteins were screened in DrugBank, and 55 of them had no interacting drugs. 55 proteins were enriched in regulation of apoptotic processes, regulation of MAPK signaling pathway and immune system process as found in PPIN proteins and PPI hubs bottlenecks. 18 of them were reported as either potential biomarkers or therapeutic targets. These 55 proteins need to be further investigated to be considered as novel drug targets.

5.2. Recommendations

For a more tissue-specific PPI network, the parent terms of Gene Ontology can be discarded. That way, the interactome size will be smaller and more prostate specific. The parent terms such as “binding” are too general, and it may cause two irrelevant molecular functions to act like they are related and increase the number of elimination GO terms.

Another way to improve the prediction of drug targets is to compare the gene expression of drug target encoding genes across other tissue and cancer types. As it stands, it is unrevealed whether the differential expressions of the identified drug targets’ genes are characteristic to prostate cancer.

Co-expression genes in one cluster may be regulated by the same transcriptional factors. Co-expressed modules found in this work can be screened for transcription factors and new transcription factor targets may be possibly identified.

Network decomposition can be implemented so that the PPI networks are decomposed to linear paths which may give us insights about unknown members of known pathways or yet not identified signaling pathways.

REFERENCES

1. Siegel, R. L., K. D. Miller and A. Jemal, “Cancer statistics, 2018”, *CA: A Cancer Journal for Clinicians*, Vol. 68, No. 1, pp. 7–30, 2018.
2. Kidd, B. A., L. A. Peters, E. E. Schadt *et al.*, “Unifying immunology with informatics and multiscale biology.”, *Nature Immunology*, Vol. 15, No. 2, pp. 118–27, 2014.
3. Vital-Lopez, F. G., V. Memišević and B. Dutta, “Tutorial on biological networks”, Vol. 2, No. 4, pp. 298–325, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2012.
4. Sharan, R. and T. Ideker, “Modeling cellular machinery through biological network comparison”, *Nature Biotechnology*, Vol. 24, No. 4, pp. 427–433, 2006.
5. Ferlay, J., I. Soerjomataram, R. Dikshit *et al.*, “Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012”, *International Journal of Cancer*, Vol. 136, No. 5, pp. E359–E386, 2015.
6. Gültekin, M., G. Boztaş, E. Ş. Utku *et al.*, *Türkiye Kanser İstatistikleri*, Tech. rep., T.C. Sağlık Bakanlığı, Türkiye Halk Sağlığı Kurumu, Ankara, 2016, [http://www.onkoloji.gov.tr/attachments/article/8653/Ana%20Rapor%202016%20\(v01.2\).pdf](http://www.onkoloji.gov.tr/attachments/article/8653/Ana%20Rapor%202016%20(v01.2).pdf), accessed at July, 2018.
7. Gültekin, M., G. Boztaş, E. Ş. Utku *et al.*, *Ölüm Nedeni İstatistikleri*, Tech. rep., Türkiye İstatistik Kurumu, 2017, http://www.tuik.gov.tr/PreTablo.do?alt_id=1083, accessed at July, 2018.

8. Çalışkan, S. and M. Sungur, "Scientific contribution of Turkey in prostate literature. Where are we? Where are we going?", *Journal of Oncological Sciences*, 2017.
9. Robinson, D., E. M. Van Allen, Y. M. Wu *et al.*, "Integrative clinical genomics of advanced prostate cancer", *Cell*, 2015.
10. Rivera-Calderón, L. G., C. E. Fonseca-Alves, P. E. Kobayashi *et al.*, "Alterations in PTEN, MDM2, TP53 and AR protein and gene expression are associated with canine prostate carcinogenesis", *Research in Veterinary Science*, 2016.
11. Singh, D., P. G. Febbo, K. Ross *et al.*, "Gene expression correlates of clinical prostate cancer behavior", *Cancer Cell*, 2002.
12. Lotan, T. L., F. Carvalho, S. B. Peskoe *et al.*, "PTEN loss is associated with up-grading of prostate cancer from biopsy to radical prostatectomy", *Modern Pathology*, 2015.
13. McCubrey, J. A., K. Lertpiriyapong, T. L. Fitzgerald *et al.*, "Roles of TP53 in determining therapeutic sensitivity, growth, cellular senescence, invasion and metastasis", *Advances in Biological Regulation*, 2017.
14. Heinlein, C. A. and C. Chang, "Androgen receptor in prostate cancer", *Endocrine Reviews*, 2004.
15. Stanbrough, M., G. J. Bubley, K. Ross *et al.*, "Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer", *Cancer Research*, Vol. 66, No. 5, pp. 2815–2825, 2006.
16. Liu, Y., X. Hu, H. Liu *et al.*, "A comparative study of radiolabeled bombesin analogs for the PET imaging of prostate cancer", *Journal of Nuclear Medicine*, Vol. 54, No. 12, pp. 2132–2138, 2013.

17. Ananias, H. J. K., Z. Yu, H. D. Hoving *et al.*, “Application of 99mTechnetium-HYNIC(tricine/TPPTS)-Aca-Bombesin(7-14) SPECT/CT in prostate cancer patients: A first-in-man study”, *Nuclear Medicine and Biology*, Vol. 40, pp. 933–938, 2013.
18. Schroeder, R. P. J., C. Müller, S. Reneman *et al.*, “A standardised study to compare prostate cancer targeting efficacy of five radiolabelled bombesin analogues”, *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 37, No. 7, pp. 1386–1396, 2010.
19. Mankoff, D. A., J. M. Link, H. M. Linden *et al.*, “Tumor receptor imaging”, *Journal of Nuclear Medicine*, Vol. 49, No. Suppl.2, pp. 149S–163S, 2008.
20. Leach, D. and G. Buchanan, “Stromal androgen receptor in prostate cancer development and progression”, *Cancers*, Vol. 9, No. 1, p. 10, 2017.
21. Wu, W. F., L. Maneix, J. Insunza *et al.*, “Estrogen receptor β , a regulator of androgen receptor signaling in the mouse ventral prostate”, *Proceedings of the National Academy of Sciences*, Vol. 114, No. 19, pp. E3816–E3822, 2017.
22. Wojciuk, G. and M. Kruszewski, “[DTPA-(PABn)-Leu5]-des-acyl ghrelin(1–5) as a new carrier of radionuclides and potential precursor of radiopharmaceuticals”, *Nuclear Medicine Communications*, Vol. 39, No. 2, 2018.
23. Morichetti, D., R. Mazzucchelli, D. Stramazzotti *et al.*, “Immunohistochemical expression of somatostatin receptor subtypes in prostate tissue from cystoprostatectomies with incidental prostate cancer”, *BJU International*, 2010.
24. Couture, F., C. Levesque, V. Dumulon-Perreault *et al.*, “PACE4-based molecular targeting of prostate cancer using an engineered 64Cu-radiolabeled peptide inhibitor”, *Neoplasia*, Vol. 16, No. 8, pp. 634–643, 2014.

25. Wang, M., S. Wey, Y. Zhang *et al.*, “Role of the unfolded protein response regulator GRP78/BiP in development, cancer, and neurological disorders”, *Antioxidants & Redox Signaling*, Vol. 11, No. 9, pp. 2307–2316, 2009.
26. Ferrara, F., D. I. Staquicini, W. H. P. Driessen *et al.*, “Targeted molecular-genetic imaging and ligand-directed therapy in aggressive variant prostate cancer”, *Proceedings of the National Academy of Sciences*, Vol. 113, No. 45, pp. 12786–12791, 2016.
27. Demirkol, M. O., O. Acar, B. Uçar *et al.*, “Prostate-specific membrane antigen-based imaging in prostate cancer: impact on clinical decision making process”, *Prostate*, Vol. 75, No. 7, pp. 748–757, 2015.
28. Flores, O., S. Santra, C. Kaittanis *et al.*, “PSMA-targeted theranostic nanocarrier for prostate cancer”, *Theranostics*, Vol. 7, No. 9, pp. 2477–2494, 2017.
29. Maurer, B. J., L. S. Metelitsa, R. C. Seeger *et al.*, “Increase of ceramide and induction of mixed apoptosis/necrosis by N-(4-Hydroxyphenyl)- retinamide in neuroblastoma cell lines”, *Journal of the National Cancer Institute*, 1999.
30. Subramanian, N., P. Torabi-Parizi, R. A. Gottschalk *et al.*, “Network representations of immune system complexity”, *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, Vol. 7, No. 1, pp. 13–38, 2015.
31. Lacroix, V., L. Cottret, P. Thebault *et al.*, “An introduction to metabolic networks and their structural analysis”, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 5, No. 4, pp. 1–44, 2008.
32. Misra, B. B. and J. J. van der Hooft, “Updates in metabolomics tools and resources: 2014-2015”, *Electrophoresis*, Vol. 37, No. 1, pp. 86–110, 2016.
33. Garcia, A. and C. Barbas, “Gas chromatography-mass spectrometry (GC-MS)-based metabolomics.”, *Methods in Molecular Biology*, Vol. 708, pp. 191–204, 2011.

34. Zhou, B., J. F. Xiao, L. Tuli *et al.*, “LC-MS-based metabolomics”, *Molecular Biosystems*, Vol. 8, No. 2, pp. 470–481, 2012.
35. Kirwan, J. A., R. J. M. Weber, D. I. Broadhurst *et al.*, “Direct infusion mass spectrometry metabolomics dataset: a benchmark for data processing and quality control.”, *Scientific Data*, Vol. 1, p. 140012, 2014.
36. Fagerer, S. R., S. Nielsen, A. Ibanez *et al.*, “Matrix-assisted laser desorption/ionization matrices for negative mode metabolomics”, *European Journal of Mass Spectrometry*, Vol. 19, No. 1, pp. 39–47, 2013.
37. Hölscher, D., R. Shroff, K. Knop *et al.*, “Matrix-free UV-laser desorption/ionization (LDI) mass spectrometric imaging at the single-cell level: Distribution of secondary metabolites of *Arabidopsis thaliana* and *Hypericum* species”, *Plant Journal*, Vol. 60, No. 5, pp. 907–918, 2009.
38. Lee, D. S., H. Burd, J. Liu *et al.*, “Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets”, *Journal of Bacteriology*, Vol. 191, No. 12, pp. 4015–4024, 2009.
39. Fang, K., H. Zhao, C. Sun *et al.*, “Exploring the metabolic network of the epidemic pathogen *Burkholderia cenocepacia* J2315 via genome-scale reconstruction.”, *BMC Systems Biology*, Vol. 5, No. 1, p. 83, 2011.
40. Mazumdar, V., E. S. Snitkin, S. Amar *et al.*, “Metabolic network model of a human oral pathogen”, *Journal of Bacteriology*, Vol. 91, No. 1, pp. 74–90, 2009.
41. Wilczynski, B. and E. E. Furlong, “Challenges for modeling global gene regulatory networks during development: Insights from *Drosophila*”, *Developmental Biology*, Vol. 340, No. 2, pp. 161–169, 2010.

