

## Research Article

# Interactions between Germline and Somatic Mutated Genes in Aggressive Prostate Cancer

Tarun Karthik Kumar Mamidi , Jiande Wu, and Chindo Hicks 

*Department of Genetics, Louisiana State University Health Sciences Center, School of Medicine, 533 Bolivar St., New Orleans, LA 70112, USA*

Correspondence should be addressed to Chindo Hicks; [chick3@lsuhsc.edu](mailto:chick3@lsuhsc.edu)

Received 18 December 2018; Revised 29 January 2019; Accepted 15 February 2019; Published 17 March 2019

Academic Editor: Cristina Magi-Galluzzi

Copyright © 2019 Tarun Karthik Kumar Mamidi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Prostate cancer (PCa) is the most common diagnosed malignancy and the second leading cause of cancer-related deaths among men in the USA. Advances in high-throughput genotyping and next generation sequencing technologies have enabled discovery of germline genetic susceptibility variants and somatic mutations acquired during tumor formation. Emerging evidence indicates that germline variations may interact with somatic events in carcinogenesis. However, the possible oncogenic interactions and cooperation between germline and somatic variation and their role in aggressive PCa remain largely unexplored. Here we investigated the possible oncogenic interactions and cooperation between genes containing germline variation from genome-wide association studies (GWAS) and genes containing somatic mutations from tumor genomes of 305 men with aggressive tumors and 52 control samples from The Cancer Genome Atlas (TCGA). Network and pathway analysis were performed to identify molecular networks and biological pathways enriched for germline and somatic mutations. The analysis revealed 90 functionally related genes containing both germline and somatic mutations. Transcriptome analysis revealed a 61-gene signature containing both germline and somatic mutations. Network analysis revealed molecular networks of functionally related genes and biological pathways including P53, STAT3, NKX3-1, KLK3, and Androgen receptor signaling pathways enriched for germline and somatic mutations. The results show that integrative analysis is a powerful approach to uncovering the possible oncogenic interactions and cooperation between germline and somatic mutations and understanding the broader biological context in which they operate in aggressive PCa.

## 1. Introduction

Despite remarkable progress in patient screening using the prostate specific antigen (PSA) and improved patient care, prostate cancer (PCa) remains a major public health problem [1, 2]. PCa is the most diagnosed cancer and the second leading cause of cancer-related death among men in the US [1, 2]. The majority of patients present with localized tumors that will remain indolent and pose no harm even without treatment [3]. However, a significant proportion of patients ~ 20% to 30% will develop aggressive tumors, which progress rapidly to metastatic form and is often lethal if not detected and treated early [3, 4]. Therefore, there is an urgent need for the discovery of molecular markers for early detection of aggressive PCa.

Advances in high-throughput genotyping and reduction in genotyping costs have enabled discovery of genetic variants, primarily single nucleotide polymorphisms (SNPs) herein called germline mutations associated with an increased risk of developing PCa using genome-wide association studies GWAS [5]. These findings are providing foundational knowledge about the genetic susceptibility landscape of PCa. Most notably, these discoveries are being incorporated into risk prediction models for identification of patients at high risk of developing aggressive disease [6, 7]. For example, recently, Seibert et al. developed and validated a polygenic hazard score to guide screening for aggressive PCa [7]. However, despite this remarkable success, establishing the link between genetic predisposition and tumorigenesis remains a challenge. This knowledge gap has hampered

translation of genetic susceptibility variants from GWAS studies into clinical practice to improve human health.

The recent surge of next generation sequencing of the cancer genomes has led to an expanded molecular classification of PCa and discovery of somatic driver mutations acquired during tumorigenesis [8]. Large multicenter projects such as The Cancer Genome Atlas (TCGA) [9] and the International Cancer Genome Consortium (ICGC) [10] have performed a series of detailed analyses of the somatic alterations affecting tumor genomes in sporadic PCa. Discoveries from these large-scale studies are providing valuable information about the genomic landscape of the tumor genomes. However, to date, somatic mutation information has not been maximally leveraged and integrated with information on germline genetic susceptibility variants to infer the possible oncogenic interactions and mechanisms of cooperation between germline and somatic variation in aggressive PCa. Emerging evidence indicates that germline variations may interact with somatic events in carcinogenesis [11, 12]. Recently Wang et al. reported interactions between germline genetic susceptibility loci and somatic alterations in lung cancer [11]. Feigelson et al. proposed approaches for integrating germline and tumor genomic data in cancer research [12]. However, the possible oncogenic interactions between germline and somatic alterations in aggressive PCa have not been reported. The objective of this exploratory study was to investigate the possible oncogenic interactions and cooperation among and between genes containing germline mutations and genes containing somatic mutations in aggressive PCa. Our working hypothesis was that genes containing germline and somatic mutations are functionally related and interact in gene regulatory networks and biological pathways driving aggressive PCa. To address this hypothesis we combined germline mutation information derived from GWAS with somatic mutation information from TCGA using gene expression data on aggressive PCa from the TCGA. Our evaluation focuses on genes containing germline and somatic mutated rather than individual mutations to understand the broader biological context in which they cooperate and to establish putative functional bridges between germline-somatic interactions and the pathways they control.

## 2. Materials and Methods

**2.1. Germline Mutations and Associated Genes.** We have developed and published a comprehensive catalogue of genetic susceptibility variants (SNPs) and genes reported to be associated with an increased risk of developing PCa from GWAS [5]. For this study, we used an updated version of this catalogue. The information in this catalogue was supplemented with information from the GWAS catalogue [13] to ensure completeness of the information used in this study. Details about methods used in collection of genetic susceptibility variants and genes from GWAS have been described in our earlier publication [5]. The methods for data collection were based on the guidelines proposed by the Human Genome Epidemiology Network for systematic review of genetic associations [14–18], which we have

described elsewhere [5]. Here, we provide a brief, but detailed description of the data used in this study as well as the inclusion and exclusion criteria used.

We reviewed over 200 published reports on GWAS on PCa. The reports were screened by title, abstract, and full-text review to identify the studies meeting our eligibility criteria. Because most of the GWAS reports have not been clinical phenotype-specific (i.e., indolent or aggressive PCa specific), we reviewed all GWAS studies available on PCa. After screening, 150 studies that met our eligibility criteria were selected and subjected to further detailed review. The exclusion criteria for the 50 studies included removal of studies with insufficient or incomplete information, reviews, studies reporting only intergenic regions, and studies with very small sample sizes (i.e., studies containing <500 subjects in cases and controls). For the remaining 150 studies used in this study, they were considered eligible if they met the following criteria: (1) the studies must have been based on a case-control study design using unrelated individuals, (2) publications must have been of full length and published in peer-reviewed journals or online in English language before June 2018, (3) PCa must have been diagnosed by histological examination, (4) the sample sizes must be more than 500 for the cases and more than 500 for the controls to reduce sampling errors, (5) the study must have provided sufficient information such that genotype frequencies for both PCa and controls can be discerned without ambiguity, and (6) the studies must have used the appropriate and recommended statistical methods to infer the associations by taking into account the covariates and accounting for population structure and genetic background [14].

