npg

ARTICLE

Comprehensive evaluation of the association between prostate cancer and genotypes/haplotypes in *CYP17A1*, *CYP3A4*, and *SRD5A2*

Anu Loukola[1], Monica Chadha[1], Sharron G Penn[1], David Rank[1], David V Conti[2], Deborah Thompson[3], Mine Cicek[4], Brad Love[1], Vesna Bivolarevic[1], Qiner Yang[1], Yalin Jiang[1], David K Hanzel[1], Katherine Dains[1], Pamela L Paris[4], Graham Casey[4] and John S Witte*[2,3]

[1]Amersham Biosciences, Sunnyvale, CA 94085, USA; [2]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA; [3]International Agency for Cancer Research, Lyon, France; [4]Department of Cancer Biology, Cleveland Clinic Foundation, Cleveland, OH 44195, USA

Genes involved in the testosterone biosynthetic pathway – such as *CYP17A1*, *CYP3A4*, and *SRD5A2* – represent strong candidates for affecting prostate cancer. Previous work has detected associations between individual variants in these three genes and prostate cancer risk and aggressiveness. To more comprehensively evaluate *CYP17A1*, *CYP3A4*, and *SRD5A2*, we undertook a two-phase study of the relationship between their genotypes/haplotypes and prostate cancer. Phase I of the study first searched for single-nucleotide polymorphisms (SNPs) in these genes by resequencing 24 individuals from the Coriell Polymorphism Discovery Resource, 92–110 men from prostate cancer case–control sibships, and by leveraging public databases. In all, 87 SNPs were discovered and genotyped in 276 men from case–control sibships. Those SNPs exhibiting preliminary case–control allele frequency differences, or distinguishing (ie, 'tagging') common haplotypes across the genes, were identified for further study (24 SNPs in total). In Phase II of the study, the 24 SNPs were genotyped in an additional 841 men from case–control sibships. Finally, associations between genotypes/haplotypes in *CYP17A1*, *CYP3A4*, and *SRD5A2* and prostate cancer were evaluated in the total case–control sample of 1117 brothers from 506 sibships. Family-based analyses detected associations between prostate cancer risk or aggressiveness and a number of *CYP3A4* SNPs (*P*-values between 0.006 and 0.05), a *CYP3A4* haplotype (*P*-values 0.05 and 0.009 in nonstratified and stratified analysis, respectively), and two *SRD5A2* SNPs in strong linkage disequilibrium (*P* = 0.02). Undertaking a two-phase study comprising SNP discovery, haplotype tagging, and association analyses allowed us to more fully decipher the relation between *CYP17A1*, *CYP3A4*, and *SRD5A2* and prostate cancer.
European Journal of Human Genetics (2004) **12**, 321–332. doi:10.1038/sj.ejhg.5201101
Published online 15 October 2003

Introduction

Prostate cancer is the most common noncutaneous neoplasm among men. According to the American Cancer Society, 189 000 US men were diagnosed with prostate cancer in the year 2002, and approximately 30 000 US men died from this disease. A clear risk factor for prostate cancer is age, suggesting an important role of exposure to endogenous and/or exogenous factors (eg, hormones or environmental agents). Another known risk factor is ethnicity: prostate cancer is almost two times more common in African Americans than in Caucasians, whereas Asians have the lowest risk.[1,2] Finally, there is an

increased risk among men with a positive family history of prostate cancer.[3] These factors suggest that both genes and environment predispose men to this disease.

Substantial evidence from linkage studies supports the existence of genetic factors for prostate cancer development and progression. A number of chromosomal regions have been linked to prostate cancer, including 1q24–q25, 1q42.2–q43, 1p36, 8p22–23, 16q23, 17p11, 19q, 20q13, and Xq27–28.[4] On 17p11, *ELAC2* was the first prostate cancer-susceptibility gene cloned.[5] Common allelic variants of *ELAC2* were shown associated with prostate cancer in some studies,[6,7] but not all.[8–11] Another candidate prostate cancer-susceptibility gene, *RNASEL*, was identified in the HPC1 locus on 1q24–25.[12] Subsequent studies support a role both for inactivating mutations and variants in *RNASEL* and prostate cancer risk.[13–15]

Since prostate growth depends on active testosterone,[16] and testosterone administration leads to the development of prostate adenocarcinomas in rats,[17] studies of genes involved in androgen metabolism have provided further insight into the genetic basis of this disease. Three such genes are *CYP17A1*, *CYP3A4*, and *SRD5A2*. *CYP17A1* is involved with the biosynthesis of testosterone in the gonads and adrenals, and is involved with catalyzing steroid 17α-hydroxylase and 17,20-lyase activities at key points in this process.[18] *CYP3A4* catalyzes the 6β-hydroxylation of testosterone, suggesting that it may be involved in the metabolism and oxidative deactivation of this steroid.[19] The *SRD5A2* gene product catalyzes the conversion of testosterone to dihydrotestosterone,[20] a highly active form of this androgen.

Numerous previous studies have detected associations between variants in *CYP17A1*, *CYP3A4*, and *SRD5A2* and prostate cancer.[21–31] These studies have focused only on specific individual variants in the candidate genes. However, fully deciphering the impact of these genes on prostate cancer may require the evaluation of multiple variants or haplotypes. Therefore, we present here a two-phase comprehensive analysis of the association between multiple single-nucleotide polymorphism (SNP) genotypes/haplotypes in *CYP17A1*, *CYP3A4*, and *SRD5A2* and prostate cancer.

Materials and methods
Study population
A total of 1117 men (637 cases and 480 controls) from 506 sibships were included in the study. In total, 92 sibships were concordant (ie, where all included men were diagnosed with prostate cancer, and there was no control sibling) and contained altogether 197 men. The remaining 414 sibships were discordant (ie, at least one control sibling) and contained 440 cases and 480 controls. Altogether, 52 discordant sibships had more than one control sib. Hence, while all of the 1117 study subjects

contributed information for calculating allele frequencies, only the 440 cases and 480 controls (920 men total) from discordant families were informative for the conditional logistic regression analysis of association.

The study population was recruited between January 1998 and January 2001 from the major medical institutions in the greater Cleveland area and from the Henry Ford Health System in Detroit. The study was approved by the collaborating institution's Review Boards, and informed consent was obtained from all participating men. Characteristics of the study population have been described elsewhere.[13]

Men diagnosed with histologically confirmed prostate cancer at age 73 or younger were invited to join the study if they had a living unaffected brother who was either older than the proband, or at most 8 years younger than the age at diagnosis of the proband. This age restriction was selected in an attempt to increase the potential for finding a genetic association with disease and to substantiate that the controls were not unaffected simply due to being of a younger age. To help confirm that the controls were not diseased, we tested the prostate-specific antigen (PSA) levels in their blood (see discussion). Information on the cases' Gleason score and tumor stage (TNM) was determined from their medical records. The study population was comprised of 90% Caucasians (European Americans), and the remainder primarily African American (9%).

SNP discovery
A complementary approach was used to identify SNPs within *CYP17A1*, *CYP3A4*, and *SRD5A2*. We performed SNP discovery by sequencing individuals from two populations: (1) the Coriell Polymorphism Discovery Resource (Collins *et al*[32]; Coriell Cell Repositories, Camden, NJ, USA); and (2) prostate cancer case–control sibships. The Coriell population includes 24 individuals: six European Americans, six African Americans, three Hispanic Americans, three Native Americans, and six Asian Americans. The prostate cancer case–control sibships used for SNP discovery were two randomly selected subsets of our entire study population, and included 67 cases and 43 controls for *CYP17A1* and *CYP3A4*, and 51 cases and 41 controls for *SRD5A2*. There was no overlap between the two subgroups. Of the 110 individuals sequenced for *CYP17A1* and *CYP3A4*, 106 were European American, two were Hispanic American, and two were African American. Of the 92 individuals sequenced for *SRD5A2*, 84 were European American and eight were African American.

