Research Article

Muhammad Zubair Mahboob, Arslan Hamid, Nada Mushtaq, Sana Batool, Hina Batool, Nadia Zeeshan, Muhammad Ali, Kalsoom Sughra* and Naeem Mahmood Ashraf*

Data-mining approach for screening of rare genetic elements associated with predisposition of prostate cancer in South-Asian populations

Güney Asya Popülasyonlarında Prostat Kanseri Predispozisyonuyla İlişkili Nadir Genetik Elementlerin Taranmasında Veri Madenciliği Yaklaşımı

https://doi.org/10.1515/tjb-2018-0454
Received November 9, 2018; accepted May 23, 2019; previously published online August 30, 2019

Abstract

Objective: Prostate cancer (PCa) is a complex heterogeneous disease and a major health risk to men throughout the world. The potential tumorigenic genetic hallmarks associated with PCa include sustaining proliferative signaling, resisting cell death, aberrant androgen receptor signaling, androgen independence, and castration resistance. Despite numerous comprehensive genome-wide association studies (GWAS), certain genetic elements associated with PCa are still unknown. This situation demands more systematic GWAS studies in different populations. This study presents a computational strategy for identification of novel and uncharacterized genetic factors associated with incidence of PCa in South Asian populations.

Materials and methods: Genome-wide association studies (GWAS) catalog and Gene Expression Omnibus (GEO) furnished PCa-related genetic studies. Database for Annotation, Visualization and Integrated Discovery (DAVID) functionally annotated these genes and wANNOVAR separated South Asian (SAS) populations – specific genetic factors at MAF threshold <0.05.

Results: The study reports 195 genes as potential contributors to prostate cancer in SAS populations. Some of identified genes are PYGO2, RALBP1, RFX5, SLC22A3, VPS53, HMCN1 and KIF1C.

Conclusion: The identified genetic elements may assist in development of population-specific screening and management strategies for PCa. Moreover, this approach may also be used to retrieve potential genetic elements associated with other types of cancers.

Keywords: Prostate cancer (PCa); South Asians populations (SAS); Genome wide associations (GWAS); Microarray expressions; Minor allele frequency (MAF).

Öz

Amaç: Prostat kanseri (PCa) karmaşık heterojen bir hastalıktır ve dünyadaki erkekler için önemli bir sağlık riskidir. PCa ile ilişkili potansiyel tümörijenik genetik işaretler proliferatif sinyalleşmeyi sürdürmeyi, hücre ölümüne direnmeyi, anormal androjen reseptör sinyalini ve androjen bağımsızlığını ve kastrasyon direncini içerir. Çok sayıda kapsamlı genom çapında ilişkilendirme çalışmasına (GWAS) rağmen, PCa ile ilişkili bazı genetik unsurlar hala bilinmemektedir. Bu durum, farklı popülasyonlarda daha sistematik GWAS çalışmaları gerektirmektedir. Bu çalışma Güney Asya popülasyonlarında PCa insidansıyla ilişkili yeni ve daha önce karakterize edilmemiş genetik faktörlerin tanımlanması için hesaplamalı bir strateji sunmaktadır.

***Corresponding authors: Kalsoom Sughra and Naeem Mahmood Ashraf,** University of Gujrat – Hafiz Hayat Campus, Department of Biochemistry and Biotechnology, Gujrat, Pakistan, e-mail: kalsoom.sughra@uog.edu.pk (K. Sughra); naeem.mahmood@uog.edu.pk. https://orcid.org/0000-0003-3614-0702 (N. M. Ashraf)
Muhammad Zubair Mahboob and Nada Mushtaq: University of Gujrat, Department of Biochemistry and Biotechnology, Gujrat, Pakistan
Arslan Hamid: University of Stuttgart, Department of Sciences, Stuttgart, Germany
Sana Batool and Hina Batool: University of the Punjab, School of Biological Sciences, Lahore, Pakistan
Nadia Zeeshan: University of Gujrat – Hafiz Hayat Campus, Department of Biochemistry and Biotechnology, Gujrat, Pakistan
Muhammad Ali: Department of Biotechnology, Abbottabad Campus, COMSATS Institute of Information Technology, Abbottabad, Pakistan

