NEOPLASIA
www.neoplasia.com

Volume 15 Number 4 April 2013 pp. 448–453 448

Genome-wide Association Study Identifies Genetic Determinants of Urine PCA3 Levels in Men[1,2]

Zhuo Chen[*,†,3], Jielin Sun[*,†,3], Seong-Tae Kim[*,†], Jack Groskopf[‡], Junjie Feng[*,†], William B. Isaacs[§], Roger S. Rittmaster[¶], Lynn D. Condreay[#], Siqun Lilly Zheng[*,†] and Jianfeng Xu[*,†]

*Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, NC; [†]Departments of Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston-Salem, NC; [‡]Gen-Probe Incorporated, San Diego, CA; [§]Johns Hopkins Medical Institutions, Baltimore, MD; [¶]Oncology, GlaxoSmithKline, Research Triangle Park, NC; [#]Genetics Development, GlaxoSmithKline, Research Triangle Park, NC

Abstract

Prostate cancer gene 3 (*PCA3*) is a non-coding gene specifically overexpressed in prostate cancer (PCa) that has great potential as a clinical biomarker for predicting prostate biopsy outcome. However, genetic determinants of PCA3 expression level remain unknown. To investigate the association between genetic variants and PCA3 mRNA level, a genome-wide association study was conducted in 1371 men of European descent in the REduction by DUtasteride of prostate Cancer Events trial. First-voided urine specimens containing prostate cells were obtained after digital rectal examination. The PROGENSA PCA3 assay was used to determine PCA3 score in the urinary samples. A linear regression model was used to detect the associations between (single nucleotide polymorphisms) SNPs and PCA3 score under an additive genetic model, adjusting for age and population stratification. Two SNPs, rs10993994 in *β-microseminoprotein* at 10q11.23 and rs10424878 in *kallikrein-related peptidase 2* at 19q13.33, were associated with PCA3 score at genome-wide significance level ($P = 1.22 \times 10^{-9}$ and 1.06×10^{-8}, respectively). Men carrying the rs10993994 "T" allele or rs10424878 "A" allele had higher PCA3 score compared with men carrying rs10993994 "C" allele or rs10424878 "G" allele ($\beta = 1.25$ and 1.24, respectively). This is the first comprehensive search for genetic determinants of PCA3 score. The novel loci identified may provide insight into the molecular mechanisms of PCA3 expression as a potential marker of PCa.

Neoplasia (2013) 15, 448–453

Abbreviations: GWAS, genome-wide association study; REDUCE, REduction by DUtasteride of prostate Cancer Events; PCa, prostate cancer; PSA, prostate-specific antigen; *PCA3*, *prostate cancer gene 3*; MSMB, β-microseminoprotein; KLK2, kallikrein-related peptidase 2; KLK3, kallikrein-related peptidase 3; IPSS, International Prostate Symptom Score

Address all correspondence to: Dr Jianfeng Xu or Dr Siqun Lilly Zheng, Center for Cancer Genomics, Medical Center Blvd, Winston-Salem, NC 27157. E-mail: jxu@wakehealth.edu, szheng@wakehealth.edu

[1]This work was supported by a National Cancer Institute RC2 grant (CA148463) and a research contract by GlaxoSmithKline (GSK) to J.X. L.D.C. and R.S.R. were GSK employees during this study and hold stock in GSK. J.G. is an employee of Hologic Gen-Probe, which supported PROGENSA PCA3 Assay testing under contractual agreement with GSK. J.X. certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (e.g., employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), have been disclosed.

[2]This article refers to supplementary materials, which are designated by Tables W1 to W3 and Figures W1 to W3 and are available online at www.neoplasia.com.

[3]These authors contributed equally.

Received 19 December 2012; Revised 6 February 2013; Accepted 7 February 2013

DOI 10.1593/neo.122144

Introduction

Prostate cancer (PCa) is the most prevalent noncutaneous malignancy and the second leading cause of cancer mortality in the United States [1]. Prostate-specific antigen (PSA) is the most commonly used screening tool for PCa. PSA screening has led to progress in early detection [2]. However, elevated serum PSA is regularly detected in benign prostate disorders such as prostatitis and benign prostatic hyperplasia (BPH), resulting in a relatively high rate of unnecessary PSA-triggered biopsies [3].

Compared to PSA, *prostate cancer gene 3* (*PCA3*) is more specific for PCa. PCA3 is a prostate-specific non-coding mRNA that is overexpressed in prostate tumor tissue and metastases and is not observed in conjunction with benign prostate disorders such as prostatitis or enlarged prostate [4]. Using PCA3, prostate tissue specimens containing less than 10% of cancer cells can be accurately discriminated from noncancer tissue [4]. PCA3 score is generated from the ratio of PCA3 mRNA level to PSA mRNA level, through which *PCA3* expression is normalized with the PSA expression used as a housekeeping gene. Recently, several independent studies have shown that PCA3 score has better performance than PSA in predicting prostate biopsy outcome in subjects with elevated PSA (>2.5 ng/ml) [5–8]. PCA3 score was also observed to be indicative of biopsy Gleason score in the REduction by DUtasteride of prostate Cancer Events (REDUCE) trial [9] and another study in a multidescent population by Nakanishi et al. [10]. Moreover, PCA3 was correlated with prostate tumor size [10].

Despite the well-characterized clinical use of PCA3, our knowledge of the genetic determinants of PCA3 expression levels is negligible. In this study, we conducted a genome-wide association study (GWAS) to identify SNPs that are associated with the urine level of PCA3 in the REDUCE trial.

Materials and Methods

Study Population

Subjects in our GWAS were originally enrolled in the REDUCE trial, a 4-year multicenter, randomized, double-blind, placebo-controlled clinical study to evaluate the effect of dutasteride on the risk of PCa in men. Individuals participating in this trial had to meet the following criteria: 1) age between 50 and 75 years; 2) prostate volume (PV) $\leq$80 cm^3; 3) PSA level of 2.5 to 10 ng/ml for subjects <60 years old and 3 to 10 ng/ml for those $\geq$60 years; 4) negative biopsy (within 6 months of enrollment) for PCa, high-grade prostatic intraepithelial neoplasia, or atypical small acinar proliferation [11].

