J Hum Genet (2005) 50: 507–515
DOI 10.1007/s10038-005-0285-1

ORIGINAL ARTICLE

**Go Shioji · Yoichi Ezura · Toshiaki Nakajima
Kenji Ohgaki · Hiromichi Fujiwara · Yoshinobu Kubota
Tomohiko Ichikawa · Katsuki Inoue · Taro Shuin
Tomonori Habuchi · Osamu Ogawa · Taiji Nishimura
Mitsuru Emi**

Nucleotide variations in genes encoding plasminogen activator inhibitor-2 and serine proteinase inhibitor B10 associated with prostate cancer

Received: 2 May 2005 / Accepted: 25 July 2005 / Published online: 20 September 2005

Abstract Genes encoding the serine proteinase inhibitor B family (*SERPINB*s) are mainly clustered on human chromosome 18 (18q21). Several serpins are known to affect malignant phenotypes of tumor cells, so aberrant genetic variants in this molecular family are candidates for conferring susceptibility for risk of cancer. We investigated whether eight selected non-synonymous variations within *SERPINB* loci at 18q21 might be associated with risk of prostate cancer in Japanese men. A case-control study involving 292 prostate-cancer patients and 384 controls revealed significant differences in regard to distribution of four missense variations in genes encoding plasminogen activator inhibitor 2 (*PAI2*) and *SERPINB10*. The most significant association was detected for the N120D polymorphism in the *PAI2* gene ($P = 5.0 \times 10^{-5}$); men carrying the 120-N allele (120-N/N and 120-N/D genotypes) carried a 2.4-fold increased risk of prostate cancer (95% confidence interval 1.45–4.07). Associations were also detected for three other missense polymorphisms in those two genes. Strong linkage disequilibrium in the region encompassing *PAI2* and *SERPINB10* extended to about 50 kbp. The results suggested that missense variations in one or both of these genes confer important risks for prostate cancer, and may be themselves tumorigenic. Although confirmative replication studies on larger cohorts are awaited, clinical examination of these variations may become useful for identifying individuals at high risk for prostate cancer.

Keywords Serine proteinase family · *PAI2* · *SERPINB10* · Missense polymorphism · Linkage disequilibrium

G. Shioji · Y. Ezura · T. Nakajima · K. Ohgaki
H. Fujiwara · M. Emi
Department of Molecular Biology, Institute of Gerontology, Nippon Medical School, Kawasaki, Japan

G. Shioji · K. Ohgaki · H. Fujiwara · T. Nishimura
Department of Urology, Nippon Medical School, Tokyo, Japan

Y. Ezura (✉)
Department of Molecular Pharmacology,
Medical Research Institute,
Tokyo Medical and Dental University,
2-3-10 Kanda-Surugadai, Chiyoda-ku,
Tokyo 101-0062, Japan
E-mail: ezura.mph@mri.tmd.ac.jp
Tel.: +81-35280-8067
Fax: +81-35280-8067

Y. Kubota
Department of Urology, Yokohama City University Medical School, Yokohama, Japan

T. Ichikawa
Department of Urology, Graduate School of Medicine, Chiba University, Chiba, Japan

K. Inoue
Department of Urology, Showa University Fujigaoka Hospital, Yokohama, Japan

T. Shuin
Department of Urology, Kochi Medical University, Kochi, Japan

T. Habuchi
Department of Urology, Akita University, Akita, Japan

O. Ogawa
Department of Urology, University of Kyoto Medical School, Kyoto, Japan

Introduction

Prostate cancer is the most frequently observed malignancy among elderly men, especially in advanced nations. In order to better understand the etiology of this prevalent disease, epidemiological approaches have established certain factors for susceptibility that include age, ethnicity, country of origin, and family history. A hereditary component is indicated by considerable evidence (Lichtenstein et al. 2000; Ostrander and Stanford 2000). Five chromosomal regions were defined by familial mapping studies (Smith et al. 1996; Berthon

et al. 1998; Xu et al. 1998; Gibbs et al. 1999), from which the ribonuclease L gene (*RNASEL*) (Carpten et al. 2002) and the elaC homologue 2 gene (*ELAC2*) (Tavtigian et al. 2001) were later identified as the genetic elements conferring susceptibility at hereditary prostate cancer loci 1 (HPC1) and 2 (HPC2). However, since aberrant alleles of those two genes may account only for 5–10% of prostate cancers (Carter et al. 1992), additional, low-penetrance susceptibility genes must be sought.

