Abstract: The incidence of prostate cancer (PCA) varies by ethnicity. This study aimed to provide insights into the genetic cause of PCA, which can result in differences in incidence among individuals of diverse ancestry. We collected data on PCA-associated single-nucleotide polymorphisms (SNPs) from a genome-wide association study catalog. Fisher’s exact tests were used to analyze the significance of enrichment or depletion of the effect on the allele at a given SNP. A network analysis was performed based on PCA-related SNPs that showed significant differences among ethnicities. The SNP-based polygenic risk score (PRS) was calculated, and its correlation with PCA incidence was evaluated. European, African, and East Asian populations had different heatmap patterns. Calculated PRS from the allele frequencies of PCA was the highest among Africans, followed by Europeans, and was the lowest among East Asians. PRS was positively correlated with the incidence and mortality of PCA. Network analysis revealed that AR, CDKN1B, and MAD1L1 are genes related to ethnic differences in PCA. The incidence and mortality of PCA showed a strong correlation with PRS according to ethnicity, which may suggest the effect of genetic factors, such as the AR gene, on PCA pathogenesis.

Keywords: prostate cancer; polygenic risk score; allele frequency; single nucleotide polymorphism; androgen receptor gene

1. Introduction

Prostate cancer (PCA) is the second most frequently diagnosed cancer in men and the fifth leading cause of death worldwide [1]. Epidemiological data showed that PCA incidence and mortality showed disparities across ethnic groups; men of African ancestry had the highest incidence and mortality, followed by those with European and Asian ancestries [2,3]. Comparison of epidemiological data from the United Kingdom and the United States [2] revealed the prevalence of PCA is the highest in those with African ethnicity and the lowest in those with East Asian ethnicity [2,4]. These findings suggest that genetics have an important role in PCA. To date, the reasons for ethnic differences of PCA incidence and mortality are not fully understood. A closer look at risk factors for PCA, such as advancing age, race, positive family history of PCA, and Western diet, might help elucidate the reason into consideration [3,5]. Hormone levels, such as those of dihydrotestosterone (DHT) and testosterone (T), and the DHT:T ratio have been studied in patients with PCA, and DHT:T ratios were higher in African Americans than in Europeans and Asians, further supporting that these hormone effects may have a role in PCA [6]. Moreover, higher 25-hydroxyvitamin D (Vit D) levels were associated with reduced PCA mortality [7]; however, one study reported that Europeans have higher Vit D levels than Asians, thus contradicting the protective theory of Vit D in PCA [8]. Nevertheless, ethnic differences in disease prevalence suggest that studies on genetic factor analysis are warranted.
Population-based studies, such as genome-wide association studies (GWAS) are purposed to find common variants for common diseases. Despite this, GWAS have revealed a substantial familial relative risk (FRR) (28.4%) for PCa [9]. Furthermore, hereditary tumor suppressor genes such as BRCA1, BRCA2, and homologous recombination deficiency (HRD) genes including ATM, CDK12, FANCA, RAD51, RAD51C, CHEK2, and PALB2, are known to be involved in PCa [10–13]. However, these well-known tumor suppressor genes can only account for approximately 30% of PCa cases and cannot explain the ethnic variance across populations, where the incidence rate of PCa is 167 and 52.2 per age-adjusted 100,000 population for African Americans and Asians in America, respectively [2,3]. These studies indicate that there may be some common genes that we regarded with less significance that may be involved in the pathogenesis of PCa, which can explain the differences in PCa incidence and mortality across ethnicities.

Researchers have suggested using whole-genome sequencing data of healthy subjects to identify disease phenotypes [14]. By combining data from the GWAS catalog of the National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) and the incidence and mortality rates from various ethnic groups with PCa, it may be possible to model risk single-nucleotide polymorphisms (SNPs) related to PCa in different ethnic groups. We have previously analyzed differences of SNP-related models to two ophthalmic diseases and one Vit D deficiency and published significant results regarding the strong link between disease incidence and genetic risk [15,16]. Using this technique, it is possible that genetic impact based on SNPs associated with PCa are responsible for the difference in PCa incidence and mortality across ethnic groups. When considering cancer through an epidemiologic approach, two end values need to be considered: incidence and mortality. By identifying the causes of incidence, prophylactic measures can be implemented to reduce the incidence of PCa. Furthermore, by determining the causes of mortality, we can find more effective treatment options.

Herein, this study aimed to identify the most important contributing factors, including genetics, hormone levels, such as DHT:T ratio, and environmental factors, such as Vit D, to the incidence and mortality of PCa.

2. Materials and Methods

2.1. Ethical Considerations

This study was approved by the Institutional Review Board of the Veterans Health Service Medical Center, Korea (approval number 2021-09-004), and the requirement for informed consent was waived owing to the use of de-identified data.