Brought to you by | Turkish Biochemical Society (TBD)
Authenticated
Download Date | 12/7/19 2:07 PM

Gereç ve Yöntem: PCa ile ilgili genetik çalışmalarda GWAS kataloğu ve Gene Expression Omnibus (GEO) kullanıldı. Açıklama, Görselleştirme ve Entegre Keşif Veri Tabanı (DAVID), bu genleri fonksiyonel olarak açıkladı ve wANNOVAR, Güney Asya (SAS) popülasyonlarını, MAF eşiğinde <0.05 eşdeğer genetik faktörleri ayırdı.
Bulgular: Çalışma, SAS popülasyonlarında prostat kanserine potansiyel katkıda bulunan 195 gen olduğunu bildirmektedir. Tanımlanan genlerin bazıları PYGO2, RALBP1, RFX5, SLC22A3, VPS53, HMCN1 ve KIF1C'dir.
Sonuç: Tanımlanan genetik unsurlar PCa için popülasyona özgü tarama ve yönetim stratejilerinin geliştirilmesine yardımcı olabilir. Ayrıca, bu yaklaşım, diğer kanser türleriyle ilişkili potansiyel genetik elementleri elde etmek için de kullanılabilir.

Anahtar Sözcükler: Prostat kanseri (PCa); Güney Asya Popülasyonları (SAS); Genom Çapında İlişkiler (GWAS); Mikroarray Ekspresyonları; Minör Alel Frekansı (MAF).

Introduction

Prostate gland, a part of the male reproductive system is responsible for proper nourishment of sperms. Prostate-related disorders especially prostate cancer (PCa) is known to succumb the majority of men in the world. Although the incidence and mortality of PCa vary among different populations, in Western societies it is the second leading cause of cancer death in men [1, 2]. Similarly, men of Caribbean, African and Saharan African origin have been found to be at higher risk of PCa [3, 4]. Formerly Asian men were at the lower risk of PCa, however, during the past few decades, it is on a steady increase in the Asian countries. Currently, the PCa is the sixth leading cause of mortality in Asian men [5, 6]. Prostate cancer is, therefore, becoming a real public health issue in many populations of the world.

Due to the vivid increase of the prostate cancer in the South Asian (SAS) populations, there is a considerable concern to identify the genetic causes of this high prevalence. Till now, prostate-specific antigen (PSA) represents the gold standard biomarker for diagnosis of PCa [7]. However, multiple genome-wide association studies have shown that several genetic elements may also play a critical role in the development of the complex disease in various populations and therefore are useful for the efficient diagnosis of the diseases within these populations. In fact, numerous genetic elements associated with PCa-risk in distinct populations are also made known. Among these genetic elements, TMPRSS2-ERG fusion has a prevalence of about 50% [8]. Another prominent genetic factor is PTEN whose inactivation has reported in 70% of Caucasians but only 34% in Chinese patients [9].

Similarly, several other genetic loci including MTA-1, MYBL2, FLS353, BRCA1, BRCA2, HOXB13, NKX3.1, APPL2, TPD52, LTC4S, ALDH1A3 and AMD1 have also been reported multiple times as the risk factors for PCa in the various populations of the world [10, 11]. Although some of the genetic elements including IRX4, FOXP4, RFX6, C2orf43, TLR-4, MMP2, TIMP2, SRD5A2, SMARCA2 and FAM111A are reported frequently in the South Asian populations, however, many underlying genetic risk factors still need to be divulged in these populations [12–15]. Since the past few years, the extracellular vesicles from the body fluids are being analyzed to discover the novel cancer biomarkers [16]. All available methods for the identification of culprit genetic elements corresponding to a particular disease are, however, time-consuming, arduous, and also needs enormous funding.

Recent advancements in the field of computational biology have made it possible to predict the genetic association of complex diseases. The present study, therefore, make use of the multiple computational tools for the efficient identification of the population-specific genetic elements associated with the prostate cancer in the SAS populations. Genome-wide association and microarray expression data constitute the primary data used in this study. Total five South Asians populations registered in the 1000 Genomes browsers have been considered relevant in this study. These include Bengali from Bangladesh (BEB), Gujarati Indian from Houston Texas (GIH), Indian Telugu from the UK (ITU), Punjabi from Lahore Pakistan (PJL), Sri Lankan Tamil from the UK (STU).