Case-control studies of sporadic cases have examined common variations in a number of candidate genes; examples are the androgen receptor (*AR*) gene and the kallikrein 3 gene (also known as prostate specific antigen). Several independent studies have shown that variations in *ELAC2*, *AR*, steroid-5-alpha-reductase, alpha polypeptide 2 (*SRD5A2*), and a cytochrome P450 gene (*CYP17*) show relatively good reproducibility in terms of conferring susceptibility to prostate cancer, although the significance of the results has not been fully validated. Additional candidates should be examined because having a reasonable number of reliable susceptibility genes may enable us to evaluate the combined contribution of multiple risk factors for this disease.

Among the important candidates are matrix proteases and their inhibitors, which are responsible for degradation of the extracellular matrix and thus would be related to tumor invasion. Serine proteinase inhibitors (SERPINs), a large superfamily of proteinase inhibitors (family A-I), are involved in multiple fundamental processes such as angiogenesis, inflammation, activation of complement, and fibrinolysis. Among the nine serpin families, many members of class B (SERPINBs; also called ovalbumine-like serpins) are often implicated in cancer etiology (Silverman et al. 2001). Genes encoding SERPINBs are clustered in two chromosomal regions, 18q21 and 6q25. Moreover, all of the *SERPINB* genes that appear to be involved in cancer etiology, i.e., plasminogen activator inhibitor-2 (*PAI-2*), squamous cell carcinoma antigens 1 and 2 (*SCCA1* and *SCCA2*), and maspin (*SERPINB5*), localize together at 18q21. Therefore, intense examination of genetic variations in that region should be fruitful.

In the work reported here, we investigated nucleotide variations in *SERPINB* genes at chromosome 18q21. By focusing on several single-nucleotide polymorphisms (SNPs) with non-synonymous coding sequences (missense cSNPs) present in *SERPINB10* and *PAI2*, we analyzed potential associations between SNP genotypes and the risk of prostate cancer among 676 Japanese men. We also examined linkage disequilibrium (LD) in the region encompassing those variations.

Subjects and methods

Subjects

Prostate-cancer patients over the age of 45 ($n = 292$) were recruited over the period 1995–2001 from seven university hospitals in Japan: Nippon Medical School Hospital (Tokyo), Yokohama City University Hospital, Akita University Hospital, Kyoto University Hospital, Kochi Medical School Hospital, Nagoya City University Hospital, and Chiba University Hospital. Diagnoses were ascertained by histo-pathological validation of material from trans-rectal needle biopsies or resected tumor tissues. The age at diagnosis, available for 249 of the patients, was 66.1 ± 6.76 [0.43] (mean $\pm$ SD [SE]), ranging from 45–90 years; tumor-grading scores (Gleason's scores) for 208 subjects ranged from 1–9 (mean $\pm$ SD $= 6.4 \pm 1.9$). Clinical staging of tumors from 244 patients according to the Jewett Staging System indicated that 7 were Stage A, 99 were Stage B, 73 were Stage C, and 65 were Stage D. Healthy control subjects ($n = 384$) were recruited from health check programs held in three different areas in eastern Japan, basically matching the sampling areas. All participants gave written, informed consent prior to the study, which was approved by the Institutional Review Boards of the Research Consortium. We regarded these subjects as representative of the general population and thus the predictive prevalence of prostate-cancer risks at age 60–70 years was expected to be at a negligible level (about 0.1%). Therefore, although the mean age of the control subjects (58.4 ± 8.58 [0.44]; range 32–69 years) was significantly different from that of prostate-cancer patients ($P < 0.0001$, Mann-Whitney test), contamination by individuals who have prostate-cancer risk at age 66 would be negligible. The possibility for an area-dependent difference in genotype frequency was tested by genotyping these subjects for 100 different, randomly selected SNPs. No significant difference was detected between any combinations of SNPs and areas when Bonferroni's approximation was considered; among 100 SNPs, genotype frequency of five to seven SNPs distributed differentially per area at the level of $P < 0.05$, but not at the adjusted level of $P < 0.0005$ ($P > 0.003$).