42. Arda, H. E. and A. J. Walhout, “Gene-centered regulatory networks”, *Briefings in Functional Genomics and Proteomics*, Vol. 9, No. 1, pp. 4–12, 2009.
43. Park, P. J., “ChIP-seq: advantages and challenges of a maturing technology.”, *Nature Reviews. Genetics*, Vol. 10, No. 10, pp. 669–80, 2009.
44. Chai, L. E., S. K. Loh, S. T. Low *et al.*, “A review on the computational approaches for gene regulatory network construction”, *Computers in Biology and Medicine*, Vol. 48, No. 1, pp. 55–65, 2014.
45. Calvano, S. E., W. Xiao, D. R. Richards *et al.*, “A network-based analysis of systemic inflammation in humans”, *Nature*, Vol. 437, No. 7061, pp. 1032–1037, 2005.
46. Seok, J., W. Xiao, L. L. Moldawer *et al.*, “A dynamic network of transcription in LPS-treated human subjects”, *BMC Systems Biology*, Vol. 3, No. 1, p. 78, 2009.
47. Ramsey, S. A., S. L. Klemm, D. E. Zak *et al.*, “Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics”, *PLoS Computational Biology*, Vol. 4, No. 3, 2008.
48. De Las Rivas, J. and C. Fontanillo, “Protein-protein interaction networks: unraveling the wiring of molecular machines within the cell.”, *Briefings in Functional Genomics*, Vol. 11, No. 6, pp. 489–96, 2012.
49. Zoraghi, R. and N. E. Reiner, “Protein interaction networks as starting points to identify novel antimicrobial drug targets”, *Current Opinion in Microbiology*, Vol. 16, No. 5, pp. 566–572, 2013.
50. Rao, V. S., K. Srinivas, G. N. Sujini *et al.*, “Protein-protein interaction detection: Methods and analysis”, *International Journal of Proteomics*, Vol. 2014, No. ii, pp. 1–12, 2014.

51. Gavin, A.-C., M. Bösch, R. Krause *et al.*, “Functional organization of the yeast proteome by systematic analysis of protein complexes.”, *Nature*, Vol. 415, No. 6868, pp. 141–147, 2002.
52. Urh, D. S., M. and K. Zhao, “Affinity chromatography: general methods.”, *Methods in Enzymology*, Vol. 463, No. 09, pp. 417–38, 2009.
53. Lee, C., “Coimmunoprecipitation assay.”, *Methods in Molecular Biology*, Vol. 362, pp. 401–6, 2007.
54. Michnick, S. W., P. H. Ear, C. Landry *et al.*, “Protein-fragment complementation assays for large-scale analysis, functional dissection and dynamic studies of protein-protein interactions in living cells.”, *Methods in Molecular Biology*, Vol. 756, pp. 395–425, 2011.
55. Rohila, J. S., M. Chen, R. Cerny *et al.*, “Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants”, *Plant Journal*, Vol. 38, No. 1, pp. 172–181, 2004.
56. Tong, A. H., M. Evangelista, A. B. Parsons *et al.*, “Systematic genetic analysis with ordered arrays of yeast deletion mutants.”, *Science*, Vol. 294, No. 5550, pp. 2364–8, 2001.
57. Uetz, P., L. Giot, G. Cagney *et al.*, “A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*.”, *Nature*, Vol. 403, No. 6770, pp. 623–627, 2000.
58. Ooi, S. L., X. Pan, B. D. Peyser *et al.*, “Global synthetic-lethality analysis and yeast functional profiling”, *Trends in Genetics*, Vol. 22, No. 1, pp. 56–63, 2006.
59. Valente, G. T., M. L. Acencio, C. Martins *et al.*, “The development of a universal in silico predictor of protein-protein interactions”, *PLoS ONE*, Vol. 8, No. 5, 2013.

60. Lu, L., H. Lu and J. Skolnick, "Multiprospector: An algorithm for the prediction of protein-protein interactions by multimeric threading", *Proteins: Structure, Function and Genetics*, Vol. 49, No. 3, pp. 350–364, 2002.
61. Pazos, F. and A. Valencia, "In silico two-hybrid system for the selection of physically interacting protein pairs", *Proteins: Structure, Function and Genetics*, Vol. 47, No. 2, pp. 219–227, 2002.
62. Sato, T., Y. Yamanishi, M. Kanehisa *et al.*, "Improvement of the Mirrortree Method by Extracting Evolutionary Information", *Sequence and Genome Analysis: Method and Applications*, pp. 129–139, 2011.
63. Srinivas, K., A. A. Rao, G. R. Sridhar *et al.*, "Methodology for phylogenetic tree construction", *Journal of Proteomics & Bioinformatics*, Vol. 1, pp. 5–11, 2008.
64. Grigoriev, A., "A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*", *Nucleic Acids Research*, Vol. 29, No. 17, pp. 3513–3519, 2001.
65. Elzawahry, A., A. Patil, Y. Kumagai *et al.*, "Innate immunity interactome dynamics", *Gene Regulation and Systems Biology*, Vol. 2014, No. 8, 2014.
66. Acuner-Ozbabacan, E., B. Engin, E. Guven-Maiorov *et al.*, "The structural network of Interleukin-10 and its implications in inflammation and cancer.", *BMC genomics*, Vol. 15 Suppl 4, No. Suppl 4, p. S2, 2014.
67. Jin, S., Y. Li, R. Pan *et al.*, "Characterizing and controlling the inflammatory network during influenza A virus infection.", *Scientific Reports*, Vol. 4, p. 3799, 2014.
68. Mani, R., R. P. St Onge, J. L. Hartman *et al.*, "Defining genetic interaction.", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 105, No. 9, pp. 3461–6, 2008.

69. Tong, A. H., G. Lesage, G. D. Bader *et al.*, “Global mapping of the yeast genetic interaction network”, *Science*, Vol. 303, No. 5659, pp. 808–813, 2004.
70. Ooi, S. L., D. D. Shoemaker and J. D. Boeke, “DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray”, *Nature Genetics*, Vol. 35, No. 3, pp. 277–286, 2003.
71. Schuldiner, M., S. R. Collins, N. J. Thompson *et al.*, “Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile”, *Cell*, Vol. 123, No. 3, pp. 507–519, 2005.
72. Bommi-Reddy, A., I. Almeciga, J. Sawyer *et al.*, “Kinase requirements in human cells: III. Altered kinase requirements in VHL-/- cancer cells detected in a pilot synthetic lethal screen.”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 105, No. 43, pp. 16484–16489, 2008.
73. Fedeles, S. V., X. Tian, A.-R. Gallagher *et al.*, “A genetic interaction network of five genes for human polycystic kidney and liver diseases defines polycystin-1 as the central determinant of cyst formation.”, *Nature Genetics*, Vol. 43, No. 7, pp. 639–647, 2011.
74. Ragni, E., H. Piberger, C. Neupert *et al.*, “The genetic interaction network of CCW12, a *Saccharomyces cerevisiae* gene required for cell wall integrity during budding and formation of mating projections.”, *BMC Genomics*, Vol. 12, p. 107, 2011.
75. Hyduke, D. R. and B. Ø. Palsson, “Towards genome-scale signalling network reconstructions.”, *Nature Reviews. Genetics*, Vol. 11, No. 4, pp. 297–307, 2010.
76. Papin, J. A., T. Hunter, B. Ø. Palsson *et al.*, “Reconstruction of cellular signalling networks and analysis of their properties”, *Nature Reviews. Molecular Cell Biology*, Vol. 6, No. 2, pp. 99–111, 2005.

77. Ideker, T., V. Thorsson, J. A. Ranish *et al.*, “Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.”, *Science*, Vol. 292, No. 5518, pp. 929–934, 2001.
78. Lum, L., S. Yao, B. Mozer *et al.*, “Identification of Hedgehog pathway components by RNAi in *Drosophila* cultured cells.”, *Science*, Vol. 299, No. 5615, pp. 2039–45, 2003.
79. Nielsen, U. B., M. H. Cardone, A. J. Sinskey *et al.*, “Profiling receptor tyrosine kinase activation by using Ab microarrays.”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 100, No. 16, pp. 9330–5, 2003.
80. Meyer, T. and M. N. Teruel, “Fluorescence imaging of signaling networks”, *Trends in Cell Biology*, Vol. 13, No. 2, pp. 101–106, 2003.
81. Li, S., L. Wang, M. Berman *et al.*, “Mapping a dynamic innate immunity protein interaction network regulating type I interferon production”, *Immunity*, Vol. 35, No. 3, pp. 426–440, 2011.
82. Chevrier, N., P. Mertins, M. N. Artyomov *et al.*, “Systematic discovery of TLR signaling components delineates viral-sensing circuits”, *Cell*, Vol. 147, No. 4, pp. 853–867, 2011.
83. Lee, M. N., M. Roy, S.-E. Ong *et al.*, “Identification of regulators of the innate immune response to cytosolic DNA and retroviral infection by an integrative approach.”, *Nature Immunology*, Vol. 14, No. 2, pp. 179–85, 2013.
84. Barabási, A.-L., N. Gulbahce and J. Loscalzo, “Network medicine: a network-based approach to human disease”, *Nature Reviews Genetics*, 2011.
85. Goh, K.-I., M. E. Cusick, D. Valle *et al.*, “The human disease network”, *Proceedings of the National Academy of Sciences of the United States of America*, 2007.

86. Gault, C. R., L. M. Obeid and Y. A. Hannun, "An overview of sphingolipid metabolism: From synthesis to breakdown", *Advances in Experimental Medicine and Biology*, Vol. 688, pp. 1–23, 2010.
87. Maceyka, M. and S. Spiegel, "Sphingolipid metabolites in inflammatory disease.", *Nature*, Vol. 510, No. 7503, pp. 58–67, 2014.
88. Sukocheva, O. and C. Wadham, "Role of sphingolipids in oestrogen signalling in breast cancer cells: An update", *Journal of Endocrinology*, 2014.
89. Beckham, T. H., C. J. Cheng, S. T. Marrison *et al.*, "Interdiction of sphingolipid metabolism to improve standard cancer therapies", *Advances in Cancer Research*, Vol. 117, pp. 1–36, 2014.
90. Xiao, C. and S. Ghosh, "NF-kappaB, an evolutionarily conserved mediator of immune and inflammatory responses", *Advances in Experimental Medicine and Biology*, Vol. 560, pp. 41–45, 2005.
91. Presa, N., A. Gomez-Larrauri, I. G. Rivera *et al.*, "Regulation of cell migration and inflammation by ceramide 1-phosphate", *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, Vol. 1861, No. 5, pp. 402–409, 2016.
92. Nixon, G. F., "Sphingolipids in inflammation: Pathological implications and potential therapeutic targets", *British Journal of Pharmacology*, Vol. 158, No. 4, pp. 982–993, 2009.
93. Modrak, D. E., D. V. Gold and D. M. Goldenberg, "Sphingolipid targets in cancer therapy", *Molecular Cancer Therapeutics*, 2006.
94. Ségui, B., N. Andrieu-Abadie, J.-P. Jaffrézou *et al.*, "Sphingolipids as modulators of cancer cell death: potential therapeutic targets.", *Biochimica et Biophysica Acta*, 2006.