In the REDUCE trial, a total of 1649 individuals of European descent (885 in the placebo and 764 in the dutasteride treatment arms) had available PCA3 measurement. Among them, 162 of 885 subjects (18.3%) in the placebo and 116 of the 764 subjects (15.2%) in the dutasteride arm were diagnosed with PCa in the 4-year follow-up. The remaining 1371 individuals (723 in the placebo arm and 648 in the dutasteride arm) with negative prostate biopsy were included for further GWAS analysis. Characteristics of the 1371 individuals are summarized in Table W1.

Measurement of PCA3 and PSA mRNA Levels

First-voided urine specimens containing prostate cells were obtained after digital rectal examination before year 2 and 4 biopsies. The PROGENSA PCA3 Assay was used to measure PCA3 and PSA mRNA levels in the urinary samples at an independent laboratory (Caris Life Sciences/Molecular Profiling Institute, Phoenix, AZ) [12]. The first available PCA3 and PSA value (at year 2 or year 4) was used in our analysis.

Genotyping, Imputation, and Quality Control

Genotyping was conducted using an Illumina HumanOmniExpress BeadChip system (729,755; single nucleotide polymorphisms [SNPs] included) at the Center for Cancer Genomics, Wake Forest University School of Medicine. Samples with a genome-wide call rate $\geq$95% were included in subsequent GWAS analysis. For SNPs, the quality control criteria were as follows: minor allele frequency >0.01, genotype call rate >95%, and Hardy-Weinberg Equilibrium >0.001. The remaining 586,462 SNPs were used to impute the genotypes of SNPs that were not directly genotyped in the genome, by IMPUTE software [13] using combined data of the 1000 Genomes low-coverage pilot project and HapMap3 data as references (www.1000genomes.org and www.hapmap.org). The same quality control and a posterior probability of >90% was applied to imputed SNP calls.

Personalized PCA3 Score and Cutoffs

Personalized PCA3 score and cutoff were calculated for each individual. First, a linear regression model was fitted, where log-transformed PCA3 scores were treated as outcomes and the number of alleles associated with higher PCA3 scores was treated as covariates. Genetic effects were calculated for three genotypes (aa, Aa, and AA), using the fitted values. Relative allelic effects (percentage of increase per allele) were calculated by dividing the fitted values of aa and Aa. The relative risk to the population for each genotype was then computed on the basis of the relative allelic effect and genotypic frequency. Second, assuming a multiplicative model, the combined

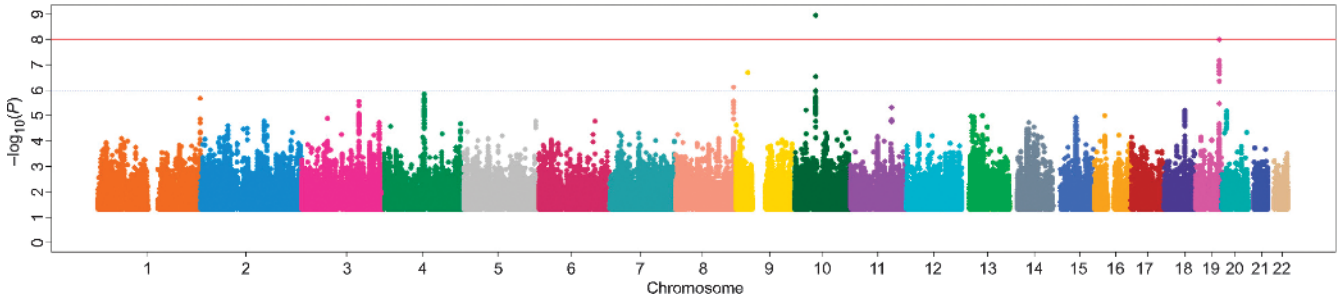

Figure 1. Manhattan plot of genome-wide association analyses for PCA3 score in 1371 subjects with negative prostate biopsy. The *x*-axis represents chromosomal position; the *y*-axis represents $-\log_{10} P$ values from linear regression. The red horizontal solid line indicates the preset threshold of $P = 1 \times 10^{-8}$.

Table 1. SNPs Associated with PCA3 Score in 1371 Subjects with Negative Prostate Biopsy.

SNP	Chr	BP*	Gene	Alleles	Risk Allele		Mean of PCA3 Score†			β (SE)‡	P‡
					Allele	Freq	LL	LH	HH		
rs10993994	10	51,219,502	*MSMB*	C/T	T	0.41	13.35	16.82	20.11	1.25 (0.037)	1.22×10^{-9}
rs10424878	19	56,066,779	*KLK2*	G/A	A	0.40	13.68	16.61	20.76	1.24 (0.038)	1.06×10^{-8}

*BP genomic position is based on NCBI build 36.
†The mean values of PCA3 score are back log-transformed. LL indicates homozygous carriers of the allele that is associated with a lower value of PCA3 score; LH indicates heterozygous carrier of the allele that is associated with a lower value of PCA3 score; HH indicates homozygous carriers of the allele that is associated with a higher value of PCA3 score.
‡The *P* values are based on linear regression analysis on the log-transformed PCA3 score and adjusted for age and the top eigenvector assuming an additive model; β is back log-transformed, while SE is log-transformed.

relative genetic effect was calculated by multiplying the relative genotypic effect for each SNP relative to the general population. Third, a personalized PCA3 score for each subject was determined by dividing the raw PCA3 score by the combined relative genetic effect, while the personalized cutoff for each subject was generated by multiplying a uniform cutoff of 35 by the combined relative genetic effect.

Statistical Analysis

We calculated PCA3 score as follows: PCA3 mRNA copy number/PSA mRNA copy number × 1000. PSA mRNA copy number here was used to correct for the number of prostate cells present in urine sample. PCA3 score was log transformed to approximate a normal distribution for further analysis (Figure W1).

A linear model implemented in PLINK software package [14] was used to detect the association of each SNP (genotyped and imputed) with PCA3 score under an additive genetic model in 1371 subjects negative for PCa biopsy. Principal component analysis was conducted to detect potential population stratification by EIGENSTRAT software [15]. Age and the top eigenvector that indicates ancestral heterogeneity within a group of individuals were adjusted as covariates in multiple regression analysis. Genome-wide significance level was set at a *P* value of 1×10^{-8}. For regions containing more than one significant variant ($P < 10^{-6}$), multiple linear regression analysis was applied to test the independence of each SNP, adjusting for the most significant SNPs, as well as all other covariates.