Selection of SNPs and predictive analysis of amino acid changes

Within a cluster of eight *SERPINB* genes at 18q21, 25 missense nucleotide variations have been archived in the dbSNP database (Table 1). To select the most likely candidates for risk of prostate cancer from among these variations, we evaluated the likely effect of each amino acid alteration, using the "Sorting Intolerant From Tolerant" computer program (SIFT: http://blocks.fhcrc.org/sift/SIFT.html; Ng and Henikoff 2001). Amino acid sequences encoded by these eight *SERPINB* genes were obtained from the RefSeq database of NCBI, and each of the entire sequences in FASTA format was imputed to run the SIFT program under its default settings. The SIFT algorithm estimates differences between original and altered sequences, based on the assumption that amino acid positions important for the correct biological function of a protein are conserved across the protein family and/or across evolutionary history. As output, a

Table 1 Summary of SIFT analysis of missense SNPs localized in *SERPINB* loci

No.	dbSNP ID[a]	Gene symbol	Nucleotide	Allele frequency	Codon	Amino acid	SIFT score[b]	Selection[c]
1	rs1020694	*SERPINB13*	c.877A > G	0.80:0.20	Agt/Ggt	S293G	0.24	
2	rs1395268	*SERPINB11*	c.152C > A	0.85:0.15	gCa/gAa	E51A	0.96	
3	rs4940595	*SERPINB11*	c.268T > G	0.53:0.47	Taa/Gaa	E90X	NA	Yes
4	rs1506418	*SERPINB11*	c.541A > G	0.56:0.44	Acc/Gcc	T181A	0.57	
5	rs1506419	*SERPINB11*	c.562T > A	0.55:0.45	Tgg/Agg	W188R	0.04**	Yes
6	rs1395266	*SERPINB11*	c.878T > C	0.74:0.26	aTt/aCt	I293T	0.23	
7	rs1395267	*SERPINB11*	c.907C > T	0.81:019	Ccu/Tcu	P303S	0.62	
8	rs1944270	*SERPINB8*	c.203G > A	0.71:0.29	cGa/cAa	R68Q	0.36	
9	rs1648493	*SERPINB8*	c.474G > A	0.98:0.02	aaG/aaC	K158N	0.02**	Yes
10	rs3169983	*SERPINB8*	c.910A > G	0.89:0.11	Act/Gct	T304A	0.31	
11	rs3826616	*SERPINB8*	c.1075G > A	0.54:0.46	cGc/cAc	R359H	0.61	
12	rs17782413	*SERPINB7*	c.797G > A	0.84:0.16	cGa/cAa	R266Q	0.38	
13	rs2289520	*SERPINB5*	c.559G > C	0.72:0.18	Gtc/Ctc	V187L	0.37	
14	rs1455555	*SERPINB5*	c.955A > G	0.54:0.46	Atc/Gtc	I319V	0.19*	Yes
15	rs1065205	*SERPINB3*	c.1069A > G	0.86:0.14	Acu/Gcu	T357A	0.30	
16	rs3180227	*SERPINB3*	c.1052G > C	0.87:0.13	gGa/gCa	G351A	0.73	
17	rs6098	*PAI2*	c.358A > G	0.66:0.34	Aat/Gat	N120D	0.16*	Yes
18	rs6103	*PAI2*	c.1212C > G	0.67:0.33	aaC/aaG	N404K	0.30	
19	rs6104	*PAI2*	c.1238C > G	0.56:0.44	tCc/tGc	S413C	0.02**	Yes
20	rs17072097	*SERPINB10*	c.7T > G	0.69:0.31	Tct/Gct	S3A	0.87	
21	rs8097425	*SERPINB10*	c.123A/G	0.50:0.50	atA/atG	I41M	0.12*	Yes
22	rs724528	*SERPINB10*	c.296T > C	0.78:0.22	aTc/aCc	I99T	0.43	
23	rs17072146	*SERPINB10*	c.404G > A	0.83:0.17	gGt/gAt	G135D	0.68	
24	rs9967382	*SERPINB10*	c.418C > T	0.52:0.48	Cct/Tct	P140S	0.64	
25	rs963075	*SERPINB10*	c.736C > T	0.65:0.35	Cgc/Tgc	R246C	0.06*	Yes