95. Spiegel, S. and S. Milstien, “The outs and the ins of sphingosine-1-phosphate in immunity.”, *Nature Reviews. Immunology*, Vol. 11, No. 6, pp. 403–15, 2011.
96. Rivera, J., R. L. Proia and A. Olivera, “The alliance of sphingosine-1-phosphate and its receptors in immunity”, *Nature Reviews. Immunology*, Vol. 8, No. 10, pp. 753–763, 2008.
97. Chatr-Aryamontri, A., R. Oughtred, L. Boucher *et al.*, “The BioGRID interaction database: 2017 update”, *Nucleic Acids Research*, Vol. 45, No. D1, pp. D369–D379, 2017.
98. Orchard, S., M. Ammari, B. Aranda *et al.*, “The MIntAct project - IntAct as a common curation platform for 11 molecular interaction databases”, *Nucleic Acids Research*, Vol. 42, No. D1, 2014.
99. Prasad, T. S. K., R. Goel, K. Kandasamy *et al.*, “Human Protein Reference Database — 2009 update”, *Nature Biotechnology*, Vol. 37, pp. 767–772, 2009.
100. Franceschini, A., D. Szklarczyk, S. Frankild *et al.*, “STRING v9.1: Protein-protein interaction networks, with increased coverage and integration”, *Nucleic Acids Research*, Vol. 41, No. D1, 2013.
101. Shaul, Y. D., B. Yuan, P. Thiru *et al.*, “MERAV: A tool for comparing gene expression across human tissues and cell types”, *Nucleic Acids Research*, Vol. 44, No. D1, pp. D560–D566, 2016.
102. Carithers, L. J. and H. M. Moore, “The Genotype-Tissue Expression (GTEx) Project”, *Biopreservation and Biobanking*, Vol. 13, No. 5, pp. 307–308, 2015.
103. Pontén, F., K. Jirström and M. Uhlen, “The Human Protein Atlas—a tool for pathology.”, *The Journal of Pathology*, Vol. 216, No. 4, pp. 387–93, 2008.

104. Tomczak, K., P. Czerwińska and M. Wiznerowicz, “The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge”, *Współczesna Onkologia*, Vol. 1A, pp. A68–A77, 2015.
105. Services, H., “GEO : The Gene Expression Omnibus”, *Gene Expression*, Vol. 23, pp. 2–3, 2007.
106. Pachter, L., *Models for transcript quantification from RNA-Seq*, Apr 2011.
107. Hamosh, A., A. F. Scott, J. S. Amberger *et al.*, “Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders”, *Nucleic Acids Research*, Vol. 33, No. DATABASE ISS., 2005.
108. Beck, T., R. K. Hastings, S. Gollapudi *et al.*, “GWAS Central: A comprehensive resource for the comparison and interrogation of genome-wide association studies”, *European Journal of Human Genetics*, Vol. 22, No. 7, pp. 949–952, 2014.
109. Ramos, E. M., D. Hoffman, H. A. Junkins *et al.*, “Phenotype-genotype integrator (PheGenI): Synthesizing genome-wide association study (GWAS) data with existing genomic resources”, *European Journal of Human Genetics*, 2014.
110. Kanehisa, M., Y. Sato, M. Kawashima *et al.*, “KEGG as a reference resource for gene and protein annotation”, *Nucleic Acids Research*, Vol. 44, No. D1, pp. D457–D462, 2016.
111. Breuer, K., A. K. Foroushani, M. R. Laird *et al.*, “InnateDB: Systems biology of innate immunity and beyond - Recent updates and continuing curation”, *Nucleic Acids Research*, Vol. 41, No. D1, 2013.
112. Subramanian, A., R. Narayan, Rajiv, S. M. Corsello *et al.*, “A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles”, *Cell*, 2017.

113. Wright, G. L., C. Haley, M. L. Beckett *et al.*, “Expression of prostate-specific membrane antigen in normal, benign, and malignant prostate tissues”, *Urologic Oncology: Seminars and Original Investigations*, Vol. 1, No. 1, pp. 18–28, 1995.
114. Van Dam, S., U. Vösa, A. van der Graaf *et al.*, “Gene co-expression analysis for functional classification and gene–disease predictions”, *Briefings in Bioinformatics*, 2017.
115. Langfelder, P. and S. Horvath, “WGCNA: An R package for weighted correlation network analysis”, *BMC Bioinformatics*, Vol. 9, 2008.
116. Ritchie, M. E., B. Phipson, D. Wu *et al.*, “Limma powers differential expression analyses for RNA-sequencing and microarray studies”, *Nucleic Acids Research*, 2015.
117. Zhang, B. and S. Horvath, “A general framework for Weighted Gene Co-Expression Network Analysis”, *Statistical Applications in Genetics and Molecular Biology*, Vol. 4, No. 1, 2005.
118. Yip, A. M. and S. Horvath, “Gene network interconnectedness and the generalized topological overlap measure”, *BMC Bioinformatics*, 2007.
119. Öztürk, H., E. Özkirimli and A. Özgür, “Classification of beta-lactamases and Penicillin Binding Proteins using ligand-centric network models”, *PLoS ONE*, 2015.
120. Bhowmick, S. S. and B. S. Seah, “Clustering and summarizing protein-protein interaction networks: A survey”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 3, pp. 638–658, 2016.
121. Bader, G. D. and C. W. Hogue, “An automated method for finding molecular complexes in large protein interaction networks”, *BMC Bioinformatics*, Vol. 4, p. 2, 2003.

122. Altaf-Ul-Amin, M., Y. Shinbo, K. Mihara *et al.*, “Development and implementation of an algorithm for detection of protein complexes in large interaction networks”, *BMC Bioinformatics*, Vol. 7, p. 207, 2006.
123. Li, M., J. Chen, J. Wang *et al.*, “Modifying the DPCLus algorithm for identifying protein complexes based on new topological structures.”, *BMC bioinformatics*, Vol. 9, No. 1, p. 398, 2008.
124. King, A. D., N. Pržulj and I. Jurisica, “Protein complex prediction via cost-based clustering”, *Bioinformatics*, Vol. 20, No. 17, pp. 3013–3020, 2004.
125. Nepusz, T., H. Yu and A. Paccanaro, “Detecting overlapping protein complexes in protein-protein interaction networks”, *Nature Methods*, Vol. 9, No. 5, pp. 471–472, 2012.
126. Enright, A. J., “An efficient algorithm for large-scale detection of protein families”, *Nucleic Acids Research*, Vol. 30, No. 7, pp. 1575–1584, 2002.
127. Satuluri, V., S. Parthasarathy and D. Ucar, “Markov clustering of protein interaction networks with improved balance and scalability”, pp. 247–256, 2010.
128. Shih, Y. K. and S. Parthasarathy, “Identifying functional modules in interaction networks through overlapping Markov clustering”, *Bioinformatics*, Vol. 28, No. 18, 2012.
129. Pereira-Leal, J. B., A. J. Enright and C. A. Ouzounis, “Detection of functional modules from protein interaction networks.”, *Proteins*, Vol. 54, No. 1, pp. 49–57, 2004.
130. Cho, Y.-R., W. Hwang, M. Ramanathan *et al.*, “Semantic integration to identify overlapping functional modules in protein interaction networks.”, *BMC bioinformatics*, Vol. 8, p. 265, 2007.

131. Spirin, V. and L. A. Mirny, “Protein complexes and functional modules in molecular networks”, *Proceedings of the National Academy of Sciences*, Vol. 100, No. 21, pp. 12123–12128, 2003.
132. Adamcsek, B., G. Palla, I. J. Farkas *et al.*, “CFinder: Locating cliques and overlapping modules in biological networks”, *Bioinformatics*, Vol. 22, No. 8, pp. 1021–1023, 2006.
133. Liu, G., L. Wong and H. N. Chua, “Complex discovery from weighted PPI networks”, *Bioinformatics*, Vol. 25, No. 15, pp. 1891–1897, 2009.
134. Georgii, E., S. Dietmann, T. Uno *et al.*, “Enumeration of condition-dependent dense modules in protein interaction networks”, *Bioinformatics*, Vol. 25, No. 7, pp. 933–940, 2009.
135. Spielman, D. A. and S.-H. Teng, “A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning”, *SIAM Journal on Computing*, Vol. 42, No. 1, pp. 1–26, 2013.
136. Macropol, K., T. Can and A. Singh, “RRW: repeated random walks on genome-scale protein networks for local cluster discovery”, *BMC Bioinformatics*, Vol. 10, No. 1, 2009.
137. Frey, B. J. and D. Dueck, “Clustering by passing messages between data points.”, *Science*, Vol. 315, No. 5814, pp. 972–976, 2007.
138. Chen, J. and B. Yuan, “Detecting functional modules in the yeast protein-protein interaction network”, *Bioinformatics*, Vol. 22, No. 18, pp. 2283–2290, 2006.
139. Luo, F., Y. Yang, C.-F. Chen *et al.*, “Modular organization of protein interaction networks.”, *Bioinformatics*, Vol. 23, No. 2, pp. 207–214, 2007.

140. Dotan-Cohen, D., A. A. Melkman and S. Kasif, “Hierarchical tree snipping: Clustering guided by prior knowledge”, *Bioinformatics*, Vol. 23, No. 24, pp. 3335–3342, 2007.
141. Navlakha, S., J. White, N. Nagarajan *et al.*, “Finding biologically accurate clusterings in hierarchical tree decompositions using the variation of information”, *Lecture Notes in Computer Science*, Vol. 5541 LNBI, pp. 400–417, 2009.
142. Mete, M., F. Tang, X. Xu *et al.*, “A structural approach for finding functional modules from large biological networks”, *BMC Bioinformatics*, Vol. 9, 2008.
143. Rivera, C. G., R. Vakil and J. S. Bader, “NeMo: Network Module identification in Cytoscape.”, *BMC Bioinformatics*, Vol. 11 Suppl 1, p. S61, 2010.
144. Asur, S., D. Ucar and S. Parthasarathy, “An ensemble framework for clustering protein-protein interaction networks”, *Bioinformatics*, Vol. 23, 2007.
145. Navlakha, S. and C. Kingsford, “Exploring biological network dynamics with ensembles of graph partitions.”, *Pacific Symposium on Biocomputing*, Vol. 177, pp. 166–77, 2010.
146. Greene, D., G. Cagney, N. Krogan *et al.*, “Ensemble non-negative matrix factorization methods for clustering protein-protein interactions”, *Bioinformatics*, Vol. 24, No. 15, pp. 1722–1728, 2008.
147. Maraziotis, I. A., K. Dimitrakopoulou and A. Bezerianos, “Growing functional modules from a seed protein via integration of protein interaction and gene expression data.”, *BMC Bioinformatics*, Vol. 8, No. 1, p. 408, 2007.
148. Lu, H., B. Shi, G. Wu *et al.*, “Integrated analysis of multiple data sources reveals modular structure of biological networks”, *Biochemical and Biophysical Research Communications*, Vol. 345, No. 1, pp. 302–309, 2006.

