



Original Research

A network-based integration for understanding racial disparity in prostate cancer

Baoyi Zhang^a, Kevin Yao^b, Chao Cheng^{c,d,e,*}^a Department of Chemical and Biomolecular Engineering, Rice University, Houston, TX 77030, United States^b Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, United States^c Department of Medicine, Baylor College of Medicine, Houston, TX 77030, United States^d Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX 77030, United States^e Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX 77030, United States

ARTICLE INFO

Keywords:

Prostate cancer
African American
Network

ABSTRACT

Compared to Caucasians (CAs), African Americans (AAs) have a higher rate of incidence and mortality in prostate cancer and are prone to be diagnosed at later stages. To understand this racial disparity, molecular features of different types, including gene expression, DNA methylation and other genomic alterations, have been compared between tumor samples from the two races, but led to different disparity associated genes (DAGs). In this study, we applied a network-based algorithm to integrate a comprehensive set of genomic datasets and identified 130 core DAGs. Out of these genes, 78 were not identified by any individual dataset but prioritized and selected through network propagation. We found DAGs were highly enriched in several critical prostate cancer-related signaling transduction and cell cycle pathways and were more likely to be associated with patient prognosis in prostate cancer. Furthermore, DAGs were over-represented in prostate cancer risk genes identified from previous genome wide association studies. We also found DAGs were enriched in kinase and transcription factor encoding genes. Interestingly, for many of these prioritized kinases their association with racial disparity did not manifest from the original genomic/transcriptomic data but was reflected by their differential phosphorylation levels between AA and CA prostate tumor samples. Similarly, the disparity relevance of some transcription factors was not reflected at the mRNA or protein expression level, but at the activity level as demonstrated by their differential ability in regulating target gene expression. Our integrative analysis provided new candidate targets for improving prostate cancer treatment and addressing the racial disparity problem.

Introduction

Prostate cancer is the second leading cause of cancer-related death among men in the US [1]. The incidence and mortality of prostate cancer vary greatly among races. Specifically, African Americans (AAs) have about 75% higher morbidity and over twice the mortality rate of prostate cancer than Caucasians (CAs) [2]. Additionally, AAs with prostate cancer are usually diagnosed at a late stage and their disease progresses quickly [3,4]. Although external factors such as socioeconomic status, health insurance coverage, and household income have been found to contribute to racial disparity in prostate cancer [5–7], more research has focused on the impact of internal factors, the molecular features such as

somatic mutation and gene expression between AA and CA prostate cancers [8–10]. Previous studies have shown distinct genetic profiles among AA prostate tumors, including PTEN loss, MYC amplification, and SPINK1 overexpression [10–12].

Although previous studies have extensively compared molecular features between AAs and CAs [8–10,13–16], their results are not always consistent and may even be conflicting. For example, some studies found AAs had higher frequencies of SPOP mutations [10,15], while others found low SPOP mutation frequencies in AAs compared with CAs [8]. Furthermore, there are still debates regarding the clinical implications of these differential molecular features between AA and CA, so it is inconclusive whether these molecular differences are causally related to

Abbreviations: CA, Caucasian; AA, African American; DAG, racial disparity associated gene; CNV, copy number variation; PRAD, prostate adenocarcinoma; GEO, gene expression omnibus; RPPA, reverse phase protein array; FDR, false discovery rate; GWAS, genome-wide association study; CRPC, castration-resistant prostate cancer; GO, Gene Ontology.

* Corresponding author at: Baylor College of Medicine, One Baylor Plaza, ICTR D100.1, Houston, TX 77030, United States.

E-mail address: chao.cheng@bcm.edu (C. Cheng).

<https://doi.org/10.1016/j.tranon.2021.101327>

Received 20 December 2021; Accepted 21 December 2021

Available online 5 January 2022

1936-5233/© 2022 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

racial disparity in prostate cancer. As an example, SPINK1 was found to be overexpressed in AA across many studies [8,10,12,15], but its role in the development of prostate cancer remains unclear. Several studies reported no association between SPINK1 expression and poorer clinical outcomes [17–19], while others found it could predict disease-specific and recurrence-free survival in prostate cancer [20,21].

During tumorigenesis, genetic and epigenetic aberrations of genes will integrate at the pathway level to determine the ultimate phenotypes. Considering race as a phenotype variation, the seemingly inconsistent observations at the gene level between different studies might be explained at the pathway or regulatory network level. Using the SPOP mutation as an example, it is possible that other genes in related pathways also contribute to racial disparity, e.g., the onco-protein AR and SRC in the AR signaling pathway that are regulated by SPOP-mediated degradation [22]. Indeed, it has been reported that differentially expressed genes between AA and CA prostate cancer are enriched in AR-related pathways [23]. Interestingly, differential expression of AR itself between the two races are still under debate [13,24]. Thus, in this study we propose to integrate different types of data (genomic, epigenomic, and transcriptomic) to better understand the inconsistency between them and to identify a core set of genes associated with racial disparity in prostate cancer.

Network-based methods have been widely used to integrate different data types. For example, hotnet2 [25] integrates CNV and somatic mutation data through gene interaction networks and uses a network diffusion approach to identify dysregulated pathways in certain cancer types. Nevertheless, hotnet2 only considers somatic aberration data and ignores other molecular aberrations such as DNA methylation and RNA expression. Therefore, NetICS [26] integrates more data types such as DNA methylation and RNA expression to identify mediator genes within networks, which are affected by upstream genetic aberrations and lead to altered downstream gene expressions. This approach has been proved to accurately predict cancer genes based on TCGA data and might be applied to other scenarios as well.

In this study, we assume that there is a core set of genes that receive the effects of upstream genetic alterations and trigger downstream abnormal gene expression through these networks, thereby driving the racial disparity in prostate cancer. Based on this assumption, we performed a series of comparative analyses between AAs and CAs using multiple genomic data types including copy number variation, somatic mutation, DNA methylation and RNA expression to identify differential altered genes between the two races. Detected differential alterations were further mapped into a protein-protein interaction network on gene level. By applying a random walk strategy to propagate influence of these alterations through the network, we prioritized genes associated with racial disparity and identified a set of 130 core disparity associated genes (DAGs). These DAGs had important clinical implications in prostate cancer, including prognosis, progression and development. Furthermore, we found that these DAGs dysregulated several pathways in prostate cancer, including MAPK signaling and AR signaling pathways. Potential targets for precision therapy may lie in these pathways and DAGs, providing new insights into handling the racial disparity in prostate cancer.

Materials and methods

Datasets

Somatic mutation, copy number variation (CNV), DNA methylation, RNA expression and clinical information for TCGA prostate adenocarcinoma (PRAD) were downloaded through Firehose portal (<http://gdac.broadinstitute.org/>). Specifically, mutation annotation files were downloaded, which contained mutation status of 24,058 genes for 332 samples. The original CNV data provided segmental chromosome CNVs. We calculated relative CNV for each segment by $\log_2(\text{copy-number}/2)$. A gene's CNV was then determined by averaging the

segment CNVs with which it overlaps, weighted by the ratio of overlapping length to the gene's length. Using such method, we obtained relative copy numbers of 23,311 genes for 492 samples. DNA methylation level in genes was represented by the averaged β values of their promoters, covering 21,855 genes in 498 samples. RNA-seq data included 20,502 genes for 497 samples. From the clinical data, we extracted 58 AAs and 415 CAs. Samples other than AA and CAs and those of unknown race were excluded. Protein expression data for prostate cancer were downloaded from TCGA from reverse phase protein array (RPPA) experiment, which included the expression level of 195 proteins in 352 prostate cancer samples (32 AA and 301 CA). For some protein, the dataset provides their abundance in both phosphorylated and unphosphorylated states.

Four processed microarray datasets were downloaded from the Gene Expression Omnibus (GEO) database with accession IDs: GSE6956 [27], GSE17356 [28], GSE41969 [29], GSE91037 [30]. Specifically, the four datasets contained 87 (40 AA and 47 CA), 27 (14 AA and 13 CA), 802 (350 AA and 452 CA) and 26 samples (13 AA and 13 CA) respectively. Gene expression level was represented by expression level of corresponding probesets. For genes with multiple probesets, we selected the probe with maximum average intensity across all samples to represent the gene. Except for GSE91037, the other three provide Gleason scores of patients. Besides, GSE41969 provides age while GSE6956 and GSE17356 provide tumor stage information. A summary of these datasets can be found in Table S1.

