Association Between Genetic Polymorphisms in the Prostate-Specific Antigen Gene Promoter and Serum Prostate-Specific Antigen Levels


Background: Recent evidence suggests that genetic variation in the promoter of the prostate-specific antigen (PSA) gene may contribute to individual variation in serum PSA levels. However, polymorphisms associated with variations in PSA levels have not been identified. Methods: We used the polymerase chain reaction to amplify the promoter region of the PSA gene (nucleotide positions −3873 to −5749 with respect to the start of transcription) of 409 healthy white men at risk for lung disease. Polymerase chain reaction products were sequenced to identify polymorphisms in the PSA gene promoter and to genotype the men for common single nucleotide polymorphisms (SNPs) and were cloned into luciferase reporter constructs to assay PSA promoter activity in human LNCaP prostate cancer cells. Analysis of variance was used to test the association of polymorphism frequencies with mean serum PSA levels. All statistical tests were two-sided. Results: The −4643G/A SNP (G allele) had a 21.2% prevalence and was associated with increases in serum PSA levels (P = .017) and PSA promoter activity (P < .001). The −5412C/T SNP (C allele) had a 22.0% prevalence and was associated with an increase in serum PSA levels (P = .0015). The −5429T/G SNP (G allele) had a 23.0% prevalence, was associated with an increase in serum PSA levels (P = .021), and was in linkage disequilibrium with the −5412C/T SNP. The promoter activity of the −5412 C−5429 G haplotype was higher than that of the −5412 T−5429 T haplotype (P < .001). Conclusions: Genetic variations in the PSA promoter are associated with serum PSA levels in men without prostatic disease. PSA promoter genotype information may help to refine models of PSA cutoff values. [J Natl Cancer Inst 2003; 95:1044–53]

Prostate-specific antigen (PSA) is an androgen-regulated serine protease produced by secretory epithelial cells that line the lumen of normal prostatic glands and by most prostate cancers (1–3). The prostate is the major site of PSA expression in men (1,3–5). Consequently, PSA expression has become the most widely used marker for prostate cancer screening and for patients’ responses to therapeutic interventions. Many clinicians consider a serum PSA concentration greater than 4 ng/mL to be an indicator of a potential prostatic abnormality, and they recommend that men with such PSA levels undergo further screening by prostate needle biopsy. However, PSA testing has a low sensitivity and specificity for detecting prostate cancer (6). Factors that may contribute to this low sensitivity and specificity include the presence of any non-cancerous prostatic disease (i.e., prostatitis or benign prostatic hyperplasia), age, and race. The contributions of age and race to serum PSA levels have led some to propose the use of age- and race-specific cutoff values for serum PSA testing (7,8). However, there remains considerable controversy regarding the use of such cutoffs as criteria for further testing because some studies have reported a decreased sensitivity for prostate cancer detection when such cutoffs are used (9,10). Indeed, identifying which patients with PSA levels between 2 and 4 ng/mL should undergo needle biopsy screening is one current area of focus (11,12).

The PSA gene contains a 6-kb promoter in the 5′ region that contributes to tissue and hormone specificity of PSA expression (13–15). This promoter contains androgen-responsive elements (AREs) that regulate promoter activity by binding to androgen receptors. ARE I and ARE II are located in the proximal region of the PSA promoter and are centered at −70 base pairs (bp) and −394 bp, respectively, with respect to the transcription start site (16). ARE III is located in the 5′ upstream enhancer region and is centered at −4200 bp with respect to the transcription start site (13–15). ARE I and ARE III have high affinities for the androgen receptor (13,14,16–20), whereas ARE II has a low affinity for the androgen receptor (13). Recent data demonstrate the presence of additional high-, medium-, and low-affinity AREs within the 5′ upstream enhancer region of the PSA promoter between −3870 bp and −4366 bp with respect to the transcription start site (20). Other areas of the 5′ upstream region of the PSA gene may be important for PSA expression, but they are poorly characterized. Indeed, few reports have evaluated the contributions of sequences upstream of a unique XbaI restriction
site located at –5322 bp with respect to the start of transcription of the PSA gene to PSA promoter activity, largely because that site has been used to clone promoter constructs.