The novel genetic loci identified through this computational study can assist in the designing the genotyping arrays for early diagnosis and management of the PCa and therefore may help in the development of population-specific targeted therapy for PCa in the SAS populations. The study, therefore, would undoubtedly help to curtail the gap in the rapid disease progression and control in the SAS populations.

Material and methods

The main steps used to carry out this study were as follows;

Selection of relevant studies from GWAS and GEO

Two literature-based databases Genome-wide association studies (GWAS Catalog), and Gene Expression Omnibus

Brought to you by | Turkish Biochemical Society (TBD)
Authenticated
Download Date | 12/7/19 2:07 PM

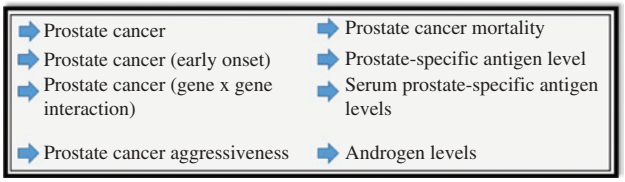


Figure 1: Terms assigned to retrieve PCa-associated studies through GWAS and GEO.

(GEO) were initially queried. This search was intended to select only those studies which report genetic factors likely to associate with PCa-risk in the various populations. Among these databases, GWAS is a catalog of the National Human Genome Research Institute (NHGRI) and the European Bioinformatics Institute (EMBL-EBI). This database is the hub of disease-related genes and variants [17]. Contrary to this, GEO is a functional genomics data repository that offers retrieval of the heterogeneous datasets from the high-throughput gene expression and genomic hybridization experiments [18]. To search relevant studies from the GWAS catalog and GEO a total of eight PCa-related search terms selected from GWAS catalog. These search terms used as the query in both of the mentioned databases (Figure 1). In case of the GEO database from all studies retrieved through these terms only those comparing tumor and normal tissues or treated with untreated tissues were considered significant. These studies commonly carried on Affymetrix and Illumina platform. Remaining studies done on cell lines or reporting methylation profiling have not examined. The selected studies published from 2009 to 2015.

Final gene lists generation

The first gene list, "GWAS genes list" includes all genes retrieved from the all selected genome-wide association studies. However, for retrieval of genes from the microarray expression studies, all these studies were analyzed using R 3.3.2. [19]. Multiple Bioconductor packages such as GEOquery, limma, Affy, gcrma, illuminaHumanv4.db, hgu133plus2.db were used for the analysis of expression studies in R [20–24]. These packages were used to download the GEO data files, normalize, read, annotate, and analyze the microarray data. This results in the second gene list "differentially expressed genes (DEGs)" list. It includes all genes expressing differentially between disease and healthy subjects at Benjamini-Hochberg (BH) p-value ≤0.05. This p-value adjusted for the multiple testing with the BH method to control the false discovery rate [25].

Functional annotation of gene lists

"GWAS genes list" and "Differentially Expressed Genes list" were functionally annotated using Database for Annotation, Visualization, and Integrated Discovery (DAVID). This server correlates genes and protein lists with the biological denotations and confirms their functional association with disease [26]. For analysis of gene lists in the DAVID, multiple categories were selected (Table 1). p-Value threshold of <0.05 was applied to reduce false positive results [27]. Functionally annotated genes from both lists at these selected parameters were combined for downstream analysis and referred as "core genes list".

Retrieval of variants associated with core genes list

Chromosomal coordinates of core list genes retrieved from Ensemble Biomart server. The attributes retrieved through this server include; chromosomal number, start and end position of the gene, and associated gene names [28]. The attributes were then used to download variant cell format (VCF) files from 1000 Genomes browser [29]. All VCF files were merged through the combine-VCF tool to produce a single VCF file containing variants associated with genes in "core gene list".

Selection of SAS specific genes and variants

Finally, the implementation of population specificity filter to the core gene list results in the retrieval of genetic elements purely associated with the risk of PCa in SAS populations. For the assessment of SAS specific variants, VCF files uploaded in wANNOVAR [30]. wANNOVAR is a tool to annotate single nucleotide variants, deletions and insertions. The potential SAS specific exonic-variants selected at minor allele frequency (MAF) threshold of <0.05 [31]. This MAF threshold filtered those variants which are rare in the SAS populations and likely to associate with the diseases. Thus all variants genetic elements retrieved at this step would

Table 1: List of options used in DAVID to target specific terms and IDs.