Results

Urine mRNA levels of PCA3, PSA (copies per milliliter, expressed as PCA3CPM and PSACPM, respectively), as well as the resulting PCA3 scores are summarized in Table W1. Although the median values of PCA3CPM and PSACPM in the dutasteride arm differed from those in the placebo arm ($P < .0001$), PCA3 scores were not significantly different between dutasteride and placebo arms (Wilcoxon rank sum test, $P = .72$).

In the present GWAS, a total of 3,076,666 SNPs (genotyped and imputed) were analyzed for association with log-transformed PCA3 score in 1371 subjects (placebo and dutasteride arms; Figure 1). The inflation factor of GWAS results based on a quantile-quantile (Q-Q) plot analysis was modest ($\lambda = 1.01$; Figure W2). Therefore, the reported *P* values are not corrected for genomic inflation.

Two SNPs, rs10993994 in the *β-microseminoprotein* (*MSMB*) gene at 10q11.23 and rs10424878 in the *kallikrein-related peptidase 2* (*KLK2*) gene at 19q13.33, reached a genome-wide significant *P* value of 1.22×10^{-9} and 1.06×10^{-8}, respectively (Table 1), adjusting for age and the top eigenvector for potential population stratification in an additive model. Seventeen SNPs with a *P* value smaller than 1×10^{-6} are presented in Table W2. Several other SNPs in the *MSMB* gene region or *KLK2/kallikrein-related peptidase 3* (*KLK3*) gene cluster were significant in the *P* value range of 1×10^{-6} to 1×10^{-8} (Figure 2, *A* and *B*). After adjusting for the most significant SNPs and all other covariates, none of the other SNPs in the *MSMB* or *KLK2/KLK3* regions remained independently associated with PCA3 score. No significant association between PCa risk and the 183 SNPs within or surrounding the *PCA3* gene at 9q21 (78,519,174-78,642,285, NCBI build 36) was detected.

The effects of rs10993994 and rs10424878 on PCA3 score were then evaluated. As shown in Table 1, the least square mean values of PCA3 score were 13.35, 16.82, and 20.11 for the rs10993994 CC, CT, and TT carriers, respectively ($\beta = 1.25$ for "T" allele). The least square mean values of PCA3 score were 13.68, 16.61, and 20.76 for rs10424878 GG, AG, and AA carriers, respectively ($\beta = 1.24$ for "A" allele).

In addition, we examined the associations of the two top SNPs with PCa risk, baseline PV, and International Prostate Symptom Score (IPSS; Table W3). The SNP rs10993994 was significantly associated with PCa risk (odds ratio [OR] = 1.20, $P = .02$), baseline PV ($\beta = 0.03$, $P = .01$), and IPSS ($\beta = -0.27$, $P = .05$). The SNP rs10424878 was not significantly associated with PCa risk (OR = 1.08, $P = .35$) or baseline IPSS ($\beta = 0.04$, $P = .76$) but was associated with baseline PV ($\beta = -0.02$, $P = .02$).

Moreover, a personalized cutoff value of PCA3 score was estimated for each individual. The estimated personalized cutoff ranged from 24.9 to 60.6. The estimated personalized cutoff was 24.9 for subjects carrying both the rs10993994 CC genotype and the rs10424878 GG genotype, while it was 60.6 for subjects carrying both the rs10993994 TT genotype and the rs10424878 AA genotype (Figure W3).

Discussion

In the present GWAS, conducted among 1371 European descent men with negative prostate biopsies, the strongest overall effects on PCA3 score were observed for two independent genetic variants, rs10993994 in *MSMB* at 10q11.21 ($P = 1.22 \times 10^{-9}$) and rs10424878 in *KLK2* at 19q13.33 ($P = 1.06 \times 10^{-8}$). Presently, this is the first discovery of genetic determinants based on a genome-wide study for PCA3, one of the most valuable early detection biomarkers for PCa.

SNP rs10993994 is located 57 bp upstream of exon 1 of the *MSMB* gene, which encodes MSMB, one of the major constituents of seminal plasma [16]. MSMB is a member of immunoglobulin binding factor family and can act as a tumor suppressor by binding to cell surface receptors and regulating prostate cell apoptosis through mitogen-activated protein (MAP) kinase/AKT signaling [17]. Previous studies discovered that the expression of *MSMB* is much higher in normal and benign prostate tissues than in tumor tissues [18] and urinary MSMB level decreases during tumorigenesis [19].

As the most thoroughly studied genetic variant in the *MSMB* gene region, rs10993994 has been reported to be associated with PCa [20]