We previously identified a specific genetic polymorphism in ARE I (21) that was subsequently found to be associated with serum PSA level (22). This single nucleotide polymorphism (SNP), –158 G/A, is a G to A change at position –158 bp with respect to the start of transcription; the two alleles are found at approximately equal frequencies among whites (21). Xue et al. (22) reported that the A allele is associated with increased serum PSA levels in healthy men. This polymorphism has also been associated with an increased risk for the development of prostate cancer (23,24). These data suggest that the –158 G/A polymorphism directly contributes to differences in PSA gene promoter activity. However, we recently found that this polymorphism was not associated with serum PSA level in two separate groups of men without prostate cancer (25,26). We also assessed the in vitro activity of PSA gene promoter constructs that differed only by the –158 G/A polymorphism and found no contributions of this SNP to differences in PSA gene promoter activity (26). Those data suggest that previous associations of the –158 G/A polymorphism with serum PSA level reported by others are likely to be due to linkage disequilibrium (the dependence of an allele at one locus on alleles at another locus) of the –158 G/A polymorphism with other polymorphisms in the PSA gene and its promoter. In this study, we further characterized the PSA gene for polymorphisms and examined the associations of these sequence variations with serum PSA levels and PSA gene promoter activity.

**Subjects and Methods**

**Human Subjects**

The subjects in this study were a previously described subset of a population of asbestos workers who were recruited for a study of the interaction of asbestos exposure with genetic and environmental factors in the induction of asbestos-induced lung diseases (25). This subset consisted of 518 male painters, plumbers, pipe fitters, heavy-equipment operators, and electricians whose PSA levels were assessed to examine their risk of prostate cancer after asbestos exposure. All subjects gave written informed consent, received a physical examination, and provided complete medical and occupational histories. Whole blood collected from each subject at the physical examination was used for DNA isolation (25) and to determine serum PSA levels (25). We excluded the 49 African-American subjects in that subset from our study because of insufficient numbers to make reliable statistical predictions about SNP associations with serum PSA. We also excluded 27 subjects who had been diagnosed with prostate cancer and 14 subjects who had undetectable levels of PSA (≤0.1 ng/mL). DNA samples from the remaining 428 subjects were previously amplified by polymerase chain reaction (PCR) and sequenced to genotype them for SNPs in ARE I (17). The frequency of this SNP is included here for reference. Nineteen DNA samples failed to yield a PCR product for the target region or had insufficient DNA for amplification. The analyses of genotypes and associations with serum PSA level were conducted on the remaining 409 samples. The 409 study subjects from whom those samples were obtained had a mean (± standard deviation) age of 63.7 (±9.1) years and a median serum PSA concentration of 1.01 ng/mL (range = 0.14–20.7 ng/mL). Three hundred thirty-eight subjects had a serum PSA concentration less than 2.5 ng/mL, and 71 subjects had a PSA concentration of 2.5 ng/mL or higher. Thirty-six subjects had a PSA concentration of 4 ng/mL or higher, and 11 of these subjects had a PSA concentration of 9 ng/mL or higher. The research protocol was approved by the Saint Louis University and Wake Forest University Institutional Review Boards. Additional information on this study set was reported previously (25).

**PCR Amplifications**

We used nested PCR to amplify a 1.9-kb region of the PSA gene encompassing nucleotides –3875 to –5749 relative to the transcription start site [all numbering of the PSA gene is as reported by Schuur et al. (14)]. We used 25–70 ng of genomic DNA extracted from each study subject’s peripheral lymphocytes as template in a 100-μL PCR volume. In the first set of reactions, each tube contained 1× Thermophilic DNA Polymerase buffer (Promega, Madison, WI), 2.5 mM MgCl2, 100 μM of each dNTP (Promega), 150 nM of each oligonucleotide primer, 5 U of Taq DNA Polymerase (Promega), and 0.2 U of Vent DNA polymerase (New England Biolabs, Beverly, MA). The sequences of the 5′ and 3′ primers were 5′-TTTGGCAGTGAGT GCTGC-3′ and 5′-GCTTGGGAATATCCCTGCAGG-3′, respectively. In the first set of reactions, the samples were heated to 94°C for 5 minutes and then to 80°C for 10 minutes. The polymerases were added to the reaction after the first minute at 80°C. The reactions were then subjected to 30 cycles of 95°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute. A final extension was performed at 72°C for 7 minutes. In the second (i.e., nested) set of PCRs, 10 μL of the first reaction was used as template with internal 5′ primer (5′-ATGAATTCGTCGACCACA GTGTAATGCCATCCAGG-3′) and 3′ primer (5′-ATAGGATCC AGACTGTCCCTGCAGACAAGG-3′), which introduced unique SalI and BamHI restriction sites (underlined), respectively, into the PCR products. All reaction conditions were identical for the nested amplifications, except that after the 10-minute incubation at 80°C, we subjected the reactions to an initial three cycles of 94°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute, which were followed by 27 cycles of 94°C for 1 minute, 58°C for 1 minute, and 72°C for 1 minute. Amplified samples were stored at 4°C and then used for genotyping by DNA sequencing or for constructing luciferase reporter constructs, as described below.