We manually extracted the information from the 150 studies meeting our eligibility criteria and the accompanying Web sites containing supplementary data. The extracted information included SNP identification number (rs-ID); evidence of association as determined by the GWAS  $P$  value; a composite of strong ( $P \leq 10^{-8}$ ), moderate ( $P = 10^{-5}$ – $10^{-7}$ ), and weak ( $P = 10^{-2}$ – $10^{-4}$ ) association; gene name; and associated chromosome position to which the genes map as determined by the dbSNP database [19] and the Human Genome Nomenclature database [20]. This search yielded more than 500 SNPs in introns and exons mapped to 266 genes from a population of more than 250,000 cases and more than 250,000 controls. The GWAS data set included 61 genes containing genetic variants directly associated with an increased risk of developing aggressive PCa. The database was cross-referenced with the GWAS catalogue database for validation and to ensure that all reported genetic variants and associated genes are represented in the data set. A complete list of genetic variants and associated genes along with sources or published reports from which they were derived is presented in Table SA provided as supplementary data to this report.

**2.2. Somatic Mutation Information and Gene Expression Data.** Today, treatment decisions for PCa patients are guided by various stratification algorithms [21]. Among these parameters, the most potent predictor of PCa mortality is the Gleason

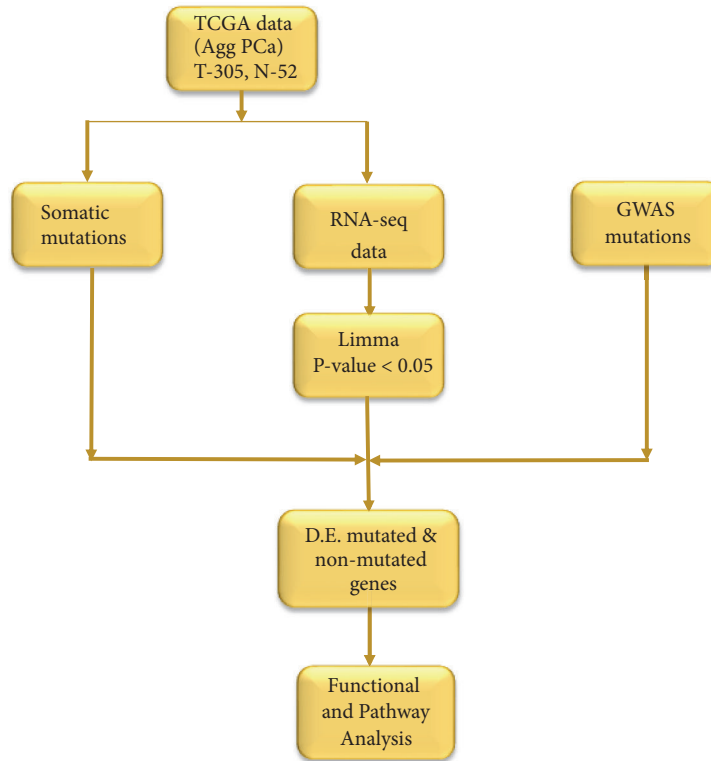


FIGURE 1: Design and workflow for integrative analysis combining germline with somatic mutation information using gene expression data. RNA-seq read count data and somatic information were downloaded from The Cancer Genome Atlas (TCGA) via the Genomics Data Commons. Germline mutation information was manually curated from GWAS studies. Limma (R) package was used for the discovery of differentially expressed (DE) mutated and nonmutated genes. Ingenuity Pathway Analysis (IPA) was used for functional analysis and discovery of molecular networks and biological pathways enriched for germline and somatic mutations.

grade which ranges from 6 to 10 in the modern era [21]. The presence of Gleason grade 6 is associated with very low cancer-specific mortality rates, even in the absence of intervention. Intermediate grade disease (Gleason grade 7) has a much more variable course. High Gleason grade 8-10 are aggressive and often lethal. For this study, we used somatic mutation information and associated genes derived from 188 patients with aggressive tumors (i.e., tumors with Gleason grade 8-10) from TCGA. In addition, because Gleason grade 7 follows a variable clinical course we used 4 + 3 score used by the American Urological Association to assign this group of patients to aggressive PCa [21]. For classification of tumors as aggressive, we used the clinical information provided by the TCGA and the classification protocols of the American Urological Association [21]. Somatic mutation information and genes were downloaded from the Genomics Data Commons and is available at <https://gdc.cancer.gov/> [22].

Gene expression data derived from RNA-seq was downloaded from TCGA using Genomics Data Commons (GDC) data transfer tool along with clinical information at <https://gdc.cancer.gov/> [22]. The gene expression data set included 305 patients with aggressive tumors and 52 normal control samples. Note that the number of patients with gene expression data is higher than the number of sequenced patients. The data matrix was filtered to remove rows with missing data, such that each row has at least  $\geq 30\%$  data using

cpm (counts per million) filter ( $>0.5$ ) implemented in R [23]. The resulting data set was normalized by TMM (The trimmed mean of M-values) normalization method and then transformed by Voom, using Limma package implemented in R [23]. The normalized data contained 18,428 probes and was used in the analysis. The probe IDs and gene symbols and names were matched for interpretation using the Ensemble database, a database used for gene annotation of sequencing experiments and sequencing technology platforms.

**2.3. Data Analysis.** The project design and data analysis workflow are presented in Figure 1. We performed unbiased whole genome analysis comparing gene expression levels between patients diagnosed with aggressive tumors and matched control samples using the Limma package implemented in R [23] to identify all significant differentially expressed genes distinguishing tumors from control samples. This unbiased approach was carried out to discover, germline and somatically mutated genes as well as nonmutated genes associated with PCa. We used the false discovery rate (FDR) procedure to correct for multiple hypothesis testing [24]. The genes were ranked on  $P$  values and the FDR. A gene was defined as PCa susceptibility gene if it belongs to protein-coding genes and contained the genetic susceptibility variant(s) and its expression in PCa tissue was associated

with PCa risk SNPs. A gene was considered as PCa driver gene if it contained somatic mutations and there was evidence to be PCa related or the gene contained PCa-related driver mutations. A gene was considered to be involved in the germline and somatic genomes if it contained both germline and somatic mutations.

We performed hierarchical clustering using Morpheus [25] using sets of differentially expressed genes to identify clusters of coregulated genes and to assess similarity in patterns of gene expression profiles, among the germline and somatically mutated and nonmutated genes. Prior to clustering the data was standardized, normalized, and centered. For clustering, we used the Pearson correlation coefficient as the measure of distance between pairs of genes and the complete linkage as the clustering method. We performed enrichment analysis using Ingenuity Pathway Analysis (IPA) software [26] to identify molecular networks and biological pathways enriched for germline and somatic mutations. Using IPA, the most highly significantly differentially expressed genes distinguishing patients with tumors from control samples were mapped onto networks and canonical pathways. The probability scores and the log  $P$  values were calculated to assess the likelihood and reliability of correctly assigning the genes to the correct molecular networks and biological pathways. A false discovery rate was used to correct for multiple hypothesis testing in pathway analysis. The predicted molecular networks and biological pathways were ranked based on  $z$ -scores and log  $P$  values; respectively. Gene ontology (GO) [27] analysis as implemented in IPA was performed to characterize the functional relationships among genes in the networks as implemented in IPA. Genes were classified according to the molecular functions, biological processes and cellular components in which they are involved. Genes were considered interacting or cooperating if they were involved in the same molecular functions, biological process, cellular components, molecular networks, or biological pathways.