DAVID categories	Selected options
Disease	OMIM Disease, Genetic Association DB
Gene ontology	GOTERM_BP_FAT
Pathways	BBID, BIOCARTA, KEGG

Brought to you by | Turkish Biochemical Society (TBD)
Authenticated
Download Date | 12/7/19 2:07 PM

likely to be associated with PCa susceptibility in SAS population.

Expression of potential genes in prostate tissue

The involvement of all these potent genes in the disease susceptibility was confirmed using expression profiling of the prostate tissue. The Human Protein Atlas, a database well-known to report tissue-restricted variable and stable expression of proteome and transcriptome of the human body, was used for this purpose [32]. All of the genes have a significant expression in the prostate tissues considered as functionally important genetic elements that may help to determine the etiology of the disease.

Results

Selection of PCa related studies

To search the primary databases, specific PCa related search terms chosen from the GWAS catalog. Total thirty-seven studies in the GWAS catalog matched to the eight search terms used as a query (Supplementary file 1). All of the retrieved studies were reporting PCa and its associated genetic factors in the various populations of the world. Total 10 studies from the GEO found as related to prostate cancer (Supplementary file 3).

Gene lists generation

From thirty-seven studies selected from the GWAS catalog, 322 genes were the part of "GWAS gene list" (Supplementary file 2). On the other hand, for retrieval of differentially expressed genes from 10 studies selected retrieved through the GEO database, it is mandatory to read all these studies through the R package [33]. Among these 10 studies, eight were conducted on Affymetrix while the other two were on the Illumina platform. Analysis of these studies through R results in the accurate comparison of genetic elements among cancerous and control samples (Supplementary file 4). This discriminative analysis initially results in the 14,367 differentially expressed genes at BH p-value of ≤0.05 (Table 2). To further enhance the confidence that these DEGs are purely related to PCa-risk, an additional criterion applied to this existent differentially expressed gene list. Only those genes which were appearing in more than one of the selected studies considered for further analysis. For this purpose, all these studies divided into five different groups based on the chip similarity and the intersecting genes from these groups selected. As a result, only 654 genes were finally the part of "GEO gene list" (Supplementary file 5).

Table 2: Differentially expressed genes (DEGs) retrieved by comparison between tumor and benign tissue.

Sr#	GEO accession	Analyzed subjects	DEGs
1	[a]GSE 17906	11 PCa – 12 controls	485
2	GSE 62872	264 PCa – 160 controls	6726
3	GSE 46602	35 PCa – 15 controls	1890
4	GSE 55945	13 PCa – 8 controls	185
5	GSE 28403	9 PCa – 4 controls	0
6	GSE 30521	17 PCa – 5 controls	389
7	GSE 29079	47 PCa – 48 controls	1483
8	GSE 32269	51 PCa – 4 controls	3193
9	GSE 28680	7 Responding – 4 controls, 13 not responding – 4 controls	4
10	GSE 72920	15 Treated – 15 controls	12

[a]A GEO Series (GSExxxxx) corresponds to a record that summarizes a study.

Functional annotation of gene lists

GWAS gene list as previously stated have a total of 322 genes, out of which only 274 mapped to multiple DAVID's terms and pathways, and as a result, only 206 genes from GWAS gene list were confirmed to have the functional association with PCa. Contrarily, all of the 654 DEGs from GEO gene list mapped to DAVID and 450 genes from GEO gene list were functionally relevant. The functionally annotated genes from both lists were then combined and redundancies removed. Finally, about 644 functionally annotated genes considered for the downstream analysis. Thus about 32% of the starting genes have been eliminated at this step. This list of 644 genes referred to as "core gene list" (Supplementary file 6).

Selection of SAS specific genes and variants

The variant files associated with the core genes obtained from the 1000 Genomes browser and files were read using wANNOVAR. The wANNOVAR provided allele frequencies of all these variants for the populations reported in the 1000 Genomes browser. The MAF threshold of <0.05 applied to the allelic frequencies concerning to South Asian populations. A total of 312 genes and 361 exonic

Brought to you by | Turkish Biochemical Society (TBD)
Authenticated
Download Date | 12/7/19 2:07 PM

variants associated with these genes were found relevant in these populations at applied MAF threshold (Supplementary file 7).