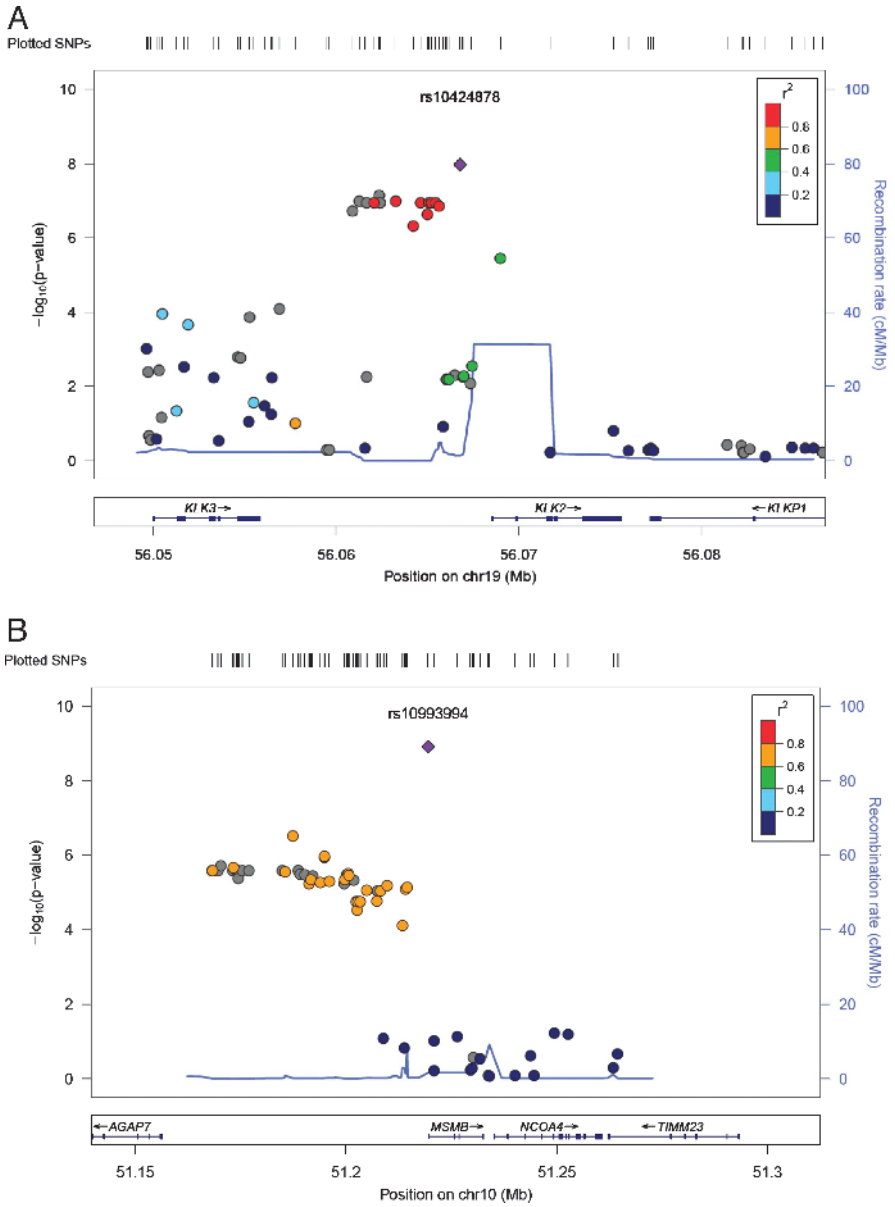

Figure 2. Association of PCA3 score at 10q11.23 (A) and 19q13.33 (B) in 1371 subjects with negative prostate biopsy. The *x*-axis represents chromosomal position; the *y*-axis represents $-\log_{10} P$ values from linear regression. Solid purple lines represent recombination rate in cM/MB. Each plotted point represents one SNP.

and this association appears to be independent of age at diagnosis or tumor grade [21]. Functional investigation revealed that the minor allele "T" of rs10993994 was causally associated with lower expression of *MSMB*, because replacement of wild-type C to T altered a cAMP response element–binding protein transcription binding site, resulting in reduced promoter activity [22]. Fine mapping of the *MSMB* promoter region did not identify other SNPs having major effects regulating MSMB expression [22]. Taken together, these findings suggest rs10993994 as a potential independent causal genetic variant for PCa.

In our study, rs10993994 in *MSMB* was associated with PCA3 score, a risk factor for PCa. The T allele, associated with a decrease in MSMB promoter activity and an increase in PCa risk, was associated with an increase in urinary PCA3. Since rs10993994 was associated with PCa risk in the placebo arm of REDUCE trial (OR = 1.20, P = .02), one potential scenario is, even in this population of men screened for PCa, men with the T allele would be more likely to have (undetected) PCa and thus have a higher urinary PCA3. However, if this was the explanation, one would expect other PCa risk SNPs (e.g., SNPs at 8q24 and *HNF1B*) to be associated with urinary PCA3. Except for rs10993994, the majority of the other 32 known PCa risk–associated SNPs [23] were not significantly associated with PCA3 score ($P > .0016$, Bonferroni corrected for 32 tests), and only two of these documented PCa risk SNPs, rs10486567 and rs1512268, have P values slightly smaller than .0016 (.0005 and .0010, respectively). Therefore, it is unlikely that rs10993994 affects PCA3 score through this mechanism.

Given that rs10993994 in MSMB is also associated with serum PSA levels, we also considered the possibility that rs10993994 affects PCA3 score indirectly through PSA. Additional association analysis of PCA3 score in which urinary PSA level was adjusted as a covariate showed that urinary PSA level had little effect on PCA3 score (P = .84). Moreover, rs10993994 is not statistically related to serum PSA

level among the 1371 subjects in our study ($P = .37$, adjusted for age and the top eigenvector). Therefore, the association between rs10993994 and PCA3 score found in our study is most likely to be independent of PSA level in urine or in serum.

Since rs10993994 is located within an androgen receptor binding region [24] and androgen signaling plays an important role in the pathogenesis of PCa [25], it is possible that rs10993994 has an effect on PCA3 score through the androgen action pathway. In the present study, 648 of the total 1371 subjects were in the dutasteride arm. Dutasteride is a 5α-reductase inhibitor that reduces systemic and tissue levels of 5α-dihydrotestosterone, the principal androgen in the prostate [25]. Therefore, we compared the effects of rs10993994 on PCA3 score in the placebo and dutasteride arms separately. The β values of the T allele in an additive model (adjusted for age and the top eigenvector) in 723 placebo subjects and 648 dutasteride subjects were 1.27 and 1.24, respectively, with P values of 2.5×10^{-6} and 6.7×10^{-5}. The effects of rs10993994 on PCA3 score were quite similar in the two arms. On the basis of these results, rs10993994 may not affect PCA3 score through the androgen pathway.

There is still another scenario that may explain the association between rs10993994 and PCA3 score. There may be some currently unknown downstream relationship whereby the level of MSMB expression directly affects the production of PCA3 mRNA or its ability to get into the urine. However, *in vitro* studies with PCA3-producing cell lines and other functional studies are required to address this hypothesis.

The second SNP associated with PCA3 score, rs10424878, is about 2 kb upstream of *KLK2*, the coding gene of KLK2. As a member of kallikrein gene family, KLK2 is a serine protease and can activate urokinase-type plasminogen activator, cleave binding proteins of insulin-like growth factor, and cleave the preform of PSA to the mature enzyme, participating in a proteolytic cascade, which has been associated with tumorigenesis and metastasis of PCa [26]. In addition, PCa tumor tissues have a higher KLK2 expression level than benign tissues [27,28] and patients with PCa were also found to have higher serum KLK2 [29,30].