For sequencing, PCR primers amplifying most of the coding regions, splice sites, 5′ and 3′ regions, and parts of introns of *CYP17A1*, *CYP3A4*, and *SRD5A2* were designed using the Primer3 program (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi). PCR products were sequenced using energy DYEnamic™ ET Terminator Kit on the MegaBACE™ DNA Analysis System (Amersham

Biosciences, Sunnyvale, CA, USA) by standard protocols. Sequence analysis was performed by assigning quality values (Phred, University of Washington, Seattle, Washington), assembling contigs (Phrap, University of Washington), automated identification of candidate heterozygote SNPs (PolyPhred, University of Washington), automated identification of candidate homozygote SNPs (High Quality Mismatch, Amersham Biosciences), and by operator confirmation (Consed, University of Washington). All polymorphisms were confirmed by single-nucleotide primer extension.

In addition, we searched the following databases for known SNPs: dbSNP (http://www.ncbi. nlm.nih.gov/SNP/), GeneSNPs (http://www.genome.utah.edu/genesnps), the Human Cytochrome *P*450 Allele Nomenclature Committee (HCANC) (http://www.imm.ki.se/CYPalleles/), the Human Gene Mutation Database (HGMD) (http://archive.uwcm. ac.uk/uwcm/mg/hgmd0.html), and HGVbase (http:// hgvbase.cgb.ki.se/) (formerly HGBASE at http://hgbase. interactiva.de/).

Genotyping

For Phase I of the study, we genotyped the SNPs discovered in *CYP17A1*, *CYP3A4*, and *SRD5A2* among 276 men (including the 92 and 110 men used in SNP discovery) from case–control sibships. Altogether, 276 men were selected for phase I to allow for detection of SNPs with low allele frequencies, and due to the convenience of fitting them all into three 96-well plates along with controls. These men included 153 cases and 123 brother controls, 70% European American and 30% African American. We then used the information from the 276 men to determine initial case–control allele frequency differences and haplotype-tagging SNPs. From these results, we determined which SNPs should be genotyped in the remainder of our study population (ie, Phase II of the study).

Genotyping of SNPs was primarily performed utilizing the MegaBACE SNuPe™ Genotyping Kit (Amersham Biosciences) on the MegaBACE™ DNA Analysis System (Amersham Biosciences). The Primer3 program was used to design PCR primers to amplify regions containing the SNPs of interest (see the Online Supplementation (Table 4) for PCR primers and conditions). PCR fragments were purified with 0.5 U of Shrimp Alkaline Phosphatase (Amersham Biosciences) and 10 U of Exonuclease I (Amersham Biosciences) at 37°C for 40 min and at 85°C for 15 min. The single base extension (SBE) reaction was set with 1 pmol of HPLC-purified SBE primer, 2–4 μl of SNuPe Premix (Amersham Biosciences), 2–4 μl of sterile water, and 1 μl of purified PCR fragment, and incubated at 25 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 10 s. For phase I of the study, SNuPe reactions were set in 96-well plates at 10 μl volume and purified with AutoSeq™96 Plates (Amersham Biosciences) prior to injecting into the Mega-BACE1000 system. For phase II of the study, SNuPe

reactions were set in 384-well plates at 5–6 μl volume, diluted with 3–4 μl of sterile water and purified with 1 U of Shrimp Alkaline Phosphatase (Amersham Biosciences) at 37°C for 45 min and at 85°C for 15 min prior to injecting into the MegaBACE4000 system. In cases where low signal was anticipated (due to faint PCR), SNuPe reactions were desalted using a custom 384-well filter plate incorporating modified size-exclusion technology (Millipore Corporation, Billerica, MA, USA). Scierra Genotyping LWS™ system (Amersham Biosciences) was utilized for the tracking and management of samples and laboratory activity for the Phase II of the study.

Specific software (SNPriDe; Amersham Biosciences) was developed for the automated design of SBE primers. Using a purified PCR fragment containing the SNP of interest as a template, a third, internal primer was designed so that the 3′ end annealed adjacent to the polymorphic base pair, and during the SBE reaction a fluorescently labeled dideoxynucleotide (terminator) was added onto the primer. The signal data were automatically processed, outputting the maximum-likelihood SNP genotypes (MegaBACE SNP Profiler, Amersham Biosciences). The system includes a user interface for editing and verification. Three SNPs, SRD5A2_SNP20 (V89L), SRD5A2_SNP22 (A49T), and CYP17_SNP29 (−34T>C) were analyzed by restriction enzyme digestion (see Online Supplementation for details).

Haplotype estimation

Alleles within each of the three candidate genes were in strong linkage disequilibrium with one another. Thus, for each gene, haplotypes were estimated using the resulting genotypes, by disease status and within major ethnic groups using the software PHASE.[33] This program uses Markov chain Monte Carlo to estimate haplotypes, imputes information for missing genotypes, and incorporates a statistical model for the distribution of unresolved haplotypes based on coalescent theory.[33] For the estimation of haplotypes, PHASE uses individual level information and provides the most likely haplotypes. Haplotype-tagging SNPs were determined by custom perl scripts (Amersham Bioscience). We first determined haplotypes and haplotype-tagging SNPs among the 276 men genotyped for Phase I of the study, where tagging SNPs were those necessary to define the most common haplotypes (eg, ⩾5%). After completing genotyping on the entire study population (Phase II of the study), we used the resulting data to estimate haplotypes.

Association analyses

Altogether, 414 discordant sibships (440 cases and 480 controls) were included in the analyses. We first compared case *versus* control allele frequencies within major ethnic groups. Then we evaluated the association between the resulting genotypes/haplotypes and prostate cancer risk by calculating odds ratios (OR, estimates of relative risk) and

95% confidence intervals from conditional logistic regression with family as the matching variable, and a robust variance estimator that incorporates familial correlations. This is a standard approach for analyzing sibling-matched case–control data, although sibling sets without any controls do not contribute any information (197 cases total here).[34] In our analyses of *CYP17A1*, *CYP3A4*, and *SRD5A2*, we used a dominant coding, which assumes that the relative risk of carrying one or two polymorphisms (or haplotype) is equivalent. The most likely haplotypes were used in the association analyses. The nontagging SNPs (included due to interesting initial results in phase I) were included in both the genotype- and haplotype-level analyses. All our analyses were stratified by ethnicity.

To control for potential confounding, age was adjusted for in all regression models. In addition to looking at the main effects of each SNP or haplotype, we also stratified the analyses by the case's disease aggressiveness, where high aggressiveness was defined by TNM stage ⩾T2B or Gleason score ⩾7, and low aggressiveness by TNM stage <T2B and Gleason score <7. All statistical analyses were undertaken with the S+ software (version 6.0, Insightful Corp, 2001). No correction was made for multiple comparisons. Note, however, that SNPs within each gene were in linkage disequilibrium and thus the corresponding tests for association within a gene are correlated.

Results

Phase I: SNP discovery
We detected 34 novel SNPs: 11 in *CYP17A1*, 18 in *CYP3A4*, and five in *SRD5A2*. In addition, we 'rediscovered' 11 SNPs from the public databases. Including these 11 SNPs, we selected a total of 53 SNPs from the databases: 18 in *CYP17A1*, 15 in *CYP3A4*, and 20 in *SRD5A2*. These were chosen based on the intention to obtain an even distribution of SNPs across the genes and the availability in the databases at that time (January–April 2001). In all, 21 SNPs were chosen from dbSNP, 27 from GeneSNPs, 12 from HGMD, eight from HGVbase, and two from HCANC (the total number of SNPs listed here exceeds 53 as several SNPs were present in multiple databases). Table 1 lists all the 87 SNPs (34 novel, 53 from databases), with their origins, exact locations, and allele frequencies.