As is evident with applied MAF threshold that these genetic elements were rare in the SAS population, therefore, they are likely to be potential contributors for PCa.

Expression of significant genes in prostate tissue

Expression analysis revealed that some of the selected genes not expressed in the prostate tissues. It may because of the absence of any studies describing the association of these genes with the PCa, however, 195 genes have shown profound expression in prostate tissue (Supplementary file 8). Out of these 195, one gene CYP17A1 was reported previously in the SAS populations, and 194 genes are newly identified PCa-risk factors for SAS populations [15, 34].

These genes considered as the potentially significant PCa-risk factors in the SAS populations because all these genetic elements not only have the required MAF but also have a significant expression in the prostate tissue (Table 3).

Table 3: List of genes identified as the potential PCa-risk factors in the South Asian populations.

ACSL1	CAND1	GCAT	LMTK2	PLA2G16	SLC2A3
ACTN1	CAPN9	GGCX	LRRFIP1	PMVK	SLC2A5
ADAM15	CCNC	GIPR	LRRN2	POU2F1	SLC41A1
ADAMTS4	CEP57L1	GLMP	MAPK8IP2	PPFIA4	SLC7A6
ADAR	CFI	GNPNAT1	MARK1	PRKAA2	SMARCA2
ADCY3	CLCA4	GOLPH3L	MASP2	PRKCI	SRRM1
ADCY6	COA6	GOT2	MDM4	PRPF31	SS18L1
ADNP	CRTC2	GPR88	MFSD10	PSD	TCEB3
ADRM1	CTBP2	HBP1	MLPH	PYGO2	TRAPPC3L
AFP	CYP17A1	HDLBP	MMADHC	RALBP1	TRIM31
AGA	CYTH2	HENMT1	MMP10	RBBP7	TRIM67
AGER	DDX6	HFE	MMP8	RERE	TTLL7
AGMAT	DEAF1	HLA-DQA1	MTHFD1L	RFX5	UBAP2L
AK7	DIAPH1	HMCN1	MUC1	RGL1	UBE2W
AKT2	DLST	HMGCS2	MYC	RHOU	ULK2
ANK3	DNAH14	HNF1B	MYSM1	RNF126	UTS2
ANKRD17	DTYMK	HS2ST1	NCOA1	RNF2	VAMP5
AP4B1	EDNRA	HSPB11	NEDD9	RRM2	VPS11
APLP2	EIF4B	HTR3B	NFATC3	RRN3	VPS53
ARHGEF16	EPHA8	IGF2BP1	NGFR	SAP30	WNT4
ARL3	ERN1	IRF4	NHLH2	SDCCAG8	YOD1
ATF6B	ETV3	ITGA6	NOL3	SDF4	ZBTB12
ATF7IP	F5	ITIH4	NPHS1	SEMA4D	ZBTB17
ATG4C	FERMT2	JMJD1C	OSBPL2	SERINC2	ZBTB7B
ATP6V1A	FGF10	KATNB1	PBX2	SERPINF1	ZFP36L1
ATXN10	FKBPL	KCNN3	PBXIP1	SHC1	ZNF24
B4GALNT2	FKTN	KIF1C	PDE4A	SIAH1	ZNF268
B9D1	FLOT1	KLK15	PDXDC1	SIX1	ZNF652
C16or f93	FOXO3	KLK3	PHACTR4	SKIL	ZRANB1
C1or f116	FOXP4	KRT18	PHB	SKIV2L	ZSWIM5
C1or f167	GALNS	KRT8	PI4KB	SLA	
C1or f43	GARS	LAMA3	PIK3R2	SLC16A4	
CADM1	GATA3	LAMC1	PKP1	SLC22A3	

Discussion

In the recent era, multiple GWA studies have been conducted around the globe to trace the roots of complex diseases including cancer. These studies have identified the participation of numerous culprit genes in the disease susceptibility in various populations. Several genes and their associated variants are known for PCa susceptibility [6, 12, 35]. Till to date only a few candidate genes and variants are reported to associate with PCa risk in SAS. The high incidence of PCa in South Asians and lack of adequate information about genetic participants, demands exploration of genetic factors primarily associated with South Asians population. By knowing the genetic foundations of the PCa in the SAS, efficient genotype-based screening, as well as management approaches, could be designed which will promisingly reduce the progression of this cancer form in SAS men.