149. Zheng, H., H. Wang and D. H. Glass, “Integration of genomic data for inferring protein complexes from global protein-protein interaction networks”, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 38, No. 1, pp. 5–16, 2008.
150. Ulitsky, I. and R. Shamir, “Identifying functional modules using expression profiles and confidence-scored protein interactions”, *Bioinformatics*, Vol. 25, No. 9, pp. 1158–1164, 2009.
151. Shi, L., X. Lei and A. Zhang, “Detecting protein complexes with semi-supervised learning in protein interaction networks”, *Proteome Science*, Vol. 9, No. 5, 2011.
152. Mason, O. and M. Verwoerd, “Graph Theory and Networks in Biology”, *Work*, p. 52, 2006.
153. Pavlopoulos, G. A., M. Secrier, C. N. Moschopoulos *et al.*, “Using graph theory to analyze biological networks.”, *BioData Mining*, Vol. 4, No. 1, p. 10, 2011.
154. Boccaletti, S., V. Latora, Y. Moreno *et al.*, “Complex networks: Structure and dynamics”, *Physics Reports*, Vol. 424, No. 4–5, pp. 175–308, 2006.
155. Subramanian, A., P. Tamayo, P., V. K. Mootha *et al.*, “Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles”, *Proceedings of the National Academy of Sciences*, Vol. 102, No. 43, pp. 15545–15550, 2005.
156. Yu, G., L.-G. Wang, Y. Han *et al.*, “clusterProfiler: an R package for comparing biological themes among gene clusters”, *OMICS: A Journal of Integrative Biology*, 2012.
157. McCormack, T., O. Frings, A. Alexeyenko *et al.*, “Statistical assessment of crosstalk enrichment between gene groups in biological networks”, *PLoS ONE*, Vol. 8, No. 1, 2013.

158. Zielinski, R., P. F. Przytycki, J. Zheng *et al.*, “The crosstalk between EGF, IGF, and Insulin cell signaling pathways - computational and experimental analysis”, *BMC Systems Biology*, Vol. 3, No. 1, p. 88, 2009.
159. Glaab, E., A. Baudot, N. Krasnogor *et al.*, “EnrichNet: network-based gene set enrichment analysis”, *Bioinformatics*, Vol. 28, No. 18, pp. i451–i457, 2012.
160. Alexeyenko, A. W. Lee, M. Pernemalm *et al.*, “Network enrichment analysis: extension of gene-set enrichment analysis to gene networks”, *BMC Bioinformatics*, Vol. 13, No. 1, p. 226, 2012.
161. Ogris, C., D. Guala, T. Helleday *et al.*, “A novel method for crosstalk analysis of biological networks: Improving accuracy of pathway annotation”, *Nucleic Acids Research*, Vol. 45, No. 2, p. e8, 2017.
162. Chan, J. N. Y., C. Nislow and A. Emili, “Recent advances and method development for drug target identification”, *Trends in Pharmacological Sciences*, Vol. 31, No. 2, pp. 82–88, 2009.
163. Katsila, T., G. A. Spyroulias, G. P. Patrinos *et al.*, “Computational approaches in target identification and drug discovery”, *Computational and Structural Biotechnology Journal*, Vol. 14, pp. 177–184, 2016.
164. Salwinski, L., “The Database of Interacting Proteins: 2004 update”, *Nucleic Acids Research*, 2004.
165. Pagel, P., S. Kovac, M. Oesterheld *et al.*, “The MIPS mammalian protein-protein interaction database”, *Bioinformatics*, 2005.
166. Kar, G., A. Gursoy and O. Keskin, “Human cancer protein-protein interaction network: A structural perspective”, *PLoS Computational Biology*, 2009.

167. Watts, D. J. and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks”, *Nature*, Vol. 393, No. 6684, pp. 440–442, 1998.
168. Yu, H., D. Greenbaum, H. X. Lu *et al.*, “Genomic analysis of essentiality within protein networks”, *Trends in Genetics*, Vol. 20, No. 6, pp. 227–231, 2004.
169. Barabási, A. L., “Scale-free networks: A decade and beyond”, *Science*, Vol. 325, No. 5939, pp. 412–413, 2009.
170. Goh, K. I., B. Kahng and D. Kim, “Graph theoretic analysis of protein interaction networks of eukaryotes”, *Physica A: Statistical Mechanics and its Applications*, Vol. 357, No. 3-4, pp. 501–512, 2005.
171. Feng, Y., Q. Wang and T. Wang, “Drug target protein-protein interaction networks: A systematic perspective”, *BioMed Research International*, 2017.
172. Xu, J. and Y. Li, “Discovering disease-genes by topological features in human protein-protein interaction network.”, *Bioinformatics*, 2006.
173. Yu, H., P. M. Kim, E. Sprecher *et al.*, “The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics”, *PLoS Computational Biology*, Vol. 3, No. 4, pp. 713–720, 2007.
174. Capra, M., P. G. Nuciforo, S. Confalonieri *et al.*, “Frequent alterations in the expression of serine/threonine kinases in human cancers”, *Cancer Research*, 2006.
175. Festuccia, C., “Investigational serine/threonine kinase inhibitors against prostate cancer metastases”, *Expert Opinion on Investigational Drugs*, 2017.
176. Ma, T., C. A. Schreiber, G. J. Knutson *et al.*, “Effects of oxygen on the antigenic landscape of prostate cancer cells”, *BMC Research Notes*, 2015.

177. P., B. S., “AR, the cell cycle, and prostate cancer”, *Nuclear Receptor Signaling*, 2007.
178. Barboro, P., N. Ferrari and C. Balbi, “Emerging roles of heterogeneous nuclear ribonucleoprotein K (hnRNP K) in cancer progression”, *Cancer Letters*, 2014.
179. Matta, A., S. C. Tripathi, L. V. DeSouza *et al.*, “Heterogeneous ribonucleoprotein K is a marker of oral leukoplakia and correlates with poor prognosis of squamous cell carcinoma”, *International Journal of Cancer*, 2009.
180. Fei, T., Y. Chen, T. Xiao *et al.*, “Genome-wide CRISPR screen identifies HN-RNPL as a prostate cancer dependency regulating RNA splicing”, *Proceedings of the National Academy of Sciences*, 2017.
181. Abbas, A. and S. Gupta, “The role of histone deacetylases in prostate cancer.”, *Epigenetics*, 2008.
182. Singh, V., M. Ram, R. Kumar *et al.*, “Phosphorylation: Implications in cancer.”, *The Protein Journal*, Vol. 36, No. 1, pp. 1–6, 2017.
183. Gross, S., R. Rahal, Rami, N. Stransky *et al.*, “Targeting cancer with kinase inhibitors”, *Journal of Clinical Investigation*, 2015.
184. Fitzgerald, K. A., J. C. Evans, J. McCarthy *et al.*, “The role of transcription factors in prostate cancer and potential for future RNA interference therapy.”, *Expert Opinion on Therapeutic Targets*, 2014.
185. Altuwaijri, S. , D. K. Lee, K.-H. Chuang *et al.*, “Androgen receptor regulates expression of skeletal muscle-specific proteins and muscle cell types.”, *Endocrine*, Vol. 25, No. 1, 2004.
186. Ocana, A., C. Nieto-Jiménez, A. Pandiella *et al.*, “Neutrophils in cancer: Prognostic role and therapeutic strategies”, *Molecular Cancer*, 2017.

187. Weber, M. J. and D. Gioeli, “Ras signaling in prostate cancer progression”, *Journal of Cellular Biochemistry*, 2004.
188. Brizzolara, A., R. Benelli, R. Venè *et al.*, “The ErbB family and androgen receptor signaling are targets of Celecoxib in prostate cancer”, *Cancer Letters*, 2017.
189. Farrow, J. M., J. C. Yang and C. P. Evans, “Autophagy as a modulator and target in prostate cancer”, *Nature Reviews. Urology*, 2014.
190. Kibbe, W. A., C. Arze, V. Felix *et al.*, “Disease Ontology 2015 update: An expanded and updated database of Human diseases for linking biomedical knowledge through disease data”, *Nucleic Acids Research*, 2015.
191. Zhang, Q., H. Wang, L. Ren *et al.*, “Primary synovial sarcoma of the prostate metastatic to the liver and lung: A case report”, *World Journal of Surgical Oncology*, 2014.
192. Hoption Cann, S. A., Z. Qiu and C. Van Netten, “A prospective study of iodine status, thyroid function, and prostate cancer risk: Follow-up of the first national health and nutrition examination survey”, *Nutrition and Cancer*, 2007.
193. Liberzon, A., C. Birger, H. Thorvaldsdóttir *et al.*, “The Molecular Signatures Database Hallmark Gene Set Collection”, *Cell Systems*, Vol. 1, No. 6, pp. 417–425, 2015.
194. Hanahan, D. and R. A. Weinberg, “Hallmarks of cancer: The next generation”, *Cell*, Vol. 144, No. 5, pp. 646–674, 2011.
195. Bhowmick, N. A., E. G. Neilson and H. L. Moses, “Stromal fibroblasts in cancer initiation and progression”, *Nature*, Vol. 432, No. 7015, pp. 332–337, 2004.
196. Davies, M. A. and Y. Samuels, “Analysis of the genome to personalize therapy for melanoma”, *Oncogene*, Vol. 29, No. 41, pp. 5545–5555, 2010.

197. Jiang, B. H. and L. Z. Liu, "PI3K/PTEN signaling in tumorigenesis and angiogenesis", *Biochimica et Biophysica Acta - Proteins and Proteomics*, Vol. 1784, No. 1, pp. 150–158, 2008.
198. Sudarsanam, S. and D. E. Johnson, "Functional consequences of mTOR inhibition.", *Current Opinion in Drug Discovery & Development*, Vol. 13, No. 1, pp. 31–40, 2010.
199. Sherr, C. J. and F. McCormick, "The RB and p53 pathways in cancer", *Cancer Cell*, Vol. 2, No. 2, pp. 103–112, 2002.
200. Fridman, J. S. and S. W. Lowe, "Control of apoptosis by p53", *Oncogene*, Vol. 22, No. 56, pp. 9030–9040, 2003.
201. Blasco, M. A., "Telomeres and human disease: Ageing, cancer and beyond", *Nature Reviews. Genetics*, Vol. 6, No. 8, pp. 611–622, 2005.
202. Durant, S. T., "Telomerase-independent paths to immortality in predictable cancer sub-types", *Journal of Cancer*, Vol. 3, No. 1, pp. 67–82, 2012.
203. Hanahan, D. and J. Folkman, "Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis", *Cell*, Vol. 86, No. 3, pp. 353–364, 1996.
204. Hsu, P. P. and D. M. Sabatini, "Cancer cell metabolism: Warburg and beyond", *Cell*, Vol. 134, No. 5, pp. 703–707, 2008.
205. Grivennikov, S. I., F. R. Greten and M. Karin, "Immunity, inflammation, and cancer", *Cell*, Vol. 140, No. 6, pp. 883–899, 2010.
206. Uhlen, M., P. Oksvold, L. Fagerberg *et al.*, "Towards a knowledge-based Human Protein Atlas", *Nature Biotechnology*, Vol. 28, No. 12, pp. 1248–1250, 2010.