Among the 34 novel SNPs, 26 (76%) were discovered in both the Coriell and case–control populations. Three SNPs were only observed in the Coriell data, and the remaining five were found only in the prostate cancer sibships. Of these five, three were relatively rare (allele frequencies 0.2–1.5%), suggesting that they may not have been discovered in the Coriell population simply due to its small sample size ($n = 24$). Nevertheless, the other two SNPs that were only found in the prostate cancer sibships (CYP3A4_SNP12 and CYP17_SNP42) showed higher allele frequencies (7.5

and 21.8%, respectively), suggesting that they might be specific to the prostate cancer case–control population.

Phase I: genotyping and haplotyping
The 87 SNPs were geneotyped in a total of 276 males from prostate cancer sibships (29 in *CYP17A1*, 33 in *CYP3A4*, and 25 in *SRD5A2*). Altogether 11 SNPs gave ambiguous genotyping results. This might have been due to unoptimized genotyping reactions or primer self-priming due to secondary structures and unspecificity of PCR and/or SNuPe primers, especially within the cytochrome *P*450 gene family. Of the remaining 76 SNPs, a similar percentage of those novel (41%, or 12/29) and public (38%, or 18/47) had allele frequencies ⩾10%. However, 19/47 (40%) of the public SNPs were found to be monoallelic in the 276 men, suggesting that they are either extremely rare, population specific, or artifacts.

In the light of these results, we excluded the 11 SNPs with ambiguous genotype results, the 19 SNPs that appeared monoallelic in all samples tested, and additional four (three novel and one public SNP) that were seen only in the Coriell Diversity Set but not in the prostate cancer sibships. We also excluded one SNP because >15% of data were missing (due to a low success rate for PCR and SNuPe reaction). Finally, we excluded 12 SNPs because their minor allele frequencies were less than 5% in all of the following four subgroups: European American controls, European American cases, African American controls, and African American cases (see Table 1 for details). Following these exclusions, a total of 40 SNPs remained for consideration in the Phase II association study (14 in *CYP17A1*, 16 in *CYP3A4*, and 10 in *SRD5A2*) (Table 1).

Using the preliminary genotype information, haplotypes estimated with a frequency ⩾5% in at least one of the four major subgroups (ie, European American controls, European American cases, African American controls, African American cases) were identified. Each gene had a single 'common' haplotype, with a frequency ranging between 42 and 51% (not shown). Haplotype-tagging SNPs were identified and used as a basis for inclusion in Phase II of the study. In addition, nontagging SNPs exhibiting suggestive case *versus* control allele frequencies were considered (Table 1). Altogether, 24 SNPs were selected for Phase II.

Phase II
The 24 tagging and suggestive SNPs were genotyped in an additional 841 men, giving information on a total of 1117 individuals for Phase II. Case and control allele frequencies by ethnic groups are presented in Table 1. Haplotypes estimated with a frequency ⩾3% in at least one of the four major subgroups of the study population were identified. The major haplotypes for *CYP17A1*, *CYP3A4*, and *SRD5A2* along with their frequencies are presented in Figure 1.

In our association analyses, we did not detect any associations between *CYP17A1* genotypes/haplotypes and

Table 1 The origins, nucleotide changes and allele frequencies of single nucleotide polymorphisms (SNPs) in *CYP17A1*, *CYP3A4*, and *SRD5A2* observed in the Coriell Diversity set (CDS), European Americans, and African Americans

| | | | | | | Allele frequency | | | |
| | | | | | | European Americans | | African Americans | |
SNP	Origin[a]	Nucleotide change[b]	[c]	[d]	CDS	Ctrls	Cases	Ctrls	Cases
CYP17A1									
SNP18	Novel (C+C)	−1488C>G	B	I	0.26	NA[e]	NA	NA	NA
SNP19	Novel (C+C)	−1204C>T	B	I	0.10	NA	NA	NA	NA
SNP29	dbSNP/HGVbase (−)	−34T>C	1,2,3	II	0.44	0.40	0.38	0.33	0.38
SNP30	GeneSNPs/dbSNP (−)	C22W (66C>G)	B	I	—	NA	NA	NA	NA
SNP31	GeneSNPs/dbSNP/HGVbase (−)	H46H (138C>T)		I	0.43	0.43	0.36	0.41	0.35
SNP32	GeneSNPs/dbSNP/HGVbase (−)	S65S (195G>T)		I	0.46	0.44	0.33	0.40	0.32
SNP4	Novel (C+C)	IVS1 +426G>A	1	II	0.27	0.40	0.39	0.30	0.36
SNP20	Novel (C+C)	IVS1 +466G>A	4	II	0.06	0.03	0.01	0.02	0.02
SNP8	Novel (C+C)	IVS1 −700C>G		I	0.19	0.14	0.15	0.21	0.08
SNP26	GeneSNPs/dbSNP (+<)	IVS1 −679C>T	D	I	—	0.06	0.02	0.04	0.01
SNP11	Novel (CDS)	IVS1 −565G>A	A	I	0.04	—	—	—	—
SNP1	Novel (C+C)	IVS1 −271A>C		I	0.44	0.46	0.40	0.40	0.43
SNP6	Novel (C+C)	IVS1 −99C>T	2	II	0.38	0.29	0.28	0.11	0.15
SNP23	HGMD (*)	S106P (316T>C)	A	I	—	—	—	—	—
SNP25	HGMD (*)	IVS2 +5G>T	A	I	—	—	—	—	—
SNP7	dbSNP (R)	IVS2 +105A>C	1	II	0.46	0.29	0.28	0.13	0.15
SNP22	dbSNP (+<)	IVS2 −83C>T	1	II	0.04	0.002	0.0008	0.06	0.09
SNP24	HGMD (+<)	E194X (580G>T)	D	I	—	—	—	—	0.01
SNP5	dbSNP (R)	IVS3 +35T>C		I	0.06	0.16	0.16	0.20	0.07
SNP12	Novel (C+C)	IVS3 +141A>T		I	0.04	0.04	0.02	—	0.01
SNP21	GeneSNPs/dbSNP (*)	D283D (849C>T)	A	I	—	—	—	—	—
SNP3	Novel (C+C)	IVS5 +75C>G	1	II	0.33	0.40	0.39	0.20	0.23
SNP33	HGMD (−)	IVS7 +5G>A	D	I	—	—	0.02	—	—
SNP34	HGMD (*)	F417C (1250T>G)	A	I	—	—	—	—	—
SNP35	GeneSNPs/dbSNP/HGVbase (−)	P428P (1284C>A)	B	I	—	NA	NA	NA	NA
SNP36	HGMD (*)	R440H (1319G>A)	A	I	—	—	—	—	—
SNP37	HGMD (*)	R496C (1486C>T)	A	I	—	—	—	—	—
SNP42	Novel (CAP)	Stop +712G>A		I	—	0.28	0.18	0.18	0.21
SNP16	GeneSNPs (R)	Stop +2074G>A	D	I	0.06	0.01	0.02	—	0.03
CYP3A4									
SNP48	GeneSNPs (*)	−8086G>A	A	I	—	—	—	—	—
SNP49	GeneSNPs (−)	−6790G>A	B	I	0.50	NA	NA	NA	NA
SNP47	dbSNP (R)	−1232C>T	1	II	0.19	0.05	0.04	0.56	0.56
SNP22	Novel (CDS)	−847A>T		I	0.06	—	—	0.20	0.15
SNP12	Novel (CAP)	−747C>G	1,2	II	—	0.08	0.08	0.01	0.04
SNP11	HCANC (+)	−392A>G	3	II	0.13	0.04	0.04	0.58	0.54
SNP45	HGVbase (*)	−290A>G	A	I	—	—	—	—	—
SNP50	GeneSNPs (*)	−26G>A	A	I	—	—	—	—	—
SNP21	Novel (CDS)	IVS1 −868C>T		I	NA	—	0.009	0.18	0.17
SNP20	dbSNP (R)	IVS2 +671T>A		I	0.15	0.07	0.07	0.42	0.44
SNP19	Novel (CDS)	IVS2 −132C>T	D	I	0.02	—	0.009	—	—
SNP26	Novel (C+C)	IVS3 +1992T>C	B	I	0.40	NA	NA	NA	NA
SNP32	Novel (CDS)	IVS3 −734G>A	B	I	NA	NA	NA	NA	NA
SNP33	GeneSNPs (+<)	IVS4 −172G>A	A	I	0.02	—	—	—	—
SNP17	HCANC (*)	S222P (664T>C)	A	I	—	—	—	—	—
SNP1	dbSNP (R)	IVS7 +34T>G	1	II	0.17	0.06	0.05	0.62	0.56
SNP14	dbSNP (R)	IVS7 +526C>T		I	0.02	—	0.02	0.11	0.11
SNP13	Novel (C+C)	IVS7 −202C>T	1	II	0.31	0.14	0.13	0.66	0.72
SNP2	Novel (C+C)	IVS9 +187C>G	B	I	0.08	NA	NA	NA	NA
SNP27	Novel (C+C)	IVS9 +841T>G		I	0.06	—	0.01	0.07	0.08
SNP46	HGVbase (+<)	M318I (954G>A)	B	I	—	NA	NA	NA	NA
SNP10	dbSNP (R)	IVS10 +12G>A		I	0.42	0.16	0.14	0.67	0.66
SNP34	GeneSNPs/dbSNP (*)	I431T (1292T>C)	A	I	—	—	—	—	—
SNP18	HCANC (*)	M445T (1334T>C)	A	I	—	—	—	—	—
SNP29	Novel (CDS)	IVS12 +581C>T	A	I	0.02	—	—	—	—
SNP30	Novel (CDS)	IVS12 +586G>A	D	I	0.02	0.04	0.01	—	0.01
SNP31	Novel (CDS)	IVS12 +646C>A	A	I	0.02	—	—	—	—
SNP28	Novel (C+C)	IVS12 −473T>G		I	0.08	0.006	0.01	0.24	0.27