2.2. Comparison of PCa SNPs among Global Populations

NHGRI-EBI GWAS catalog (https://www.ebi.ac.uk/gwas/home, accessed on 20 April 2021) was used to find the SNPs associated with PCa traits (EFO 0000305, EFO 0000708, EFO 0000714, EFO0001071, EFO 0001075, EFO 0001360, EFO 0001663, EFO 0005842, EFO 0006999, EFO 0007000, EFO1000650, EFO 1001515, and EFO 1001516). The catalog included 36 studies and 623 associations when searched for the keyword of “Prostate carcinoma” in the mapped disease trait column of the GWAS catalog. After eliminating repetitive SNPs and discarding the data not found in the database of the 1000 Genome Project (1000 GP), 600 SNPs from the GWAS catalog were used for the analysis of allele frequencies associated with PCa.

We determined the PCa genetic risk model by examining the $\beta$ coefficients and odds ratios for the affected allele. We also analyzed the text descriptions in the primary GWAS reports. The details and the advantages of this method have been described by Mao et al., [17] and Dudbridge et al., [18]. The population-level allele frequencies of SNPs were from the 1000 GP phase 3 data, which surveyed genetic variations in 2504 individuals from 26 populations worldwide grouped from African (AFR), East Asian (EAS), European (EUR), South Asian (SAS), and American (AMR) categories based on their geographical locations and ancestry [19].
These data were retrieved from: ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ (accessed on 15 January 2020). The variant coordinates were based on the human genome assembly GRCh37. After statistical analysis, we performed a network analysis of the genes we studied in the GWAS catalog.

2.3. Calculation of Polygenic Genetic Risk Scores Using SNPs Related to PCa

The polygenic risk score (PRS) was calculated as:

\[
\text{Polygenic risk score} = \sum_{i=1}^{I} \frac{\beta_i X_i}{2I}
\]

where \( I \) is the number of PCa-related SNPs, \( X_i \) is the copies of risk alleles \((X_i \in (0,1,2))\) at the \( i \)th SNP, and \( \beta_i \) is the average odds ratio of the \( i \)th SNP reported in GWAS studies \[17,18\].

If an individual has two copies of the risk allele at each PCa-related SNP, the risk score was set to 1. However, the risk score was 0 if he had no risk allele. Thus, a male with a PRS of 1 had the highest possible genetic risk of PCa, whereas one with a score of 0 had the lowest. If the copies of effect alleles \((0/1/2)\) were randomly assigned to each SNP, the expected value of the risk score was set to 0.5. We used the average PRS to determine the correlations between age-adjusted PCa incidence and mortality data \[2,3\]. We used the open database of Korean Reference Genome Database (KRGDB) and the Korean PCa incidence and mortality epidemiology data to assess how well this genetic model can predict the incidence and mortality of Korean ethnic groups \[20\].

2.4. Comparison of Allele Frequencies

Fisher’s exact test was used to assess whether the effect on the allele at a given SNP was significantly higher or lower than the global population frequency in the 1000 GP database. The heatmap visualizes the allele patterns in different populations: red and blue colors were used to indicate higher and lower frequencies compared to the global average, respectively. If the effect on the allele was enriched in a population, then the negative \( \log_{10} \) of the \( p \)-value (a positive number) was used to represent the SNP associated with that population in the heatmap. In contrast, if it was depleted, the \( \log_{10} \) of the \( p \)-value (a negative number) was used. In addition, SNPs with different frequencies among the EAS and AFR populations (log-adjusted \( p \)-value of Fisher’s exact test in \(|AFR|+|EAS|\) with a cutoff of >60 were used as the mean sum of the absolute value was approximately <60, which in turn selects the genes most relevant to PCa.

2.5. Validation of PRS Modeling for PCa Epidemiology

To find the relationship between PCa incidence and mortality between ethnic groups, the per age-adjusted by 100,000 incidences and mortality statistical data of the United States census of 2012–2015 \[2\]. This data was compared with that of the United Kingdom from 2008 to 2010, which had similar results \[4\]. The per age-adjusted by 100,000 incidences and mortality PCa data of Korea was from the Korean census of 2007–2013 \[20\]. Linear regression analysis was performed using PRS for PCa versus PCa epidemiology in similar periods.

The DHT:T values were measured for Black, Hispanic, and White males by Litman et al., and the Korean DHT:T values were measured by Yoon et al., \[21,22\]. Vit D level was computed from the average Vit D levels from the global Vit D data, which were the same values used in Yoon et al., \[8,23\]. Vitamin D was measured globally and had little variance. Although there may be some variance and differences due to different methods used to measure DHT or T, these values were used, as there were no other common data.

2.6. Network Analysis

From the derived gene set obtained from the GWAS catalog, we used the network analysis Phenolyzer (Wang Genomics Lab) to extract the relationship among PCa genes \[24\]. As the 1000 GP database recruited normal patients without known active diseases nor malignancies,
well-known tumor suppressor genes, such as \textit{BRCA1} and \textit{BRCA2}, and HRD genes, such as \textit{ATM}, \textit{CHECK2}, \textit{BARD1}, \textit{RAD51}, \textit{RAD51C}, \textit{NBN}, \textit{PALB2}, and \textit{BRIP1}, were not well seen in the GWAS catalog, which is why this was excluded from our study; the purpose of GWAS studies is to find common variants for common diseases and not pathogenic genes for cancer. Although all the SNPs for PCa were included from the GWAS catalog in this analysis, it is difficult to determine whether the intron parts of the pathogenic gene are indeed related to PCa. Therefore, we first derived the network of PCa using the terminology “prostate cancer” for all known genes associated with PCa, and collected only the SNPs related to PCa in the GWAS, whose absolute value were $\geq 100$ from the \log_{10} p$-value. This was fed into the Phenolyzer to assess how the extracted SNPs were related to each other in PCa [24].