207. Ghiassian, S. D., J. Menche and A. L. Barabási, “A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome”, *PLoS Computational Biology*, Vol. 11, No. 4, 2015.
208. Bubendorf, L., A. Schöpfer, U. Wagner *et al.*, “Metastatic patterns of prostate cancer: An autopsy study of 1,589 patients”, *Human Pathology*, 2000.
209. Gandaglia, G., P. I. Karakiewicz, A. Briganti *et al.*, “Impact of the site of metastases on survival in patients with metastatic prostate cancer”, *European Urology*, 2015.
210. Norris, J. S., A. Bielawska, T. Day *et al.*, “Combined therapeutic use of AdGFP-FasL and small molecule inhibitors of ceramide metabolism in prostate and head and neck cancers: A status report”, *Cancer Gene Therapy*, 2006.
211. Seelan, R. S., C. Qian, A. Yokomizo *et al.*, “Human acid ceramidase is overexpressed but not mutated in prostate cancer”, *Genes Chromosomes and Cancer*, 2000.
212. French, K. J., R. S. Schrecengost, B. D. Lee *et al.*, “Discovery and evaluation of inhibitors of human sphingosine kinase”, *Cancer Research*, 2003.
213. Malavaud, B., D. Pchejetski, C. Mazerolles *et al.*, “Sphingosine kinase-1 activity and expression in human prostate cancer resection specimens.”, *European Journal of Cancer*, 2010.
214. Zhu, Y.-J., H. You, J.-X. Tan *et al.*, “Overexpression of sphingosine kinase 1 is predictive of poor prognosis in human breast cancer”, *Oncology Letters*, 2017.
215. Beljanski, V., C. Knaak and C. D. Smith, “A novel sphingosine kinase inhibitor induces autophagy in tumor cells.”, *The Journal of Pharmacology and Experimental Therapeutics*, 2010.

216. Antoon, J. W., W. D. Meacham, M. R. Bratton *et al.*, “Pharmacological inhibition of sphingosine kinase isoforms alters estrogen receptor signaling in human breast cancer”, *Journal of Molecular Endocrinology*, 2011.
217. Venant, H., M. Rahmaniyan, E. E. Jones *et al.*, “The sphingosine kinase 2 inhibitor ABC294640 reduces the growth of prostate cancer cells and results in accumulation of dihydroceramides in vitro and in vivo.”, *Molecular Cancer Therapeutics*, 2015.
218. Meshki, J., M. C. Caino, V. A. von Burstin *et al.*, “Regulation of prostate cancer cell survival by protein kinase Cepsilon involves bad phosphorylation and modulation of the TNFalpha/JNK pathway.”, *The Journal of Biological Chemistry*, 2010.
219. Lin, H.-P., C.-Y. Lin, C. Huo *et al.*, “AKT3 promotes prostate cancer proliferation cells through regulation of Akt, B-Raf, and TSC1/TSC2.”, *Oncotarget*, 2015.
220. Jiang, Z. and Y. Zhou, “Using gene networks to drug target identification”, *Journal of Integrative Bioinformatics*, 2005.
221. Wishart, D. S., Y. D. Feunang, A. C. Guo *et al.*, “DrugBank 5.0: A major update to the DrugBank database for 2018”, *Nucleic Acids Research*, 2018.
222. Ozen, M., O. F. Karatas, S. Gulluoglu *et al.*, “Overexpression of miR-145-5p inhibits proliferation of prostate cancer cells and reduces SOX2 expression”, *Cancer Investigation*, Vol. 33, No. 6, pp. 251–258, 2015.
223. Gaulton, A., A. Hersey, Anne, M. L. Nowotka *et al.*, “The ChEMBL database in 2017”, *Nucleic Acids Research*, 2017.

APPENDIX A: TOPOLOGY OF CORE PROTEINS

Table A.1. Topology of Core Proteins.

Protein Name	Degree	Betw. Cent.	Norm. Betw. Cent.	Eig. Cent.	Avg. Sh. Pt. Lng.	Ecc.	Coreness
A2M	93	51652	0.021	0.060	2.52	4	34
ACPP	15	910	0.000	0.017	2.86	5	13
ACTG2	14	335	0.000	0.021	2.86	5	13
AKT1	323	131043	0.052	0.329	2.22	5	45
AKT2	62	15200	0.006	0.070	2.51	5	35
AKT3	22	563	0.000	0.024	2.86	5	18
ALB	149	115927	0.046	0.084	2.50	5	38
AR	357	161684	0.064	0.345	2.24	5	45
ARAF	76	7171	0.003	0.098	2.55	5	41
ATF4	71	19323	0.008	0.057	2.57	5	34
AZGP1	16	951	0.000	0.018	2.87	5	15
BAD	57	6515	0.003	0.064	2.64	5	35
BCL2	125	31169	0.012	0.127	2.46	5	44
BRAF	77	9137	0.004	0.100	2.54	5	39
C8orf4	1	0	0.000	0.005	3.03	6	1
CASP9	51	3538	0.001	0.054	2.65	5	32
CCNB1	132	17944	0.007	0.148	2.44	5	44
CCND1	114	18551	0.007	0.162	2.38	5	45
CCNE1	73	3849	0.002	0.090	2.57	5	39
CCNE2	20	679	0.000	0.033	2.69	5	19
CDC20	122	19928	0.008	0.154	2.40	5	45
CDC42EP5	3	6	0.000	0.004	3.17	6	3
CDK2	492	208286	0.083	0.554	2.20	5	45
CDKN1A	281	120320	0.048	0.327	2.23	5	45
CDKN1B	108	18145	0.007	0.161	2.39	5	45
CDKN3	21	709	0.000	0.019	2.88	5	18
CEP55	78	19612	0.008	0.081	2.51	5	38
CHUK	146	23790	0.009	0.221	2.32	5	45
CNN1	15	358	0.000	0.014	2.93	6	14
COL3A1	19	722	0.000	0.010	2.99	5	14
CREB1	142	29322	0.012	0.119	2.49	5	42
CREB3	126	105847	0.042	0.037	2.66	5	27

Table A.1. Topology of Core Proteins (cont.)

Protein Name	Degree	Betw. Cent.	Norm. Betw. Cent.	Eig. Cent.	Avg. Sh. Pt. Lng.	Ecc.	Coreness
CREB3L1	78	40531	0.016	0.011	2.89	6	19
CREB3L3	24	3889	0.002	0.023	2.82	6	19
CREB5	65	10667	0.004	0.039	2.67	5	32
CREBBP	341	131443	0.052	0.302	2.28	5	45
CTNNB1	387	257956	0.103	0.417	2.15	4	45
DEPDC1B	23	142	0.000	0.010	3.07	6	22
DLGAP5	17	392	0.000	0.032	2.66	5	17
E2F1	130	22272	0.009	0.154	2.44	5	45
E2F2	27	1139	0.000	0.029	2.71	5	23
E2F3	57	3980	0.002	0.061	2.62	5	35
EGF	24	1642	0.001	0.031	2.68	5	20
EGFR	899	1035840	0.413	0.702	2.04	5	45
EP300	479	297189	0.119	0.456	2.18	5	45
ERBB2	247	92357	0.037	0.210	2.34	4	45
FGFR1	79	43162	0.017	0.063	2.61	5	34
FGFR2	72	26776	0.011	0.046	2.61	5	31
FN1	553	449261	0.179	0.577	2.12	4	45
FOXO1	67	10690	0.004	0.097	2.54	5	39
GRB2	649	652867	0.260	0.508	2.10	5	45
GSK3B	267	130230	0.052	0.258	2.25	5	45
GSTP1	50	6384	0.003	0.090	2.53	5	34
HSP90AA1	672	550853	0.220	0.637	2.09	5	45
HSP90AB1	564	390296	0.156	0.562	2.14	5	45
HSP90B1	140	24089	0.010	0.216	2.38	5	45
IGF1	19	1112	0.000	0.006	3.05	5	12
IGF1R	96	16245	0.006	0.136	2.41	5	42
IGFBP7	15	3281	0.001	0.008	3.00	5	12
IKBKB	167	29727	0.012	0.225	2.34	5	45
IKBKG	300	105589	0.042	0.307	2.26	5	45
INS	26	3948	0.002	0.019	2.74	5	19
JCHAIN	2	7	0.000	0.003	3.25	6	2
KIF11	90	18625	0.007	0.091	2.52	5	39
KLK3	30	2901	0.001	0.032	2.77	5	22
KRAS	263	223734	0.089	0.171	2.29	5	43
LEF1	49	2938	0.001	0.047	2.72	5	33

Table A.1. Topology of Core Proteins (cont.)

Protein Name	Degree	Betw. Cent.	Norm. Betw. Cent.	Eig. Cent.	Avg. Sh. Pt. Lng.	Ecc.	Coreness
LUM	13	1526	0.001	0.009	2.89	5	12
MAP2K2	42	2962	0.001	0.052	2.64	5	29
MAPK1	291	106887	0.043	0.285	2.26	5	45
MAPK3	224	71967	0.029	0.194	2.35	5	45
MDM2	396	189184	0.075	0.427	2.19	5	45
MGP	1	0	0.000	0.000	4.11	6	1
MSMB	1	0	0.000	0.005	3.01	6	1
MTOR	139	34140	0.014	0.158	2.39	5	45
NEFH	19	1395	0.001	0.018	2.84	6	16
NFKB1	267	116901	0.047	0.309	2.28	5	45
NFKBIA	170	49589	0.020	0.216	2.36	5	45
NKX3-1	28	5710	0.002	0.024	2.87	5	23
NRAS	47	5328	0.002	0.029	2.73	5	27
NUF2	57	4015	0.002	0.042	2.70	5	34
OGN	8	237	0.000	0.008	3.04	5	8
OLFM4	22	2940	0.001	0.008	2.87	5	16
PBK	28	1879	0.001	0.033	2.70	5	24
PDGFA	22	1377	0.001	0.012	2.90	5	18
PDGFB	41	9473	0.004	0.022	2.84	5	23
PDGFC	7	173	0.000	0.005	3.04	5	7
PDGFD	12	141	0.000	0.006	3.06	6	12
PDGFRA	59	6534	0.003	0.065	2.58	5	35
PDGFRB	114	24874	0.010	0.098	2.49	5	40
PDPK1	76	8177	0.003	0.073	2.53	5	38
PIK3CA	85	15790	0.006	0.080	2.54	5	39
PIK3CB	37	2918	0.001	0.045	2.66	5	27
PIK3CD	16	940	0.000	0.011	2.99	5	13
PIK3R1	277	123360	0.049	0.209	2.33	5	44
PIK3R2	114	18394	0.007	0.155	2.38	5	43
PTEN	249	130547	0.052	0.248	2.31	5	45
RAF1	176	37681	0.015	0.206	2.35	5	45
RB1	252	77987	0.031	0.250	2.33	5	45
RELA	354	185576	0.074	0.401	2.21	4	45
RRM2	46	2489	0.001	0.064	2.56	5	34
SERPINA1	43	5170	0.002	0.036	2.71	5	25

Table A.1. Topology of Core Proteins (cont.)