Identification of differential molecular alterations between AA and CA samples

To identify gene mutations with significant different distribution between AA and CA, we used Fisher's exact test to compare their gene mutation frequencies. A P value under 0.05 was considered as significant. Since the majority of genes have relative low mutation frequency and the number of samples are small, the statistical power for detecting variant mutation is low. Thus, we used unadjusted p-value ($p < 0.05$) directly without considering confounding variables to ensure high sensitivity. To identify differential CNV, DNA methylation and mRNA expression between AA and CAs, we used R package "limma" [31]. Specifically, relative copy number, \log_2 transformed expression level and promoter methylation of genes were taken as inputs. To avoid potential impacts of factors other than race, we also adjusted clinical variables including age and Gleason score with "limma". Adjusted p values under 0.05 were considered as significant. Since there are more than one mRNA expression dataset, meta-analysis was used to integrate results across datasets and to calculate a meta-p value for each gene based on Fisher's method. In total, we identified 39, 466, 2876 and 1340 genes with differential mutation, CNV, DNA methylation and RNA expression between AA and CAs.

Network propagation

To integrate pre-identified differential altered genes between AA and CAs, we mapped these identified genes into a protein-protein interaction network and applied a network-based approach to prioritize genes associated with racial disparity in prostate cancer. A network with 13,096 genes (nodes) and 137,686 interactions (edges) was built on human interactome data from a previous study [32]. R package "igraph" was used to construct the network [33]. We pruned duplicated edges between nodes and only kept the largest connected components within the network, resulting in 12,972 nodes and 134,800 edges remaining. Pre-identified differentially altered genes between AA and CAs were then mapped to the network as seeds with label "1", while the rest were labelled "0". To prioritize genes associated with racial disparity, we simplified the network propagation method developed by Dimitrakopoulos et al. [26], which was based on an insulated heat diffusion process [25]. Specifically, we normalized the adjacency matrix of the

network by multiplying the inverse diagonal matrix of node degrees. A diffusion matrix was then defined as:

$$F = \beta(I - (1 - \beta)W)^{-1}$$

where W denotes the normalized adjacency matrix and β denotes the restart probability. F represents the diffusion matrix, in which F_{ij} describes the influence of node i applying to node j . To find optimized β , we selected several nodes with different centrality to study the effect of β value on the network. In practice, we selected 5 nodes according to their centrality (minimum, 25% quantile, median, 75% quantile, and maximum) in the network after excluding nodes with zero centrality. As the centrality metric, we used betweenness centrality since it provides better insight into the connectivity of nodes, which plays a major role in propagation scores. We calculated the influence of the 5 nodes applying to their neighbors, and selected β values that maximize the sum of influence of the 5 nodes. Based on our calculations, the influence was maximized when β equals 0.45 (Fig. S1). The final propagation score was calculated as:

$$E = FS$$

where S denotes the vector that contains the initial labels assigned to the nodes.

To determine the statistical significance of each gene's contribution to racial disparity rather than simply using its propagation score, we constructed 100 random networks to determine a false discovery rate (FDR) for each gene. Specifically, each random network was constructed by 100 random edge swaps of the original network, so that node degrees can be preserved. For each gene, its FDR was calculated by taking the proportion of propagation scores from the 100 random networks that were higher than its propagation score in the original network. Using a cut-off as $FDR < 0.01$, we identified 130 genes as core DAGs. To visualize the interactions among the 130 genes, we created a subnetwork of them by extracting the 130 nodes and related edges between them in the original network using Cytoscape [34].

Disease associated gene sets

Disease associated gene sets were downloaded from DisGeNET [35]. This database not only curated gene-disease associations from UNIPROT, CGI, ClinGen, Genomics England, CTD (human subset), PsyGenET, and Orphanet, but also text-mined several gene-disease associations from MEDLINE abstracts. In total, 11,181 curated and 17,993 text-mined disease associated gene sets were retrieved. Fisher's exact tests were used to evaluate the enrichment differences of these gene sets between DAGs and non-DAGs. To correct p values from multiple tests, Benjamini-Hochberg method was used.

Gene set enrichment analysis

To investigate associated pathways of DAGs, we performed enrichment analysis of DAGs using KEGG pathways and GO terms. Specifically, 186 KEGG pathways and 10,185 GO terms were downloaded from C2 and C5 collection from MsigDB database [36]. Fisher's exact test was used to calculate statistical significance of enrichment of KEGG pathways and GO terms in DAGs. Benjamini-Hochberg method was used to correct p values from multiple tests.

Gene essentiality analysis

To examine the importance of the 130 DAGs in prostate cell line growth, we investigated gene essentialities of them. Gene essentiality data of prostate cancer cell lines was downloaded from DepMap database [37]. CERES score and dependency probability of 18,119 genes for VCAP cell line was retrieved [38]. The CERES score describes cell line growth rates after knock-out of given genes. Therefore, a low CERES

score or a high dependency probability indicates a high essentiality of a gene. Gene essentialities between DAGs and non-DAGs were compared by Wilcoxon rank-sum test.

SNPs associated with prostate cancer

We downloaded a total of 616 risk SNPs for prostate cancer identified by genome-wide association study (GWAS) from NHGRI-EBI GWAS Catalog [39] on 02/14/2020. Associated studies are summarized in Table S2.

The minor allele frequencies of these SNPs in different populations from the 1000 Genomes Project [40] were obtained through The Ensembl Variant Effect Predictor (<http://www.ensembl.org/vep>) [41]. There are five super populations in the 1000 Genomes Project dataset: African, Ad Mixed American, East Asian, European, South Asian. As African and European populations are genetically closest to AA and CA, we considered minor allele frequencies in the two populations as those in AA and CA. Fisher's exact test was used to compare minor allele frequencies of SNPs between the two races. In total, we compared minor allele frequencies of 6 SNPs (corresponding to TP53, SMAD2, ADAMTSL4, SAMD9, AR, MAP3K1), whose associated genes were within the 130 DAGs.

Survival analysis

To identify a gene list associated with prognosis, we downloaded a dataset (Accession ID: GSE16560) [42] from the GEO database, which contains expression profiles of 6100 genes for 281 samples, and clinical variables including disease-specific survival. We used univariate Cox regression models to assess the prognostic value of each gene based on its continuous expression levels. Because the statistical power is relatively low, with only 13 genes having p values below 0.05 after Benjamini-Hochberg adjustments, we considered unadjusted p values without multi-test corrections below 0.05 as significant to identify more prognostic genes. Specifically, 878 prognostic genes were identified. To examine whether the 130 DAGs are enriched for prognostic genes compared to non-DAGs, Fisher's exact tests were performed. We selected two DAGs (YWHAZ and JUN) as examples to visualize their prognostic value by Kaplan-Meier plots. Specifically, the whole cohort (Accession ID: GSE16560) [42] was dichotomized according to the mean expression value of each gene, and log-rank tests were used to compare survival differences between high and low expression groups.

We used R package "survival" to perform survival analysis. Cox regression models were built using the function "coxph". Comparisons between two survival curves were made by using the "survdiff" function.

Kinase-substrate interactions

To investigate which kinase are enriched for DAGs in its substrates, we downloaded Kinase-substrate dataset PhosphoSitePlus [43]. 12,228 kinase-substrate interactions from humans were retrieved. We examined which kinase substrates were enriched with the identified DAGs using Fisher's exact tests. Kinases with fewer than 20 substrates were excluded. Benjamini-Hochberg adjusted p values below 0.05 were considered significant.

Inference of transcription factor regulatory activity

Transcription factor binding profiles were downloaded from the ChEA dataset [44]. A previously developed rank-based algorithm, BASE, was used to calculate regulatory activities for transcription factors based on the expression profile of their target genes [45]. The Wilcoxon rank-sum test was used to compare transcription factor regulatory activities between AA and CA.