**Genotyping by DNA Sequencing**

Sequence variants were identified by sequencing 20 randomly selected samples of PCR-amplified DNA. DNA sequencing was performed with the use of a BigDye Terminator sequencing kit (Applied Biosystems, Foster City, CA). Each sequencing reaction plate contained PCR-amplified DNA from subjects with various ranges of serum PSA levels as well as two DNA samples of known genotypes, present in duplicate, and two blanks (no template). PCR products were purified using a Quickstep 96-well PCR purification kit (Edge Biosystems, Gaithersburg, MD) and stored in water at –20°C for later sequencing. Each 10-μL sequencing reaction contained 10–50 ng of purified PCR product, 1.5 pmol of sequencing primer (one of the 10 primers listed below), 1 μL of BigDye Terminator mix, and 1.5 μL of 5× sequencing dilution buffer (400 mM Tris–HCl [pH 9.0], 10 mM MgCl2). Cycling conditions were 94°C for 1 minute, followed by 25 cycles of 94°C for 30 seconds, 50°C for 30 seconds, and 60°C for 2 minutes.

Journal of the National Cancer Institute, Vol. 95, No. 14, July 16, 2003 ARTICLES 1045
for 30 seconds, and 60 °C for 4 minutes, and ending with a single
72 °C extension step for 5 minutes. Sequencing products were
ethanol precipitated, air dried, resuspended in 25 μL of H2O,
and analyzed on a 3700 DNA Analyzer (Applied Biosystems).
DNA sequence data were aligned and polymorphisms were
identified using Sequencher DNA analysis software (Gene
Codes Corporation, Ann Arbor, MI). Oligonucleotide primers
used for sequencing were 5'-CCTTCAGGTGAAACAGG-3',
5'-AGACCGAGGACACTCTGG-3', 5'-TACATTAGTACC
TTGCCC-3', 5'-TAGACTGCTCTGTCAACCCC-3', 5'-GAAGAC
GGACATCAGGCC-3', and 5'-GCTTTGGAATATCCTGCCAG
-3'. We used the internal PCR primers to directly sequence the 5'
and 3' ends of the PCR products, and T3 (5'-AATTAACCTC
ACTAAGGG-3') and T7 (5'-GTAATACGACTCATACTAG
GGA-3') primers to sequence the 5' and 3' ends, respectively, of
subcloned PCR products (see below).

**Luciferase Reporter Constructs**

Genomic DNA from individuals homozygous for specific
SNPs (identified by DNA sequencing as described above) were
used as templates to amplify and clone specific haplotypes using
the primers and conditions described above for the nested PCR.
PCR products were digested sequentially with SalI and BamHI
(Promega) and cloned into the SalI and BamHI sites of pBlue-
script SKII (Promega). Plasmid DNA was isolated from positive
clones (i.e., clones containing an insert of the expected size)
and analyzed by DNA sequencing as described above. The nucleo-
tide sequence from the cloned product was compared with the
deduced sequence from the PCR product. Only clones that were
identical to the genomic sequence were used for subsequent steps.
Positive clones were digested with KpnI and SacI (Promega) to release
1.9-kb inserts containing the PSA gene. The inserts were subcloned into the
KpnI and SacI sites of the lucif-
erase reporter vector pGL3Basic (Promega). pGL3Basic vector used to make the constructs was modified as previously de-
scribed (26). This modified vector had at the HindIII site, a
525-bp fragment of the proximal PSA promoter, including ARE
I (158 G allele), ARE II, and the transcription start site driving
a 525-bp fragment of the proximal PSA promoter, including ARE
used to make the constructs was modified as previously de-
**Luciferase and β-Galactosidase Assays**

All experiments were conducted using the human LNCaP
prostate cancer cell line (American Type Culture Collection,
Manassas, VA) as previously described (26). Briefly, the cells
were plated at 1.5 × 10^5 cells per well in six-well tissue culture
plates in RPMI-1640 medium supplemented with 10% fetal bo-
vine serum (FBS) (Sigma Chemical Co., St. Louis, MO). Forty-
eight hours later, the medium was removed from each well and
1 mL of transfection cocktail—1.25 μL of pCMV-β-gal plasmid DNA
(Pro-

ducer) plasmid DNA, 0.25 μg of pCMV-β-gal plasmid DNA (Pro-

gal vector was not changed by treatment with androgens (data
not shown). Each experimental condition was performed in six
replicate wells (two wells on each of three separate culture
plates). The experiments were repeated twice. Results are ex-
pressed as the mean number of luminometer units per unit of
β-galactosidase activity with 95% confidence intervals.