Protein Name	Degree	Betw. Cent.	Norm. Betw. Cent.	Eig. Cent.	Avg. Sh. Pt. Lng.	Ecc.	Coreness
SERPINA3	24	5755	0.002	0.014	2.82	5	16
SERPINA5	27	3753	0.001	0.014	2.89	5	18
SOS1	69	4941	0.002	0.076	2.57	5	38
SOS2	29	681	0.000	0.045	2.68	5	26
SPARCL1	8	253	0.000	0.009	2.94	6	8
TCF7	4	8	0.000	0.005	3.10	5	4
TCF7L1	4	4	0.000	0.006	3.08	5	4
TCF7L2	55	4133	0.002	0.065	2.59	5	34
TGFA	16	1799	0.001	0.016	2.79	5	14
TOP2A	90	9131	0.004	0.168	2.43	5	45
TP53	896	954313	0.381	0.829	2.04	5	45

APPENDIX B: TOPOLOGY OF DISEASE PROTEINS

Table B.1. Topology of Disease Proteins.

Protein Name	Degree	Betw. Cent.	Norm. Betw. Cent.	Eig. Cent.	Avg. Sh. Pt. Lng.	Ecc.	Coreness
ADNP	39	3143	0.001	0.050	2.66	5	30
AR	357	161684	0.064	0.345	2.24	5	45
ASCL2	10	41	0.000	0.014	2.96	6	10
ASXL1	29	2464	0.001	0.048	2.59	5	24
ATM	166	50900	0.020	0.169	2.37	5	45
AZGP1	16	951	0.000	0.018	2.87	5	15
BABAM1	74	6373	0.003	0.078	2.59	6	39
BIK	45	28620	0.011	0.015	2.79	5	19
BRCA1	426	195150	0.078	0.431	2.21	5	45
BRCA2	80	8842	0.004	0.092	2.54	5	40
CD82	31	4927	0.002	0.023	2.82	5	19
CDH1	493	351284	0.140	0.374	2.20	5	45
CDH13	22	2050	0.001	0.014	2.93	6	16
CDKN1B	108	18145	0.007	0.161	2.39	5	45
CHD1	27	487	0.000	0.053	2.65	6	27
CHEK2	103	19522	0.008	0.127	2.43	5	45
CHUK	146	23790	0.009	0.221	2.32	5	45
CLDN11	4	54	0.000	0.000	3.60	6	4
CPNE3	26	1559	0.001	0.044	2.68	5	23
CTBP2	102	33852	0.014	0.080	2.48	5	38
CYP3A4	13	9689	0.004	0.012	2.94	6	11
DMRT1	1	0	0.000	0.000	4.51	7	1
DPF1	12	14	0.000	0.019	2.95	5	12
EBF2	3	22	0.000	0.003	3.18	6	3
EPHB2	68	6994	0.003	0.064	2.58	5	36
ERG	87	10057	0.004	0.149	2.44	5	45
ESR2	124	18693	0.007	0.127	2.47	5	42
ETV1	25	536	0.000	0.038	2.76	5	23
EZH2	240	57920	0.023	0.264	2.32	5	45
FERMT2	43	4314	0.002	0.059	2.57	5	30
FGFR2	72	26776	0.011	0.046	2.61	5	31
FOXA1	36	4548	0.002	0.035	2.71	6	26

Table B.1. Topology of Disease Proteins (cont.)

Protein Name	Degree	Betw. Cent.	Norm. Betw. Cent.	Eig. Cent.	Avg. Sh. Pt. Lng.	Ecc.	Coreness
FRY	4	44	0.000	0.010	2.97	6	4
GJB1	9	236	0.000	0.015	2.81	5	9
GNMT	17	2463	0.001	0.016	2.86	6	14
HIP1	32	1949	0.001	0.040	2.67	5	27
HIST1H4A	151	28364	0.011	0.183	2.46	5	45
HNF1B	13	323	0.000	0.008	3.01	6	13
HPN	7	204	0.000	0.001	3.39	6	7
ITGA6	33	9668	0.004	0.028	2.64	5	20
KDM6A	33	2225	0.001	0.032	2.74	5	23
KLF6	17	955	0.000	0.024	2.75	5	15
KLK15	6	42	0.000	0.005	3.10	6	6
KLK3	30	2901	0.001	0.032	2.77	5	22
KMT2D	33	812	0.000	0.037	2.78	5	26
KRT8	69	7895	0.003	0.094	2.53	5	39
LMTK2	17	266	0.000	0.012	2.99	6	16
MAD1L1	72	7387	0.003	0.070	2.56	5	36
MAP3K1	168	27205	0.011	0.231	2.34	5	45
MDM4	68	7132	0.003	0.113	2.48	5	44
MED12	71	4605	0.002	0.067	2.55	5	36
MMP7	31	1902	0.001	0.010	2.97	5	20
MSMB	1	0	0.000	0.005	3.01	6	1
MSR1	7	247	0.000	0.007	3.03	6	7
MUC1	40	2500	0.001	0.071	2.59	5	34
MXI1	14	1140	0.000	0.021	2.74	5	13
NEDD9	62	8688	0.003	0.055	2.54	5	36
NKX3-1	28	5710	0.002	0.024	2.87	5	23
NLGN3	46	9144	0.004	0.023	2.80	5	27
NOTCH4	20	1870	0.001	0.029	2.79	5	17
OPHN1	4	16	0.000	0.003	3.18	5	4
PALB2	39	2580	0.001	0.034	2.74	6	27
PIM1	42	8614	0.003	0.059	2.54	5	32
PLXNB1	22	1696	0.001	0.018	2.87	5	17
PRKCI	84	13479	0.005	0.099	2.41	5	39
PTEN	249	130547	0.052	0.248	2.31	5	45
RAD23B	99	10042	0.004	0.107	2.50	6	44

Table B.1. Topology of Disease Proteins (cont.)

Protein Name	Degree	Betw. Cent.	Norm. Betw. Cent.	Eig. Cent.	Avg. Sh. Pt. Lng.	Ecc.	Coreness
RAD51B	11	288	0.000	0.010	2.96	5	10
RAD51D	39	5961	0.002	0.028	2.70	5	27
RASA1	98	11873	0.005	0.107	2.45	5	40
RFX6	27	707	0.000	0.011	2.96	6	23
RGS17	13	314	0.000	0.007	3.00	6	13
RNASEL	8	147	0.000	0.013	2.93	5	8
SALL3	6	14	0.000	0.006	3.15	6	6
SHROOM2	7	17	0.000	0.015	2.92	6	7
SIX1	17	650	0.000	0.008	2.96	5	16
SKIL	84	19356	0.008	0.071	2.60	5	37
SPOP	47	3877	0.002	0.046	2.69	5	32
TBX5	8	48	0.000	0.009	3.01	5	8
TERT	72	14671	0.006	0.111	2.52	5	43
TNS3	22	1299	0.001	0.031	2.71	5	20
TRIM8	60	4569	0.002	0.054	2.59	5	34
TUBA1C	151	25814	0.010	0.232	2.38	5	45
UBE3A	134	51177	0.020	0.159	2.36	5	45
USH2A	6	250	0.000	0.011	2.95	5	6
VAMP8	31	1189	0.000	0.022	2.86	5	26
VPS53	42	1338	0.001	0.022	2.83	5	31
XIAP	136	24507	0.010	0.126	2.43	5	43
ZFHX3	24	1148	0.000	0.028	2.69	6	23
ZNF652	2	5	0.000	0.002	3.34	6	2

APPENDIX C: PPIN CLUSTERS

Table C.1. Number of proteins, interactions and core proteins in clusters.

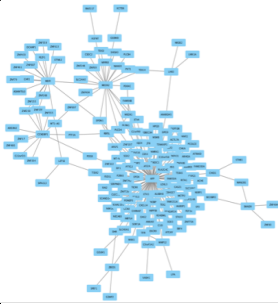
Cluster	No of Proteins	No of PPI	Core Proteins	
1	167	188	MSMB	
2	127	380	DLGAPS5	
3	81	87	C8orf4 MAPK	
4	70	148	GRB2 PIK3R2 SOS1 SOS2	

Table C.1. Number of proteins, interactions and core proteins in clusters (cont.)

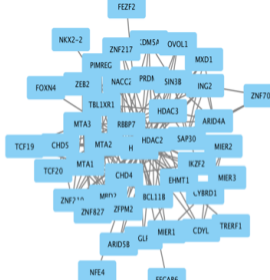
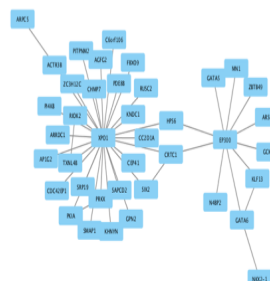
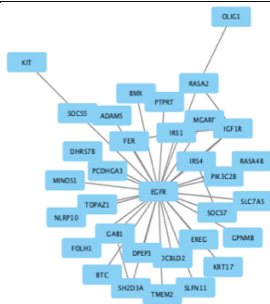
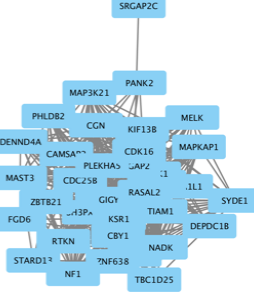
Cluster	No of Proteins	No of PPI	Core Proteins	
5	58	146	CDC42EP5 PTEN	
6	55	124	HSP90AA1 HSAP90AB1	
7	45	171	-	
8	40	44	EP300	
9	38	41	TP53	

Table C.1. Number of proteins, interactions and core proteins in clusters (cont.)

Cluster	No of Proteins	No of PPI	Core Proteins	
10	33	38	EGFR IGF1R	
11	37	480	DEPDC1B	

APPENDIX D: UNFAVORABLE PROGNOSTIC GENES

Table D.1. Unfavorable prognostic genes for prostate cancer [206].

Gene	Gene description	Diseases associated with the gene
MBD1	Methyl-CpG binding domain protein 1	Liver (U)
TEX30	Testis expressed 30	Cervical (U)
ZNF330	Zinc finger protein 330	Renal (F)
USP53	Ubiquitin specific peptidase 53	Renal (F), Colorectal (F)
ZNF654	Zinc finger protein 654	
FBXO45	F-box protein 45	Renal (U), Liver (U), Lung (U)
ARRB2	Arrestin beta 2	Cervical (F)
RPAIN	RPA interacting protein	Pancreatic (F)
DVL3	Dishevelled segment polarity protein 3	Renal (U), Colorectal (U), Liver (U)
RNF8	Ring finger protein 8	Liver (U), Melanoma (U)
HKR1	HKR1, GLI-Kruppel zinc finger family member	Pancreatic (F)
DNAJC9	DnaJ heat shock protein family (Hsp40) member C9	Lung (U)
HOOK2	Hook microtubule tethering protein 2	Renal (F)
RAN	RAN, member RAS oncogene family	Liver (U), Renal (U)
LMNB2	Lamin B2	Renal (U), Liver (U), Stomach (F), Melanoma (U)
CEP70	Centrosomal protein 70	Cervical (F)
MTF2	Metal response element binding transcription factor 2	Renal (U), Liver (U)
ZNF587B	Zinc finger protein 587B	Melanoma (U)
RBM8A	RNA binding motif protein 8A	
COPS3	COP9 signalosome subunit 3	Renal (U)
ATG3	Autophagy related 3	Liver (U), Cervical (F)
MPHOSPH6	M-phase phosphoprotein 6	Ovarian (F), Liver (U)
RUBCN	RUN and cysteine rich domain containing beclin 1 interacting protein	Liver (U), Colorectal (U)

Table D.1. Unfavorable prognostic genes for prostate cancer [206] (cont.)