2.7. Statistical Analysis
Statistical analyses were performed using R software version 4.0.1 (R Foundation, Vienna, Austria), and statistical significance was set at $p < 0.05$.

3. Results
3.1. PCa-Related SNPs in the Global Population
We collected 600 PCa-associated SNPs from 36 GWASs using the latest NHGRI-EBI catalog (April 2021). We extracted the effect of allele frequencies (EAFs) for each of the ethnic groups from the 1000 GP (Supplemental Table S1). The heatmap shows significantly enriched and depleted effect on alleles across these populations (Supplemental Figure S1). The hierarchical clustering tree showed four clusters, where the EAS and AFR were in opposite directions, with EUR as the reference SNPs in the middle. Moreover, SNPs had significantly different frequencies among the EAS and AFR populations. The log-adjusted $p$-value of Fisher’s exact test in $\mid \text{AFR} \mid + \mid \text{EAS} \mid > 60$ are summarized in Table S2 and is in Supplemental Figure S2. As it is hard to see the forest from the trees, we have selected $\mid \text{AFR} \mid + \mid \text{EAS} \mid > 100$ for Figure 1. Several SNPs, such as rs5919393 (AR) and rs10486567 (JAZF1), were overexpressed in EAS, but depleted in AFR.

3.2. Genetic Risk Scores Calculated using SNPs Related to PCa Incidence & Mortality
We calculated the PRS based on the alleles enrichment or depletion value and the odd’s risk score from the 600 PCa SNPs, with the assumption that allelic associations from most GWAS-identified variants could be replicated in non-European populations. The genetic score of PCa was the highest among AFR men, followed by AMR, SAS, EUR, and EAS men in Figure 2. This was highly associated with PCa incidence and PCa mortality in Figure 3, with high coefficients of determination ($R^2$) of 0.900 and 0.946, respectively.

3.3. DHT to T Ratios and Vitamin D Related to PCa Incidence & Mortality
Linear regression of PCa incidence and mortality versus DHT:T ratio showed a positive correlation (Figure 4a and Supplemental Figure S3a). However, $R^2$ was less than that of the PRS. In addition, PCa incidence and mortality were negatively correlated with Vit D levels (Figure 4b and Supplemental Figure S3b). The data from the Korean DHT:T ratio were calculated and plotted with the Korean PCa epidemiology data to assess the model in Figure 4a, where the real-world data are in the blue targeted circle [20,22] and the real-world data for Vit D.

3.4. Network Analysis Using PCa Related Gene Analysis
Supplemental Figure S4 shows all the genes identified when the phenolyzer was run with the diseases/phenotype keyword “Prostate Cancer”. Figure 5 presents the SNPs showing significant difference patterns between the AFR and EAS groups. From these network analyses, we found that the most significant SNPs that differed among the populations were androgen receptor (AR), \textit{MAD1L1} (mitotic arrest deficient 1 like 1), \textit{CDKN1B} (cyclin-dependent kinase inhibitor 1 B), \textit{SMAD2} (SMAD family member 2), and \textit{MNAT1} (MNAT1 component of CDK activating kinase).
Figure 1. Heatmap showing log-adjusted $p$-value of Fisher’s exact test for significant single-nucleotide polymorphisms (SNPs) related to prostate cancer in global populations (AMR: American, EUR: European, SAS: South Asian, AFR: African, EAS: East Asian). Each row shows an SNP, and each column is a population group. The red color indicates enrichment of the allele, whereas the blue color indicates the depletion. The log-adjusted $p$-value of Fisher’s exact test in $|AFR| + |EAS| > 100$ for visibility.
Figure 2. Polygenetic risk score (PRS) calculations of prostate cancer (PCa) using related single-nucleotide polymorphisms. The genetic risk scores calculated from allele frequencies for PCa were the highest in Africans, followed by Europeans and East Asians. (ACB: African Caribbean in Barbados; ASW: African ancestry in the Southwest USA; BEB: Bengali in Bangladesh; CDX: Chinese Dai in Xishuangbanna; CEU: Utah residents with Northern and Western European ancestry; CHB: Han Chinese in Beijing, China; CHS: Southern Han Chinese, China; CLM: Colombian in Medellin, Colombia; ESN: Esan in Nigeria; FIN: Finnish in Finland; GBR: British in England and Scotland; GIH: Gujarati Indian in Houston, TX, USA; GWD: Gambian in Western Division, Gambia; IBS: Iberian populations in Spain; ITU: Indian Telugu in the UK; JPT: Japanese in Tokyo, Japan; KOR: Korean in the Republic of Korea; KHV: Kinh in Ho Chi Minh City, Vietnam; LWK: Luhya inWebuye, Kenya; MSL: Mende in Sierra Leone; MXL: Mexican ancestry in Los Angeles, CA, USA; PEL: Peruvian in Lima, Peru; PJL: Punjabi in Lahore, Pakistan; PUR: Puerto Rican in Puerto Rico; STU: Sri Lankan Tamil in the UK; TSI: Toscani in Italy; YRI: Yoruba in Ibadan, Nigeria).