Results

Overview of the study

In this study, we collected a comprehensive list of genomic data that were generated to identify differential molecular features between AA and CAs in prostate cancer, including copy number variation (CNV), somatic mutation, DNA methylation, and gene expression. To better understand racial disparity, we employed a network-based method to integrate these data to identify a set of core DAGs using the pipeline shown in Fig. 1. First, for each type of data we performed comparative analyses between AA and CA prostate cancer patients to identify significantly differential molecular features on gene level. For gene expression analysis, multiple datasets were available and meta-analysis was performed to obtain combined results. Next, we integrated these results by projecting them into a curated protein-protein interaction network [32] and then adopted network propagation to prioritize genes. Statistical significance was determined by referring a null distribution obtained from topology-preserved network permutations. We identified a total of 130 DAGs that were significantly associated with racial disparity at the significance level of false discovery rate (FDR) < 0.01. Finally, we performed a series of analyses to investigate the 130 DAGs in the context of their associated pathways, their contribution to genetic risk of prostate cancer and their prognostic value. In particular, we found kinases and transcription factors were highly enriched in these DAGs, and thus we investigated their potential functions in prostate cancer.

Identification of molecular alterations of different types between AA and CA

We compared molecular difference between AA and CA prostate cancer in four types of molecular features: somatic mutation, CNV, DNA methylation and RNA expression. We first examined differential somatic mutation events between AA and CAs in the TCGA prostate adenocarcinoma dataset (TCGA-PRAD). We compared number of point mutations of each gene between the two races and identified 39 genes with significantly differential mutation frequency at the significance level of p

< 0.05 using fisher's exact test. Out of them, 38 were enriched in AA and 1 was enriched in CA (Fig. 2A). For example, SPOP mutation was found to be enriched in AA ($p = 0.008$), while TP53 mutation was found to be enriched in CA ($p = 0.04$). For CNV, we identified 466 differential CNV events between AA and CA at FDR < 0.05 from TCGA-PRAD dataset, of which 139 gene deletions and 327 gene amplifications were enriched in AA after adjusting for age and Gleason score using "limma" [31] (Fig. 2B).

To identify differentially methylated genes between AA and CA, for each gene we calculated the average methylation level (β values) of measured CpG sites that were located in its promoter. Using the TCGA-PRAD data, we identified 2876 differentially methylated genes (1786 hyper- and 1110 hypomethylated in AA) at FDR < 0.05 after adjusting for clinical variables (Fig. 2C).

Using the TCGA-PRAD dataset, we identified 845 differentially expressed genes, with 542 up-regulated and 303 down-regulated in AA (Fig. 2D). In addition to the TCGA dataset, we examined four datasets with expression profiles of AA and CAs [27–30]. We identified 487, 113, 21 differential expressed genes in GSE17356, GSE41969 and GSE6956 at FDR < 0.05 (Figs. 2D and S2). We didn't find any differential expressed genes in GSE91037, potentially due to its small sample size. Only a few differentially expressed genes were shared across datasets. Specifically, there were only 17 differentially expressed genes shared between TCGA-PRAD and GSE17356, the two datasets with the highest number of identified differentially expressed genes (Fig. S2). These results indicated high inconsistency across these datasets. Because of this, we performed a meta-analysis to combine the results from these datasets. As shown in Fig. 2E, we identified more differential expressed genes with meta-analysis. At the same threshold, 1340 differential expressed genes were identified, with 748 up-regulated and 592 down-regulated in AA across the five datasets.

Taking together, we identified 39, 466, 2878 and 1340 genes associated with differential mutation, CNV, methylation and expression between AA and CA. Specifically, 159 genes showed both differential methylation and expression, 46 genes showed both differential methylation and CNV, and 28 genes showed both differential CNV and expression between AA and CA (Fig. 2F and G, Table. S3). There were only seven genes showing differential methylation and one gene with

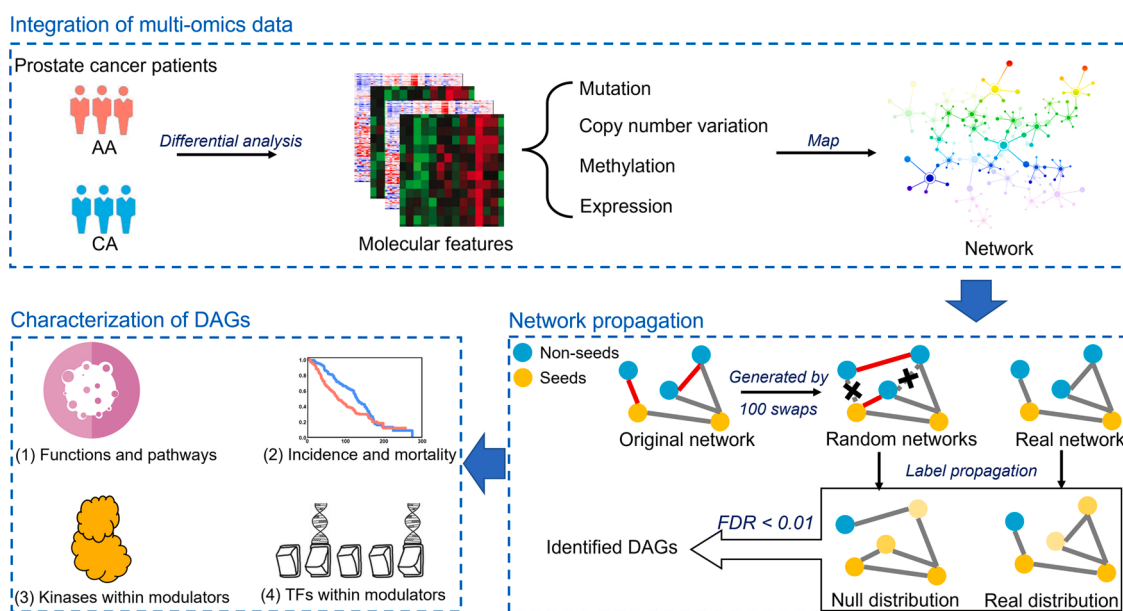


Fig. 1. The diagram of this study. We performed comparative analysis to identify four types of differential molecular alterations (somatic mutation, CNV, DNA methylation, mRNA expression) between AA and CA. The detected alterations were then mapped to an interaction network on gene level. We next applied network propagation to prioritize genes associated with racial disparity in prostate cancer. Based on FDR < 0.01, we identified 130 DAGs underlying racial disparity and characterized them in terms of function and pathways, prognostic value as well as kinases and transcription factors within them.

differential expression within differentially mutated genes because of their small number.

Network-based data integration leads to the identification of 130 core genes associated with racial disparity

To identify core DAGs, we mapped previously identified genes as seeds into a human protein-protein network and applied a label propagation method to prioritize all genes within the network. Using $FDR < 0.01$, we identified 130 DAGs (Table S4). These 130 DAGs formed a subnetwork (Fig. 3A). Within them, 52 were seeds and 78 were non-seeds. Seeds and non-seeds had similar distributions of propagation scores, non-seeds did not necessarily have lower scores than seeds (Fig. 3B). Specifically, the AR gene, which drives tumorigenesis and progression of prostate cancer [46], was within the 130 DAGs, even though it's not a seed. Indeed, AR was previously reported to differ in related signaling pathways [23] and to correlate differently with carcinogen [24] between AA and CA, despite no differences in expression levels. These results suggested that network-based methods identified latent genes associated with racial disparity, which were not readily apparent from comparative analyses alone.

The functions and pathways of DAGs

We then examined the function of these DAGs in known human diseases. To this end, we retrieved disease associated gene sets from DisGeNet [35]. As shown, gene sets associated with prostate cancer such as "Prostatic neoplasms" and "Malignant neoplasms of prostate cancer" were highly enriched in DAGs, four times compared to non-DAGs (Fig. 3C). Furthermore, gene sets associated with cancer severity such as "Neoplasm metastasis", "Neoplasm invasiveness" and "Undifferentiated carcinoma" were also enriched in DAGs, 8–15 times more than

non-DAGs (Fig. 3C). Specifically, gene sets related to castration-resistant prostate cancer (CRPC), the most lethal subtype of prostate cancer, such as "Hormone refractory prostate cancer", "Androgen independent prostate cancer" and "Metastatic castration-resistant prostate cancer", were all significantly enriched in DAGs (Fig. 3C). As previous studies reported AA and CA patients responded differently to CRPC treatment [47], DAGs might explain such racial disparity in CRPC treatments.