Table 1 Continued.

SNP	Origin[a]	Nucleotide change[b]	[c]	[d]	CDS	Allele frequency European Americans		African Americans	
						Ctrls	Cases	Ctrls	Cases
SNP24	Novel (C+C)	Stop +766delT; T>G	1	I	0.33	0.14	0.13	0.52	0.53
SNP6	Novel (CAP)	Stop +945A>T	D	I	—	0.02	0.02	—	—
SNP25	Novel (CDS)	Stop +1454C>T	1	II	0.08	0.003	0.006	0.23	0.28
SNP5	Novel (C+C)	Stop +1639A>T	1	II	0.63	0.17	0.16	0.61	0.62
SNP15	Novel (C+C)	Stop +2204G>C	1	II	0.13	0.13	0.11	0.24	0.20
SRD5A2									
SNP17	GeneSNPs (R)	−8029C>T	1,2	II	0.33	0.46	0.46	0.46	0.37
SNP18	GeneSNPs (*)	−7819G>C	A	I	—	—	—	—	—
SNP26	GeneSNPs (+)	−3001G>A	1	II	0.30	0.29	0.30	0.27	0.39
SNP28	GeneSNPs (*)	−2851A>T	A	I	—	—	—	—	—
SNP31	Novel (C+C)	−2036(A)7-8, A>T	C	I	NA	0.29	0.28	0.43	0.33
SNP5	GeneSNPs (R)	−1971G>A	B	I	0.48	NA	NA	NA	NA
SNP30	Novel (CAP)	−870G>A	D	I	—	0.01	0.02	—	0.01
SNP21	HGMD (*)	G34R (100G>A)	A	I	—	—	—	—	—
SNP22	GeneSNPs/dbSNP/HGVbase (−)	A49T (145G>A)	1,3	II	NA	0.04	0.04	0.01	0.03
SNP20	GeneSNPs/dbSNP/HGVbase (−)	V89L (265G>C)	1,3	II	NA	0.29	0.29	0.32	0.34
SNP23	GeneSNPs/dbSNP (−)	IVS1 +15C>T	B	I	0.46	NA	NA	NA	NA
SNP11	GeneSNPs/dbSNP (−)	IVS1 +24664G>T		I	0.48	0.24	0.27	0.19	0.22
SNP12	GeneSNPs/dbSNP (−)	IVS2 +626C>T	1,2	II	0.48	0.41	0.40	0.27	0.30
SNP7	HGMD (*)	G183S (547G>A)	A	I	—	—	—	—	—
SNP8	HGMD (*)	N193S (578A>G)	A	I	—	—	—	—	—
SNP9	HGMD (*)	P212R (635C>G)	A	I	—	—	—	—	—
SNP10	HGMD (*)	IVS4 +1G>T	A	I	—	—	—	—	—
SNP32	Novel (CAP)	Stop +545T>C	D	I	—	—	0.005	—	—
SNP4	Novel (C+C)	Stop +849A>G		I	0.13	0.11	0.12	0.16	0.23
SNP2	Novel (C+C)	Stop +1356A>C	D	I	0.02	0.006	0.009	—	—
SNP1	GeneSNPs (R)	Stop +1552G>A	1	II	0.16	0.12	0.12	0.19	0.23
SNP13	GeneSNPs (+)	Stop +3059G>A	1	II	0.13	0.09	0.09	0.13	0.14
SNP14	GeneSNPs (−)	Stop +5179A>C	D	I	0.02	0.01	0.005	—	—
SNP15	GeneSNPs (−)	Stop +9301G>C	1	II	0.46	0.26	0.27	0.21	0.23
SNP16	GeneSNPs (−)	Stop +9502C>T	D	I	—	0.006	—	—	—

[a](*), SNP did not show up in our study population; (R), rediscovered; (+), we had sequence coverage but did not rediscover the SNP; (+<), we had sequence coverage but did not rediscover the SNP, most likely due to the low minor allele frequency; (−), we did not have sequence coverage explaining why we did not rediscover the SNP; (CDS), novel SNP discovered originally in the Coriell Diversity Set; (CAP), novel SNP discovered originally in the prostate cancer sibships; (C+C), novel SNP discovered originally in both populations.
[b]Underlined bases indicate the allele for which frequencies are given.
[c]Excluded from haplotyping in Phase I and from consideration for Phase II based on (a) being monoallelic in the prostate cancer sibships, (b) yielding ambiguous genotyping results, (c) low success rate, (d) allele frequency <5%. Included in Phase II association analyses based on (1) being a haplotype-tagging SNP, (2) case–control difference in Phase I, (3) previous publications supporting association, (4) SNP conveniently located within the same PCR fragment as another included SNP.
[d]I, allele frequencies based on 276 samples; II, allele frequencies based on 1117 samples.
[e]NA, data not available.

prostate cancer. When looking at *CYP3A4*, we found that SNP1 was inversely associated with prostate cancer (OR = 0.53, 95% CI = 0.29–0.99; *P*-value = 0.05) (Table 2a). Furthermore, our haplotype analysis revealed an inverse association with CYP3A4_Hap4 and prostate cancer (OR = 0.46, 95% CI = 0.21–1.02; *P*-value = 0.05) (Table 3a). We also found that two SNPs in SRD5A2 were positively associated with prostate cancer: SRD5A2_SNP26 (OR = 1.57, 95% CI = 1.08–2.30; *P*-value = 0.02), and SRD5A2_SNP20 (V89L) (OR = 1.56, 95% CI = 1.08–2.25; *P*-value = 0.02) (Table 2A). These SNPs, however, were in strong linkage disequilibrium.