Figure 3. Correlation between population average polygenetic risk score (PRS) and PCa incidence and mortality. (a) PRS versus PCa incidence, with linear regression in the grey dashed line ($R^2 = 0.900$). The real-world Korean incidence is represented by a blue open rhombus, which deviate from the linear regression line, whereas the estimated Korean incidence is represented by a blue solid circle. (b) PRS versus PCa mortality, with the linear regression line shown as a gray dashed line ($R^2 = 0.946$). The estimated Korean mortality is represented by a blue open rhombus, which deviates from the linear regression line, whereas the estimated Korean mortality is represented by the blue solid circle. (AMR: American, EUR: European, AFR: African, EAS: East Asian, KOR: Korean).
Figure 4. (a) Correlation plot of dihydrotestosterone (DHT) to testosterone (T) ratio and prostate cancer (PCa) incidence. Linear regression is indicated by the gray dashed line ($R^2 = 0.765$), and the real-world Korean incidence is represented by the blue open rhombus. (b) Correlation plot of vitamin D concentration and prostate cancer (PCa) mortality. Linear regression is in the gray dashed line ($R^2 = 0.426$), and the real-world Korean PCa mortality is represented by the blue open rhombus, which deviates from the linear regression, whereas the estimated Korean mortality is in the blue shaded circle. (AMR: American, EUR: European, AFR: African, EAS: East Asian, KOR: Korean).

Figure 5. Network analysis of prostate cancer (PCa)-related single nucleotide polymorphisms (SNPs), for log-adjusted $p$-value of Fisher’s exact test in $|AFR| + |EAS|$ with a cutoff over 60. The most relevant genes are $AR$ (Androgen Receptor), $MAD1LI$ (Mitotic Arrest Deficient 1 Like 1), and $CDKN1B$ (Cyclin Dependent Kinase Inhibitor 1B). The dark blue circles are the seed gene, and the light blue are genes interacting with the seed gene. The army green circles connected to the seed genes are the protein-protein interactions. Detailed legend information is provided in https://phenolyzer.wglab.org/download/Phenolyzer_manual.pdf (accessed on 1 December 2021).
4. Discussion

We have demonstrated that the correlation analysis to find that the PRS calculated from SNPs from the GWAS catalog had low values in EAS while having high values in AFR. This had a very high correlation with PCa incidence and mortality, while DHT:T ratio and Vit D levels may have less influence on PCa incidence and mortality than the calculated PRS. In addition, we found the SNPs that differed among the PCa populations include AR, JAZF1 (juxtaposed with another zinc finger protein 1), MAD1L1, CDKN1B, and SMAD2, where AR had the most significant role, as validated from the heatmap and network analysis.

Using the genes from the GWAS catalog, the genetic risk model based on PCa-SNPs showed that the highest risk was in AFR ancestry, followed by EUR ancestry, and that the lowest was in EAS ancestry. Moreover, we found that the PRS was the lowest in EAS men, the highest in AFR men, was intermediate in EUR men, implying a correlation between genes involved in prostate-related SNPs and PCa. In particular, the high correlation coefficient between PRS and PCa incidence and PCa mortality suggest a significant causal relationship between ethnicity and PCa incidence and mortality. This is similar to the results of Conti et al., who estimated a mean PRS of 2.18-fold higher in AFR men and 0.73-fold lower in EAS men than EUR men. Their results suggest that germline variation contributes to population differences in PCa risk, where the calculation of PRS offers an approach for personalized risk prediction [25]. Our analysis using the KRGDB and Korean PCa epidemiological data showed that although the real-world data may deviate from the predicted model, the deviation was well within the region of the EAS population. Moreover, this PRS model highly correlated with the reported Korean epidemiology incidence and mortality data of PCa.

Similar research has been done, where genes were extracted to correlate epidemiology of prostate cancer via each SNP genomic variant, where the genes would predict mortality or incidence [26]. The SNPs that overlapped were rs6983267 (CCAT2), rs2066827 (CDKN1B). The difference is that we have used the PRS instead of each SNP that correlates mortality and incidence. Furthermore, we have compared PRS and other factors that may be confounding, such as DHT:T and vitamin D levels, which all may be converted to androgen.