Genetic variants in *KLK2* were found to be associated with PCa in candidate gene association studies [31], although no SNPs in *KLK2* have been confirmed or identified by GWAS. In our GWAS on PCA3 score, rs10424878 and another nine SNPs in *KLK2* and five SNPs in *KLK3* reached a P value of 1×10^{-6}. However, the associations between all of the other 14 *KLK2*/*KLK3* SNPs and PCA3 score no longer existed after adjusting for rs10424878 ($P > .05$), indicating that rs10424878 is the only independent SNP in *KLK*s that is associated with PCA3 score. Unfortunately, little is known about the pathways or functional relationships between the *KLK2* and *PCA3* gene. Additional association analysis of PCA3 score in which urinary PSA level was adjusted as a covariate showed that urinary PSA level had little effect on PCA3 score ($P = .89$). We also assessed serum PSA level with rs10424878, and no association was found ($P = .12$, adjusted for age and the top eigenvector). Overall, the effect of rs10424878 on PCA3 scores was unlikely to be mediated through serum PSA levels.

Another possible impact of rs10424878/*KLK2* on PCA3 score could be through the androgen action pathway, since *KLK2* is highly responsive to androgens. As mentioned, androgen signaling is important for PCa pathogenesis [25] and dutasteride could greatly reduce 5α-dihydrotestosterone levels [25]. Similar to rs10993994, we assessed the effect of rs10424878 on PCA3 score separately in the placebo and dutasteride arms. The β values of "A" allele in an additive model (adjusted for age and the top eigenvector) in 723 placebo subjects only and 648 dutasteride subjects only are 1.22 and 1.26, respectively, with P values of 2.06×10^{-4} and 1.13×10^{-5}. The effects of rs10424878 on PCA3 score were quite similar in the two arms. Therefore, it is unlikely that rs10424878 affects PCA3 score through the androgen pathway. Additional biologic studies are needed to elucidate the mechanism through which rs10424878 impacts urine PCA3 scores.

On the basis of genetic discovery, we estimated personalized PCA3 scores for 1649 subjects with PCA3 measurements, including 278 subjects with positive prostate biopsy results and 1371 with negative biopsy results. We compared the area under curve (AUC) between the model with the raw PCA3 score and the model with the personalized PCA3 score for predicting prostate biopsy outcome. The AUCs for both models were 0.681 and 0.682, respectively. We did not observe a statistically significant difference in AUCs between the two models, which may due to the limited number of genetic determinants identified for the PCA3 score and the relatively small effect of each variant. Although the personalized PCA3 scores showed limited improvement in AUC compared with the raw PCA3 scores, the detection rate of positive prostate biopsy was higher for men with 35 and higher for the personalized PCA3 score (31.0%, 120/388) compared with the detection rate for men with 35 and higher for the raw PCA3 score (28.3%, 119/421). Overall, improvement in clinical use of the personalized PCA3 score was limited in the current study. However, it may improve if more genetic variants for PCA3 score are identified in future studies.

Two advantages of this study were its prospective design and the stringent inclusion and exclusion criteria. Nevertheless, several limitations need to be addressed. Although the most stringent Bonferroni correction was applied to control false positive findings, our study did not include a confirmation stage in an independent population, due to the difficulty in identifying additional studies with available germline DNA, PCA3 measurements, and prostate biopsy outcomes. Another limitation is that only two SNPs were identified to be associated with PCA3 score at a genome-wide level in this study. Therefore, further discoveries and validations across populations are warranted.

In conclusion, this study represents the first effort to investigate genetic determinants of PCA3 score, which may advance the understanding of PCA3 score. More importantly, the genetic determinants identified might be helpful to facilitate the utilization of personalized cutoff of PCA3 score as a marker for PCa prediction.

Acknowledgments

We thank the patients enrolled in REDUCE (sponsored by GSK) who provided consent and genetic samples that enabled this study, and the clinicians who contributed their expertise in recruiting study patients for the REDUCE clinical study. Dave Pulford, Jennifer Aponte, Jon Charnecki, and Mary Ellyn Volk participated in consent reconciliation and sample management to enable genetic sample selection for inclusion and genotype determination. Karen King provided data management support for this project. We appreciate the assistance of Lauren Marmor in coordinating the support of the Avodart Collaborative Research Team.

References

[1] Jemal A, Siegel R, Ward E, Hao Y, Xu J, and Thun MJ (2009). Cancer statistics, 2009. *CA Cancer J Clin* **59**, 225–249.

[2] Welch HG and Albertsen PC (2009). Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986–2005. *J Natl Cancer Inst* **101**, 1325–1329.

[3] Pannek J and Partin AW (1997). Prostate-specific antigen: what's new in 1997. *Oncology (Williston Park)* **11**, 1273–1278; discussion 1279-1282.

[4] Hessels D, Klein Gunnewiek JM, van Oort I, Karthaus HF, van Leenders GJ, van Balken B, Kiemeney LA, Witjes JA, and Schalken JA (2003). DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *Eur Urol* **44**, 8–15; discussion 15-16.

[5] Aubin SM, Reid J, Sarno MJ, Blase A, Aussie J, Rittenhouse H, Rittmaster R, Andriole GL, and Groskopf J (2010). PCA3 molecular urine test for predicting repeat prostate biopsy outcome in populations at risk: validation in the placebo arm of the dutasteride REDUCE trial. *J Urol* **184**, 1947–1952.

[6] Ochiai A, Okihara K, Kamoi K, Iwata T, Kawauchi A, Miki T, and Fors Z (2011). Prostate cancer gene 3 urine assay for prostate cancer in Japanese men undergoing prostate biopsy. *Int J Urol* **18**, 200–205.

[7] Roobol MJ, Schroder FH, van Leeuwen P, Wolters T, van den Bergh RC, van Leenders GJ, and Hessels D (2010). Performance of the prostate cancer antigen 3 (PCA3) gene and prostate-specific antigen in prescreened men: exploring the value of PCA3 for a first-line diagnostic test. *Eur Urol* **58**, 475–481.

[8] Crawford ED, Rove KO, Trabulsi EJ, Qian J, Drewnowska KP, Kaminetsky JC, Huisman TK, Bilowus ML, Freedman SJ, Glover WL Jr, et al. (2012). Diagnostic performance of PCA3 to detect prostate cancer in men with increased prostate specific antigen: a prospective study of 1,962 cases. *J Urol* **188**, 1726–1731.