[a]SNPs were extracted from SNP database organized by NCBI (dbSNP: http://www.ncbi.nlm.nih.gov/SNP)
[b]Scores were given by the SIFT program (http://blocks.fhcrc.org/sift/SIFT.html). When the values were less than 0.05, substitution was regarded as intolerant (* score < 0.2, ** score < 0.05)
[c]Eight SNPs were selected for association analysis because of relatively low SIFT scores
NA: not applicable

calculated score represents estimated tolerability ranging from 0.00–1.00; when the calculated score is less than 0.05, the impact of the amino acid change is considered deleterious. On the basis of this information, we decided to focus on one nonsense SNP and seven missense cSNPs whose given scores were less than 0.2 (Table 1), in view of the specificity and sensitivity of the scoring system. We designated the selected cSNPs by combining their gene symbols and amino acid changes with an underline; for example, a nucleotide substitution at position 358 in the coding sequence of the *PAI2* gene (c.358A > G; no. 17 in Table 1) was designated *PAI2*_N120D (Tables 2 and 3).

For analysis of LD, we selected 11 SNPs from within the *PAI2* and *SERPINB10* genes whose heterozygosity was reported to exceed 10% in the database. Nucleotide variation, location, and referenced database ID for each SNP are summarized in Table 3.

Genotyping methods

Each subject was genotyped either by Invader assay (Ohnishi et al. 2001), the SD-PCR method (Iwasaki et al. 2003), or a TaqMan assay (Livak 1999). For *PAI2*_S413C, all subjects were genotyped by SD-PCR according to a protocol described previously (Iwasaki et al. 2003). For the other six of the seven missense cSNPs selected from the *PAI2*, *SERPINB11*, *SERPINB8*, and *SERPINB5* genes, the Invader assay was applied according to a published protocol (Haga et al. 2002; Iida et al. 2002), using reagents and probes provided by the supplier (Third Wave Technologies, Madison, WI, USA). Genomic DNA flanking each SNP was amplified by PCR before the assay. For LD analysis of the selected SNPs we performed TaqMan Assays-on-Demand (Applied Biosystems, Foster City, CA) according to the manufacturer's protocol. Reagents, probes, and primers were provided by the manufacturer.

Haplotype construction and LD analysis

To analyze LD in the *PAI2* and *SERPINB10* loci, we constructed haplotypes using 11 SNPs and estimated their frequencies among healthy subjects ($n = 384$) by means of an EM algorithm (SNPAlyze v3.1; DYNACOM, Chiba, Japan). Indices of LD, i.e., D, D' and r^2 (Miller et al. 2000) were analyzed for all possible two-way combinations of the 11 SNPs. To evaluate a broader range of LD, we used archived genotyping data on 48 Japanese for 27 SNPs (from rs6094 to rs963075) in surrounding genes; data were obtained from a database provided through the "SNP browser" (Applied Biosystems). Haplotype frequencies, D' and

Table 2 Summary of contingency-table analysis of the missense cSNPs

No.	Gene symbol	SNP name	Subject group[a]	Genotype[b]			Total	HWE[c]	*P* value[d]
				Maj	Het	Min			
1	*SERPINB11*	E90X	Cont	115 30.3%	194 51.2%	70 18.5%	379	0.46	0.56
			Pca	96 33.7%	136 47.7%	53 18.6%	285	0.69	
2	*SERPINB11*	W188R	Cont	111 29.4%	196 51.9%	71 18.8%	378	0.32	0.31
			Pca	98 34.8%	133 47.2%	51 18.1%	282	0.62	
3	*SERPINB8*	K158N	Cont	372 100.0%	0 0%	0 0%	372	1.00	ND
			Pca	270 100%	0 0%	0 0%	270	1.00	
4	*SERPINB5*	I319V	Cont	113 30.0%	195 51.7%	69 18.3%	377	0.34	0.65
			Pca	73 26.0%	161 57.3%	47 16.7%	281	0.01	
5	*PAI2*	N120D	Cont	136 37.5%	166 45.7%	61 16.8%	363	0.70	0.00005
			Pca	144 50.2%	121 42.2%	22 7.7%	287	0.89	
	PAI2	N404K	Cont	139 36.2%	182 47.4%	63 16.4%	384	0.97	0.001
			Pca	127 43.9%	141 48.8%	21 7.3%	289	0.10	
6	*PAI2*	S413C	Cont	122 35.2%	166 47.8%	59 17.0%	347	0.98	0.0009
			Pca	103 51.5%	72 36.0%	25 12.5%	200	0.10	
7	*SERPINB10*	I41M	Cont	139 37.5%	172 46.4%	60 16.2%	371	0.86	0.006
			Pca	129 48.3%	107 40.1%	31 11.6%	267	0.48	
8	*SERPINB10*	R246C	Cont	138 36.6%	178 47.2%	61 16.2%	377	0.96	0.0006
			Pca	131 47.8%	119 43.4%	24 8.8%	274	0.92	