Next, we investigated which pathways the DAGs were involved in. As expected, we observed the DAGs were enriched for some cancer pathways including prostate cancer from KEGG pathways (Fig. 3D). Besides, several kinases-involved signaling pathways such as MAPK, TGF- β , ERBB and WNT pathways were found to be enriched in DAGs, 10–25 times compared to non-DAGs. In addition, cell cycle related pathways were also enriched in DAGs. We found similar results using enrichment analysis of Gene Ontology (GO) terms. Signaling-related GO terms (positive regulation of signaling and kinase binding) and cell cycle related GO terms (apoptotic process and regulation of cell death) were enriched in DAGs 3 to 6 times more than non-DAGs (Fig. 3E).

Association of DAGs with incidence and mortality in prostate cancer

To elucidate potential mechanisms by which these DAGs contribute to racial disparity in prostate cancer, we characterized them in several clinical implications.

First, we examined gene essentiality of these DAGs in prostate cancer cell lines. The gene essentialities were downloaded from DepMap database [37], represented by CERES score and dependency probability. Low CERES scores or high dependency probabilities imply high essentiality of a gene for prostate cancer cell line growth. We observed that the DAGs tend to have lower CERES scores and higher dependency probabilities compared to non-DAGs (Figs. 4A and S3). Our results indicated that DAGs were more likely to be essential in prostate cancer

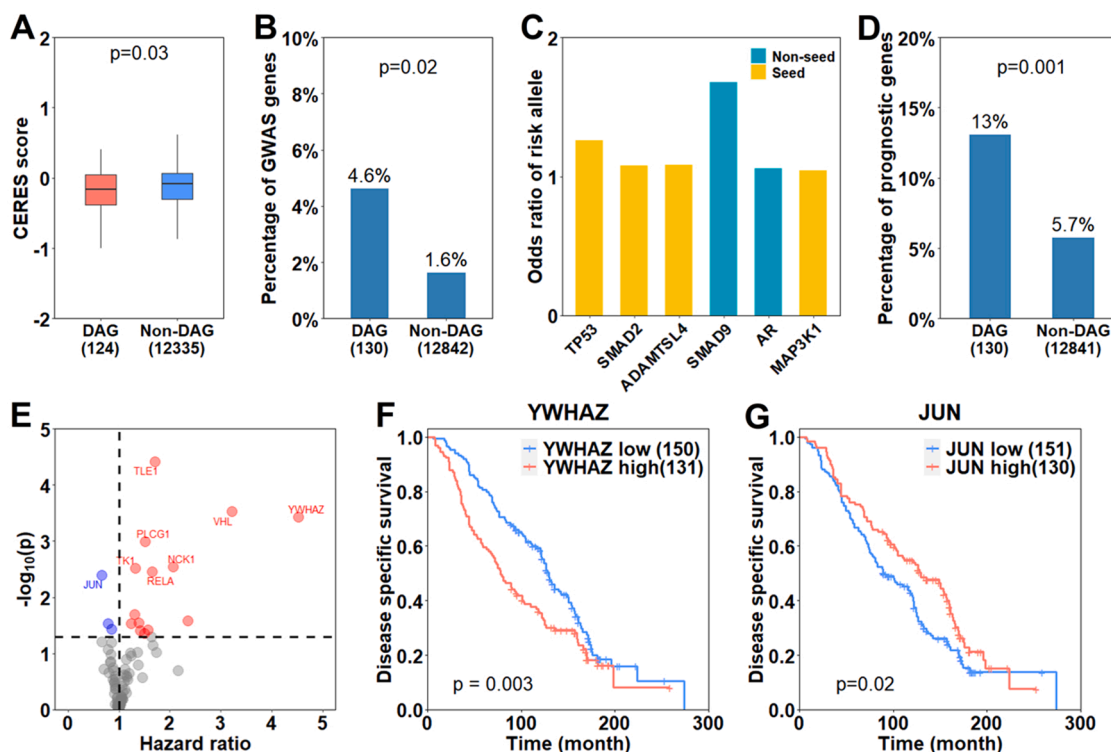


Fig. 4. Association of DAGs with incidence risk and mortality of prostate cancer. (A): DAGs have higher essentialities compared to non-DAGs, represented by lower CERES score (Wilcoxon rank-sum test). (B): DAGs were enriched for GWAS genes compared to non-DAGs (Fisher's exact test). GWAS gene: genes corresponding to prostate cancer risk SNPs in GWAS studies. (C): Genetic risk of the six GWAS genes within a DAG as measured by corresponding risk allele's odds ratio. (D): The DAGs were enriched for prognostic genes compared to non-DAGs (Fisher's exact test). (E): Hazard ratio and corresponding p values for the prognostic DAGs. (F and G): Survival difference between patients with high and low expression of YWHAZ (F) and JUN (G) with mean as cut-off (log-rank test).

compared to non-DAGs.

Second, we investigated association of the DAGs with prostate cancer incidence. To this end, we retrieved 559 genes (GWAS genes) associated with risk SNPs from genome-wide association studies (GWAS) [39]. We found that the DAGs were enriched with GWAS genes nearly three times to non-DAGs ($p = 0.02$, Fig. 4B), indicating potential roles of DAGs in prostate cancer incidence. Totally, there were six GWAS genes in DAGs, containing both seed and non-seed genes. The incidence risk of corresponding risk alleles of these six genes were represented by odds ratio as shown in Fig. 4C. Given AA and CA have different incidence rates of prostate cancers [2], we next examined if those DAGs, which were also GWAS genes, exhibited different allele frequencies between the two races. As shown, the six genes being both DAGs and GWAS genes significantly differed in their minor allele frequencies between AA and CA (Fig. S4A–F).

Third, we examined prognostic values of these DAGs. Specifically, we retrieved 878 genes prognostic for disease specific survival in prostate cancer from the GSE16560 dataset [42] at $p < 0.05$. As shown in Fig. 4D, the DAGs were enriched for more than twice as many prognostic genes as non-DAGs ($p = 0.001$). Out of the 17 prognostic DAGs, 3 were protective and 14 were hazardous (Fig. 4E). Among them, there are both seeds and non-seeds. For example, a seed gene, YWHAZ, which is associated with prostate cancer progression and cell proliferation [48–50], was in the 14 hazardous DAGs (Fig. 4F). Furthermore, non-seed genes such as JUN, an onco-gene regulates apoptosis [51], were also within the prognostic DAGs (Fig. 4G).

Taken together, these results suggested that the DAGs can affect incidence and mortality, which in turn might contribute to the racial disparity in prostate cancer.

Differential activity of disparity associated kinases between AA and CAs

Kinases play important roles in signaling transduction and regulate a variety of cellular processes. Several dysregulated kinases and associated pathways have been observed in prostate cancer, including over-expression of SRC and elevated PI3K-AKT-mTOR pathway activity [52, 53]. Therefore, we examined potential roles of kinases within the DAGs in racial disparity. There were 26 kinases as DAGs, with 8 seeds and 18 non-seeds (Fig. 5A). Using TCGA reverse phase protein array (RPPA) data, we compared protein levels of these kinases between AAs and CAs. As shown, AA has significantly higher levels of phosphorylated MAPK8 than CAs (Fig. 5B). In addition to phosphorylated kinases, TCGA RPPA includes some unphosphorylated kinases, allowing comparison of kinase protein levels both in the phosphorylated (activated) and unphosphorylated (inactivated) states between AA and CA. We found some kinases with comparable levels of inactivated states between AA and CA, but significant differences of activated states, such as EGFR and SRC (Fig. 5C and D). Furthermore, there were also some kinases that differed in AA and CA in opposite directions between inactivated and activated states, such as MAPK1 and GSK3B (Fig. 5E and F). To investigate whether the differences of kinase levels between AA and CA would remain significant after adjusting for clinical variables, we constructed a generalized linear model using kinase levels, age and Gleason score. Our results indicated the differences in protein levels of aforementioned kinases between AA and CA remain significant, even after taking into account other clinical variables (Table S5).