[9] Rittmaster R, Aubin SMJ, Reid J, Sarno MJ, Blase A, Aussie J, Rittenhouse H, Andriole GL, and Groskopf J (2010). Validation of the PCA3 molecular urine test for prediction repeat prostate biopsy outcome in the placebo arm of the dutasterided REDUCE trial. *Eur Urol* **184**, 1947–1952.

[10] Nakanishi H, Groskopf J, Fritsche HA, Bhadkamkar V, Blase A, Kumar SV, Davis JW, Troncoso P, Rittenhouse H, and Babaian RJ (2008). PCA3 molecular urine assay correlates with prostate cancer tumor volume: implication in selecting candidates for active surveillance. *J Urol* **179**, 1804–1809; discussion 1809–1810.

[11] Andriole G, Bostwick D, Brawley O, Gomella L, Marberger M, Tindall D, Breed S, Somerville M, and Rittmaster R (2004). Chemoprevention of prostate cancer in men at high risk: rationale and design of the reduction by dutasteride of prostate cancer events (REDUCE) trial. *J Urol* **172**, 1314–1317.

[12] Groskopf J, Aubin SM, Deras IL, Blase A, Bodrug S, Clark C, Brentano S, Mathis J, Pham J, Meyer T, et al. (2006). APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. *Clin Chem* **52**, 1089–1095.

[13] Marchini J, Howie B, Myers S, McVean G, and Donnelly P (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906–913.

[14] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575.

[15] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, and Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909.

[16] Lilja H and Abrahamsson PA (1988). Three predominant proteins secreted by the human prostate gland. *Prostate* **12**, 29–38.

[17] Shukeir N, Arakelian A, Kadhim S, Garde S, and Rabbani SA (2003). Prostate secretory protein PSP-94 decreases tumor growth and hypercalcemia of malignancy in a syngenic *in vivo* model of prostate cancer. *Cancer Res* **63**, 2072–2078.

[18] Vanaja DK, Cheville JC, Iturria SJ, and Young CY (2003). Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. *Cancer Res* **63**, 3877–3882.

[19] Whitaker HC, Kote-Jarai Z, Ross-Adams H, Warren AY, Burge J, George A, Bancroft E, Jhavar S, Leongamornlert D, Tymrakiewicz M, et al. (2010). The rs10993994 risk allele for prostate cancer results in clinically relevant changes in microseminoprotein-beta expression in tissue and urine. *PLoS One* **5**, e13363.

[20] Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, et al. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* **40**, 316–321.

[21] Kader AK, Sun J, Isaacs SD, Wiley KE, Yan G, Kim ST, Fedor H, DeMarzo AM, Epstein JI, Walsh PC, et al. (2009). Individual and cumulative effect of prostate cancer risk-associated variants on clinicopathologic variables in 5,895 prostate cancer patients. *Prostate* **69**, 1195–1205.

[22] Lou H, Yeager M, Li H, Bosquet JG, Hayes RB, Orr N, Yu K, Hutchinson A, Jacobs KB, Kraft P, et al. (2009). Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. *Proc Natl Acad Sci USA* **106**, 7933–7938.

[23] Liu F, Hsing AW, Wang X, Shao Q, Qi J, Ye Y, Wang Z, Chen H, Gao X, Wang G, et al. (2011). Systematic confirmation study of reported prostate cancer risk-associated single nucleotide polymorphisms in Chinese men. *Cancer Sci* **102**, 1916–1920.

[24] Lu Y, Zhang Z, Yu H, Zheng SL, Isaacs WB, Xu J, and Sun J (2011). Functional annotation of risk loci identified through genome-wide association studies for prostate cancer. *Prostate* **71**, 955–963.

[25] Biancolella M, Valentini A, Minella D, Vecchione L, D'Amico F, Chillemi G, Gravina P, Bueno S, Prosperini G, Desideri A, et al. (2007). Effects of dutasteride on the expression of genes related to androgen metabolism and related pathway in human prostate cancer cell lines. *Invest New Drugs* **25**, 491–497.

[26] Williams SA, Xu Y, De Marzo AM, Isaacs JT, and Denmeade SR (2010). Prostate-specific antigen (PSA) is activated by KLK2 in prostate cancer *ex vivo* models and in prostate-targeted PSA/KLK2 double transgenic mice. *Prostate* **70**, 788–796.

[27] Darson MF, Pacelli A, Roche P, Rittenhouse HG, Wolfert RL, Young CY, Klee GG, Tindall DJ, and Bostwick DG (1997). Human glandular kallikrein 2 (hK2) expression in prostatic intraepithelial neoplasia and adenocarcinoma: a novel prostate cancer marker. *Urology* **49**, 857–862.

[28] Lintula S, Stenman J, Bjartell A, Nordling S, and Stenman UH (2005). Relative concentrations of hK2/PSA mRNA in benign and malignant prostatic tissue. *Prostate* **63**, 324–329.

[29] Becker C, Piironen T, Pettersson K, Bjork T, Wojno KJ, Oesterling JE, and Lilja H (2000). Discrimination of men with prostate cancer from those with benign disease by measurements of human glandular kallikrein 2 (HK2) in serum. *J Urol* **163**, 311–316.

[30] Nam RK, Zhang WW, Trachtenberg J, Diamandis E, Toi A, Emami M, Ho M, Sweet J, Evans A, Jewett MA, et al. (2003). Single nucleotide polymorphism of the human kallikrein-2 gene highly correlates with serum human kallikrein-2 levels and in combination enhances prostate cancer detection. *J Clin Oncol* **21**, 2312–2319.

[31] Nam RK, Zhang WW, Klotz LH, Trachtenberg J, Jewett MA, Sweet J, Toi A, Teahan S, Venkateswaran V, Sugar L, et al. (2006). Variants of the hK2 protein gene (*KLK2*) are associated with serum hK2 levels and predict the presence of prostate cancer at biopsy. *Clin Cancer Res* **12**, 6452–6458.

Table W1. Characteristics of All Subjects with PCA3 Measurement in the REDUCE Trial.