Dutasteride, a 5-a reductase, is known to reduce the incidence of PCa by suppressing the conversion of T to DHT [27]; we have assumed that the DHT:T ratio is correlated with PCa, as the DHT:T ratio is also correlated with ethnicity. Comparison with epidemiologic data showed that the correlation between the hormone levels and PCa is weaker than that between PRS and PCa, although there may be some correlation with the DHT to T ratio in PCa incidence. Interpolation with the linear regression of the Korean DHT:T ratio versus PCa incidence showed linearity, although the $R^2$ value (0.765) was low. We hypothesized that Vit D may be correlated with PCa, as the hazard ratio (HR) of PCa-specific mortality was reduced by 0.91 for every 20 nmol/L increase in circulating Vit D levels [7], and Vit D levels were also correlated with ethnicity [15]. We found a negative correlation between Vit D level and PCa incidence, although the $R^2$ value was very low compared to genetics; the Korean Vit D levels versus PCa mortality did not fit the model. We speculate that vitamin D levels may have a weaker correlation with PCa mortality, as the vitamin D level is related to various factors, such as intake, physical activity, sun exposure, and latitude. In addition, 50% of Koreans have Vit D deficiency compared to the global population [15]; thus, this may be the cause of the deviation from the fit model.

Therefore, genetics from the GWAS catalog related to PCa was the most contributing factor to PCa incidence and mortality. We found that the SNPs with the most significant difference among the populations are AR, JAZF1 (juxtaposed with another zinc finger protein 1), MAD1L1, CDKN1B, and SMAD2. Interestingly, the AR gene was the most central and important component in the network analysis. When observing the AR SNP (rs5919393) across populations, AFR men had the lowest frequency of T (14%), while EAS men had a higher frequency of T (100%) than EUR men (85%). Although this gene is intrinsic, this may lead to the following scenarios. First, testosterone is thought to be increased by the disabled
AR protein or downregulated AR protein production. An increase in testosterone levels may lead to a higher risk of PCa. Indeed, young African men have 15% higher testosterone levels than young European men [6,28]. However, epidemiological data showed that high testosterone levels do not increase the risk of PCa. Second, dihydrotestosterone (DHT), rather than testosterone, increases the incidence of PCa. Testosterone is converted to DHT by 5-α reductase in the prostate. Wu et al. reported that DHT levels are the highest in AFR men, intermediate in EUR men, and the lowest in EAS men, which may be attributed to the high activity of 5-α reductase [6]. However, there were no SNPs related to the 5-α reductase (SRD5A) genes reported in the GWAS catalog that are related to PCa. Upon further research, we found that the effects of SRD5As polymorphisms on PCa are controversial, with some studies reporting an effect [29–31], whereas others reported no effect [32–34]. Therefore, further studies are needed to verify the connection between SRD5As and PCa, which are beyond the scope of this study. In addition, analysis of the DHT:T ratios showed that they are not strongly correlated to PRS. Thus, we speculate that the AR SNP (rs5919393) does not affect the levels of DHT or T, and the AR gene itself may influence transcription, causing an increase in PCa. This is explained by the regulation of the transcriptome of PCa cells by AR signaling via modulation of global alternative splicing [35,36]. In fact, there may be other unknown mechanisms that play a role in the association between AR gene mutation and PCa, which may explain why PCa becomes resistant to hormone therapy.

A thorough comparison of network analysis of all the genes known for PCa and the results of our study revealed differences due to the difference in the PCa patient group and the normal population, as 1000 GP and KRGDB studies have been performed primarily on healthy subjects. In our network analysis, the most important genes were AR, CDKN1B, and MAD1L1, where AR and CDK1B have been previously identified as important in GWAS studies by Farashi et al. [37]. CDKN1B, a modulator of cyclin, is involved in genitalia differentiation and may be involved in the anomaly of the testis, and mutations in this gene are associated with multiple endocrine neoplasia type IV (a disease involved in reproductive organ tumors) [38]. However, Tsukasaki et al. reported that MAD1L1, a checkpoint gene whose dysfunction is associated with chromosomal instability, is also involved in neoplasms that affect the male reproductive system, which may account for the differences in PCa incidence and mortality between EAS and AFR men [39].

This study had some limitations. First, this analysis may have systemic bias and needs additional validation. The best scenario is having clinical data of the participants of the 1000 GP, such as PCa diagnosis status, Vit D levels, DHT:T ratio, and PCa-related mortality data. However, as the 1000 GP was conducted on a random and currently healthy population of diverse ethnicities, the limitation is that PRS of PCa cannot directly correlate to the PCa incidence or mortality for the individual. However, the strength is that the PRS score can be used to calculate the PCa incidence or mortality of the general ethnic population, and possibly for the individual if he has sequenced these genes. Second, the GWAS catalog contained data where the risk allele was not clearly defined according to the minor allele frequency (MAF). We did not exclude these from our study because most MAFs were likely to be risk alleles. Thus, the results of subgroup analysis may be inaccurate. To address this issue, risk allele curation is necessary for the GWAS catalog based on the results of additional large population studies with PCa patient cohorts. Third, SNPs pertaining to HRD genes were missing from the GWAS catalog as the patients enrolled in the 1000 GP were healthy; thus, those with active PCa need to be included for future research or studies need to be designed with clinical information on PCa with genetic data. Despite these limitations of not having definite, well-known pathogenic genes related to PCa, this does not change common variants according to ethnicity; indicating that AR, CDKN1B, and MAD1L1 are more highly associated with incidence and mortality in ethnic-dependent PCa.