### 3. Results

**3.1. Discovery of Somatic Mutated and Nonmutated Gene Signatures.** To discover and characterize the somatic mutated and nonmutated gene signatures, we performed whole genome analysis comparing gene expression levels between patients with aggressive tumors and the controls samples. After controlling for multiple hypothesis testing, the analysis revealed two gene signatures. One gene signature consisted of 2,613 significantly ( $P < 0.05$ ) differentially expressed genes containing somatic mutations, of which 2,298 genes were highly significantly ( $P < 0.01$ ) differentially expressed. The other signature consisted of 10,192 significantly ( $P < 0.05$ ) differentially expressed genes containing no somatic mutations, of which 8,913 genes were highly significantly ( $P < 0.01$ ) differentially expressed. The results showing a signature of the top 34 most highly somatically mutated genes ( $\geq 5$  mutation events) that were significantly differentially expressed are presented in Table 1. Also presented in the table are the number of somatic mutation events per gene and the differential expression  $P$  value.

TABLE 1: List of the 34-gene signature containing the most highly somatically mutated genes ( $\geq 5$  Mutation events) that were significantly differentially expressed between aggressive tumors and controls.

<i>Genes</i>	Region	Mutation events	DE $P$ value
<i>SPOP</i>	17q21.33	29	2.13E-25
<i>ATM</i> **	11q22.3	13	0.004526
<i>FOXA1</i>	14q21.1	12	1.41E-31
<i>KMT2C</i>	7q36.1	12	0.04799
<i>LRPIB</i>	2q22.1	10	5.18E-31
<i>FAT3</i>	11q14.3	9	1.69E-39
<i>OBSCN</i>	1q42.13	9	7.66E-05
<i>FREM2</i>	13q13.3	8	4.60E-12
<i>CSMD3</i>	8q23.3	8	0.000312
<i>KIF16B</i>	20p12.1	8	0.020567
<i>PCDH18</i>	4q28.3	7	2.00E-15
<i>SYNE1</i> **	6q25.2	7	1.12E-07
<i>AHNAK</i>	11q12.3	7	6.32E-06
<i>PLCB4</i>	20p12.3	7	0.000102
<i>FAT2</i>	5q33.1	6	2.13E-31
<i>FAM83B</i>	6p12.1	6	8.65E-31
<i>DCHS2</i>	4q31.3	6	4.08E-27
<i>CDH23</i>	10q22.1	6	1.01E-26
<i>MACF1</i>	1p34.3	6	9.90E-06
<i>FLG</i>	1q21.3	6	2.97E-05
<i>PIK3CA</i>	3q26.32	6	3.39E-05
<i>VPS13D</i>	1p36.22	6	4.42E-05
<i>HFM1</i>	1p22.2	6	0.003362
<i>PTPRD</i>	9p24.1	6	0.017381
<i>EPHBI</i>	3q22.2	5	9.49E-41
<i>KIAA1614</i>	1q25.3	5	6.36E-26
<i>AHNAK2</i>	14q32.33	5	8.04E-26
<i>MXRA5</i>	Xp22.33	5	2.48E-07
<i>SACS</i>	13q12.12	5	7.04E-07
<i>ASHIL</i>	1q22	5	6.85E-05
<i>RYR2</i> **	1q43	5	8.91E-05
<i>CTNNA1</i> **	3p22.1	5	0.000345
<i>MT-ND5</i>	21q22.3	5	0.014434
<i>FZD4</i> **	11q14.2	5	0.042185

Note. DE indicates differential expression. \*\* indicates genes associated with aggressive PCa.

The list of highly mutated genes included the genes *SYNE1*, *ATM*, *CTNNA1*, *FZD4*, and *RYR2* implicated in PCa [28–32]. A complete list of significantly differentially expressed somatic mutated and nonmutated genes including their mutation frequencies along with their estimates of  $P$  values and false discovery rates are presented in Supplementary Table S1A (somatic mutated) and Table S1B (nonsomatic mutated) provided as supplementary data.

**3.2. Germline and Somatic Mutation Gene Signatures.** To discover and characterize the germline and somatic mutation gene signatures, we evaluated the significantly differentially

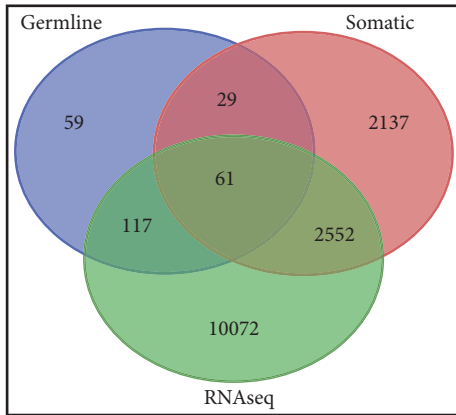


FIGURE 2: Venn diagram showing differentially and nondifferentially expressed germline and somatic mutated and nonmutated genes. Middle intersections show 61 genes containing both germline and somatic mutations and are also significantly expressed with RNA-seq dataset. Germline indicates genes with germline genetic susceptibility variants from GWAS. Somatic indicates genes with somatic mutations from TCGA. RNA-seq indicates differentially expressed genes evaluated using TCGA gene expression data.

expressed genes between cases and controls. We evaluated the 266 genes containing germline mutations associated with an increased risk of developing PCa. To get a complete picture, we performed a series of investigations. First, we investigated whether genes containing germline variation also harbor somatic mutation and whether these genes are differentially expressed between patients with aggressive tumors and control samples. Specifically, we sought to discover a gene signature with germline and somatic mutations. Secondly, we investigated whether there are any genes containing germline variations that are differentially expressed between patients with aggressive PCa and control samples, but are not somatically altered. Thirdly, we investigated whether there are somatically altered genes that are differentially expressed between patients with aggressive PCa and control samples but are not altered in the germline.

The results of these analyses are presented in a Venn diagram in Figure 2. Out of the 266 genes containing germline mutations evaluated, 90 genes contained both germline and somatic mutations (Figure 2). This analysis also revealed a 61 gene signature containing both germline and somatic mutations. The second investigation produced a 117 gene signature containing germline mutations only (Figure 2). The third investigation yielded a 2552 gene signature containing somatic mutations only (Figure 2). A total of 59 genes contained germline mutations only and were not significantly differentially expressed. A set of 29 genes contained both germline and somatic mutations and were not significantly differentially expressed. Some 2137 genes contained somatic mutations only and were not significantly differentially expressed (Figure 2). About 10072 genes were significantly differentially expressed but contained neither germline nor somatic alterations (Figure 2).