Figure 1 Major haplotypes for (**a**) *CYP17A1*, (**b**) *CYP3A4*, and (**c**) *SRD5A2*. Solid black triangles refer to the locations of novel SNPs, while white triangles denote the locations of public SNPs. All haplotypes with frequency ⩾3% in at least one of the four subgroups (European American controls, European American cases, African American controls, African American cases) are given in the center of each panel, along with their case and control frequencies. [a]EA, European American; [b]AA, African American; [c]composite haplotype refers to all the remaining rare haplotypes pooled together.

a
CYP17A1

26 865 bp

	SNP29: -34T>C	SNP04: IVS1 +426G>A	SNP20: IVS1 +466G>A	SNP06: IVS1 -99C>T	SNP07: IVS2 +105A>C	SNP22: IVS2 -83C>T	SNP03: IVS5 +75C>G	Frequency (all) Controls	Cases	Frequency (EA) [a] Controls	Cases	Frequency (AA) [b] Controls	Cases
Hap1	T	C	C	C	A	C	C	547 (.57)	527 (.60)	494 (.57)	475 (.60)	50 (.66)	47 (.62)
Hap2	C	T	.	T	C	.	G	259 (.27)	216 (.25)	253 (.29)	208 (.26)	5 (.07)	6 (.08)
Hap3	C	T	.	.	.	.	G	91 (.09)	86 (.10)	85 (.10)	79 (.10)	3 (.04)	7 (.09)
Hap4	Composite [c]							63 (.07)	51 (.06)	42 (.05)	32 (.04)	18 (.24)	16 (.21)

b
CYP3A4

37 073 bp

	SNP47: -1232C>T	SNP12: -747C>G	SNP11: -392A>G	SNP01: IVS7 +34T>G	SNP13: IVS7 -202C>T	SNP24: stop +766 delT	SNP25: stop +1454C>T	SNP05: stop +1639A>T	SNP15: stop +2204G>C	Frequency (all) Controls	Cases	Frequency (EA) [a] Controls	Cases	Frequency (AA) Controls	Cases
Hap1	C	C	A	T	C	T	C	A	G	651 (.68)	596 (.68)	629 (.72)	577 (.73)	16 (.21)	13 (.17)
Hap2	.	G	.	.	.	.	.	.	.	65 (.07)	62 (.07)	64 (.07)	59 (.07)	1 (.01)	2 (.03)
Hap3	.	.	.	.	T	G	.	T	C	63 (.07)	64 (.07)	59 (.07)	58 (.07)	4 (.05)	6 (.08)
Hap4	T	.	G	G	T	G	.	T	C	45 (.05)	25 (.03)	34 (.04)	19 (.02)	10 (.13)	5 (.07)
Hap5	Composite [c]									136 (.14)	133(.15)	88 (.10)	81 (.10)	45 (.59)	50 (.66)

c
SRD5A2

76 341 bp

	SNP17: -8029C>T	SNP26: -3001G>A	SNP22: A49T (145G>A)	SNP20: V89L (265G>C)	SNP12: IVS2 + 626C>T	SNP01: stop + 1552G>A	SNP13: stop + 3059G>A	SNP15: stop + 9301G>C	Frequency (all) Controls	Cases	Frequency (EA) [a] Controls	Cases	Frequency (AA) [b] Controls	Cases
Hap1	T	G	G	G	C	G	G	G	376 (.39)	348 (.40)	352 (.40)	321 (.40)	19 (.25)	20 (.26)
Hap2	C	A	.	C	T	.	.	C	218 (.23)	212 (.24)	206 (.24)	197 (.25)	11 (.14)	13 (.17)
Hap3	C	.	.	.	T	.	.	.	108 (.11)	84 (.10)	102 (.12)	81 (.10)	6 (.08)	3 (.04)
Hap4	C	.	.	.	.	A	A	.	73 (.08)	73 (.08)	70 (.08)	65 (.08)	-	7 (.09)
Hap5	C	A	.	C	T	.	.	.	41 (.04)	42 (.05)	41 (.05)	42 (.05)	-	-
Hap6	C	.	.	.	.	.	.	.	32 (.03)	27 (.03)	19 (.02)	20 (.03)	13 (.17)	7 (.09)
Hap7	Composite [c]								112 (.12)	94 (.10)	84 (.10)	68 (.09)	27 (.36)	26 (.34)

Table 2 (a) All nonstratified association results between *CYP17A1*, *CYP3A4*, and *SRD5A2* variants and risk of prostate cancer among cases and sibling controls;[a] (b) Statistically significant allele associations obtained from analysis stratified by aggressiveness[a]

		All subjects (n = 886–920)		European Americans (n = 781–834)		African Americans (n = 74–76)	
Genes	Genotype comparison[b]	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
(a)							
CYP17A1							
SNP29	CC or TC *vs* TT	0.91 (0.65–1.29)	0.61	0.86 (0.59–1.25)	0.42	1.96 (0.72–5.31)	0.19
SNP4	AA or GA *vs* GG	0.88 (0.62–1.25)	0.47	0.82 (0.56–1.19)	0.30	1.96 (0.72–5.31)	0.19
SNP20	AA or GA vGG	0.57 (0.25–1.31)	0.19	0.52 (0.21–1.28)	0.15	1.87 (0.55–6.35)	0.31
SNP6	TT or CT *vs* CC	0.90 (0.64–1.27)	0.56	0.81 (0.57–1.17)	0.27	2.38 (0.71–7.92)	0.16
SNP7	CC or AC *vs* AA	0.84 (0.59–1.19)	0.33	0.77 (0.53–1.11)	0.16	2.00 (0.59–6.72)	0.27
SNP22	TT or CT *vs* CC	1.99 (0.67–5.86)	0.21	NA[c]	NA	1.69 (0.43–6.68)	0.45
SNP3	GG or CG *vs* CC	0.90 (0.63–1.27)	0.54	0.81 (0.56–1.19)	0.28	2.23 (0.76–6.54)	0.14
CYP3A4							
SNP47	TT or CT *vs* CC	0.59 (0.31–1.09)	0.09	0.60 (0.29–1.23)	0.16	0.56 (0.17–1.86)	0.34
SNP12	GG or CG *vs* CC	1.51 (0.92–2.50)	0.11	1.44 (0.86–2.38)	0.16	NA	NA
SNP11	GG or AG *vs* AA	0.76 (0.43–1.36)	0.36	0.83 (0.41–1.66)	0.59	0.61 (0.23–1.63)	0.32
SNP1	GG or TG *vs* TT	0.53 (0.29–0.99)	0.05	0.57 (0.28–1.18)	0.13	0.44 (0.13–1.57)	0.21
SNP13	TT or CT *vs* CC	0.79 (0.51–1.22)	0.29	0.71 (0.45–1.12)	0.14	2.33 (0.42–12.84)	0.33
SNP24	GG or TG *vs* TT	0.95 (0.62–1.44)	0.79	0.88 (0.56–1.36)	0.56	1.81 (0.45–7.25)	0.40
SNP25	TT or CT *vs* CC	1.59 (0.58–4.39)	0.37	NA	NA	1.21 (0.37–3.98)	0.75
SNP5	TT or AT *vs* AA	0.86 (0.56–1.31)	0.47	0.74 (0.48–1.15)	0.19	4.48 (0.67–30.07)	0.12
SNP15	CC or GC *vs* GG	0.69 (0.46–1.05)	0.09	0.68 (0.44–1.05)	0.08	0.82 (0.22–3.03)	0.77
SRD5A2							
SNP17	TT or CT *vs* CC	0.87 (0.58–1.29)	0.48	0.93 (0.61–1.41)	**0.74**	0.21 (0.04–1.12)	0.07
SNP26	AA or GA *vs* GG	1.57 (1.08–2.30)	0.02	1.59 (1.08–2.34)	**0.02**	1.00 (0.19–5.31)	1.00
SNP22	AA or GA *vs* GG	0.84 (0.38–1.85)	0.66	0.90 (0.40–2.02)	0.79	NA	NA
SNP20	CC or GC *vs* GG	1.56 (1.08–2.25)	0.02	1.47 (1.00–2.16)	**0.05**	2.29 (0.81–6.50)	0.12
SNP12	TT or CT *vs* CC	1.00 (0.69–1.46)	0.98	0.98 (0.67–1.44)	0.94	0.94 (0.18–4.97)	0.94
SNP1	AA or GA *vs* GG	0.81 (0.53–1.24)	0.33	0.83 (0.53–1.31)	0.43	1.20 (0.23–6.21)	0.83
SNP13	AA or GA *vs* GG	0.94 (0.61–1.47)	0.80	0.98 (0.61–1.55)	0.92	1.64 (0.25–10.54)	0.61
SNP15	CC or GC *vs* GG	1.14 (0.79–1.63)	0.49	1.14 (0.79–1.65)	0.49	0.77 (0.15–3.94)	0.75