Since a kinase can phosphorylate many substrates including other kinases as well as itself [54–56], we next examined which kinases were enriched for DAGs in their substrate using a kinase-substrate dataset

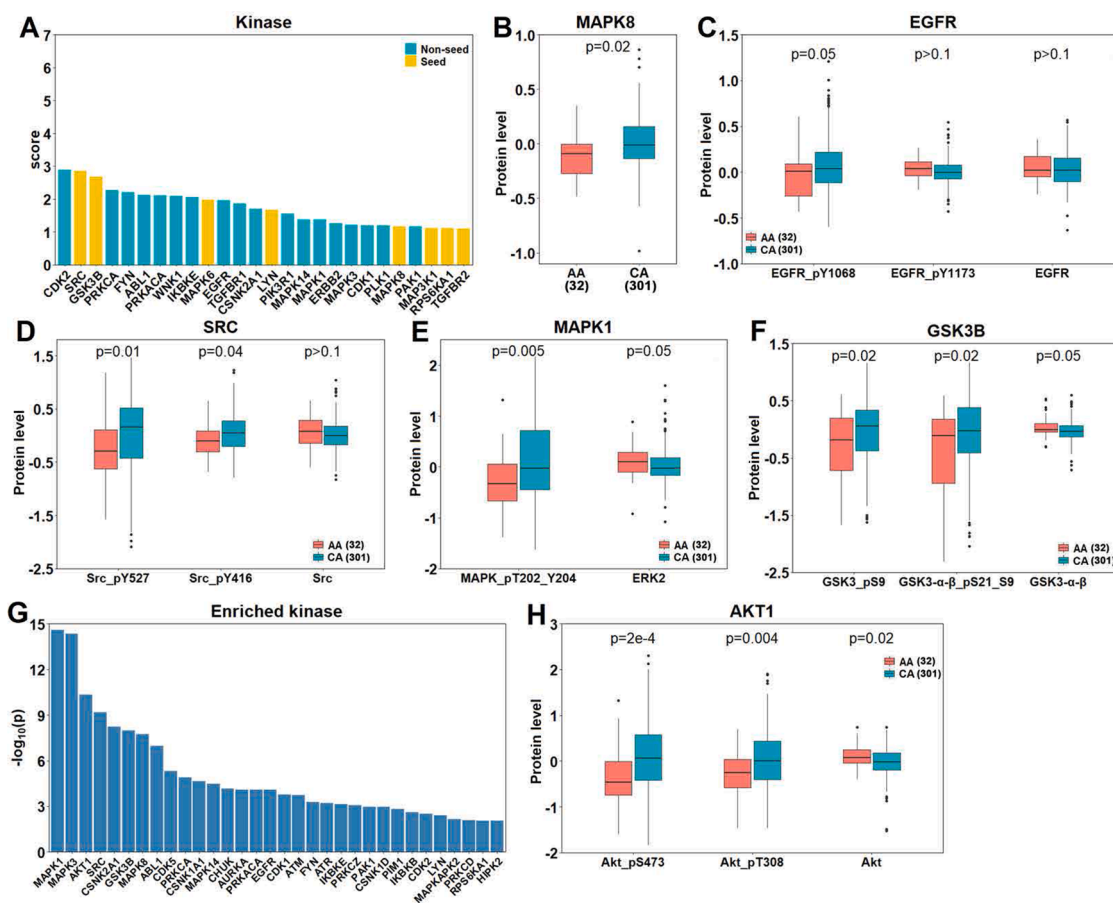


Fig. 5. Kinases within DAGs exhibited differential activity states between AA and CA. (A): Propagation score of 26 kinases within DAGs. (B–F): Protein levels of MAPK8 (B), EGFR (C), SRC (D), MAPK1 (E) and GSK3B (F) between AA and CA (Wilcoxon rank-sum test). The position of the phosphorylated amino acid in a protein is followed by "p". (G): 32 enriched kinases with the DAGs (Fisher's exact test). (H): Protein levels of AKT1 between AA and CA (Wilcoxon rank-sum test).

from PhosphoSitePlus [43]. At a threshold of adjusted p values below 0.01, we identified 32 kinases significantly enriched for the DAGs in their substrates, including 18 DAGs (Fig. 5G). Although not a DAG, AKT1 exhibited a very significant p value, ranking third, right after the two DAGs: MAPK1 and MAPK3. Indeed, AKT1 showed significantly differential activities between AA and CA, with a more elevated activity in the latter (Fig. 5H). This result remained significant after adjusting for the same clinical variables (Table S5). Taken together, our results suggested that the AA and CA differed significantly in activation states of kinases, especially those within DAGs, which may lead to differential activity of kinase-directed signaling transduction pathways between the two races, thereby contributing to the racial disparity in prostate cancer.

Differential activities of disparity associated transcription factors between AA and CA

Transcription factors regulate the expression of many genes, which in turn affects cellular growth, division and apoptosis. Dysregulation of key transcription factors (AR, ERG) has been reported to contribute to tumorigenesis and progression of prostate cancer [57,58]. Therefore, we also examined potential roles of transcription factors within the DAGs in racial disparity. A total of 22 transcription factors were found within the DAGs, with 10 seeds and 12 non-seeds (Fig. 6A). Using the TCGA RPPA dataset, we compared protein levels of 7 transcription factors within DAGs between AA and CA. AA was found with higher protein levels of TP53 ($p = 0.03$) and lower protein levels of ESR1 ($p = 0.04$), RB1 ($p = 0.06$) and MYC ($p = 0.09$) compared to CA (Fig. 6B). Interestingly, phosphorylated ESR1 along with JUN, AR and CTNNB1 had comparable protein levels between AA and CA (Fig. 6B). After adjusting for the same

clinical variables with a generalized linear model, TP53 still remained significantly different in protein levels between AA and CA (Table S6).

In addition to protein expression levels, mutations and alternative splicing can also affect the regulatory activity of transcription factors [59–61], which in turn impacts a number of downstream pathways. Hence, we subsequently compared regulatory activities of transcription factors within DAGs between AA and CA. Using a previously developed deconvolutional algorithm [45], we calculated transcription factor regulatory activities based on TCGA PRAD gene expression profile and transcription factor binding profiles from the ChEA dataset [44]. In contrast to protein levels, TP53 exhibited significantly higher regulatory activities in AA instead of CA (Figs. 6C and S5), indicating different mechanisms in dysregulation of TP53 downstream pathways between the two races. While the TCGA RPPA data didn't include the protein level of NR3C1, its regulatory activity did differ significantly between the two races (Fig. 6C). Furthermore, while AR, CTNNB1, and JUN did not differ in protein levels between the two races, we still observed significant or weakly significant differences in their regulatory activities (Fig. 6C). Specifically, the differences in AR regulatory activities between AA and CA remained significant even after adjusting for the same clinical variables as previously described (Table S7).

As a key driver in tumorigenesis and development of prostate cancer [46], AR has been investigated in multiple prostate cancer cell lines and tumor tissues in terms of its binding profiles [62–65]. The ChEA dataset includes four AR binding profiles measured from different prostate cancer cell lines or tissues (VCAP, PC3, LNCAP and tumor tissue). We calculated and compared regulatory activities in AA and CA with the four different profiles. As shown, only AR regulatory activities calculated using the corresponding binding profile from VCAP were