**Statistical Methods**

Hardy–Weinberg equilibrium tests for all genotyped SNPs
and pairwise linkage disequilibrium tests for all pairs of geno-
typed SNPs were performed using the Genetic Data Analysis
computer program (27) and SAS/Genetics software (version 2002; SAS Institute, Cary, NC). Hardy–Weinberg equilibrium
tests were based on exact tests, wherein a large number of the
possible arrays were generated by permuting the alleles among
genotypes, and the proportion of these permutted genotypic
arrays that have a smaller conditional probability than the original
data were calculated. Tests for pairwise linkage disequilibrium
were based on an exact test, assuming multinomial probability of
the multilocus genotype, conditional on the single-locus geno-
type. A Monte Carlo simulation was used to assess the statistical
significance of the observed test value by permuting the single-
locus genotypes among individuals in the sample to simulate the
null distribution. The empirical P values of both the Hardy–
Weinberg equilibrium and linkage disequilibrium tests were
based on 10,000 replicate samples. Lewontin’s D' was used to
estimate the strength of pairwise linkage disequilibrium (28).

The distribution of serum PSA levels deviated statistically
significantly from a normal distribution (Cramer–von Mises
W-Sq statistic = 11.53003; P = .005). Therefore, PSA levels
were log 10-transformed. After the transformation, the distribu-
tion approached normality but remained statistically signifi-
cantly different from a normal distribution (W-Sq statistic =
0.18; P = .011). Analysis of variance tests were therefore per-
formed to test for differences in mean values for log PSA levels among men with different genotypes for each SNP. Multiple regression models, adjusted for age, were used to estimate the effects of the genotypes by comparing men who were heterozygous or homozygous for the less frequent alleles with men who were homozygous for the more common alleles. To decrease the potential population stratification, all analyses were limited to white subjects.

Haplotype frequency was estimated using the statistical method of Devlin and Risch (28), as implemented in the computer program PHASE (http://www.stats.ox.ac.uk/mathgen/software.html). Association between the haplotypes and serum PSA level was estimated using a score test developed by Schaid et al. (29), as implemented in the computer program HAPLO.SCORE (http://www.mayo.edu/statgen) for the S-PLUS programming language or http://www.wfubmc.edu/docs/genomics for the R programming language. Age variation was modeled in the haplotype score test.

Transfection data were compared by using a two-way analysis of variance controlling for R1881 dose and haplotype of the expression construct, with post hoc analysis by the Tukey–Kramer test. All statistical tests were two-sided.

GenBank Identifiers

The reference PSA gene sequence used in this study has the GenBank accession number U37672. Unique SNP Cluster identification numbers for SNPs used in this study that are present in the SNP database are rs2569733, rs2739448, rs266868, rs266867, rs925013, and rs266882. GenBank accession numbers for SNPs and polynucleotide repeats that are not present in the SNP database are AY283612, AY283613, AY283614, AY283615, AY283616, and AY283617.

RESULTS

Association Between Sequence Variants in ARE III and Serum PSA Levels

We hypothesized that the previously reported finding that the −158 G/A SNP in ARE I was associated with serum PSA level probably reflected the linkage disequilibrium of this SNP with other polymorphisms in the PSA promoter. Our initial efforts to identify these putative polymorphisms focused on a region of the PSA gene that contains ARE III and is located from nucleotide positions −3800 to −4300 with respect to the start of transcription. Direct sequencing of PCR products amplified from the DNA of our study subjects identified two previously unreported polymorphisms in this region (Fig. 1). One of these, the −4289 A/C SNP, is located in a low-affinity, non-consensus ARE, termed ARE VI by Huang et al. (20). The C allele of this SNP had an estimated frequency of 20.9% among the subjects in our study (Table 1) and was associated with elevated PSA levels. Men with the AC or CC genotype at this SNP had statistically significantly higher PSA levels than men with the AA genotype ($P < .017$, age-adjusted model, Table 2). Excluding men who had PSA levels of 9.0 ng/mL or higher did not affect this association ($P < .028$, Table 2). The other polymorphism we identified in this region was a polycytosine (polyC) tract that varied from 8 to 9 nucleotides in length and was centered at nucleotide position −4330 in the PSA promoter (Fig. 1). We did not perform association tests for this polymorphism because it was not in Hardy–Weinberg equilibrium among our study subjects (data not shown).