Gene	Gene description	Diseases associated with the gene
SRC	SRC proto-oncogene, non-receptor tyrosine kinase	Liver (U), Urothelial (U), Renal (U)
ODF2	Outer dense fiber of sperm tails 2	Liver (U)
ZNF418	Zinc finger protein 418	
SMPD4	Sphingomyelin phosphodiesterase 4	Renal (U), Liver (U), Endometrial (U)
POLR3E	RNA polymerase III subunit E	
SRSF10	Serine and arginine rich splicing factor 10	Liver (U), Renal (U)
PUM3	Pumilio RNA binding family member 3	Liver (U)
ZNF700	Zinc finger protein 700	Renal (U), Urothelial (F), Head and neck (F), Melanoma (U)
ZNF274	Zinc finger protein 274	Glioma (F)
LETM1	Leucine zipper and EF-hand containing transmembrane protein 1	Liver (U), Renal (F)
FAM155B	Family with sequence similarity 155 member B	Breast (U), Renal (F)
MRPS22	Mitochondrial ribosomal protein S22	Endometrial (U), Pancreatic (U)
TDP2	Tyrosyl-DNA phosphodiesterase 2	Ovarian (F), Urothelial (F)
DPP4	Dipeptidyl peptidase 4	Renal (F), Thyroid (F)
PAQR6	Progestin and adipoQ receptor family member 6	Renal (U), Colorectal (U), Urothelial (F)
ZNF830	Zinc finger protein 830	
REV1	REV1, DNA directed polymerase	
ARMC8	Armadillo repeat containing 8	Liver (U), Endometrial (U)
ZNF550	Zinc finger protein 550	
ADPGK	ADP dependent glucokinase	Renal (U)
NCBP2	Nuclear cap binding protein subunit 2	Liver (U), Pancreatic (U)
WHAMM	WAS protein homolog associated with actin, golgi membranes and microtubules	Cervical (F)
CCNT2	Cyclin T2	Renal (U), Urothelial (F)
PNN	Pinin, desmosome associated protein	Liver (U)

Table D.1. Unfavorable prognostic genes for prostate cancer [206] (cont.)

Gene	Gene description	Diseases associated with the gene
TGDS	TDP-glucose 4,6-dehydratase	
PIGL	Phosphatidylinositol glycan anchor biosynthesis class L	Renal (U)
SYNJ1	Synaptojanin 1	Head and neck (F)
RUSC1	RUN and SH3 domain containing 1	Liver (U), Renal (U)
GTPBP3	GTP binding protein 3 (mitochondrial)	Renal (U), Liver (U)
EXO5	Exonuclease 5	Liver (U)
CMSS1	Cms1 ribosomal small subunit homolog (yeast)	Liver (U), Endometrial (U)
MAP6D1	MAP6 domain containing 1	
NELFCD	Negative elongation factor complex member C/D	Endometrial (U), Liver (U)
TOB1	Transducer of ERBB2, 1	Pancreatic (U)
ABCC5	ATP binding cassette subfamily C member 5	Liver (U)
RBM26	RNA binding motif protein 26	Renal (U), Urothelial (F)
ZNF503	Zinc finger protein 503	Renal (F)
SETD6	SET domain containing 6	
ZNF440	Zinc finger protein 440	Renal (F), Urothelial (F)
SIN3B	SIN3 transcription regulator family member B	Pancreatic (F)
CCT2	Chaperonin containing TCP1 subunit 2	Liver (U), Head and neck (U)
RAD18	RAD18, E3 ubiquitin protein ligase	Renal (U), Liver (U)
KIAA0907	KIAA0907	Renal (U), Liver (U), Urothelial (F)
PRPF39	Pre-mRNA processing factor 39	Liver (U), Urothelial (F), Melanoma (U), Ovarian (F)
DCUN1D5	Defective in cullin neddylation 1 domain containing 5	Liver (U), Ovarian (F), Head and neck (U)
C3orf38	Chromosome 3 open reading frame 38	Ovarian (F)
NCBP1	Nuclear cap binding protein subunit 1	Colorectal (F)

Table D.1. Unfavorable prognostic genes for prostate cancer [206] (cont.)

Gene	Gene description	Diseases associated with the gene
ZNF195	Zinc finger protein 195	Urothelial (F), Liver (U), Renal (U)
SRSF4	Serine and arginine rich splicing factor 4	Melanoma (U), Renal (U)
BCAS2	BCAS2, pre-mRNA processing factor	Liver (U), Thyroid (U), Renal (F), Ovarian (F)
ASB9	Ankyrin repeat and SOCS box containing 9	
STAU1	Staufen double-stranded RNA binding protein 1	
CEP170B	Centrosomal protein 170B	Renal (F), Liver (U)
TOP3A	Topoisomerase (DNA) III alpha	Pancreatic (F)
AKT2	AKT serine/threonine kinase 2	Endometrial (U)
VCPKMT	Valosin containing protein lysine methyltransferase	Lung (F), Renal (U)
MBD4	Methyl-CpG binding domain 4, DNA glycosylase	Renal (U), Pancreatic (U)
CCDC174	Coiled-coil domain containing 174	Urothelial (F)
CCDC14	Coiled-coil domain containing 14	Renal (U), Urothelial (F), Liver (U), Head and neck (F)
METTL10	Methyltransferase like 10	Thyroid (U)
EXOG	Exo/endonuclease G	Pancreatic (F)
NCBP3	Nuclear cap binding subunit 3	Melanoma (U)
SPAG16	Sperm associated antigen 16	Renal (F)
C1orf52	Chromosome 1 open reading frame 52	Renal (U), Liver (U)
CEP83	Centrosomal protein 83	Renal (U)
ZNF736	Zinc finger protein 736	Endometrial (U), Urothelial (F)
TP53	Tumor protein p53	Endometrial (F)
SHTN1	Shootin 1	Urothelial (U)
HNRNPA2B1	Heterogeneous nuclear ribonucleoprotein A2/B1	Liver (U), Renal (U)
NFATC2IP	Nuclear factor of activated T-cells 2 interacting protein	Renal (U)
BRPF1	Bromodomain and PHD finger containing 1	Liver (U)

Table D.1. Unfavorable prognostic genes for prostate cancer [206] (cont.)

Gene	Gene description	Diseases associated with the gene
ZRANB2	Zinc finger RANBP2-type containing 2	Renal (U), Liver (U)
POLG	DNA polymerase gamma, catalytic subunit	
RIOK3	RIO kinase 3	Pancreatic (U)
CLK3	CDC like kinase 3	Colorectal (U)
NKTR	Natural killer cell triggering receptor	Urothelial (F), Renal (U), Head and neck (F)
DHX33	DEAH-box helicase 33	
GNL3	G protein nucleolar 3	Renal (U), Liver (U)
SH2D3A	SH2 domain containing 3A	Renal (F)
ZNF668	Zinc finger protein 668	
DHX32	DEAH-box helicase 32 (putative)	Pancreatic (U)
FNBP4	Formin binding protein 4	Renal (U), Urothelial (F), Liver (U), Melanoma (U)
ZNF419	Zinc finger protein 419	Urothelial (F)
POLG2	DNA polymerase gamma 2, accessory subunit	Liver (U), Renal (U)
ZNF248	Zinc finger protein 248	Lung (F), Pancreatic (F), Renal (U)
HAUS1	HAUS augmin like complex subunit 1	Ovarian (F), Liver (U)
ZNF197	Zinc finger protein 197	
QTRT2	Queuine tRNA-ribosyltransferase accessory subunit 2	
ZNF335	Zinc finger protein 335	Renal (U), Liver (U), Head and neck (F)
ZWILCH	Zwilch kinetochore protein	Pancreatic (U), Renal (U)
DRG2	Developmentally regulated GTP binding protein 2	Endometrial (U)
ANKRD11	Ankyrin repeat domain 11	Liver (U)
RNPC3	RNA binding region (RNP1, RRM) containing 3	Lung (F), Urothelial (F), Pancreatic (F)
SPATA5L1	Spermatogenesis associated 5 like 1	
CDK5RAP1	CDK5 regulatory subunit associated protein 1	

Table D.1. Unfavorable prognostic genes for prostate cancer [206] (cont.)

Gene	Gene description	Diseases associated with the gene
RBM25	RNA binding motif protein 25	Liver (U), Renal (U)
WDR19	WD repeat domain 19	Pancreatic (F), Liver (U)
EID2	EP300 interacting inhibitor of differentiation 2	Colorectal (F)
DENND1C	DENN domain containing 1C	Renal (F), Lung (F), Urothelial (F)
SMC4	Structural maintenance of chromosomes 4	Renal (U), Pancreatic (U), Liver (U), Endometrial (U)
MXI1	MAX interactor 1, dimerization protein	Renal (U)
UMPS	Uridine monophosphate synthetase	
MINK1	Misshapen like kinase 1	Pancreatic (F), Liver (U)
TIPIN	TIMELESS interacting protein	Liver (U), Cervical (F)
PAXBP1	PAX3 and PAX7 binding protein 1	Renal (U)
PAK2	P21 (RAC1) activated kinase 2	Pancreatic (U)
FBRSL1	Fibrosin like 1	Renal (F), Urothelial (F)
FGD6	FYVE, RhoGEF and PH domain containing 6	Renal (U), Pancreatic (U)
PAK4	P21 (RAC1) activated kinase 4	Renal (F)
MDM4	MDM4, p53 regulator	Renal (U), Urothelial (F)
PGS1	Phosphatidylglycerophosphate synthase 1	Renal (U), Pancreatic (F), Liver (U)

APPENDIX E: FAVORABLE PROGNOSTIC GENES

Table E.1. Favorable prognostic genes for prostate cancer [206].

Gene	Gene description	Diseases associated with the gene
CRACR2A	Calcium release activated channel regulator 2A	
TMEM158	Transmembrane protein 158 (gene/pseudogene)	Renal (U)
EPHX1	Epoxide hydrolase 1	Lung (F)
TSKU	Tsukushi, small leucine rich proteoglycan	Renal (U), Pancreatic (U), Endometrial (F)
AOC1	Amine oxidase, copper containing 1	
TXN2	Thioredoxin 2	Renal (F)
SESN1	Sestrin 1	Head and neck (F)
AK5	Adenylate kinase 5	
GPAT4	Glycerol-3-phosphate acyltransferase 4	Ovarian (U), Urothelial (F), Endometrial (U)
RFPL2	Ret finger protein like 2	
ATP6V1E1	ATPase H ⁺ transporting V1 subunit E1	Renal (F), Liver (U)
RNF122	Ring finger protein 122	Renal (U), Thyroid (U)
COMT	Catechol-O-methyltransferase	Endometrial (F), Renal (U)
UBE2D4	Ubiquitin conjugating enzyme E2 D4 (putative)	Pancreatic (F), Renal (F)
HBEGF	Heparin binding EGF like growth factor	Head and neck (U), Pancreatic (U)
NTNG2	Netrin G2	
SLC35B2	Solute carrier family 35 member B2	Liver (U), Breast (U)
FOXQ1	Forkhead box Q1	
ARL2	ADP ribosylation factor like GTPase 2	Endometrial (U)
FOS	Fos proto-oncogene, AP-1 transcription factor subunit	
CAB39L	Calcium binding protein 39 like	Renal (F), Ovarian (F)
CSRP2	Cysteine and glycine rich protein 2	Renal (U)
MAP6	Microtubule associated protein 6	Pancreatic (F)

Table E.1. Favorable prognostic genes for prostate cancer [206] (cont.)