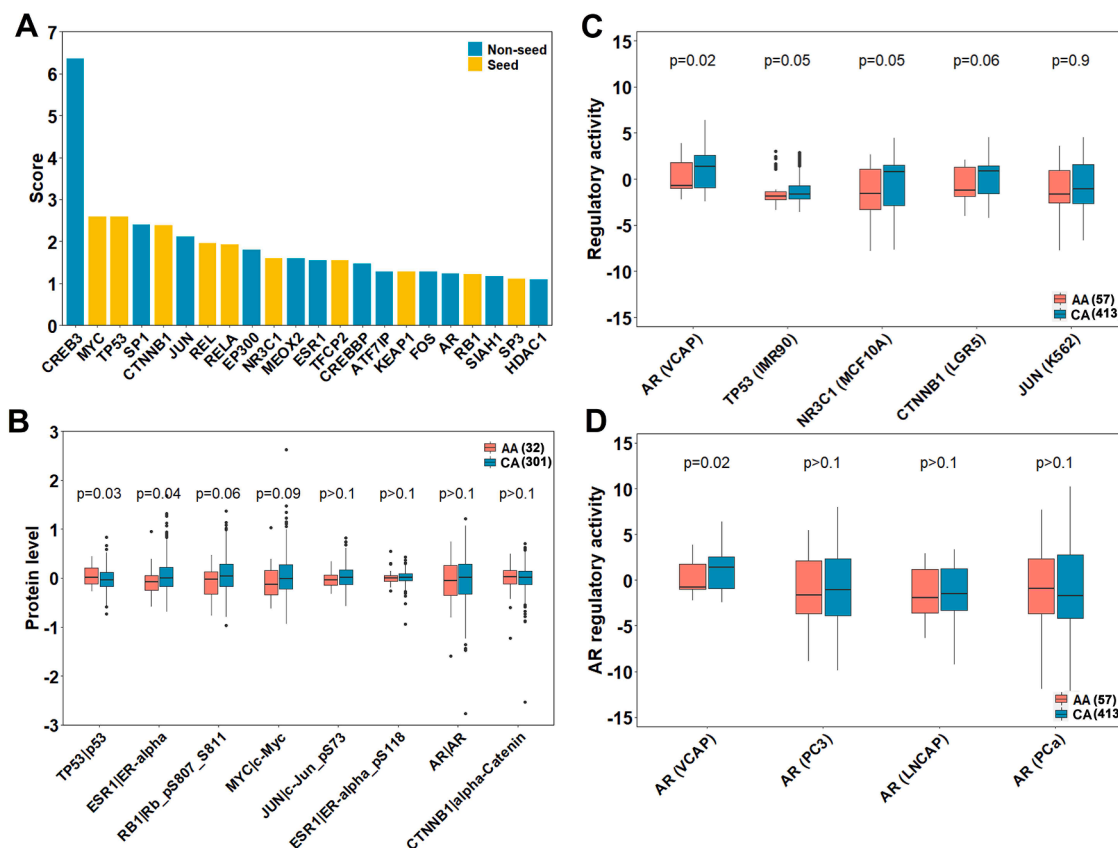


Fig. 6. Transcription factors within DAGs exhibited differential regulatory activities between AA and CA. (A): Propagation score of 22 transcription factors within DAGs. (B): Protein levels of the transcription factors within DAGs between AA and CA. Protein names of corresponding genes were after “|”. For phosphorylated proteins, positions of the phosphorylated amino acids were after “_p”. (C): Regulatory activities of the transcription factors between AA and CA. Names of cell lines from which the binding profile used to calculate the regulatory activity were indicated in parentheses. (D): AR regulatory activities calculated by different binding profiles from multiple prostate cancer cell lines between AA and CA. P values were calculated by Wilcoxon rank-sum tests.

significantly different between AA and CA (Fig. 6D). Out of the three cell lines, VCAP was the only one that can express AR-V7 protein, a splice variant of AR, which was rare in primary prostate cancer but common in CRPC and can render resistance to anti-androgen drugs [66,67]. Based on this result, there may be differences in alternative splicing events between AA and CA, resulting in differential delays to advanced prostate cancer stages.

Discussion

To investigate potential DAGs underlying racial disparity in prostate cancer, we integrated differential molecular alterations between AA and CA from different data types and applied a network-based approach to identify 130 DAGs. Subsequent characterizations revealed these DAGs were highly associated with patient prognosis as well as tumorigenesis, development and progression in prostate cancer. Furthermore, kinases and transcription factors that act as DAGs differed in activity state or regulatory activity between AA and CA, although some of them were comparable at the protein level. These results suggested that, with the network-based integration method, we detected some new racial disparity associated genes in prostate cancer that weren't detectable by common comparative analysis due to their similar expression levels.

We observed AA had significantly lower protein levels of activated AKT1, MAPK1 and GSK3B but higher levels of their inactivated state proteins compared to CA. Indeed, an elevated PI3K/AKT/mTOR signaling pathway, containing GSK3B, is observed in many prostate cancer cases [68,69]. An activated MAPK pathway is often found in CRPCs and is associated with resistance to anti-androgen drugs [69,70]. Based on these results, it might be more important to also consider kinase activity instead of only comparing mRNA or protein expression. Similarly, AR has significantly different regulatory activities regardless of similar protein levels between the two races. More importantly, we can observe the difference only by calculating regulatory activity based on the AR-binding profiles of the VCAP cell line that expresses AR-V7 isoform [66]. AR-V7 is highly expressed in CRPC and associated with resistance to anti-androgen therapy [67]. This result highlights the potential different alternative splicing of AR between the two races, which in turn leads to racial disparity.

Interestingly, we observed significant different levels of unphosphorylated ESR1 but similar levels of its phosphorylated form between AA and CA. As ESR1 exhibits different allele frequencies between AA and CA [71], there may already be effects of differential variants within races, in which case dysregulation of associated downstream pathways at phosphorylation level might not be necessary.

It should be noticed that both AA and CA populations might have diverse genetic backgrounds. Simply categorizing patients into AA and CA may overlook admixture from multiple genetic ancestries. Different proportion of other ancestries within AA population will also confound current results. Besides, the sample sizes between AA and white are imbalanced, which may reduce the statistical power. Therefore, we used meta-analysis to combine multiple datasets and network-based methods to integrate differentially altered genes between AA and CA from different data types. Although these methods can compensate to some extent for the relatively low statistical power, a larger and more balanced dataset would be more desirable. Similarly, due to low mutation frequencies of the majority of genes, we used unadjusted p values without considering other confounding variables to identify differential mutated genes. Such an approach improves sensitivity, but it also leads to the identification of some genes that are not differentially mutated. In addition, we showed the DAGs were highly associated with prognosis. However, there's no race information within that dataset. So differential association of DAGs with survival between the two races will need to be further examined.

Most importantly, our results were generated based on solely computational method. Even though the DAGs identified can be validated with high essentiality and association with prostate cancer

incidence using cell line and GWAS data, experimental validation will be needed to evaluate the specific role they play in prostate cancer. For example, knock-out animal studies can be performed to examine role of these DAGs in vivo. The clinical value of DAGs can be explored through the use of drugs that target these DAGs in cell line or animal models. Besides, by using a network-based approach, we lost the direction of differential molecular alterations. Considering the identified DAGs alone, we are unable to identify their roles in prostate cancer or the direction they contributed to racial disparity, whether they increased risk for AAs or decreased risk for CA. Furthermore, the label propagation method only considered the presence of the differential molecular alterations and overlooked how much they differ between AA and CA. Although such method allows integrating results from different data types with different distributions, it will somewhat amplify the effect of relatively less significant molecular alterations and reduce the effect of extremely significant molecular alterations. Last, the ChEA dataset [44] contains a limited number of transcription factor binding profiles, so many of the profiles we used to calculate regulatory activity were not from prostate cancer cell lines. These results would be verified and consolidated using profiles directly from prostate cancer cell lines.

Conclusion

In summary, we applied a network-based algorithm to integrate differential molecular alterations between AA and CA and identified 130 DAGs. By characterizing these DAGs, we observed significant differential downstream pathway activities of MAPK, AKT and AR. Targeting these pathways may advance precision medicine, and ultimately benefits AA as well as other ethnicities.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Data availability

The data that support the findings of this study are openly available from Gene Expression Omnibus repository (GSE6956, GSE17356, GSE41969, GSE91037 and GSE16560).

CRediT authorship contribution statement

Baoyi Zhang: Formal analysis, Investigation, Validation, Visualization, Writing – original draft. **Kevin Yao:** Writing – review & editing. **Chao Cheng:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare no potential conflicts of interest.

Acknowledgements

This work is supported by the Cancer Prevention Research Institute of Texas (CPRIT) (RR180061 to CC) and the National Cancer Institute of the National Institutes of Health (1R21CA227996 to CC). CC is a CPRIT Scholar in Cancer Research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2021.101327.