In the recent past, advancements in bioinformatics have helped to predict correspondence between genetic elements and diseases. The present study is also base on an in silico analysis. It involves probing of already existing genome-wide associations and microarray expression data for the identification of PCa – susceptible genetic elements related to SAS populations using computational tools. Both GWAS catalog and GEO databases provide a different type of data sets. GWAS specify all genes and single nucleotide polymorphism (SNPs) having an association with a particular phenotype. GEO on the other hand outline differentially expressed genes between disease and control subjects. The inclusion of data from both of these datasets would help to isolate highly relevant genetic candidates of PCa in SAS populations.

Genes sequestered from these databases were then functionally annotated through another database DAVID. Through functional analysis involvement of these genes in various pathways was evaluated. Thus although initially 322 genes from GWAS catalog and 654 genes had been considered significant after their functional exploration irrelevant genes were excluded leaving total 644

Brought to you by | Turkish Biochemical Society (TBD)
Authenticated
Download Date | 12/7/19 2:07 PM

potential genes for downstream analysis. These genetic components were examined further to check their association with SAS populations. For this purpose MAF threshold <0.05 applied to SAS-specific exonic variants in the wANNOVAR. As the variants retrieved at this threshold are rare in the population, therefore, they are expected to relate with the disease in the given population and thus considered significant. About 351 genes excluded from the core list as not being correlated with the population of interest.

Furthermore, to assure the results of the analysis, the expression of 312 relevant genes was examined in prostate tissues using Human Protein Atlas. A total of 195 genes have shown their expression in the prostate which therefore considered as the PCa susceptibility genetic elements in the SAS populations. Out of all these, only one gene (CYP17A1) is previously marked to associate with PCa susceptibility in SAS this shows the gap between disease progression and productive research on it. The CYP17A1 encodes a member of the cytochrome P450 superfamily of enzyme which is involved in drug metabolism as well as cholesterol and steroid biosynthesis. Sobti et al. have studied the association of CYP17A1 with prostate cancer in North Indian population [15, 34]. The remaining 194 genes which have been predicted as the potential PCa-risk factors in this study are previously not known to associate with PCa in these populations (Table 3). These genes could be novel genetic risk factors associated the PCa in the South Asian populations that have not been identified so far. It is therefore necessary to experimentally confirm their association with the PCa in these populations.

Conclusions

The computational approach employed in this study predicts the potential genetic elements associated with prostate cancer in the South Asian populations. The genetic markers identified through this study need further experimental validation to assure their precise incidence to the disease in the South Asian populations. The study therefore contribute significantly to development of population-specific screening, prevention and management strategies for prostate cancer. It would, therefore, facilitate to curtail the gap between the rapid disease progression and its control.

Acknowledgments: We are grateful to the public genome resource provided by the 1000 Genomes Project and all of open resource databases.

Conflict of interest statement: The authors declare no competing interests.