We next cloned representative examples of each polymorphism in the ARE III region; the haplotypes used for cloning differed only by the indicated change at the specific polymorphism (see graphical representations in Fig. 2). We used a luciferase reporter assay in human LNCaP prostate cancer cells to...
examine the promoter activities of the different haplotypes in the presence of the synthetic androgen R1881 or a vehicle control. All reporter constructs exhibited a dose-dependent increase in luciferase expression in response to increasing concentrations of the synthetic androgen R1881. However, neither the −4289 A/C SNP (Fig. 2, A) nor the −4330 polyC polymorphism (Fig. 2, B) affected PSA promoter activity in the presence or absence of R1881.

Association Between SNPs in the Far Upstream Region of the PSA Promoter and Serum PSA Levels

Our data suggested that one or more SNPs in other regions of the PSA promoter might be associated with serum PSA level. Therefore, we completely sequenced the remaining 1.2 kb of the 5' upstream region of the PSA gene using PCR-amplified DNA from 20 subjects to search for additional polymorphisms. We identified six SNPs with frequencies greater than .05 (−4643 A/G, −5217 T/A, −5307 G/A, −5412 T/C, −5429 T/G, and −5567 G/A) that we tested for associations with serum PSA level. In this region of the PSA promoter, we also identified a polyadenosine (polyA) tract at nucleotide position −5133 relative to the start of transcription that varied from 9 to 22 nucleotides in length and several other sequence variants that occurred in our subjects at a low frequency (i.e., <5%) Table 1, Fig. 1) The polyA repeat and the less frequent SNPs were not test for their association with serum PSA level. The entire spectrum of se-

### Table 1. Frequencies of single nucleotide polymorphisms (SNPs) in prostate-specific antigen (PSA) promoter region among white study subjects (N = 409)

<table>
<thead>
<tr>
<th>Location of SNP*</th>
<th>Nucleotide change</th>
<th>% chromosomes carrying SNP (N)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>−5567:AY283612</td>
<td>G to A</td>
<td>9.16 (74/808)</td>
</tr>
<tr>
<td>−5501:AY283614</td>
<td>T to C</td>
<td>2.23 (17/762)</td>
</tr>
<tr>
<td>−5466:AY283615</td>
<td>G to A</td>
<td>1.97 (15/761)</td>
</tr>
<tr>
<td>−5429:rs2569733</td>
<td>T to G</td>
<td>21.02 (196/908)</td>
</tr>
<tr>
<td>−5412:rs2739448</td>
<td>T to C</td>
<td>21.95 (176/802)</td>
</tr>
<tr>
<td>−5307:rs266868</td>
<td>G to A</td>
<td>29.78 (240/806)</td>
</tr>
<tr>
<td>−5217:rs266867</td>
<td>T to A</td>
<td>9.23 (74/802)</td>
</tr>
<tr>
<td>−4643:rs925013</td>
<td>A to G</td>
<td>21.22 (171/806)</td>
</tr>
<tr>
<td>−4289:AY283613</td>
<td>A to C</td>
<td>20.90 (171/818)</td>
</tr>
</tbody>
</table>

*Location relative to the first nucleotide of the transcription start site (nucleotide position +1): GenBank accession number or the National Center for Biotechnology Information SNP Cluster identification number.

†Number of chromosomes carrying variant/number of chromosomes tested.

### Table 2. Serum PSA levels and sequence variants in PSA gene promoter region among white study subjects*

<table>
<thead>
<tr>
<th>SNP/genotype†</th>
<th>N</th>
<th>Mean (PSA)</th>
<th>SD</th>
<th>PSA geometric mean</th>
<th>P‡</th>
<th>P§</th>
<th>P¶</th>
</tr>
</thead>
<tbody>
<tr>
<td>−158:rs266882 (ARE I)</td>
<td>109</td>
<td>−0.026</td>
<td>0.40</td>
<td>0.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−4289:AY283613 (ARE VI)</td>
<td>256</td>
<td>−0.006</td>
<td>0.39</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−4643:rs925013 (NcoI)</td>
<td>18</td>
<td>0.062</td>
<td>0.48</td>
<td>1.15</td>
<td>0.021</td>
<td>0.017</td>
<td>0.028</td>
</tr>
<tr>
<td>−5217:rs266867</td>
<td>331</td>
<td>0.049</td>
<td>0.40</td>
<td>1.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−5307:rs266868</td>
<td>198</td>
<td>0.09</td>
<td>0.42</td>
<td>1.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−5412:rs2739448 (RstUI)</td>
<td>246</td>
<td>−0.012</td>
<td>0.38</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−5429:rs2569733 (FokI)</td>
<td>66</td>
<td>0.049</td>
<td>0.42</td>
<td>1.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−5567:AY283612</td>
<td>4</td>
<td>−0.33</td>
<td>0.46</td>
<td>0.47</td>
<td>0.18</td>
<td>0.46</td>
<td>0.86</td>
</tr>
</tbody>
</table>

*PSA = prostate-specific antigen; SNP = single nucleotide polymorphism; ARE = androgen-responsive element; SD = standard deviation.