(b)

		All Subjects (n = 443–465)		European Americans (n = 394–418)		African Americans (n = 39)	
SNP	Stratification	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
CYP3A4							
SNP47	Low TNM and grade	0.19 (0.06–0.62)	**0.006**	0.07 (0.01–0.53)	0.10	0.66 (0.14–3.04)	0.59
SNP11	Low TNM and grade	0.20 (0.06–0.67)	**0.009**	0.08 (0.01–0.59)	0.13	0.66 (0.14–3.04)	0.59
SNP1	Low TNM and grade	0.21 (0.05–0.86)	**0.03**	0.16 (0.03–0.82)	**0.03**	0.65 (0.03–16.26)	0.80
SNP25	Low TNM and grade	6.54 (0.99–43.10)	**0.05**	NA[c]	NA	6.57 (1.26–34.17)	**0.03**
SNP5	Low TNM and grade	0.57 (0.30–1.10)	0.09	0.51 (0.26–0.99)	**0.05**	NA	NA
SNP15	Low TNM and grade	0.41 (0.22–0.79)	**0.007**	0.52 (0.27–1.01)	0.06	NA	NA
SRD5A2							
SNP1	Low TNM and grade	0.59 (0.32–1.10)	0.09	0.52 (0.27–1.00)	**0.05**	1.41 (0.18–10.79)	0.74

[a]From conditional logistic regression, with matching on family, and a variance estimator that incorporates sibling correlations.
[b]All results are from dominant models that compare homozygous and heterozygous carriers of variant *vs* the homozygous wildtype (OR = 1.0).
[c]NA, data not available.
Bold signifies statistically significant *P*-values.

When we stratified the study population by high and low aggressiveness of prostate cancer, several interesting associations emerged (see Tables 2b and 3b). First, five SNPs in *CYP3A4* showed statistically significant associations with low aggressiveness, four of them showing an inverse association: CYP3A4_SNP11 (CYP3A4*1B) (OR = 0.20, 95% CI = 0.06–0.67; *P*-value = 0.009), CYP3A4_SNP47 (OR = 0.19, 95% CI = 0.06–0.62; *P*-value = 0.006), CYP3A4_SNP1 (OR = 0.21, 95% CI = 0.05–0.86; *P*-value = 0.03), and CYP3A4_SNP15 (OR = 0.41, 95% CI = 0.22–0.79; *P*-value = 0.007). One of the five SNPs, CYP3A4_SNP25, showed positive association with low aggressiveness

Table 3 (a) All nonstratified haplotype association results for *CYP17A1*, *CYP3A4*, and *SRD5A2*[a] (see Figure 1 for details of haplotypes); (b) Statistically significant haplotype associations obtained from analysis stratified by high aggressiveness (ie, high TNM stage or Gleason score) and low aggressiveness (ie, low TNM stage and Gleason score)[a]

Haplotype	All subjects (n = 920)		European Americans (n = 834)		African Americans (n = 76)	
	OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
(a)						
CYP17A1						
Hap1	1.0	—	1.0	—	1.0	—
Hap2	0.83 (0.61–1.12)	0.22	0.80 (0.58–1.10)	0.17	2.63 (0.45–15.33)	0.28
Hap3	1.07 (0.67–1.70)	0.78	1.09 (0.65–1.83)	0.74	1.41 (0.49–4.08)	0.52
Hap4	0.85 (0.56–1.31)	0.47	0.84 (0.51–1.40)	0.51	1.02 (0.43–2.42)	0.97
CYP3A4						
Hap1	1.0	—	1.0	—	1.0	—
Hap2	1.25 (0.74–2.08)	0.41	1.16 (0.69–1.96)	0.57	NA[b]	NA
Hap3	1.20 (0.70–2.03)	0.51	1.07 (0.62–1.82)	0.82	3.34 (0.49–22.89)	0.22
Hap4	0.46 (0.21–1.01)	**0.05**	0.44 (0.20–0.96)	**0.04**	0.99 (0.06–16.37)	0.99
Hap5	1.08 (0.78–1.50)	0.66	1.05 (0.74–1.51)	0.77	1.86 (0.60–5.75)	0.28
SRD5A2						
Hap1	1.0	—	1.0	—	1.0	—
Hap2	1.14 (0.82–1.60)	0.43	1.12 (0.80–1.58)	0.50	2.57 (0.43–15.52)	0.30
Hap3	0.76 (0.48–1.21)	0.25	0.81 (0.51–1.30)	0.39	NA	NA
Hap4	1.13 (0.72–1.77)	0.61	1.03 (0.64–1.66)	0.90	NA	NA
Hap5	1.59 (0.78–3.24)	0.20	1.58 (0.79–3.19)	0.20	NA	NA
Hap6	1.27 (0.60–2.68)	0.52	2.16 (0.87–5.37)	0.10	0.64 (0.10–4.00)	0.63
Hap7	0.74 (0.50–1.09)	0.13	0.80 (0.51–1.23)	0.30	1.11 (0.29–4.27)	0.88

Haplotype	Stratification	All Subjects (n = 395–465)		European Americans (n = 362–418)		African Americans (n = 33–39)	
		OR (95% CI)	P-value	OR (95% CI)	P-value	OR (95% CI)	P-value
(b)							
CYP3A4							
Hap4	Low TNM and grade	0.06 (0.008–0.50)	**0.009**	0.09 (0.01–0.68)	**0.02**	NA[b]	NA
SRD5A2							
Hap3	High TNM or grade	0.52 (0.29–0.91)	**0.02**	0.53 (0.30–0.95)	**0.03**	NA	NA

[a]From conditional logistic regression, with matching on family, and a variance estimator that incorporates sibling correlation.
[b]NA, data not available.
Bold signifies statistically significant P-values.