Reference

- [1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2019, *CA Cancer J. Clin.* 69 (2019) 7–34.
- [2] C.E. DeSantis, K.D. Miller, A. Goding Sauer, A. Jemal, R.L. Siegel, Cancer statistics for African Americans, 2019, *CA Cancer J. Clin.* 69 (2019) 211–233.
- [3] V. Ravery, S. Dominique, V. Hupertan, S. Ben Rhouma, M. Toubian, L. Boccon-Gibod, et al., Prostate cancer characteristics in a multiracial community, *Eur. Urol.* 53 (2008) 533–538.
- [4] I.J. Powell, C.H. Bock, J.J. Ruterbusch, W. Sakr, Evidence supports a faster growth rate and/or earlier transformation to clinically significant prostate cancer in black than in white American men, and influences racial progression and mortality disparity, *J. Urol.* 183 (2010) 1792–1796.
- [5] S.A. Weprin, D.C. Parker, J.D. Jones, J.R. Kaplan, L.L. Giusto, J.H. Mydlo, et al., Association of low socioeconomic status with adverse prostate cancer pathology among African American men who underwent radical prostatectomy, *Clin. Genitourin. Cancer* 17 (2019) e1054–e1059.
- [6] B.A. Mahal, D.R. Ziehr, A.A. Aizer, A.S. Hyatt, J.D. Sammon, M. Schmid, et al., Getting back to equal: the influence of insurance status on racial disparities in the treatment of African American men with high-risk prostate cancer, *Urol. Oncol.* 32 (2014) 1285–1291.
- [7] D.R. Ziehr, B.A. Mahal, A.A. Aizer, A.S. Hyatt, C.J. Beard, A.V. D'Amico, et al., Income inequality and treatment of African American men with high-risk prostate cancer, *Urol. Oncol.* 33 (2015) 18.e7–18.e13.
- [8] F. Khani, J.M. Mosquera, K. Park, M. Blattner, C. O'Reilly, T.Y. MacDonald, et al., Evidence for molecular differences in prostate cancer between African American and Caucasian men, *Clin. Cancer Res.* 20 (2014) 4925–4934.
- [9] W. Liu, S.L. Zheng, R. Na, L. Wei, J. Sun, J. Gallagher, et al., Distinct genomic alterations in prostate tumors derived from African American men, *Mol. Cancer Res.* 18 (2020) 1815–1824.
- [10] Y. Koga, H. Song, Z.R. Chalmers, J. Newberg, E. Kim, J. Carrot-Zhang, et al., Genomic profiling of prostate cancers from men with African and European ancestry, *Clin. Cancer Res.* 26 (2020) 4651–4660.
- [11] W. Liu, S.L. Zheng, R. Na, L. Wei, J. Sun, J. Gallagher, et al., Distinct genomic alterations in prostate tumors derived from African American men, *Mol. Cancer Res.* 18 (2020) 1815–1824.
- [12] F.A. Faisal, D. Sondi, J.J. Tosoian, V. Choerung, M. Alshalafa, A.E. Ross, et al., Racial variations in prostate cancer molecular subtypes and androgen receptor signaling reflect anatomic tumor location, *Eur. Urol.* 70 (2016) 14–17.
- [13] K.E. Gaston, D. Kim, S. Singh, O.H. Ford, J.L. Mohler, Racial differences in androgen receptor protein expression in men with clinically localized prostate cancer, *J. Urol.* 170 (2003) 990–993.
- [14] B. Shuch, M. Mikhail, J. Satagopan, P. Lee, H. Yee, C. Chang, et al., Racial disparity of epidermal growth factor receptor expression in prostate cancer, *J. Clin. Oncol.* 22 (2004) 4725–4729.
- [15] J. Yuan, K.H. Kensler, Z. Hu, Y. Zhang, T. Zhang, J. Jiang, et al., Integrative comparison of the genomic and transcriptomic landscape between prostate cancer patients of predominantly African or European genetic ancestry, *PLoS Genet.* 16 (2020), e1008641.
- [16] F.W. Huang, J.M. Mosquera, A. Garofalo, C. Oh, M. Baco, A. Amin-Mansour, et al., Exome sequencing of African-American prostate cancer reveals loss-of-function ERF mutations, *Cancer Discov.* 7 (2017) 973–983.
- [17] R. Flavin, A. Pettersson, W.K. Hendrickson, M. Fiorentino, S. Finn, L. Kunz, et al., SPINK1 protein expression and prostate cancer progression, *Clin. Cancer Res.* 20 (2014) 4904–4911.
- [18] K.-C. Huang, A. Evans, B. Donnelly, T.A. Bismar, SPINK1 overexpression in localized prostate cancer: a rare event inversely associated with ERG expression and exclusive of homozygous PTEN deletion, *Pathol. Oncol. Res.* 23 (2017) 399–407.
- [19] F.A. Faisal, H.B. Kaur, J.J. Tosoian, A.A. Tomlins, E.M. Schaeffer, T.L. Lotan, SPINK1 expression is enriched in African American prostate cancer but is not associated with altered immune infiltration or oncologic outcomes post-prostatectomy, *Prostate Cancer Prostatic Dis.* 22 (2019) 552–559.
- [20] X. Zhang, X. Yin, P. Shen, G. Sun, Y. Yang, J. Liu, et al., The association between SPINK1 and clinical outcomes in patients with prostate cancer: a systematic review and meta-analysis, *Onco. Targets Ther.* 10 (2017) 3123–3130.
- [21] M.H. Johnson, A.E. Ross, M. Alshalafa, N. Erho, K. Yousefi, S. Glavaris, et al., SPINK1 defines a molecular subtype of prostate cancer in men with more rapid progression in an at risk, natural history radical prostatectomy cohort, *J. Urol.* 196 (2016) 1436–1444.
- [22] Z. Wang, Y. Song, M. Ye, X. Dai, X. Zhu, W. Wei, The diverse roles of SPOP in prostate cancer and kidney cancer, *Nat. Rev. Urol.* 17 (2020) 339–350.
- [23] B.-D. Wang, Q. Yang, K. Ceniccola, F. Bianco, R. Andrawis, T. Jarrett, et al., Androgen receptor-target genes in African American prostate cancer disparities, *Prostate Cancer* 2013 (2013), 763569.
- [24] C.M. Neslund-Dudas, R.B. McBride, A. Kandegedara, B.A. Rybicki, O.N. Kryvenko, D. Chitale, et al., Association between cadmium and androgen receptor protein expression differs in prostate tumors of African American and European American men, *J. Trace Elem. Med. Biol.* 48 (2018) 233–238.
- [25] M.D.M. Leiserson, F. Vandin, H.-T. Wu, J.R. Dobson, J.V. Eldridge, J.L. Thomas, et al., Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes, *Nat. Genet.* 47 (2015) 106–114.
- [26] C. Dimitrakopoulos, S.K. Hindupur, L. Häfliger, J. Behr, H. Montazeri, M.N. Hall, et al., Network-based integration of multi-omics data for prioritizing cancer genes, *Bioinformatics* 34 (2018) 2441–2448.
- [27] T.A. Wallace, R.L. Prueitt, M. Yi, T.M. Howe, J.W. Gillespie, H.G. Yfantis, et al., Tumor immunobiological differences in prostate cancer between African-American and European-American men, *Cancer Res.* 68 (2008) 927–936.
- [28] O.A. Timofeeva, X. Zhang, H.W. Resson, R.S. Varghese, B.V.S. Kallakury, K. Wang, et al., Enhanced expression of SOS1 is detected in prostate cancer epithelial cells from African-American men, *Int. J. Oncol.* 35 (2009) 751–760.
- [29] I.J. Powell, G. Dyson, S. Land, J. Ruterbusch, C.H. Bock, S. Lenk, et al., Genes associated with prostate cancer are differentially expressed in African American and European American men, *Cancer Epidemiol. Biomarkers Prev.* 22 (2013) 891–897.
- [30] B. Baumann, A.M. Costa, Z. Richards, R. Deaton, A. Sapatynska, A. Murphy, et al., Association of high miR-182 levels with low-risk prostate cancer, *Am. J. Pathol.* 189 (2019) 911–923.
- [31] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, et al., limma Powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.* 43 (2015) e47.
- [32] J. Menche, A. Sharma, M. Kitsak, S.D. Ghiassian, M. Vidal, J. Loscalzo, et al., Disease networks. Uncovering disease-disease relationships through the incomplete interactome, *Science* 347 (2015), 1257601.
- [33] Csardi, G., Nepusz, T. The igraph software package for complex network research: 9.
- [34] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (2003) 2498–2504.
- [35] J. Piñero, J.M. Ramírez-Anguita, J. Sañch-Pitarch, F. Ronzano, E. Centeno, F. Sanz, et al., The DisGeNET knowledge platform for disease genomics: 2019 update, *Nucleic Acids Res.* 48 (2020) D845–D855.
- [36] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 15545–15550.
- [37] A. Tsherniak, F. Vazquez, P.G. Montgomery, B.A. Weir, G. Kryukov, G.S. Cowley, et al., Defining a cancer dependency map, *Cell* 170 (2017) 564–576, e16.
- [38] DepMap 20Q2 public [internet]. figshare; 2020 [cited 2021 Oct 22]. Available from: https://figshare.com/articles/dataset/DepMap_20Q2_Public/12280541/4.
- [39] A. Buniello, J.A.L. MacArthur, M. Cerezo, L.W. Harris, J. Hayhurst, C. Malangone, et al., The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019, *Nucleic Acids Res.* 47 (2019) D1005–D1012.
- [40] A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H.M. Kang, 1000 Genomes Project Consortium, A global reference for human genetic variation, *Nature* 526 (2015) 68–74.
- [41] W. McLaren, L. Gil, S.E. Hunt, H.S. Riat, G.R.S. Ritchie, A. Thormann, et al., The Ensembl variant effect predictor, *Genome Biol.* 17 (2016) 122.
- [42] A. Sboner, F. Demicheli, S. Calza, Y. Pawitan, S.R. Setlur, Y. Hoshida, et al., Molecular sampling of prostate cancer: a dilemma for predicting disease progression, *BMC Med. Genomics* 3 (2010) 8.
- [43] P.V. Hornbeck, B. Zhang, B. Murray, J.M. Kornhauser, V. Latham, E. Skrzypek, PhosphoSitePlus, 2014: mutations, PTMs and recalibrations, *Nucleic Acids Res.* 43 (2015) D512–D520.
- [44] A. Lachmann, H. Xu, J. Krishnan, S.I. Berger, A.R. Mazloom, A. Ma'ayan, ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments, *Bioinformatics* 26 (2010) 2438–2444.
- [45] C. Cheng, X. Yan, F. Sun, L.M. Li, Inferring activity changes of transcription factors by binding association with sorted expression profiles, *BMC Bioinform.* 8 (2007) 452.
- [46] C. Huggins, C.V. Hodges, Studies in prostatic cancer. I. The effect of castration, of estrogen and androgen injection on serum phosphatases in metastatic carcinoma of the prostate, *CA Cancer J. Clin.* 22 (1972) 232–240.
- [47] R.L. Bittling, M. Goodman, D.J. George, Racial disparity in response to prostate cancer systemic therapies, *Curr. Oncol. Rep.* 22 (2020) 96.
- [48] C.-C. Yu, L.-C. Chen, W.-H. Lin, V.C. Lin, C.-Y. Huang, T.-L. Lu, et al., Genetic association analysis of cell cycle regulators reveals YWHAZ has prognostic significance in prostate cancer, *Cancer Genomics Proteomics* 17 (2020) 209–216.
- [49] S. Lage-Vickers, J. Bizzotto, M.P. Valacco, P. Sanchis, S. Nemirovsky, E. Labanca, et al., The expression of YWHAZ and NDRG1 predicts aggressive outcome in human prostate cancer, *Commun. Biol.* 4 (2021) 103.
- [50] T. Murata, K. Takayama, T. Urano, T. Fujimura, D. Ashikari, D. Obinata, et al., 14-3-3 ζ , a novel androgen-responsive gene, is upregulated in prostate cancer and promotes prostate cancer cell proliferation and survival, *Clin. Cancer Res.* 18 (2012) 5617–5627.
- [51] E. Bossy-Wetzel, L. Bakiri, M. Yaniv, Induction of apoptosis by the transcription factor c-Jun, *EMBO J.* 16 (1997) 1695–1709.
- [52] M. Asim, I.A. Siddiqui, B.B. Hafeez, A. Banihmad, H. Mukhtar, Src kinase potentiates androgen receptor transactivation function and invasion of androgen-independent prostate cancer C4-2 cells, *Oncogene* 27 (2008) 3596–3604.
- [53] B.S. Carver, C. Chapinski, J. Wongvipat, H. Hieronymus, Y. Chen, S. Chandralapaty, et al., Reciprocal feedback regulation of PI3K and androgen receptor signaling in PTEN-deficient prostate cancer, *Cancer Cell* 19 (2011) 575–586.
- [54] S. Maddika, S.R. Ande, E. Wiechec, L.L. Hansen, S. Wesselborg, M. Los, Akt-mediated phosphorylation of CDK2 regulates its dual role in cell cycle progression and apoptosis, *J. Cell Sci.* 121 (2008) 979–988.
- [55] X.Q. Wang, P. Sun, A.S. Paller, Inhibition of integrin-linked kinase/protein kinase B/Akt signaling: mechanism for ganglioside-induced apoptosis, *J. Biol. Chem.* 276 (2001) 44504–44511.