†Shown is the nucleotide position relative to transcription start site as defined by Schuur et al. (14): National Center for Biotechnology Information SNP Cluster identification number or GenBank accession number (unique features).

‡Analysis of variance test.

§Multivariate comparison between men being heterozygous or homozygous for least prevalent allele versus men being homozygous for most prevalent allele, adjusted for age.

¶Men with a serum PSA level ≥9.0 ng/mL were excluded.

††Data for this SNP, provided here for reference, were previously reported for the same study population used in our study plus an additional 11 samples (25).
Polymorphisms in the androgen-responsive element (ARE) III and prostate-specific antigen (PSA) promoter activity. Human LNCaP prostate cancer cells were transiently transfected with luciferase reporter constructs containing specific PSA gene haplotypes and with pCMV-β-gal and treated with the indicated concentrations of R1881 for 24 hours. At the end of the incubation, cells were harvested, and luciferase and β-galactosidase activities were determined as described in the “Subjects and Methods” section. Comparison of differences in relative light units between different promoter constructs was performed by analysis of variance; *P* < .05 was considered statistically significant. A) Comparison of the −4289 A/C single nucleotide polymorphism. B) Comparison of the −4330 polycytosine polymorphism. Left panels show the results of the luciferase assay normalized to β-galactosidase activity (in relative light units). Each point represents the mean of four to six replicates, and bars represent 95% confidence intervals. There are no statistically significant differences between curves in A or in B.

Association Between Multiple PSA Gene Promoter Polymorphisms and Serum PSA Levels

We examined associations between the six relatively common SNPs and serum PSA levels. Two SNPs (−5217T/A and −5567G/A) were not statistically significantly associated with serum PSA level. Of the remaining four SNPs, the −5307 A/G SNP was the most weakly associated with serum PSA level. Men with AG or GG genotypes at that SNP had statistically significantly higher PSA levels than men with the AA genotype (*P* = .017, age-adjusted model, Table 2). By contrast, SNPs −4643A/G, −5412T/C, and −5429T/G were strongly associated with serum PSA level (Table 2). The G allele of the −4643A/G SNP, which had an estimated frequency of 21.2% among our study subjects, was associated with increased PSA level. Men with the GA or GG genotype at this SNP had statistically significantly higher PSA levels than men with the AA genotype (*P* = .0095, age-adjusted model). In luciferase reporter assays, the G allele at −4643 displayed a statistically significantly more potent promoter activity than the A allele at all androgen concentrations tested (*P* < .001) (Fig. 3, A). The C allele of the −5412T/C SNP, which had an estimated frequency of 22.0% among our study subjects, was also associated with elevated PSA level. Men with a TC or CC genotype at this SNP had statistically significantly higher PSA levels than men with a TT genotype (*P* < .001, age-adjusted model). Similarly, the G allele of the −5429T/G SNP, which had an estimated frequency of 23.0% among our study subjects, was associated with elevated PSA levels. Men with a TG or GG genotype at this SNP had statistically significantly higher PSA levels than men with a TT genotype (*P* = .009, age-adjusted model). Excluding study subjects whose serum PSA levels were 9.0 ng/mL or higher did not substantially affect the statistical significance of the association results for any of the SNPs we evaluated (Table 2).

Haplotype analysis suggested that an estimated 98% of the haplotypes in our study subjects were either −5429G/−5412C or −5429T/−5412T (data not shown). We therefore cloned examples of each of these two haplotypes, which differed from each other at no other location, and examined their promoter activities in luciferase reporter assays. Fig. 3, B, demonstrates that the promoter activity of the −5429G/−5412C haplotype was statistically significantly higher than that of the −5429T/−5412T haplotype at all concentrations of R1881 tested (*P* < .001). In other reporter assays, we examined the interaction of these sites with other polymorphisms, including the −4289 A/C SNP and the −4330 polyC polymorphism. The −5429G/−5412C haplotype was associated with consistently higher promoter activity than the −5429T/−5412T haplotype, regardless of which other sequence variants were present (data not shown).
mated frequency of 20.0%, and the other consisted of the three alleles that were associated with reduced PSA levels and promoter activities (−5429T/−5412T/−4643A), with an estimated frequency of 77.0%. Because of this strong linkage disequilibrium, it was difficult to genetically dissect the contribution of each of the three SNPs to the association with PSA levels. Instead, we examined the association of each of the two haplotypes with serum PSA levels. As shown in Fig. 4, men who had at least one copy of the −5429G/−5412C/−4643G haplotype had statistically significantly higher PSA levels than men who were homozygous for the −5429T/−5412T/−4643A haplotype ($P$ < .004, adjusted for age). This trend was observed in each of the four age categories we examined (Fig. 4). However, only among men who were aged 51–60 or 61–70 years did those who had at least one copy of the −5429G/−5412C/−4643G haplotype have statistically significantly higher PSA levels than men who were homozygous for the −5429T/−5412T/−4643A haplotype.