[a] *Cont* Control subject group, *Pca* prostate-cancer patient group
[b] Number of the subjects and genotype frequency are indicated in each genotypical category. *Maj* homozygous major allele carriers, *Het* heterozygous subjects, *Min* homozygous minor allele carriers
[c] Hardy-Weinberg equilibrium was tested by chi-square test. Values are *P* values
[d] *P* values were calculated by chi-square test (degrees of freedom = 2)

r^2 were calculated as described above. To analyze the association between specific haplotypes involving the *PAI2* and *SERPINB10* loci and prostate-cancer risk, we constructed haplotypes using only five cSNPs and estimated their frequencies separately in the patient and control groups.

Table 3 LD analysis of analyzed polymorphism

No.	SNP name	nt.	Location	dbSNP ID[a]	Genotyping methods
1	*PAI2*_IVS2 + 36T > G	T/G	Intron2	rs6094	TaqMan
2	*PAI2*_N120D	A/G	Exon4	rs6098	Invader
3	*PAI2*_IVS4−182A > T	A/T	Intron4	rs1916658	TaqMan
4	*PAI2*_IVS5 + 130T > C	T/C	Intron5	rs1916659	TaqMan
5	*PAI2*_IVS6 + 267A > C	A/C	Intron6	rs1015416	TaqMan
6	*PAI2*_N404K	G/A	Exon8	rs6103	Invader
7	*PAI2*_S413C	C/G	Exon8	rs6104	SD-PCR
8	*SERPINB10*_−7513T > C	T/C	Promoter	rs1006961	TaqMan
9	*SERPINB10*_I41M	A/G	Exon1	rs8097425	TaqMan
10	*SERPINB10*_IVS5 + 113C > T	C/T	Intron5	rs3786340	TaqMan
11	*SERPINB10*_R246C	C/T	Exon6	rs963075	TaqMan

[a] ID numbers are from dbSNP database of NCBI (http://www.ncbi.nlm.nih.gov/SNP/)

Statistical analyses

We analyzed the distribution of genotype frequencies between distinctive study groups (prostate-cancer patients and controls), using 2×3 tables and chi-square tests to reveal trends. When dominant or recessive effects were assumed, 2×2 tables were analyzed by chi-square tests. Statistical significance was set to less than 5% ($P < 0.05$). Hardy-Weinberg equilibrium was evaluated by chi-square tests (with two degrees of freedom) in each group. The odds ratio (OR) and 95% confidence interval with respect to N120D genotypes, for example, were calculated based on the hypothesis that possession of the major A-allele (120-N) is an inherent genetic risk for prostate cancer.

Results

Selection of missense SNPs using a predictive analysis program, SIFT

To examine missense variations in the *SERPINB*-family genes clustered at 18q21, we extracted 25 SNPs from the dbSNP database and estimated the impact of amino acid changes using the SIFT algorithm (Table 1). The SIFT scores ranged from 0.02–0.96, the lowest scores being given to *PAI2*_S413C (score = 0.02) and *SERPINB8*_K158N (0.02). In addition to those SNPs, *SERPINB11*_W188R (0.03) was considered likely to be deleterious (Table 1). Relatively low SIFT scores were given to *SERPINB10*_R246C (0.06) and *PAI2*_N120D (0.16); although those would normally be regarded as "tolerant" (scores > 0.05), we decided to include these two borderline SNPs to avoid overlooking potentially causative variations whose effects might have been underestimated by the computer program. One nonsense cSNP and seven missense cSNPs were selected in this way, to be tested for association with prostate-cancer risk in our test population (Table 2).