The results showing a signature of the 61 genes containing both germline and somatic mutations are presented

in Table 2. The signature includes the genes *PGBDI*, *OXR1*, *GZFI*, *ITGAX*, *ORC3*, *BMPRI1B*, *KLK3*, *KLK2*, *NKX3-1*, *SLC22A3*, *POU5F1B*, *LMTK2*, *NAALADL2*, *LDAH*, *PDLIM5*, *SLC22A3*, *JAZF1*, *LMTK2*, *CASC8*, *DAP2IP*, *TIMM23B*, *MSMB*, *MYEOV*, *FLT1*, *SL35B4*, and *HNF1B* containing genetic variants reported to be directly associated with aggressive PCa (Table 2) (references provided in supplementary Table SA). Also presented in the table are the SNP rs-IDs, the GWAS *P* value indicating the strength of germline mutation association PCa, the frequency of germline and somatic mutation events, and the gene expression *P* value indicating the level of significance in expression levels between tumors and control samples for individual genes.

There was significant variation in the distribution of germline and somatic mutations among the genes. Overall the genes containing germline variation did not have a high frequency of somatic mutations (Table 2). The genes *FGFR2*, *HOXB13*, *RNASL*, *PDLIM5*, *NKX3-1*, *KLK2*, *POU5F1B*, *KLK3*, *SLC19A2*, *SLC22A3*, *ITGA6*, *TXB5*, *ZNF827*, and *SLC41A1* contained more than one germline mutation (Table 2). The genes *KLK3* and *KLK2* were the most highly germline mutated. The genes *LRPIB*, *BCL11A*, *SX1*, *DNAH5*, *PRDM15*, *GLI2*, *TRIM31*, *TCF7L2*, *ATF7IP*, *KIAA1211*, *ZNF827*, *FREMI*, *MDM4*, *TBX3*, *LMTK2*, *STAT3*, *PKHDI*, and *IL16* contained more than one somatic mutation (Table 2). The gene *LRPIB* was the most highly somatically altered (Table 2). Interestingly, the genes *RNASL*, *FTO*, *BMPRI1A*, *ITGA6*, *TCF7L2*, *FREMI*, *LMTK2*, and *STAT3* containing germline mutations with weak to moderate associations were found to contain germline and somatic mutations and were differentially expressed between PCa and controls (Table 2).

**3.3. Patterns of Expression Profiles for Containing Germline and Somatic Mutations.** To investigate whether the 61 genes containing both germline and somatic mutations are coregulated and have similar patterns of expression profiles, we performed hierarchical clustering. The results showing the patterns of expression profiles are presented in Figure 3. The analysis revealed that genes containing germline and somatic mutations have similar patterns of expression profiles, suggesting that they are likely to be coregulated and functionally related. As expected, there was significant variation in the patterns of gene expression profiles. The variability in patterns of expression profiles among the genes containing germline and somatic mutations (Figure 3) can partially be explained by the fact that these genes were derived from different populations and different clinical phenotypes. Under such condition the observed outcome is expected. Five genes (*KLK2*, *KLK3*, *NKX3-1*, *HOXB13*, and *SIX1*) were consistently highly expressed in both tumors and control samples (Figure 3). The genes *EPHA10*, *PDLIM5*, *POU5F1B*, *DNAH5*, *EIF2S3*, *SLC19A2*, *NAALADL2*, *ZNF652*, *MYO6*, *EBF2*, *TBX1*, and *MD4*, were upregulated in tumors and down regulated in controls (Figure 3).

Additional hierarchical clustering combining the 61 genes containing germline and somatic mutations with the 117 genes containing germline mutations only also revealed

TABLE 2: List of the 61-gene signature containing both germline and somatic mutations found to be significantly differentially expressed between aggressive tumors and controls. \*\* indicates the gene containing genetic variants directly associated with aggressive PCa.

<i>Genes</i>	Chrom	GWAS		Germline	DE	
		SNP_ID	<i>P</i> value		Somatic	<i>P</i> value
<i>PGBDI</i> **	6p22.1	rs1233708	0.004	1	2	0.02169
<i>OXRI</i> **	8q23	rs16901979	0.002	1	1	1.52E-20
<i>GZFI</i> **	20p11.21	rs6076072	0.01	1	1	2.42E-12
<i>ITGAX</i> **	16p11.2	rs8045738	0.006	1	1	1.20E-11
<i>ORC3</i> **	6q15	rs9450716	0.002	1	1	3.31E-06
<i>BMPRIB</i> **	4q22	rs17021918	0.03	1	1	3.17E-06
<i>KLK3</i> **	19q13.41	rs2739472	9.00E-186	15	1	4.43E-13
<i>KLK2</i> **	19q13.33	rs1354774	6.00E-20	5	1	4.53E-19
<i>NKX3-1</i> **	8p21.2	rs13272392	4.00E-34	4	1	3.30E-20
<i>SLC22A3</i> **	6q25.3	rs4646284	3.20E+52	4	1	6.62E-11
<i>RNASEL</i>	1q25	rs486907	0.004	3	1	8.88E-26
<i>ZNF827</i>	4q31.22	rs56935123	4.00E-09	2	3	9.16E-06
<i>FGFR2</i>	10q26.12	rs10886902	2.00E-53	2	1	8.24E-29
<i>HOXB13</i>	8q24.21	rs188140481	6.00E-34	2	1	2.36E-27
<i>PDLIM5</i>	4q22	rs17021918	4.20E-15	2	1	1.36E-21
<i>POU5F1B</i> **	8q24.21	rs16901979	1.00E-16	2	1	4.94E-17
<i>SLC19A2</i>	1q23.3	rs3765227	1.26E-04	2	1	6.46E-11
<i>ITGA6</i>	2q31.1	rs126212278	0.001	2	1	5.56E-10
<i>TBX5</i>	12q24.1	rs1270884	6.75E-11	2	1	1.53E-09
<i>SLC41A1</i>	1q32.1	rs6679073	4.00E-15	2	1	0.026695
<i>LRPIB</i>	2q21.2	rs10210358	2.00E-06	1	10	7.28E-33
<i>PKHD1</i>	6p21.2	rs10498792	3.00E-06	1	4	0.01337
<i>ATF7IP</i>	12p13.1	rs3213764	2.00E-09	1	3	1.13E-06
<i>FREMI</i>	9p22.3	rs1552895	0.002	1	3	7.30E-05
<i>BCL11A</i>	2p16.1	rs2556375	6.00E-19	1	2	2.02E-17
<i>SIX1</i>	14q23.1	rs7153648	2.00E-09	1	2	4.77E-16
<i>DNAH5</i>	5p15.2	rs4463179	2.00E-06	1	2	8.71E-16
<i>PRDM15</i>	21q22.3	rs6586243	7.79E-06	1	2	6.34E-13
<i>GLI2</i>	2q14	rs11122834	5.00E-06	1	2	1.28E-11
<i>TRIM31</i>	6p22.1	rs115457135	1.00E-07	1	2	4.61E-11
<i>TCF7L2</i>	10q25.3	rs7903146	0.009	1	2	1.43E-08
<i>KIAA1211</i>	4q12	rs629242	7.25E-07	1	2	5.48E-06
<i>MDM4</i>	1q32	rs4245739	2.01E-11	1	2	9.95E-05
<i>TBX3</i>	12q24.21	rs11067228	1.00E-14	1	2	0.000208
<i>LMTK2</i> **	7q22.1	rs6465657	0.007	1	2	0.00243
<i>STAT3</i> **	17q21	rs744166	0.03	1	2	0.002775
<i>IL16</i>	15q26.3	rs7175701	9.80E-08	1	2	0.026789
<i>EPHA10</i>	1p34.3	rs731174	5.00E-06	1	1	1.44E-35
<i>FERMT2</i>	14q22.1	rs8008270	1.78E-14	1	1	3.20E-30
<i>FTO</i>	16q12.2	rs9939609	0.04	1	1	6.96E-16
<i>BMPRI1A</i>	10q22.3	rs11597689	0.03	1	1	1.09E-13
<i>EIF2S3</i>	Xp22.11	rs6627995	1.00E-13	1	1	1.21E-12
<i>TTC7A</i>	2p16.3	rs10194115	5.00E-07	1	1	1.81E-11
<i>NAALADL2</i> **	3q26.31	rs78943174	4.00E-08	1	1	4.07E-10
<i>PPFIBP2</i>	11p15.4	rs12791447	4.00E-08	1	1	5.20E-10
<i>KCNN3</i>	1q21.3	rs1218582	1.95E-08	1	1	9.96E-10
<i>AMIGO2</i>	12q13.11	rs2711721	2.00E-06	1	1	2.17E-09
<i>ZNF652</i>	17q21.32	rs7210100	3.40E-13	1	1	5.55E-09
<i>MYO6</i>	6q14.1	rs9443189	4.00E-08	1	1	2.71E-08
<i>EBF2</i>	8p21.2	rs11135910	8.00E-11	1	1	8.98E-08