- [56] J. Beenstock, N. Mooshayef, D. Engelberg, How do protein kinases take a selfie (autophosphorylate)? *Trends Biochem. Sci.* 41 (2016) 938–953.
- [57] Z. Culig, F.R. Santer, Androgen receptor signaling in prostate cancer, *Cancer Metastasis Rev.* 33 (2014) 413–427.
- [58] P. Adamo, M.R. Ladomery, The oncogene ERG: a key factor in prostate cancer, *Oncogene* 35 (2016) 403–414.
- [59] N. Mäkinen, H.-R. Heinonen, S. Moore, I.P.M. Tomlinson, Z.M. van der Spuy, L. A. Aaltonen, MED12 exon 2 mutations are common in uterine leiomyomas from South African patients, *Oncotarget* 2 (2011) 966–969.
- [60] C.E. Barbieri, S.C. Baca, M.S. Lawrence, F. Demichelis, M. Blattner, J.-P. Theurillat, et al., Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer, *Nat. Genet.* 44 (2012) 685–689.
- [61] J. Li, Y. Wang, X. Rao, Y. Wang, W. Feng, H. Liang, et al., Roles of alternative splicing in modulating transcriptional regulation, *BMC Syst. Biol.* 11 (2017) 89.
- [62] G.-H. Wei, G. Badis, M.F. Berger, T. Kivioja, K. Palin, M. Enge, et al., Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo, *EMBO J.* 29 (2010) 2147–2160.
- [63] B. Lin, J. Wang, X. Hong, X. Yan, D. Hwang, J.-H. Cho, et al., Integrated expression profiling and ChIP-seq analyses of the growth inhibition response program of the androgen receptor, *PLoS ONE* 4 (2009) e6589.
- [64] A. Urbanucci, B. Sahu, J. Seppälä, A. Larjo, L.M. Latonen, K.K. Waltering, et al., Overexpression of androgen receptor enhances the binding of the receptor to the chromatin in prostate cancer, *Oncogene* 31 (2012) 2153–2163.
- [65] N.V. Guseva, O.W. Rokhlin, T.B. Bair, R.B. Glover, M.B. Cohen, Inhibition of p53 expression modifies the specificity of chromatin binding by the androgen receptor, *Oncotarget* 3 (2012) 183–194.
- [66] A. Sharp, I. Coleman, W. Yuan, C. Sprenger, D. Dolling, D.N. Rodrigues, et al., Androgen receptor splice variant-7 expression emerges with castration resistance in prostate cancer, *J. Clin. Invest.* 129 (2019) 192–208.
- [67] E.S. Antonarakis, C. Lu, H. Wang, B. Lubber, M. Nakazawa, J.C. Roeser, et al., AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer, *N. Engl. J. Med.* 371 (2014) 1028–1038.
- [68] M.A. Hermida, J. Dinesh Kumar, N.R. Leslie, GSK3 and its interactions with the PI3K/AKT/mTOR signalling network, *Adv. Biol. Regul.* 65 (2017) 5–15.
- [69] B.Y. Shorning, M.S. Dass, M.J. Smalley, H.B. Pearson, The PI3K-AKT-mTOR pathway and prostate cancer: at the crossroads of AR, MAPK, and WNT signaling, *Int. J. Mol. Sci.* 21 (2020) E4507.
- [70] E.G. Bluemn, I.M. Coleman, J.M. Lucas, R.T. Coleman, S. Hernandez-Lopez, R. Tharakan, et al., Androgen receptor pathway-independent prostate cancer is sustained through FGF signaling, *Cancer Cell* 32 (2017) 474–489, e6.
- [71] J. Hernández, I. Balic, T.L. Johnson-Pais, B.A. Higgins, K.C. Torkko, I. M. Thompson, et al., Association between an estrogen receptor alpha gene polymorphism and the risk of prostate cancer in black men, *J. Urol.* 175 (2006) 523–527.