(OR = 6.54, 95% CI = 0.99–43.10; *P*-value = 0.05). Second, we observed an inverse association between CYP3A4_Hap4 and low aggressiveness (OR = 0.06, 95% CI = 0.008–0.50; *P*-value = 0.009) (Table 3b). Finally, we detected an inverse association between SRD5A2_Hap3 and high aggressiveness (OR = 0.52, 95% CI = 0.29–0.91; *P*-value = 0.02) (Table 3b).

Discussion

The motivation for studying the relation between prostate cancer and the three candidate genes involved in the testosterone biosynthetic pathway (*CYP17A1*, *CYP3A4*, and *SRD5A2*) comes from several observations. First, the prostate is an androgen-dependent organ. Second, prostate

cancer typically strikes at an old age – after decades of exposure to testosterone. Third, men castrated at an early age do not develop prostate cancer. And, finally, African Americans experience earlier puberty,[35] higher serum levels of total testosterone,[36] and higher incidence of prostate cancer.

Our study did not detect any associations with *CYP17A1*, including the previously reported association between prostate cancer and the −34 T>C promoter SNP in *CYP17A1* (CYP17_SNP29 in this study).[22,25–27,29,31] There are also other negative reports for this promoter SNP,[37–39] making its role in the development of prostate cancer controversial. We did, however, detect an association with a number of *CYP3A4* genotypes and haplotypes. Interestingly, all but one of the statistically significant findings in

CYP3A4 showed protective effects with the associated minor alleles.

These included CYP3A4*1B[21] (CYP3A4_SNP11 in our study), which was inversely associated when we stratified the analysis by low disease aggressiveness, as previously identically reported from this population.[40] Note that most of the other *positive* associations for CYP3A4*1B have come from case-only studies, when comparing men with more aggressive to those with less aggressive disease.[21,24] We also found CYP3A4_Hap4 to be inversely associated with low aggressiveness. CYP3A4_Hap4 differs from the most common haplotype (Hap1) by seven SNPs, including SNP47, SNP11, SNP1, and SNP15, all of which showed inverse genotype-level associations in our stratified analyses. Moreover, the CYP3A4_Hap4 association was more statistically significant than any of the genotype-level associations. Since the associated SNPs are not necessarily exclusive to the associated haplotype, it is difficult to discern which may be the causal variant, or whether they may all simply be in linkage disequilibrium with – and on the same haplotype as – the causal, yet unknown, allele. Furthermore, since many of our SNPs are in linkage disequilibrium, some associations we saw may not represent independent effects, but rather reflect association with another linked SNP.

In addition to the findings in *CYP3A4*, we detected positive associations between prostate cancer risk and two SNPs in *SRD5A2*. One of these is a novel association (SRD5A2_SNP26), whereas the other (SRD5A2_SNP20, or V89L) has already been reported in the current and previous populations (30; personal communication). However, due to the extremely high linkage disequilibrium between the two SNPs, we cannot distinguish which (if either) may be causal for disease. We did not confirm the previously reported[23,28] association between prostate cancer and SRD5A2_SNP22 (A49T), possibly due to the low frequency (3.6%) of this SNP. Finally, we observed an inverse association between the high aggressiveness of prostate cancer and the SRD5A2_Hap3; this differs from the most common haplotype (Hap1) by two SNPs, both of which showed suggestive case–control differences in Phase I (data not shown).

The complementary approach used here to discover SNPs within the three candidate genes entailed resequencing a diversity panel and a disease-specific population, and searching in public databases. Using a diverse population in SNP discovery efforts is important because they should lead to the detection of a large number of SNPs. On the other hand, using disease-specific populations may reveal mutations not seen in a diversity set. By using both types of populations, we were able to discover 34 novel SNPs in the three genes. Moreover, studying these populations was crucial because more than one-third of the SNPs selected from public databases were monoallelic in our study group.

In our study, we included 53 SNPs from the public databases; 19 of them did not show up in our population. Furthermore, we did not have sequence coverage for 15 of the 34 biallelic public SNPs. Of the remaining 19 public SNPs, we rediscovered 11 (58%). We missed three public SNPs that had a reasonable allele frequency even though we had sequence coverage for those SNPs. In addition, we missed five public SNPs with very low minor allele frequencies. Several of these missed SNPs were located at the very ends of our contigs, making it less likely that we had sequence coverage in both directions. We typically followed through only variants that were seen in both directions.

We restricted our consideration to SNPs with allele frequencies $\geqslant 5\%$ and haplotypes with frequency $\geqslant 3\%$ in at least one of the following subpopulations: European American controls, European American cases, African American controls, and African American cases. This restriction was made because we would have severely limited statistical power to detect associations at low frequencies. We used a slightly lower frequency cutoff for haplotypes to try and distinguish those that might be carrying a causal variant.

To help confirm the controls' disease status, we determined the PSA levels in their blood. If they had PSA levels above 4 ng/ml, they were informed of this and advised that they should schedule an appointment with their primary care physician to further evaluate this test result. We retained such individuals in the study as controls unless a subsequent diagnosis of prostate cancer was made, at which time they were reclassified as cases. Keeping them in the study is important because automatically excluding men with elevated PSA levels regardless of their ultimate prostate cancer status can lead to biased estimates of association.[41,42]

Since prostate cancer risk varies by population, we adjusted our analyses by ethnicity (ie, European American and African American). Unfortunately, the relatively small number of African Americans studied here (96 total) resulted in unstable estimates of association within this ethnic group. Nevertheless, by using a sibling-based study design, we are assured that our controls have been ascertained from the cases' genetic source population, excluding the potential for bias due to population stratification.[43]

A two-phase design such as used here allows one to include a large number of candidate SNPs (87 here) and initially screen them in a relatively small sample set (276 men here). Once the relevant SNPs are identified (24 here), resources can be focused on genotyping a larger sample set (841 additional men here) in order to achieve adequate statistical power to detect the potential associations. This reveals the major haplotypes with reasonable resources, and hence may be a valuable approach for future studies. The number of subjects studied in each phase, however,

will depend upon the haplotype structure of one's chromosomal region of interest, the expected frequency of causal genotypes or haplotypes, and the magnitude of their impact on disease.

In summary, we have detected a number of intriguing associations between prostate cancer and genotypes/haplotypes in *CYP3A4* and *SRD5A2*. These results need to be confirmed in other, independent and ethnically diverse populations. If confirmed, functional studies would be valuable for deciphering the potential causal impact of these SNPs on prostate cancer. Deciphering which of these truly impact disease susceptibility would prove invaluable for screening purposes, and finding aggressiveness genes may provide important information about the most appropriate recommended course of treatment among men already diagnosed with prostate cancer.