TABLE 2: Continued.

Genes	Chrom	GWAS		DE		
		SNP_ID	P value	Germline	Somatic	P value
<i>TBX1</i>	22q11.21	rs2238776	2.00E-08	1	1	7.97E-08
<i>PTGFRN</i>	1p13.1	rs2806864	6.00E-07	1	1	2.52E-07
<i>TNC</i>	9q33.1	rs7847271	4.00E-06	1	1	3.93E-07
<i>ATP9B</i>	18q23	rs7241993	2.00E-09	1	1	4.37E-06
<i>HAPLN1</i>	5q14.3	rs4466137	3.00E-06	1	1	6.25E-05
<i>NOTCH4</i>	6p21.3	rs3096702	4.78E-09	1	1	5.81E-05
<i>VGLL3</i>	3p12.1	rs9757252	5.00E-06	1	1	7.72E-05
<i>IKZF2</i>	2q13.1	rs7569918	8.85E-04	1	1	9.51E-05
<i>KLF4</i>	9q31.2	rs817826	5.00E-14	1	1	0.002599
<i>TNS3</i>	7p12.3	rs56232506	9.00E-09	1	1	0.007265
<i>DDHD1</i>	14q22.1	rs8008270	2.00E-14	1	1	0.008911

Note. DE indicates differential expression.

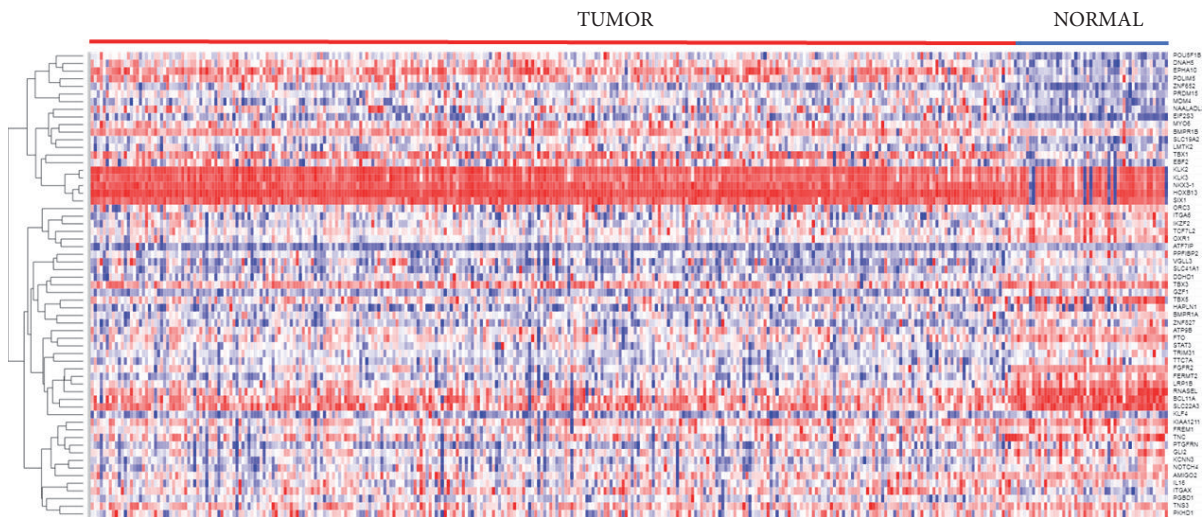


FIGURE 3: Patterns of expression profiles for the 61 genes containing both germline and somatic mutations distinguishing patients with aggressive tumors and controls generated using hierarchical clustering.

similarities in patterns of gene expression profiles (results not presented here). Further analysis combining the 61 genes containing germline and somatic mutations with highly somatic mutated genes (i.e., >5 somatic mutation events per gene) also revealed similarity in patterns of expression profiles (results not presented). This suggests that genes containing germline and somatic mutations are likely coregulated and may not only be involved in *cis* regulation but also in *trans* regulation.

**3.4. Enrichment Analysis of Molecular Networks and Biological Pathways.** To gain insights about the broader biological context through which interactions and cooperation among and between genes containing germline and somatic mutations operate, we performed network and pathway analysis using IPA. We hypothesized that genes containing germline and somatic mutations are functionally related and interact with one another in molecular networks and biological pathways enriched for germline and somatic mutations. Thus,

we sought to discover molecular networks and biological pathways enriched for germline and somatic mutations. As a first step, we performed network and pathway analysis including the 61 genes containing germline and somatic mutations from Table 2 and the 34 genes that were found to contain high somatic mutation events (>5 somatic mutation events) from Table 1. In the second step, we combined the 61 genes containing germline variation and somatic mutations and the 34 highly somatic mutated genes with the 117 genes containing only germline mutations. Here we sought to investigate the possible oncogenic interactions between the germline and the somatic genomes. In these analyses, we filtered out genes with spurious and predicted interactions, keeping only genes with  $\geq 3$  direct connections to ensure the reliability of networks and functional relationships among the genes in the networks.