Characteristics*	Placebo		Dutasteride		Total		
	PCa Neg†	PCa Pos‡	PCa Neg†	PCa Pos‡	PCa Neg†	PCa Pos‡	Total
Number	723	162	648	116	1,371	278	1649
Age (year)	61.9 ± 5.90	63.94 ± 5.83	62.50 ± 5.97	63.29 ± 5.95	62.30 ± 5.93	62.62 ± 5.97	62.45 ± 5.95
Baseline PSA (ng/ml)	5.6 (4.2–7.2)	5.8 (4.7–7.3)	5.5 (4.2–7.0)	5.6 (4.3–7.2)	5.6 (4.2–7.1)	5.7 (4.5–7.2)	5.6 (4.3–7.1)
Baseline PV (cm^3)	44.94 (35.44–56.38)	42.52 (30.74–54.85)	44.91 (35.16–58.13)	40.33 (28.89–53.18)	44.91 (35.32–57.20)	41.80 (30.47–53.78)	44.49 (34.00–56.45–257.17)
PCA3 score	15.79 (8.43–31.12)	32.84 (15.04–61.44)	15.88 (8.47–31.50)	25.16 (15.38–43.68)	15.81 (8.46–31.23)	29.09 (15.08–55.30)	17.35 (9.08–35.68)
PSACPM (copy/ml)	182,562 (46,066–582,825)	145,671 (45,529–545,953)	66,106 (7,852–21,396)	97,406 (28,802–231,258)	111,256 (29,976–395,903)	119,205 (38,627–438,594)	113,444 (31,732–396,440)
PCA3CPM (copy/ml)	2,704 (703–12,165)	4,913 (1,155–22,924)	1,152 (338–4,094)	2,215 (732–6,532)	1,758 (457–6,860)	3,459 (1,013–13,550)	1,951 (516–7,545)

*Ages are described as means ± SD; Baseline PSA, Baseline PV, PCA3 scores, PSACPM (urinary PSA mRNA copies per milliliter), and PCA3CPM (urinary PCA3 mRNA copies per milliliter) are described as medians (quantile 1–quantile 3).
†PCa Neg, subjects with negative prostate biopsy.
‡PCa Pos, subjects with positive prostate biopsy.

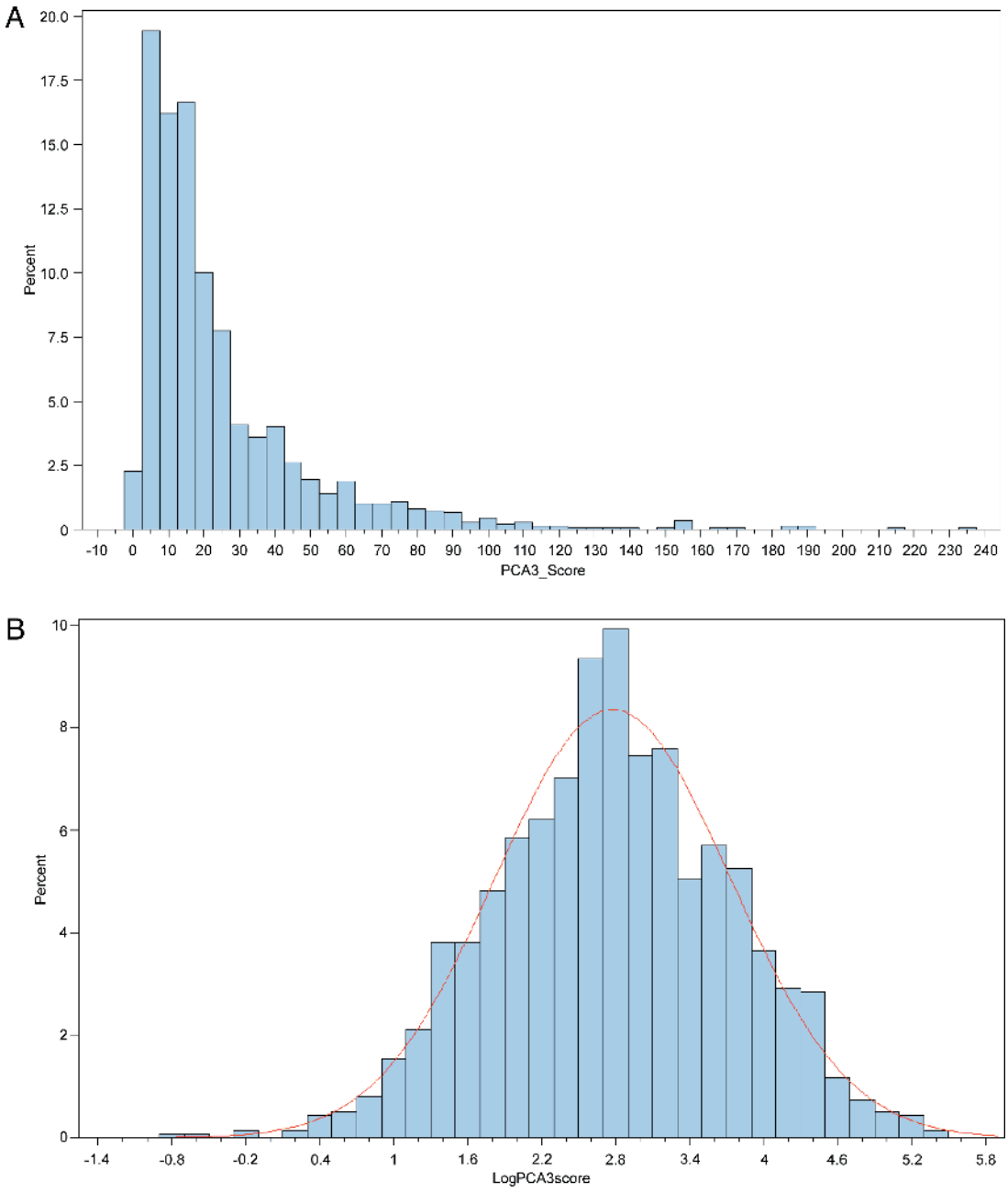

Figure W1. (A) PCA3 score distribution in subjects without prostate cancer. (B) Log-transformed PCA3 score distribution in subjects without prostate cancer.