Association of *SERPINB* SNP genotypes with risk of prostate cancer

To analyze potential associations between missense SNP genotypes and prostate-cancer risk, we determined the frequencies of individual genotypes among our cancer and control subjects (Table 2). Among the eight cSNPs selected for analysis, all except one (*SERPINB5*_K158N) were polymorphic in our test subjects, and the distribution was in Hardy-Weinberg equilibrium ($P > 0.05$) in both the prostate-cancer patient group and in the control group. In a comparison of distributions of genotype frequencies between the two groups, the most significant association was detected for *PAI2*_N120D; frequencies among individuals homozygous for the 120-N allele (N/N), heterozygous 120-N/120-D (N/D), and homozygous for the 120-D allele (D/D) were 50, 42, and 8% respectively in cancer patients vs. 37, 46, and 17% in healthy controls. A significant trend toward possession of the major allele in the prostate-cancer group was indicated by chi-square analysis using 2×3 tables ($P = 5.0\times10^{-5}$). A similar association was detected for three other SNPs from *PAI2* and *SERPINB10* (Table 2). Major alleles seemed to have a co-dominant effect, because their frequencies were shifted positively in the prostate-cancer group. Significant differences were detected when subjects were divided according to the presence or the absence of the minor allele (*PAI2*_N120D; $P = 1.4\times10^{-3}$, OR = 1.68 by chi-square test), or by the presence or absence of the major N-allele ($P = 8.0\times10^{-4}$, OR = 2.43 by chi-square test; Table 4).

Haplotype construction and analysis of linkage disequilibrium

Since we detected multiple associated SNPs within two neighboring loci, *PAI2* and *SERPINB10*, we first analyzed LD after estimating frequencies of haplotypes constructed on the basis of genotyping data for 27 SNPs

Table 4 Contingency-table analysis of the *PAI2*_N120D variant

Genotypic categorization[a]	*PAI2*_N120D	Subject groups[b]		*P* value[c]	Odds ratio	95% confidence interval
		Cont	Pca			
By minor allele possession	120-D (−)	136 (37%)	144 (50%)	0.0014	1.68	1.23–2.30
	120-D (+)	227 (63%)	143 (50%)			
	Total	363	287			
By major allele possession	120-N (+)	302 (37%)	265 (63%)	0.0008	2.43	1.45–4.07
	120-N (−)	61 (53%)	22 (47%)			
	Total	363	287			

[a]Subjects were genotypically categorized into two groups, by possession of either the minor 120-D allele or the major 120-N allele
[b]*Cont* Control group, *Pca* prostate-cancer group
[c]*P* values for chi-square test (degrees of freedom = 2)