5. Conclusions

From the public health perspective of PCa, the PRS consisting of PCa-related SNPs, including the AR SNP (rs5919393), may be suggested for high-risk populations to initiate
proactive screening and prophylactic treatment to reduce the incidence and mortality of PCa. Further in-depth research on the reported SNPs identified by the PRS model may shed more insights on targeted therapy to reduce PCa mortality.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes13112039/s1, Figure S1: Heatmap of all prostate cancer-related genes to single-nucleotide polymorphisms in the global population. Figure S2: Heatmap of all prostate cancer-related genes to single-nucleotide polymorphisms in the global population with the log-adjusted p-value of Fisher’s exact test in |AFR| + |EAS| > 60. Figure S3: (a) Correlation plot of dihydrotestosterone (DHT) to testosterone (T) ratio and prostate cancer (PCa) mortality. Linear regression is in the gray dashed line ($R^2 = 0.708$), and the real-world Korean incidence and DHT to T ratio is in the blue open rhombus. (b) Correlation plot of vitamin D concentration and prostate cancer (PCa) incidence. Linear regression is in the gray dashed line ($R^2 = 0.348$), and the real-world Korean PCa mortality and vitamin D concentration is the blue open rhombus, which deviates from the linear regression, whereas the estimated Korean mortality is in the blue solid circle. Figure S4: Network analysis of prostate cancer (PCa)-related single nucleotide polymorphisms (SNPs) from the genome-wide association studies catalog.; Table S1: Effect of allele frequencies of prostate cancer-related single-nucleotide polymorphisms among continental groups., Table S2: Effect of allele frequencies on prostate cancer-related single-nucleotide polymorphisms among continental group with Log_{10}-adjusted p-value of Fisher’s exact test in |AFR| + |EAS| > 60.

Author Contributions: Conceptualization, B.W.Y., H.-T.S. and J.H.S.; methodology, B.W.Y. and H.-T.S.; formal analysis, B.W.Y., H.-T.S. and J.H.S.; investigation, B.W.Y., H.-T.S. and J.H.S.; resources, H.-T.S.; data curation, B.W.Y., H.-T.S. and J.H.S.; writing—original draft preparation, B.W.Y., H.-T.S. and J.H.S.; writing—review and editing, B.W.Y., H.-T.S. and J.H.S.; visualization, B.W.Y.; supervision, H.-T.S. and J.H.S.; funding acquisition, J.H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Chung-Ang University Research Grants in 2022 and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. 2022R1C1C1002929).

Institutional Review Board Statement: This study was approved and monitored by the IRB of the Veterans Health Service Medical Center, Korea (IRB no. 2021-09-004).

Informed Consent Statement: Patient consent was waived due to retrospective data analysis and de-identify data, Institutional Review Board of Veterans Health Service Medical Center approved waiver of informed consent.

Data Availability Statement: The R code and datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. The allele frequency of the Korean reference genome database (KRGDB) is publicly available at [http://152.99.75.168:9090/KRGDBDN/dnKRGinput.jsp], and all three total merged sets of common variants, rare variants, and indels. The 1000 Genomes data is publicly available, all the files from the following folder were downloaded at ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502 (accessed on 15 January 2020). The genome-wide association study (GWAS) catalog data are available from the NHGRI-EBI website: https://www.ebi.ac.uk/gwas/docs/file-downloads, accessed on 20 April 2021. Phenolyzer is available at https://phenolyzer.wglab.org/, accessed on 1 December 2021.

Acknowledgments: This study was conducted with bioresources from the National Biobank of Korea, the Korea Disease Control and Prevention Agency, Republic of Korea (KBN-2019-053).

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer 2015, 136, E359–E386. [CrossRef] [PubMed]
2. Iyengar, S.; Hall, I.J.; Sabatino, S.A. Racial/Ethnic Disparities in Prostate Cancer Incidence, Distant Stage Diagnosis, and Mortality by U.S. Census Region and Age Group, 2012–2015. Cancer Epidemiol. Biomark. Prev. 2020, 29, 1357–1364. [CrossRef]
3. Kheirandish, P.; Chinegwundoh, F. Ethnic differences in prostate cancer. Br. J. Cancer 2011, 105, 481–485. [CrossRef] [PubMed]
4. Lloyd, T.; Hounsome, L.; Mehay, A.; Mee, S.; Verne, J.; Cooper, A. Lifetime risk of being diagnosed with, or dying from, prostate cancer by major ethnic group in England 2008–2010. *BMJ Med. 2015, 13, 171. [CrossRef] [PubMed]

5. Bashir, M.N. Epidemiology of Prostate Cancer. *Asian Pac. J. Cancer Prev. 2015, 16, 5137–5141. [CrossRef]

6. Wu, A.H.; Whittemore, A.S.; Kolonel, L.N.; John, E.M.; Gallagher, R.P.; West, D.W.; Hankin, J.; Teh, C.Z.; Dreon, D.M.; Paffenbarger, R.S., Jr. Serum androgens and sex hormone-binding globulins in relation to lifestyle factors in older African-American, white, and Asian men in the United States and Canada. *Cancer Epidemiol. Biomark. Prev. 1995, 4, 735–741.