Pairwise linkage disequilibrium tests demonstrated that most of the SNPs examined in the PSA gene promoter were in strong linkage disequilibrium (Table 4). This includes linkage disequilibrium between the −158 SNP and further upstream SNPs. The exception to the positive linkage disequilibrium is the −5217 T/A.
SNP with the −5567 G/A and −4289 A/C SNPs. We next determined the major haplotypes of the three SNPs that were statistically significantly associated with serum PSA levels and the −158 G/A SNP. We found that, among these four SNPs, three haplotypes accounted for 97% of the PSA gene sequences evaluated in our study population (data not shown). The A allele of the −158 G/A SNP was exclusively linked with the −5429T/−5412T/−4643A haplotype, whereas 62% of the G alleles of the −158 G/A SNP were linked with this same haplotype and the remaining 38% were linked with the −5429G/−5412C/−4643G haplotype. We found no statistically significant association between haplotypes with all four SNPs (−158 G/A, −4643 A/G, −5412 T/C, and −5429 T/G) and serum PSA levels.

**DISCUSSION**

In this study, our goal was to further characterize the PSA gene for polymorphisms and examine the associations of these polymorphisms with serum PSA levels. Using direct DNA sequencing, we identified two previously unreported sequence variants, an SNP and a polyC repeat, in the androgen-responsive upstream enhancer region previously described by Huang et al. (20). However, neither of these variants was strongly associated with serum PSA level or affected PSA promoter function in luciferase reporter assays. Our analysis of sequence variations farther upstream of the PSA promoter identified an SNP at nucleotide position −4643 with respect to the transcription start site that was statistically significantly associated with serum PSA level and affected the functional activity of PSA promoter constructs. We also identified a cluster of six SNPs in a 350-bp region immediately 5′ to the unique XbaI site at nucleotide position −5322 (see Fig. 1). Two of these SNPs (−5429 T/G and −5412 T/C) were in strong linkage disequilibrium with each other and were positively associated with serum PSA levels in healthy white men at risk for lung disease. PSA promoter assays, which showed that the promoter activity of the −5412C/−5429G haplotype was statistically significantly higher than that of the −5412T/−5429T haplotype, supported a functional role for these polymorphisms. However, the mechanism of this functional effect is not known.

Previous results by Xue et al. (22) suggested a role for the −158 G/A SNP in modulating serum PSA levels in men without prostatic disease. In that study, men with the AA genotype at the −158 G/A SNP had higher serum PSA levels than men with the AG or GG genotypes. However, we were not able to reproduce those findings in two independent study groups (25,26). Instead, our results suggest that men with the AG and GG genotypes at this SNP have higher PSA levels than men with the AA genotype (25,26). In addition, we found that the −158 G/A SNP did not affect PSA promoter activity (26).

The data presented in this study may explain the differences between our findings (25,26) and those of Xue et al. (22). Our linkage disequilibrium and haplotype analyses suggest that the A allele of the −158 G/A SNP is exclusively linked to the far upstream haplotype, −5429T/−5412T/−4643A, that was associated with reduced serum PSA levels and reduced promoter activity in vitro. By contrast, the G allele of the −158 G/A SNP was distributed between far upstream haplotypes, −5429T/−5412T/−4649A and −5429G/−5412C/−4643G, that were associated with both lower and higher serum PSA concentrations, respectively. This difference in distribution between the −158 G and A alleles with different haplotypes in the far upstream region of the PSA gene may have influenced the association of the −158 SNP with serum PSA levels observed by Xue et al. (22). Depending on the study group, this association may (22) or may not (25,26) reach statistical significance. The results of the current study suggest that genotyping SNPs in the far upstream region of the PSA gene may improve the sensitivity of PSA testing for prostate cancer. In our study group, there was no statistically significant association between the −158 G/A SNP and serum PSA level [see Table 2 and (25)], whereas the −4643 A/G, −5412 T/C, and −5429 T/G SNPs were strongly associated with serum PSA level. Restriction fragment length polymorphisms generated by both the −4643 A/G and −5412 T/C SNPs (NcoI and BsrUI restriction sites, respectively) could easily be used for genotyping subjects. Whether the −4346 A/G SNP is superior to the −5412 T/C SNP, or vice versa, as a marker for predicting serum PSA levels is unknown and will require further study in other study cohorts.