The results of network and pathway analysis for the 61 genes containing both germline and somatic mutations

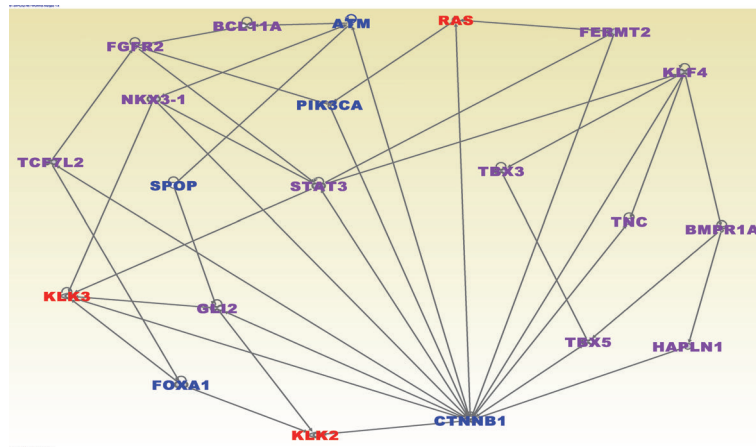


FIGURE 4: Molecular networks showing interactions among genes containing both germline and somatic mutations. Gene names (symbols) are presented in the nodes. Gene names in purple color font contain both germline and somatic mutations. Gene names in blue color fonts indicate genes with more >5 somatic mutations whereas gene names in red font indicate genes containing germline mutations directly associated with aggressive prostate cancer. Note that the networks were filtered to remove spurious connections.

and the 34 genes with high somatic mutation frequency are presented in Figure 4. In the figure, genes are presented as nodes and the vertices or solid lines indicate relationships among the genes. The font color of the gene symbol indicates the mutation events as described in the figure legend. Interestingly, out of the 95 genes evaluated, 94 were predicted to be significantly involved in cancer ( $P < 1.47 \times 10^{-22}$ ), with 68 predicted to be directly involved in PCa.

As expected, the analysis revealed genes with multiple overlapping functions. The network with the highest Z-score (Z-score = 44) contained genes *EPHBI*, *FERM2*, *FREMI*, *GLI2*, *HOXB13*, *KIFI6B*, *KLF4*, *KLK2*, *KLK3*, *MDM4*, *NKX3-1*, *PCDH18*, *PKHDI*, *RNASEL*, *SIX1*, *SPOP*, *TBX1*, *TBX3*, *TBX5*, and *VPS13D* predicted to be involved in organism and tissue development. The second networks with a Z-score = 31 revealed the genes *ATF7IP*, *CTNNB1*, *EIF2S3*, *FLG*, *FOXA1*, *FZD4*, *IKZF2*, *KCNN3*, *MACF1*, *PPFIBP2*, *PRDM15*, *PTPRD*, *SLC19A2*, *STAT3*, *TCF7L2*, and *TNS3* predicted to be involved in cancer, cellular development, and proliferation.

In the network (Figure 4) the genes in purple color font contain both germline and somatic mutations, the genes in blue color fonts indicate genes with high somatic mutation frequency >5, whereas gene names in red font indicate genes containing germline mutations directly associated with aggressive PCa. As evidenced in Figure 4, the genes containing germline and somatic mutations were found to be functionally related and interacting with one another. In addition, the genes containing germline and somatic mutations were found to be interacting with the most highly somatically mutated genes.

Pathway analysis revealed various signaling pathways enriched for both germline and somatic mutations. Among the most significant pathways included the PCa signaling pathway ( $P < 5.81 \times 10^{-6}$ ) with associated genes *ATM*, *CTNNB1*, *FGFR2*, *KLK3*, *NKX3-1*, and *PIK3CA*; the MSP-RON signaling pathway ( $P < 1.54 \times 10^{-5}$ ) with associated genes *ATM*, *FGFR2*, *KLK2*, *KLK3*, and *PIK3CA* and the P53 signaling

pathway ( $P < 1.24 \times 10^{-4}$ ) and the associated genes *ATM*, *CTNNB1*, *FGFR2*, *MDM4*, and *PIK3CA*; and the molecular mechanisms of cancer signaling pathway ( $P < 3.19 \times 10^{-4}$ ) with associated genes *ATM*, *BMPR1A*, *BMPR1B*, *CTNNB1*, *FGFR2*, *FZD4*, *PIK3CA*, and *PLCB4*. Interestingly, the discovered pathways included genes containing genetic variants directly associated with aggressive PCa such as *ATM* and *KLK3*.

The results of network and pathway analysis based on combining the 61 genes containing germline variation and somatic mutations and the 34 highly somatically mutated genes with the 117 genes containing germline mutations only are presented in Figure 5. Here we sought to investigate the possible oncogenic interactions between the germline and the somatic genomes. The analysis revealed 15 networks containing genes with multiple overlapping functions with z-scores ranging from 2 to 34. The top networks with Z-score 30 – 34 contained genes predicted to be involved in cancer and cell cycle. The results showing the functional relationships among the genes in the three top networks are shown in Figure 5. Network analysis revealed that genes containing germline and somatic mutations, high somatic mutation frequency and genes containing only germline mutations are functionally related (Figure 5). Interestingly, genes containing genetic variants directly associated with aggressive PCa (*KLK2*, *JAZF1*, *KLK3*, and *HNFBIP*) were found to be interacting with other sets of genes.

Pathway analysis revealed biological pathways enriched for both germline and somatic mutations. Interestingly, genes containing germline variations only were found to be involved in the same pathways with genes containing somatic mutations only. This suggests that germline and somatic mutations may cooperate through the pathways. The most significant pathways were the MSP-RON signaling pathway ( $P < 3.21 \times 10^{-8}$ ), molecular mechanisms of cancer ( $P < 3.31 \times 10^{-7}$ ), PDGF signaling ( $P < 4.00 \times 10^{-7}$ ), Axonal guidance signaling ( $P < 5.60 \times 10^{-7}$ ), P53 signaling ( $P < 1.38 \times 10^{-6}$ ) and STAT3 signaling pathway ( $P < 7.40 \times 10^{-6}$ ). Interestingly in



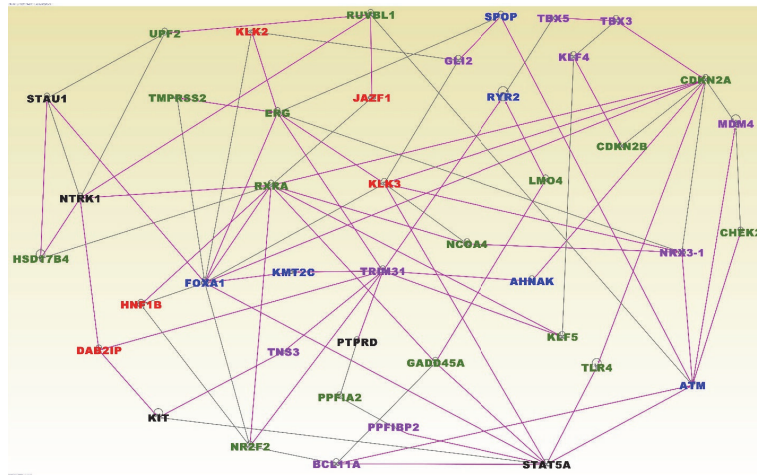


FIGURE 5: Molecular networks showing interactions among genes containing both germline and somatic mutations. Gene names (symbols) are presented in the nodes. Gene names in purple color font contain both germline and somatic mutations. Gene names in blue color fonts indicate genes with more >5 somatic mutations, gene names in green font indicate genes containing germline mutations only, and gene names in red font indicate genes containing germline mutations directly associated with aggressive prostate cancer. Note that the networks were filtered to remove spurious connections. The colors red and black represent overlap in functional relations among genes in merged networks.

both network and pathway analysis, genes containing genetic variants with strong and moderate to weak associations were found to be functionally related and interacting with genes containing somatic mutations.