Gene	Gene description	Diseases associated with the gene
TNS2	Tensin 2	Pancreatic (F)
MYB	MYB proto-oncogene, transcription factor	Stomach (F)

APPENDIX F: OMIM AND GWAS ENTRIES

Table F.1. Prostate cancer related genes form OMIM and GWAS.

GWAS				OMIM	
ADNP	FERMT2	MIR4752	RPL23AP61	AR	HPCQTL19
AFM	FGFR2	MIR6129	RPL26P19	ASXL1	HPN
AGAP7P	FLJ40194	MIR6808	RPL7P41	ATM	IL16
AQP7P4	FRY	MLPH	RPS25P10	AZGP1	JAZF1
AR	GGCX	MMP7	RPS26P30	BRCA1	KAI1
ARMC2	GJB1	MS4A5	RPS2P1	BRCA2	KDM6A
ASCL2	GLULP5	MSMB	RPS2P38	CD82	KLF6
ASS1P11	GOLPH3L	MXRA8	SALL3	CDH1	KLK3
ATP9B	HMG2N2P32	MYEOV	SGCZ	CDH13	KMT2D
BABAM1	HNF1B	NEDD9	SHROOM2	CDKN1B	LMTK2
BIK	HNRNPA1P67	NEK10	SIDT1	CHD1	MAD1L1
BMS1P9	ITGA6	NKX3-1	SIX1	CHEK2	MED12
BRCA2	JAZF1	NLGN3	SIX4	CHUK	MSMB
CASC11	KCND3	NOTCH4	SKIL	CPNE3	MSR1
CASC15	KLK15	NUDT11	SLC22A1	CTBP2	MUC1
CASC17	KLK2	OPHN1	SLC22A2	CYP3A4	MXI1
CASC8	KLK3	PARP4P2	SLC22A3	EHBP1	PALB2
CCAT1	KRT78	PCAT1	SLC7A3	EHPB2	PCAP
CCAT2	KRT8	PCAT19	SRRM1P1	ELAC2	PCGEM1
CENPVL3	LILRA5	PDLIM5	TANC1	EPHB2	PIM1
CHL1-AS1	LINC00421	PIP4K2A	TBX5	ERG	PLXNB1
CLDN11	LINC00506	PPATP1	TERT	ETV1	PRNCR1
COL6A3	LINC01153	PPP1R14A	TEX11	EZH2	PTEN
CTD-2194D22.4	LINC01266	PRKCI	TNS3	FOXA1	RAD51D
CXorf67	LMTK2	PRNCR1	TRIM8	GNMT	RNASEL
CXXC4-AS1	LOC101928075	PRPH	TUBA1C	HIP1	SLC45A3
DMRT1	LOC101929468	RAD23B	UBE3A	HIST1H4A	SPOP
DPF1	LOC102724596	RAD51B	USH2A	HIST2H3C	SRD5A1
EBF2	LOC728114	RASA1	VAMP8	HMG2N2P46	SRD5A2
EEFSEC	LUZP2	RFX6	VPS53	HNF1B	TERT
EHBP1	MAP3K1	RGS17	XAGE3	HPC3	TNSF11
ESR2	MARCH8	RNU6-66P	XPOTP1	HPC4	XAF1
FAM60DP	MDM4	RNU6ATAC5P	ZGPAT	HPC5	XIAP
FAM84B	MIR3164	RNU7-89P	ZNF652	HPC6	ZFHX3
FARP2	MIR4686				

APPENDIX G: TOP 100 DIAMOND DISEASE GENES (alpha=2)

Table G.1. Top 100 DIAMOnD disease genes for the network (alpha=2).

#rank	Gene	#rank	Gene	#rank	Gene	#rank	Gene
1	E2F1	26	NCOA6	51	KLF5	76	TOP2B
2	TP53	27	PPARG	52	TBP	77	MYOD1
3	KAT2B	28	NCOA2	53	RXRA	78	SNW1
4	MDM2	29	NR3C1	54	PARP1	79	PIAS3
5	AKT1	30	HDAC1	55	CDKN1A	80	THRA
6	ESR1	31	PML	56	UBE2I	81	CREB1
7	NCOA3	32	NR4A1	57	ZBTB16	82	MAPK14
8	PIAS1	33	SMARCA4	58	TDG	83	NFATC1
9	NCOR2	34	CEBPB	59	FOXO1	84	ATF2
10	EP300	35	STAT5A	60	HNF4A	85	JUND
11	CREBBP	36	NFKB1	61	SIRT1	86	ATF3
12	FOXO3	37	STAT3	62	SMAD4	87	HDAC4
13	SMAD3	38	HDAC3	63	MAPK8	88	STAT1
14	HIF1A	39	MYC	64	MAPK3	89	PIAS2
15	CTNNB1	40	SMARCA2	65	NR0B3	90	PPARD
16	SRC	41	JUN	66	TP73	91	PPARGC1A
17	ETS1	42	RUNX1	67	DAXX	92	NRIP1
18	SP1	43	FOS	68	TRIM24	93	CTBP1
19	NCOR1	44	VDR	69	SUMO1	94	BCL3
20	RELA	45	CEBPA	70	KAT5	95	GSK3B
21	PGR	46	RB1	71	SIN3A	96	NFE2L2
22	CCND1	47	SMARCB1	72	SMARCC1	97	SRF
23	MAPK1	48	HDAC2	73	SMARCE1	98	ARID1A
24	NCOA1	49	SMAD2	74	CARM1	99	BCL6
25	RARA	50	RUNX2	75	SMARCC2	100	SPI1

APPENDIX H: TOP 100 DIAMOND DISEASE GENES (alpha=10)

Table H.1. Top 100 DIAMOnD disease genes for the network (alpha=10).

#rank	Gene	#rank	Gene	#rank	Gene	#rank	Gene
1	AKT1	26	SMARCA2	51	NTRK1	76	HSF1
2	TP53	27	VDR	52	HSPA1A	77	PGR
3	NCOR2	28	NCOR1	53	STUB1	78	STAT5A
4	NCOA3	29	RB1	54	KLF5	79	XPO1
5	SRC	30	CEBPB	55	NCOA2	80	HDAC6
6	E2F1	31	SMARCB1	56	CDKN1A	81	HMGA1
7	UBE2I	32	YWHAZ	57	HSPA5	82	HMGB1
8	HDAC1	33	NCOA6	58	STAT3	83	CTBP1
9	ELAVL1	34	CEBPA	59	COPS5	84	DAXX
10	NFKB1	35	SMARCC1	60	HNF4A	85	SFPQ
11	CREBBP	36	NCOA1	61	APP	86	PML
12	NR3C1	37	CARM1	62	NR0B2	87	STAT1
13	SMAD3	38	FOXO1	63	APEX1	88	RUNX1
14	FOXO3	39	SMARCE1	64	YWHAQ	89	GRB2
15	EGFR	40	RBL1	65	NFATC1	90	MAPK8
16	KAT2B	41	RARA	66	DNMT1	91	CHEK1
17	SMARCC2	42	HDAC3	67	THRA	92	HNRNPA1
18	HIF1A	43	JUN	68	HCFC1	93	NR5A1
19	HDAC2	44	PRKCD	69	HSP90AB1	94	RELA
20	SMAD2	45	ABL1	70	NRIP1	95	YWHAE
21	HSPA8	46	VHL	71	THRB	96	MAPK14
22	CTNNB1	47	XRCC6	72	MAPK1	97	CDC5L
23	SMARCA4	48	SMAD4	73	MED1	98	CDK8
24	MYC	49	KAT2A	74	RUNX2	99	HNRNPK
25	ETS1	50	NONO	75	XRCC5	100	POU2F1

APPENDIX I: TARGETS FOR DRUG REPOSITIONING

Table I.1. Targets for Drug Repositioning.

AXL	EPHB6	LAMB1	NTRK2	PLD1
BLMH	GBA	MAP2K3	PDXK	RHEB
DHPS	HMMR	MAPK8IP1	PFN2	RPS6KA6
DIABLO	IGFBP3	MKNK2	PGM1	S100B
EGLN1	IL4R	NEK2	PIK3C3	TNFRSF10B
EGLN3	ITPR1	NOD2	PIK3CB	

APPENDIX J: POTENTIAL DRUG TARGETS

Table J.1. Potential drug targets.

Potential Drug Targets	Reported as cancer related target/biomarker in literature	Reported as Pca related target/biomarker in literature	Interaction with at least one ligand from ChEMBL	Total
GNA12	+	+	+	3
KAT6A	+	+	+	3
NAE1	+	+	+	3
NOTCH2	+	+	+	3
NR2F2	+	+	+	3
PAFAH1B2	+	+	+	3
PBK	+	+	+	3
PRKD2	+	+	+	3
RPS6KA2	+	+	+	3
ARID4B	+	+	-	2
ASF1A	+	-	+	2
ATF1	+	-	+	2
CBFB	-	+	+	2
CCNB2	+	-	+	2
CHAF1B	+	+	-	2
DUSP3	+	-	+	2
EHMT1	+	-	+	2
ESRRG	+	-	+	2
ID3	+	+	-	2
L1CAM	+	+	-	2
MLST8	+	+	-	2
PBRM1	+	-	+	2
PHLPP1	+	-	+	2
PTP4A1	+	-	+	2
SOX5	+	+	-	2
TNKS2	+	-	+	2
USP1	+	-	+	2
AFF4	+	-	-	1
AKAP13	+	-	-	1
EIF2AK3	-	-	+	1
ERLIN1	+	-	-	1

Table J.1. Potential drug targets (cont.)

Potential Drug Targets	Reported as cancer related target/biomarker in literature	Reported as Pca related target/biomarker in literature	Interaction with at least one ligand from ChEMBL	Total
GGA2	+	-	-	1
ING3	-	+	-	1
IRF2	+	-	-	1
IRF9	-	+	-	1
ISL1	+	-	-	1
NOD1	-	-	+	1
RAC2	-	-	+	1
TNFAIP1	+	-	-	1
USP32	+	-	-	1
ARHGEF11	-	-	-	0
BUD31	-	-	-	0
EXOC2	-	-	-	0
GLI3	-	-	-	0
IRAK2	-	-	-	0
KCTD5	-	-	-	0
NBR1	-	-	-	0
NFRKB	-	-	-	0
NUP85	-	-	-	0
POLR3C	-	-	-	0
PRKRA	-	-	-	0
RAB4A	-	-	-	0
RNF123	-	-	-	0
SEC24B	-	-	-	0
UBE2Z	-	-	-	0