7. Song, Z.-Y.; Yao, Q.; Zhuo, Z.; Ma, Z.; Chen, G. Circulating vitamin D level and mortality in prostate cancer patients: A dose–response meta-analysis. *Endocr. Connect. 2018, 7, R294–R303. [CrossRef]

8. van Schoor, N.; Lips, P. Global Overview of Vitamin D Status. *Endocrinol. Metab. Clin. N. Am. 2017, 46, 845–870. [CrossRef]

9. Schumacher, F.R.; Al Olama, A.A.; Berndt, S.I.; Benlloch, S.; Ahmed, M.; Saunders, E.J.; Dadaev, T.; Leongamornlert, D.; Anokian, E.; Cieza-Borrella, C.; et al. Association analyses of more than 140,000 men identify 65 new prostate cancer susceptibility loci. *Nat. Genet. 2018, 50, 928–936. [CrossRef]

10. Chau, V.; Madan, R.A.; Figg, W.D. Exploiting defects in homologous recombination repair for metastatic, castration-resistant prostate cancer. *Cancer Biol. Ther. 2020, 21, 884–887. [CrossRef]

11. Lofan, T.L.; Kaur, H.B.; Salles, D.C.; Murali, S.; Schaeffer, E.M.; Lanchbury, J.S.; Isaacs, W.B.; Brown, R.; Richardson, A.L.; Cussenot, O.; et al. Homologous recombination deficiency (HRD) score in germline BRCA2- versus ATM-altered prostate cancer. *Mod. Pathol. 2021, 34, 1185–1193. [CrossRef]

12. Pritchard, C.C.; Mateo, J.; Walsh, M.F.; De Sarkar, N.; Abida, W.; Beltman, H.; Garofalo, A.; Gulati, R.; Carreira, S.; Eeles, R.; et al. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N. Engl. J. Med. 2016, 375, 443–453. [CrossRef] [PubMed]

13. Rantapero, T.; Wahlfors, T.; Kähler, A.; Huitlman, C.; Lindberg, J.; Tammela, T.L.J.; Nykter, M.; Schleutker, J.; Wiklund, F. Inherited DNA Repair Gene Mutations in Men with Lethal Prostate Cancer. *Genes 2020, 11, 314. [CrossRef] [PubMed]

14. Pinese, M.; Lacaze, P.; Rath, E.M.; Stone, A.; Brion, M.-J.; Ameur, A.; Nagpal, S.; Puttick, C.; Husson, S.; Degrave, D.; et al. The Medical Genome Reference Bank contains whole genome and phenotype data of 2570 healthy elderly. *Nat. Commun. 2020, 11, 1–14. [CrossRef] [PubMed]

15. Shin, H.-T.; Yoon, B.W.; Seo, J.H. Comparison of risk allele frequencies of single nucleotide polymorphisms associated with age-related macular degeneration in different ethnic groups. *BMC Ophthalmol. 2021, 21, 1–14. [CrossRef] [PubMed]

16. Shin, H.-T.; Yoon, B.W.; Seo, J.H. Analysis of risk allele frequencies of single nucleotide polymorphisms related to open-angle glaucoma in different ethnic groups. *BMC Med. Genom. 2021, 14, 1–18. [CrossRef] [PubMed]

17. Mao, L.; Fang, Y.; Campbell, M.; Southerland, W.M. Population differentiation in allele frequencies of obesity-associated SNPs. *BMC Genom. 2017, 1, 1–16. [CrossRef]

18. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet. 2013, 9, e1003348. [CrossRef]

19. Auton, A.; Saldin, T. The 1000 genomes project. In *Assessing Rare Variation in Complex Traits*; Springer: New York, NY, USA, 2015; pp. 71–85.

20. Han, H.H.; Park, J.W.; Na, J.C.; Chung, B.H.; Kim, C.-S.; Ko, W.J. Epidemiology of prostate cancer in South Korea. *Prostate Int. 2015, 3, 99–102. [CrossRef]

21. Litman, H.J.; Bhasin, S.; Link, C.L.; Araujo, A.B.; McKinlay, J.B. Serum Androgen Levels in Black, Hispanic, and White Men. *J. Clin. Endocrinol. Metabol. 2006, 91, 4326–4334. [CrossRef]

22. Yoon, Y.-D.; Lee, C.-J.; Chun, E.-H.; Lee, J.-Y. Concentrations of Bioavailable Testosterone and Dihydrotestosterone Determined by Luminescence Immunoassay in Serum. *Clin. Exp. Reprod. Med. 1988, 15, 83–92.