#### 4. Discussion

Studies of cancer genomes have been mostly concerned with understanding the somatic mutations or transcriptome changes that arise during tumorigenesis. For many years and decades, germline and somatic mutation information have generally been analyzed and reported as separate endeavors. Hundreds of genetic variants and genes associated with an increased risk of developing PCa have been reported in GWAS [5]. Similarly, hundreds of somatic mutations and genes have been reported [9, 10]. However, emerging evidence indicates that germline variations may interact with somatic events to drive carcinogenesis [33–35]. In this exploratory study, we investigated the possible oncogenic interactions and cooperation between the genes containing germline and somatic mutations. We found evidence that genes containing germline mutations also contain somatic mutations, and that these genes are functionally related and interact with one another in molecular networks and biological pathways enriched for both genetic alterations. Our analysis reveals that oncogenic interactions and cooperation between germline and somatic mutations likely occurs through molecular networks and biological pathways. To our knowledge this is the first study to show that genes containing germline mutations also contain somatic mutations and that these genes interact with one another in molecular networks and biological pathways in aggressive PCa. The novel aspect of the study is that germline and somatic mutation may interact and cooperate through molecular networks and

biological pathways to drive aggressive PCa. The clinical significance of these results is that such information could be used for the development of new precision prevention strategies and polygenic risk scores to identify men at high risk of developing aggressive PCa [6, 7].

In this study, evaluation of somatic mutation profiles revealed that genes containing genetic susceptibility variants did not exhibit an overall increase in somatic mutation frequency. This is consistent with the results of a recent study which explored somatic mutation profiles within susceptibility regions in a range of cancers [36]. In that report, the authors found that genes in cancer susceptibility regions did not exhibit high somatic mutation frequency [36]. However, the published analysis considered overall somatic mutations, which represented a substantial proportion of passenger mutations, while our study focused on network and pathway analysis to gain insights about the broader biological context in which germline and somatic mutations may cooperate to drive aggressive PCa. Importantly, network and pathway analysis showed that germline and somatic mutations were involved in many common signaling pathways which have been implicated in PCa. Interestingly, many highly somatically mutated genes not containing germline mutations were found to interact with germline mutated genes in biological pathways. This suggests that cooperation between germline and somatic alteration may not only involve *cis* regulation but also *trans* regulation through molecular networks and biological pathways. The clinical significance is that such pathways could serve as targets for the development of novel targeted therapeutics. Although we did not investigate the regulatory mechanisms and the impact of germline variation on gene expression in this study, association of PCa risk variants with gene expression in tumor and normal tissue has been reported [37]. Moreover, *cis*-regulatory variation in

the human prostate transcriptome by using gene-level allele-specific expression has been reported [38].

The low somatic mutation frequency in germline mutated genes could be due to natural selection which is likely to confer selective advantage to specific genes [39]. For example, genomic alteration, clonal selection, and evolution of the tumor microenvironment, may contribute towards unique physiological characteristics under selection pressure contributing to increased mutations in the somatic or tumor genome as observed in this study for the most highly somatic mutated genes, a phenomenon which has been reported in PCa [39]. Although our approach focuses on PCa, the methodology and approach could be applied to other cancers as well.

This exploratory study investigated the possible oncogenic interactions and cooperation among and between genes containing germline and somatic mutations. However, limitations of the study must be acknowledged. Important limitations include the lack of consistent methodological reporting between studies examining low and high grade PCa in GWAS studies which can confound the results. Additionally, genetic variants and genes from GWAS were derived from many diverse populations. Gene expression can be population-specific and there is evidence from the literature that some genetic susceptibility variants may confer population-specific risks [40]. It is worth noting that this study focused on men of European ancestry because the population is well represented in both GWAS and TCGA. Future studies should include other ethnic populations, especially those underrepresented in genomic studies like African American men who are disproportionately affected by aggressive PCa.

## 5. Conclusions

Our findings suggest possible oncogenic interactions and cooperation between genes containing germline variation and somatic mutations. The results suggest a complex interplay between germline and somatic variation, underscoring the need for further studies to understand how the genetic variants and somatic alterations interact to drive aggressive PCa in different populations. Additionally, functional experiments are warranted to uncover the molecular mechanisms underlying these interactions and cooperation between the germline and the somatic genomes.

## Data Availability

Germline mutation information from GWAS along with sources of original information is provided in the supplementary materials in supplementary Table SA. Original data on somatic mutations, gene expression data, and clinical information is available at the TCGA via the Genomics Data Commons <https://gdc.cancer.gov/>.

## Disclosure

The content in this report is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health or any funding source.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

The authors wish to thank the school of medicine for providing funding in support of this research and the patients who volunteered and provide the tumor samples used to generate both GWAS and TCGA data. This research was supported by Louisiana State University School of Medicine in New Orleans. LSUHSC # U54 GM12254691 and UAB # ULITR001417 grants are also acknowledged.

## Supplementary Materials

Table S1B: list of significantly differentially expressed nonsomatic mutated genes associated with aggressive PCa along with estimates of  $P$  values, adjusted  $P$  values, and log fold change. Table SA: a comprehensive list of single nucleotide polymorphisms (herein called genetic variants). Table S1A: list of significantly differentially expressed somatic mutated genes associated with aggressive PCa. (*Supplementary Materials*)

## References

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 1, pp. 7–30, 2017.
- [2] American Cancer Society, *Cancer Facts and Figures 2017* (Annual Report).
- [3] M. S. Litwin and H. Tan, "The diagnosis and treatment of prostate cancer," *Journal of the American Medical Association*, vol. 317, no. 24, pp. 2532–2542, 2017.
- [4] L. M. FitzGerald, E. M. Kwon, M. P. Conomos et al., "Genome-wide association study identifies a genetic variant associated with risk for more aggressive prostate cancer," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 20, no. 6, pp. 1196–1203, 2011.
- [5] C. Hicks, L. Miele, T. Koganti, and S. Vijayakumar, "Comprehensive assessment and network analysis of the emerging genetic susceptibility landscape of prostate cancer," *Cancer Informatics*, vol. 12, pp. 175–191, 2013.
- [6] M. Aly, F. Wiklund, J. Xu et al., "Polygenic risk score improves prostate cancer risk prediction: Results from the Stockholm-1 cohort study," *European Urology*, vol. 60, no. 1, pp. 21–28, 2011.
- [7] N. Shao, Y. Zhu, and D. Ye, "Role of polygenic hazard score in prostate-specific antigen (PSA) screening for prostate cancer," *Translational Cancer Research*, vol. 7, no. S7, pp. S743–S744, 2018.
- [8] A. Abeshouse, J. Ahn, R. Kbani et al., "The molecular taxonomy of primary prostate cancer," *Cell*, vol. 163, no. 4, pp. 1011–1025, 2015.
- [9] J. N. Weinstein, E. A. Collisson, G. B. Mills et al., "The cancer genome atlas pan-cancer analysis project," *Nature Genetics*, vol. 45, no. 10, pp. 1113–1120, 2013.
- [10] The International Cancer Genome Consortium, "International network of cancer genome projects," *Nature*, vol. 464, pp. 993–998, 2010.