Our study excluded men with clinically significant prostate cancer. However, we did not exclude subjects on the basis of their serum PSA levels, which could have introduced bias in our study due to undiagnosed prostate cancer in some study subjects. Recent studies have reported that serum PSA levels below 9 ng/mL are not predictive of the volume or grade of prostate tumors and have no predictive value for prostate cancer (30,31), but that serum PSA levels above 9 ng/mL are highly predictive of prostate cancer (30,31). Therefore we repeated our analysis excluding men with serum PSA levels at or above 9 ng/mL. We found no statistically significant effect on any of our results (Table 2).

Given the current focus on improving the sensitivity and specificity of the PSA screening test for prostate cancer and the

---

**Table 4. Pairwise test of linkage disequilibrium of single nucleotide polymorphisms (SNPs) in the prostate-specific antigen (PSA) gene promoter***

<table>
<thead>
<tr>
<th>SNP</th>
<th>−5567</th>
<th>−5429</th>
<th>−5412</th>
<th>−5307</th>
<th>−5217</th>
<th>−4643</th>
<th>−4289</th>
<th>−158</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−5567</td>
<td>1.000</td>
<td>1.000</td>
<td>0.975</td>
<td>0.433</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>−5429</td>
<td>0.16</td>
<td>0.968</td>
<td>1.000</td>
<td>1.000</td>
<td>0.958</td>
<td>0.963</td>
<td>0.834</td>
<td></td>
</tr>
<tr>
<td>−5412</td>
<td>0.010</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.938</td>
<td>0.941</td>
<td>0.847</td>
<td></td>
</tr>
<tr>
<td>−5307</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.426</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−5217</td>
<td>1.000</td>
<td>0.010</td>
<td>0.004</td>
<td>0.012</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>−4643</td>
<td>0.007</td>
<td>0.006</td>
<td>0.000</td>
<td>0.040</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>−4289</td>
<td>0.018</td>
<td>0.000</td>
<td>0.000</td>
<td>0.096</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>−158</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.096</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*Upper right cells above the blank diagonal show the absolute value of Lewontin’s D’, an estimate of the strength of pairwise linkage disequilibrium (27), and lower left cells below the blank diagonal show the associated P values.

---
association of PSA promoter SNPs with statistically significant differences in mean serum PSA levels among men without prostatic disease, we propose the initiation of studies that comprehensively assess the utility of these SNPs in models that attempt to define the appropriate PSA cutoff value for determining whether a man should undergo further screening by prostate needle biopsy. For instance, the cutoff value for a man with a PSA promoter genotype associated with reduced serum PSA levels (i.e., the −5429T/−5412T/−4643A haplotype) may be lower (i.e., ≤4 ng/ml) than that for a man with a genotype that is associated with elevated serum PSA levels (i.e., the −5429G/−5412C/−4643G haplotype). Further study in a much larger and more well-defined population will be required to determine the degree of change in the cutoff value and the direction of this change.

The potential utility of polymorphisms in the far upstream region of the PSA gene as markers of prostate cancer risk is unknown. Two independent groups (23,24) have found that the −158 G/A SNP is associated with the risk of developing more aggressive prostate cancers. The strong linkage disequilibrium of the −158 G/A SNP with SNPs that were associated with increased activity in an in vitro assay of PSA promoter function suggests that PSA has a functional role in prostate cancer progression. This functional role may be due to the ability of PSA to cleave insulin-like growth factor binding protein 3 (32), transforming growth factor β (34), and perhaps other potentially important prostatic growth factors. Our results suggest that the SNPs in the far upstream enhancer region of the PSA gene may be good candidates for incorporation into a genetic model for prostate cancer risk.

REFERENCES

(11) Haese A, Dworschack RT, Partin AW. Percent free prostate specific antigen in the total prostate specific antigen 2 to 4 ng/ml. range does not substantially increase the number of biopsies needed to detect clinically significant prostate cancer compared to the 4 to 10 ng/ml. range. J Urol 2002;168:504–8.
(12) Roehl KA, Antenor JA, Catalona WJ. Robustness of free prostate specific antigen measurements to reduce unnecessary biopsies in the 2.6 to 4.0 ng/ml. J Urol 2002;168:922–5.
(31) Stamey T, Johnstone I, McNeal J, Lu A, Yemoto C. Preoperative serum prostate specific antigen levels between 2 and 22 ng/ml correlate poorly


NOTE

Manuscript received December 19, 2002; revised May 1, 2003; accepted May 12, 2003.