23. Yoon, B.-W.; Shin, H.-T.; Seo, J. Risk Allele Frequency Analysis of Single-Nucleotide Polymorphisms for Vitamin D Concentrations in Different Ethnic Group. *Genes 2021, 12, 1530. [CrossRef] [PubMed]

24. Yang, H.; Robinson, P.N.; Wang, K. Phenozyler: Phenotype-based prioritization of candidate genes for human diseases. *Nat. Methods 2015, 12, 841–843. [CrossRef]

25. Conti, D.V.; Darst, B.F.; Moss, L.C.; Saunders, E.J.; Sheng, X.; Chou, A.; Schumacher, F.R.; Al Olama, A.A.; Benlloch, S.; Dadaev, T.; et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat. Genet. 2021, 53, 65–75. [CrossRef] [PubMed]

26. Vieira, G.M.; Gellen, L.P.A.; Leal, D.F.D.V.B.; Pastana, L.F.; Vinagre, L.W.M.S.; Aquino, V.T.; Fernandes, M.R.; de Assumpção, P.P.; Burbano, R.M.R.; dos Santos, S.E.B.; et al. Correlation between Genomic Variants and Worldwide Epidemiology of Prostate Cancer. *Genes 2022, 13, 1039. [CrossRef]

27. Andriole, G.L.; Bostwick, D.G.; Brawley, O.W.; Gemella, L.G.; Marberger, M.; Montorsi, F.; Pettaway, C.A.; Tammela, T.L.J.; Teloken, C.; Tindall, D.J.; et al. Effect of Dutasteride on the Risk of Prostate Cancer. *N. Engl. J. Med. 2010, 362, 1192–1202. [CrossRef] [PubMed]

28. Ross, R.; Bernstein, L.; Pike, M.; Henderson, B.; Lobo, R.; Stanczyk, F.; Shimizu, H. 5-alpha-reductase activity and risk of prostate cancer among Japanese and US white and black males. *Lancet 1992, 339, 887–889. [CrossRef]

29. Lévesque, E.; Laverdière, J.; Lacombe, L.; Caron, P.; Rouleau, M.; Turcotte, V.; Têtu, B.; Fradet, Y.; Guillemette, C. Importance of 5-alpha-reductase gene polymorphisms on circulating and intraprostatic androgens in prostate cancer. *Clin. Cancer Res. 2014, 20, 576–584. [CrossRef]
30. Hsing, A.W.; Chen, C.; Chokkalingam, A.P.; Gao, Y.T.; Dightman, A.D.; Nguyen, H.T.; Deng, J.; Cheng, J.; Sesterhenn, A.I.; Mostofi, F.K.; et al. Polymorphic markers in the SRD5A2 gene and prostate cancer risk: A population-based case-control study. *Cancer Epidemiol. Biomark. Prev.* 2001, 10, 1077–1082.

31. Shiota, M.; Fujimoto, N.; Yokomizo, A.; Takeuchi, A.; Itsumi, M.; Inokuchi, J.; Tatsugami, K.; Uchiumi, T.; Naito, S. SRD5A gene polymorphism in Japanese men predicts prognosis of metastatic prostate cancer with androgen-deprivation therapy. *Eur. J. Cancer* 2015, 51, 1962–1969. [CrossRef]

32. Li, J.; Coates, R.J.; Gwinn, M.; Khoury, M.J. Steroid 5-{alpha}-reductase Type 2 (SRD5a2) gene polymorphisms and risk of prostate cancer: A HuGE review. *Am. J. Epidemiol.* 2010, 171, 1–13. [CrossRef] [PubMed]

33. Choi, S.Y.; Kim, H.J.; Cheong, H.S.; Myung, S.C. The association of 5-alpha reductase type 2 (SRD5A2) gene polymorphisms with prostate cancer in a Korean population. *Korean J. Urol.* 2015, 56, 19–30. [CrossRef] [PubMed]

34. Choubey, V.K.; Sankhwar, S.N.; Carlus, S.J.; Singh, A.N.; Dalela, D.; Thangaraj, K.; Rajender, S. SRD5A2 Gene Polymorphisms and the Risk of Benign Prostatic Hyperplasia but not Prostate Cancer. *Asian Pac. J. Cancer Prev.* 2015, 16, 1033–1036. [CrossRef] [PubMed]

35. Shah, K.; Gagliano, T.; Garland, L.; O’Hanlon, T.; Bortolotti, D.; Gentili, V.; Rizzo, R.; Giamas, G.; Dean, M. Androgen receptor signaling regulates the transcriptome of prostate cancer cells by modulating global alternative splicing. *Oncogene* 2020, 39, 6172–6189. [CrossRef] [PubMed]

36. Fujita, K.; Nonomura, N. Role of Androgen Receptor in Prostate Cancer: A Review. *World J. Men's Health* 2019, 37, 288–295. [CrossRef]

37. Farashi, S.; Kryza, T.; Clements, J.; Batra, J. Post-GWAS in prostate cancer: From genetic association to biological contribution. *Nat. Cancer* 2019, 19, 46–59. [CrossRef]

38. Alrezk, R.; Hannah-Shmouni, F.; Stratakis, C.A. MEN4 and CDKN1B mutations: The latest of the MEN syndromes. *Endocr. -Relat. Cancer* 2017, 24, T195–T208. [CrossRef]

39. Tsukasaki, K.; Miller, C.W.; Greenspun, E.; Eshaghian, S.; Kawabata, H.; Fujimoto, T.; Tomonaga, M.; Sawyers, C.; Said, J.W.; Koeffler, H.P. Mutations in the mitotic check point gene, MAD1L1, in human cancers. *Oncogene* 2001, 20, 3301–3305. [CrossRef]