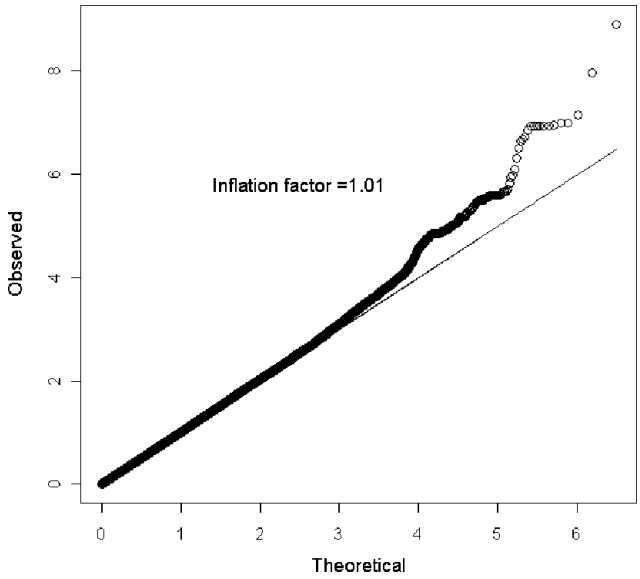

Figure W2. Q-Q plot. Q-Q plot presenting the distribution of theoretical compared to observed $-\log_{10}P$ for the association test results after adjusting for top eigenvector.

Table W2. SNPs Associated with PCA3 Score, Sorted by *P* Value (1×10^{-6} P 1×10^{-8}).

SNP	Chr	BP*	Gene	Alleles	Allele†	Freq‡	β (SE)§	*P*¶	Status
rs2739468	19	56,062,355	*KLK3*	A/C	A	0.43	1.23 (0.038)	6.94×10^{-8}	Imputed
rs1506684	19	56,063,231	*KLK2*	A/G	A	0.43	1.22 (0.038)	1.00×10^{-7}	Genotyped
rs2739461	19	56,061,263	*KLK2*	T/C	T	0.43	1.23 (0.038)	1.00×10^{-7}	Imputed
rs965537	19	56,064,609	*KLK2*	C/T	C	0.43	1.22 (0.038)	1.10×10^{-7}	Imputed
rs2739464	19	56,061,660	*KLK3*	G/A	G	0.43	1.22 (0.038)	1.13×10^{-7}	Imputed
rs2739466	19	56,062,064	*KLK3*	A/G	A	0.43	1.22 (0.038)	1.13×10^{-7}	Imputed
rs2739469	19	56,062,405	*KLK3*	T/C	T	0.43	1.22 (0.038)	1.13×10^{-7}	Imputed
rs2739472	19	56,065,091	*KLK2*	T/C	T	0.43	1.22 (0.038)	1.13×10^{-7}	Imputed
rs2739473	19	56,065,197	*KLK2*	A/G	A	0.43	1.22 (0.038)	1.13×10^{-7}	Imputed
rs2739475	19	56,065,438	*KLK2*	A/G	A	0.43	1.22 (0.038)	1.13×10^{-7}	Imputed
rs2739476	19	56,065,630	*KLK2*	A/G	A	0.43	1.22 (0.038)	1.39×10^{-7}	Imputed
rs2739459	19	56,060,886	*KLK3*	G/A	G	0.42	1.22 (0.038)	1.89×10^{-7}	Imputed
rs2889829	9	27,600,657	*C9orf72*	A/G	G	0.64	0.82 (0.038)	2.13×10^{-7}	Genotyped
rs1997563	19	56,065,002	*KLK2*	T/C	T	0.43	1.22 (0.038)	2.28×10^{-7}	Imputed
rs2926494	10	51,187,362	*MSMB*	T/C	T	0.46	1.21 (0.038)	3.00×10^{-7}	Genotyped
rs2569739	19	56,064,207	*KLK2*	C/T	C	0.43	1.21 (0.038)	4.74×10^{-7}	Imputed
rs62531686	8	1,710,268	*FLJ45872*	T/C	C	0.3	0.83 (0.038)	8.04×10^{-7}	Imputed

*BP genomic position is based on NCBI build 36.
†Allele shown here is the one with elevated PCA3 score.
‡Freq is the frequency of Allele†.
§β Values are back log-transformed, while SE is log-transformed.
¶*P* values are based on linear regression analysis on the log-transformed PCA3 score and adjusted for age and the top eigenvector assuming an additive model.

Table W3. Associations between rs10993994 and rs10424878 and PCa Risk, Baseline PV, and IPSS.

SNP	Alleles (M/m)	PCa Risk*		Baseline PV†		Baseline IPSS†	
		OR	*P*	β (SE)	*P*	β (SE)	*P*
rs10993994	C/**T**	1.20	.02	0.03 (0.01)	.01	−0.27 (0.14)	.05
rs10424878	G/**A**	1.08	.35	−0.02 (0.01)	.02	0.04 (0.14)	.76

*Association analysis for PCa risk was conducted in the placebo arm of the REDUCE trial using logistic regression.

†Association analyses for baseline PV and IPSS were conducted in both the placebo and dutasteride arms of the REDUCE trial using a linear regression model assuming additive effects of the minor alleles (0, 1, and 2). PV and IPSS were treated as continuous variables.

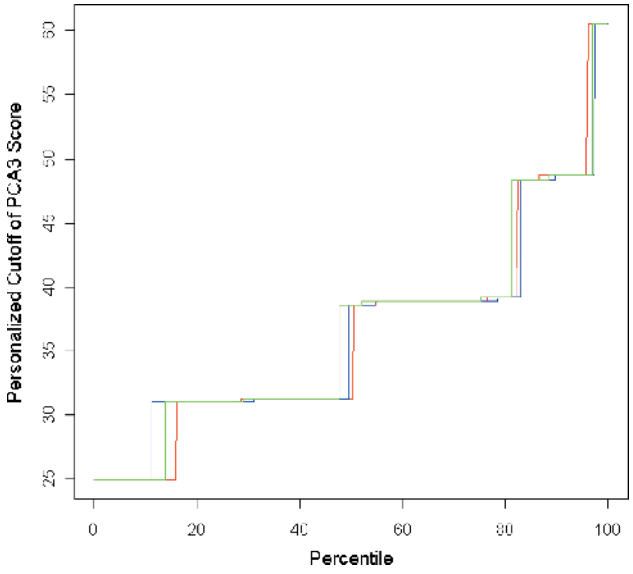

Figure W3. Cumulative distributions of the personalized cutoff of PCA3 score. The *x*-axis represents the percentile of the population and the *y*-axis represents the personalized cutoff of PCA3 score, which is adjusted by the genotypes of rs10993994 and rs10424878. Red lines stand for subjects with positive prostate biopsy; blue lines stand for subjects with negative prostate biopsy; green lines stand for all the subjects.