- [11] Y. Wang, C. Wang, J. Zhang et al., "Interaction analysis between germline susceptibility loci and somatic alterations in lung cancer," *International Journal of Cancer*, vol. 143, no. 4, pp. 878–885, 2018.
- [12] H. S. Feigelson, K. A. B. Goddard, C. Hollombe et al., "Approaches to integrating germline and tumor genomic data in cancer research," *Carcinogenesis*, vol. 35, no. 10, pp. 2157–2163, 2014.
- [13] D. Welter, J. MacArthur, J. Morales et al., "The NHGRI GWAS Catalog, a curated resource of SNP-trait associations," *Nucleic Acids Research*, vol. 42, no. 1, pp. D1001–D1006, 2014.
- [14] J. P. A. Ioannidis, P. Boffetta, J. Little et al., "Assessment of cumulative evidence on genetic associations: Interim guidelines," *International Journal of Epidemiology*, vol. 37, no. 1, pp. 120–132, 2008.
- [15] M. J. Khoury, L. Bertram, P. Boffetta et al., "Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases," *American Journal of Epidemiology*, vol. 170, no. 3, pp. 269–279, 2009.
- [16] G. S. Sagoo, J. Little, and J. P. Higgins, "Systematic reviews of genetic association studies," *PLoS Medicine*, vol. 6, no. 3, p. e28, 2009.
- [17] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *Journal of Clinical Epidemiology*, vol. 151, pp. 264–269, 2009.
- [18] A. Liberati, D. G. Altman, J. Tetzlaff et al., "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration," *PLoS Medicine*, vol. 6, Article ID e1000100, 2009.
- [19] dbSNP, <http://www.ncbi.nlm.nih.gov/SNP/>.
- [20] Human Genome Nomenclature Committee (HGNC), <http://www.genenames.org/>.
- [21] M. S. Cookson, G. Aus, A. L. Burnett et al., "Variation in the definition of biochemical recurrence in patients treated for localized prostate cancer: the American urological association prostate guidelines for localized prostate cancer update panel report and recommendations for a standard in the reporting of surgical outcomes," *The Journal of Urology*, vol. 177, no. 2, pp. 540–545, 2007.
- [22] National Cancer Institute, Genomic Data Commons, <https://gdc.cancer.gov/>.
- [23] M. E. Ritchie, B. Phipson, D. Wu et al., "Limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, p. e47, 2015.
- [24] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society B: Methodological*, vol. 57, no. 1, pp. 289–300, 1995.
- [25] Morpheus, <https://software.broadinstitute.org/morpheus/>.
- [26] Ingenuity Pathways Analysis (IPA) System, Redwood, CA: Ingenuity Systems, <http://www.ingenuity.com/>.
- [27] M. Ashburner, C. A. Ball, J. A. Blake et al., "Gene ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, no. 1, pp. 25–29, 2000.
- [28] Y. Cheng, P. Yu, X. Duan et al., "Genome-wide analysis of androgen receptor binding sites in prostate cancer cells," *Experimental and Therapeutic Medicine*, vol. 9, no. 6, pp. 2319–2324, 2015.
- [29] H. H. Cheng, C. C. Pritchard, B. Montgomery, D. W. Lin, and P. S. Nelson, "Prostate cancer screening in a new era of genetics," *Clinical Genitourinary Cancer*, vol. 15, no. 6, pp. 625–628, 2017.
- [30] M. Gillard, J. Lack, A. Pontier et al., "Integrative genomic analysis of coincident cancer foci implicates CTNNB1 and PTEN alterations in ductal prostate cancer," *European Urology Focus*, 2017.
- [31] S. Gupta, K. Iljin, H. Sara et al., "FZD4 as a mediator of ERG oncogene-induced WNT signaling and epithelial-to-mesenchymal transition in human prostate cancer cells," *Cancer Research*, vol. 70, no. 17, pp. 6735–6745, 2010.
- [32] P. Mariot, N. Prevarskaya, M. M. Roudbaraki et al., "Evidence of functional ryanodine receptor in apoptosis of prostate cancer (LNCaP) cells," *The Prostate*, vol. 43, no. 3, pp. 205–214, 2000.
- [33] H. Carter, R. Marty, M. Hofree et al., "Interaction landscape of inherited polymorphisms with somatic events in cancer," *Cancer Discovery*, vol. 7, no. 4, pp. 410–423, 2017.
- [34] N. Bonifaci, B. Górski, B. Masojć et al., "Exploring the Link between Germline and Somatic Genetic Alterations in Breast Carcinogenesis," *PLoS ONE*, vol. 5, no. 11, p. e14078, 2010.
- [35] T. G. Grünewald and O. Delattre, "Cooperation between somatic mutations and germline susceptibility variants in tumorigenesis – a dangerous liaison," *Molecular & Cellular Oncology*, vol. 3, no. 3, p. e1086853, 2015.
- [36] M. J. Machiela, B. M. Ho, V. A. Fisher, X. Hua, and S. J. Chanock, "Limited evidence that cancer susceptibility regions are preferential targets for somatic mutation," *Genome Biology*, vol. 16, no. 1, article no. 193, 2015.
- [37] K. L. Penney, J. A. Sinnott, S. Tyekucheva et al., "Association of prostate cancer risk variants with gene expression in normal and tumor tissue," *Cancer Epidemiology Biomarkers & Prevention*, vol. 24, no. 1, pp. 255–260, 2015.
- [38] N. Larson, S. McDonnell, A. French et al., "Comprehensively evaluating cis -regulatory variation in the human prostate transcriptome by using gene-level allele-specific expression," *American Journal of Human Genetics*, vol. 96, no. 6, pp. 869–882, 2015.
- [39] C. Song, "Darwinian selection in prostate cancer and medical treatment," *International Journal of Clinical Medicine*, vol. 08, no. 06, pp. 353–367, 2017.
- [40] D. V. Conti, K. Wang, X. Sheng, and J. T. Bensen, "Two novel susceptibility loci for prostate cancer in men of african ancestry," *Journal of the National Cancer Institute*, vol. 9, no. 8, 2017.