References
1 Smith DS, Bullock AD, Catalona WJ, Herschman JD: Racial differences in a prostate cancer screening study. *J Urol* 1996; **156**: 1366–1369.
2 Hsing AW, Tsao L, Devesa SS: International trends and patterns of prostate cancer incidence and mortality. *Int J Cancer* 2000; **85**: 60–67.
3 McLellan DL, Norman RW: Hereditary aspects of prostate cancer. *CMAJ* 1995; **153**: 895–900.
4 Neville PJ, Casey G, Witte JS: Hereditary prostate cancer and genetic risk; in Klein (ed): *Management of Prostate Cancer*. New Jersey: Humana Press, Inc., 2003, 2nd edn.
5 Tavtigian SV, Simard J, Teng DH *et al*: A candidate prostate cancer susceptibility gene at chromosome 17p. *Nat Genet* 2001; **27**: 172–180.
6 Rebbeck TR, Walker AH, Zeigler-Johnson C *et al*: Association of HPC2/ELAC2 genotypes and prostate cancer. *Am J Hum Genet* 2000; **67**: 1014–1019.
7 Suarez BK, Gerhard DS, Lin J *et al*: Polymorphisms in the prostate cancer susceptibility gene HPC2/ELAC2 in multiplex families and healthy controls. *Cancer Res* 2001; **61**: 4982–4984.
8 Rokman A, Ikonen T, Mononen N *et al*: ELAC2/HPC2 involvement in hereditary and sporadic prostate cancer. *Cancer Res* 2001; **61**: 6038–6041.
9 Vesprini D, Nam RK, Trachtenberg J *et al*: HPC2 variants and screen-detected prostate cancer. *Am J Hum Genet* 2001; **68**: 912–917.
10 Wang L, McDonnell SK, Elkins DA *et al*: Role of HPC2/ELAC2 in hereditary prostate cancer. *Cancer Res* 2001; **61**: 6494–6499.
11 Xu J, Zheng SL, Carpten JD *et al*: Evaluation of linkage and association of HPC2/ELAC2 in patients with familial or sporadic prostate cancer. *Am J Hum Genet* 2001; **68**: 901–911.
12 Carpten J, Nupponen N, Isaacs S *et al*: Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. *Nat Genet* 2002; **30**: 181–184.
13 Casey G, Neville PJ, Plummer SJ *et al*: RNASEL Arg462Gln variant is implicated in up to 13% of prostate cancer cases. *Nat Genet* 2002; **32**: 581–583.
14 Rokman A, Ikonen T, Seppala EH *et al*: Germline alterations of the RNASEL gene, a candidate HPC1 gene at 1q25, in patients and families with prostate cancer. *Am J Hum Genet* 2002; **70**: 1299–1304.
15 Wang L, McDonnell SK, Elkins DA *et al*: Analysis of the RNASEL gene in familial and sporadic prostate cancer. *Am J Hum Genet* 2002; **71**: 116–123.
16 Colombel M, Olsson CA, Ng PY, Buttyan R: Hormone-regulated apoptosis results from reentry of differentiated prostate cells onto a defective cell cycle. *Cancer Res* 1992; **52**: 4313–4319.
17 Gupta S, Ahmad N, Mohan RR, Husain MM, Mukhtar H: Prostate cancer chemoprevention by green tea: *in vitro* and *in vivo* inhibition of testosterone-mediated induction of ornithine decarboxylase. *Cancer Res* 1999; **59**: 2115–2120.
18 Carey AH, Waterworth D, Patel K *et al*: Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. *Hum Mol Genet* 1994; **3**: 1873–1876.
19 Waxman DJ, Lapenson DP, Aoyama T, Gelboin HV, Gonzalez FJ, Korzekwa K: Steroid hormone hydroxylase specificities of eleven cDNA-expressed human cytochrome P450s. *Arch Biochem Biophys* 1991; **290**: 160–166.
20 Makridakis N, Ross RK, Pike MC *et al*: A prevalent missense substitution that modulates activity of prostatic steroid 5alpha-reductase. *Cancer Res* 1997; **57**: 1020–1022.
21 Rebbeck TR, Jaffe JM, Walker AH, Wein AJ, Malkowicz SB: Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. *J Nat Cancer Inst* 1998; **90**: 1225–1229.
22 Lunn RM, Bell DA, Mohler JL, Taylor JA: Prostate cancer risk and polymorphism in 17 hydroxylase (CYP17) and steroid reductase (SRD5A2). *Carcinogenetics* 1999; **20**: 1727–1731.
23 Makridakis NM, Ross RK, Pike MC *et al*: Association of missense substitution in SRD5A2 gene with prostate cancer in African-American and Hispanic men in Los Angeles, USA. *Lancet* 1999; **354**: 975–978.
24 Paris PL, Kupelian PA, Hall JM *et al*: Association between a CYP3A4 genetic variant and clinical presentation in African-American prostate cancer patients. *Cancer Epidemiol Biomarkers Prev* 1999; **8**: 901–905.
25 Wadelius M, Andersson AO, Johansson JE, Wadelius C, Rane E: Prostate cancer associated with CYP17 genotype. *Pharmacogenetics* 1999; **9**: 635–639.
26 Gsur A, Bernhofer G, Hinteregger S *et al*: A polymorphism in the CYP17 gene is associated with prostate cancer risk. *Int J Cancer* 2000; **87**: 434–437.
27 Habuchi T, Liqing Z, Suzuki T *et al*: Increased risk of prostate cancer and benign prostatic hyperplasia associated with a CYP17 gene polymorphism with a gene dosage effect. *Cancer Res* 2000; **60**: 5710–5713.
28 Jaffe JM, Malkowicz SB, Walker AH *et al*: Association of SRD5A2 genotype and pathological characteristics of prostate tumors. *Cancer Res* 2000; **60**: 1626–1630.
29 Kittles RA, Panguluri RK, Chen W *et al*: CYP17 promoter variant associated with prostate cancer aggressiveness in African Americans. *Cancer Epidemiol Biomarkers Prev* 2001; **10**: 943–947.
30 Nam RK, Toi A, Vesprini D *et al*: V89L polymorphism of type-2,5-alpha reductase enzyme gene predicts prostate cancer presence and progression. *Urology* 2001; **57**: 199–204.

31 Stanford JL, Noonan EA, Iwasaki L *et al*: A polymorphism in the CYP17 gene and risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev* 2002; **11**: 243–247.

32 Collins FS, Brooks LD, Chakravarti A: A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res* 1998; **8**: 1229–1231.

33 Stephens M, Smith NJ, Donnelly P: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**: 978–989.

34 Breslow NE, Day NE: Statistical methods in cancer research. Vol I – The analysis of case–control studies. *IARC Sci Publ* 1980; **32**: 5–338.

35 Herman-Giddens M, Wang L, Koch G: Secondary sexual characteristics in boys: estimates from the national health and nutrition examination survey III, 1988–1994. *Arch Pediatr Adolescent Med* 2001; **155**: 1022–1028.

36 Ross R, Bernstein L, Judd H, Hanisch R, Pike M, Henderson B: Serum testosterone levels in healthy young black and white men. *J Natl Cancer Inst* 1986; **76**: 45–48.

37 Chang B, Zheng SL, Isaacs SD *et al*: Linkage and association of CYP17 in hereditary and sporadic prostate cancer. *Int J Cancer* 2001; **95**: 354–359.

38 Haiman CA, Stampfer MJ, Giovannucci E *et al*: The relationship between a polymorphism in CYP17 with plasma hormone levels and prostate cancer. *Cancer Epidemiol Biomarkers Prev* 2001; **10**: 743–748.

39 Ntais C, Polycarpou A, Ioannidis JP: Association of the CYP17 gene polymorphism with the risk of prostate cancer: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2003; **12**: 120–126.

40 Plummer SJ, Conti DV, Paris PL, Curran AP, Casey G, Witte JS: CYP3A4 and CYP3A5 genotypes, haplotypes and risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev* 2003; **12**: 928–932.

41 Lubin JH, Hartge P: Excluding controls: misapplications in case–control studies. *Am J Epidemiol* 1984; **120**: 791–793.

42 Poole C: Controls who experienced hypothetical causal intermediates should not be excluded from case–control studies. *Am J Epidemiol* 1999; **150**: 547–551.

43 Thomas DC, Witte JS: Population stratification: a problem for case–control studies of candidate gene associations? *Cancer Epidemiol Biomarkers Prev* 2002; **11**: 505–512.