References

1. Mann T, Lutwak-Mann C. Male reproductive function and semen: themes and trends in physiology, biochemistry and investigative andrology, 1 ed. Berlin: Springer, 1981:17–499.
2. Miller DC, Hafez K, Stewart A, Montie J, Wei J. Prostate carcinoma presentation, diagnosis, and staging: an update from the National Cancer Data Base. Cancer 2003;98:1169–78.
3. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69–90.
4. Stewart BW, Wild CP, editors. World Cancer Report. International Agency for Research on Cancer, World Health Organization, 2014;1:16–52.
5. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87–108.
6. Center MM, Jemal A, Lortet-Tieulent J, Ward E, Ferlay J, Brawley O, et al. International variation in prostate cancer incidence and mortality rates. Eur Urol 2012;61:1079–92.
7. Knipe DW, Evans D, Kemp J, Eeles R, Easton D, Kote-Jarai Z, et al. Genetic variation in prostate-specific antigen–detected prostate cancer and the effect of control selection on genetic association studies. Cancer Epidemiol Biomarkers Prev 2014;23:1356–65.
8. Demichelis F, Fall K, Perner S, Andrén O, Schmidt F, Setlur S, et al. TMPRSS2: ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene 2007;26:4596–9.
9. Cairns P, Okami K, Halachmi S, Halachmi N, Esteller M, Herman JG, et al. Frequent inactivation of PTEN/MMAC1 in primary prostate cancer. Cancer Res 1997;57:4997–5000.
10. Dhanasekaran SM, Barrette T, Ghosh D, Shah R, Varambally S, Kurachi K, et al. Delineation of prognostic biomarkers in prostate cancer. Nature 2001;412:822–6.
11. Ali HE, Lung PY, Sholl AB, Gad SA, Bustamante JJ, Ali HI, et al. Dysregulated gene expression predicts tumor aggressiveness in African-American prostate cancer patients. Sci Rep 2018;8:16335.
12. Priyadarshini A, Chakraborti A, Mandal AK, Singh SK. Asp299Gly and Thr399Ile polymorphism of TLR-4 gene in patients with prostate cancer from North India. Indian J Urol 2013;29:37–41.
13. Srivastava P, Lone T, Kapoor R, Mittal RD. Association of promoter polymorphisms in MMP2 and TIMP2 with prostate cancer susceptibility in North India. Arch Med Res 2012;43:117–24.
14. Chen R, Ren S, Yiu M, Fai N, Cheng W, Ian L, et al. Prostate cancer in Asia: a collaborative report. Asian J Urol 2014;1:15–29.
15. Sobti R, Onsory K, Al-Badran A, Kaur P, Watanabe M, Krishan A, et al. CYP17, SRD5A2, CYP1B1, and CYP2D6 gene polymorphisms with prostate cancer risk in North Indian population. DNA Cell Biol 2006;25:287–94.
16. Nawaz M, Camussi G, Valadi H, Nazarenko I, Ekström K, Wang X, et al. The emerging role of extracellular vesicles as biomarkers for urogenital cancers. Nat Rev Urol 2014;11:688–701.
17. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 2013;42:D1001–6.
18. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30:207–10.

Brought to you by | Turkish Biochemical Society (TBD)
Authenticated
Download Date | 12/7/19 2:07 PM

19. Ihaka R, Gentleman R. R: a language for data analysis and graphics. J Comput Graph Stat 1996;5:299–314.
20. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. Bioinformatics 2007;23:1846–7.
21. Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S, editors. Bioinformatics and computational biology solutions using R and bioconductor. New York: Springer, 2005;1:397–420.
22. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004;20:307–15.
23. Wu J, Irizarry R, MacDonald J, Gentry J. Gcrma: background adjustment using sequence information. R package version, 2012;2200:3–10.
24. Dunning M, Lynch A, Eldridge M. illuminaHumanv4. db: Illumina HumanHT12v4 annotation data (chip illuminaHumanv4). R Package version, 2015;1:1–4.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 1995;57:289–300.
26. Huang DW, Sherman BT, Lempicki R. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2008;4:44–57.
27. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. Br Med J 1995;310:170.
28. Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G, et al. BioMart–biological queries made easy. BMC Genomics 2009;10:22.
29. Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, et al. The 1000 Genomes Project: data management and community access. Nat Methods 2012;9:459–62.
30. Chang X, Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. J Med Genet 2012;49:433–6.
31. Sidore C, Busonero F, Maschio A, Porcu E, Naitza S, Zoledziewska M, et al. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nat Genet 2015;47:1272–81.
32. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science 2015;347:1260419.
33. Dudoit S, Gentleman RC, Quackenbush J. Open source software for the analysis of microarray data. Biotechniques 2003;34(S3):S45–51.
34. Madigan MP, Gao YT, Deng J, Pfeiffer RM, Chang BL, Zheng S, et al. CYP17 polymorphisms in relation to risks of prostate cancer and benign prostatic hyperplasia: a population-based study in China. Int J Cancer 2003;107:271–5.
35. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 2008;40:310–5.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/tjb-2018-0454).

Brought to you by | Turkish Biochemical Society (TBD)
Authenticated
Download Date | 12/7/19 2:07 PM