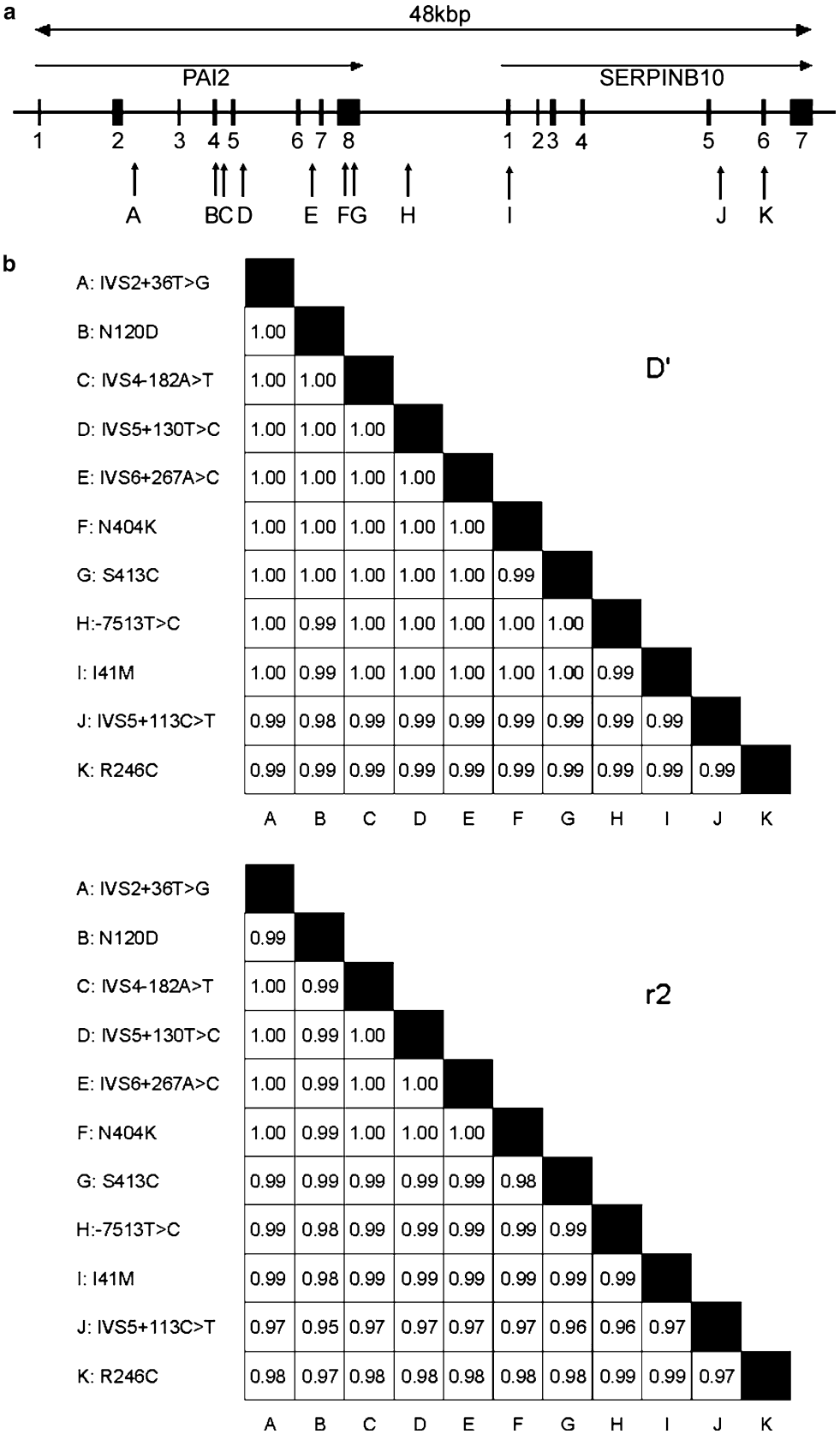

Fig. 1 a, b Analysis of haplotypes and linkage disequilibrium (LD) for 11 variations within the *PAI2* and *SERPINB10* genes. **a** Schematic diagram of the genomic structure containing *PAI2* and *SERPINB10*. *Vertical and horizontal lines* indicate exons and introns, respectively. Locations of the 11 tested variations are indicated by *upward arrows*. Each variation is designated by a letter (*A*– *K*); *A PAI2*_IVS2 + 36T > G, *B PAI2*_N120D, *C PAI2*_IVS4−182A > T, *D PAI2*_IVS5 + 130T > C, *E PAI2*_IVS6 + 267A > C, *F PAI2*_N404K, *G PAI2*_S413C, *H SERPINB10*_−7513T > C, *I SERPINB10*_I41M, *J SERPINB10*_IVS5 + 113C > T, and *K SERPINB10*_R246C. **b** Indices of LD (*D′* and r^2) calculated for every possible pair among the 11 variations, shown in tabular form

selected from "SNP Browser" (Applied Biosystems) among 48 healthy Japanese individuals. As expected, significant LD was verified over the entire *PAI2–SERPINB10* locus, indicated by high D' and r^2 scores within the chromosomal region of about 50 kb between markers rs6094 and rs963075 (data not shown). The nucleotide variations in the other genes in the cluster were independent of the LD block; no individual cSNP examined in our study subjects showed significant LD with any of the variations in the *PAI2* or *SERPINB10* loci ($D' < 0.12$, $r^2 < 0.004$).

To test if a specific haplotype of markers in this chromosomal region would present a significant association with prostate-cancer risk, haplotype and LD analyses were conducted on all 384 control subjects. By genotyping 11 SNPs within the extended locus, significant LD was verified over the entire *PAI2* and *SERPINB10* genes (Fig. 1). In an effort to resolve a specific variation contributing to a haplotype for cancer risk, we estimated maximum-likelihood frequencies of haplotypes constructed using only five cSNPs. However, because LD was extremely strong between each pair of the missense SNPs and only two major haplotypes accounted for 96–99% of the chromosomes (Table 5), no specific variation responsible for the cancer-risk haplotype could be determined.

Discussion

For undertaking the study reported here, we hypothesized that variations in *SERPINB* genes might give rise to individual differences in susceptibility for promotion and/or progression of prostate tumors, because four *SERPINB* genes (*PAI2*, *MASPIN*, *SCCA1*, *SCCA2*) had been implicated in the etiology of various cancers. Using the SIFT program to evaluate potential functional consequences of various SNPs, we focused on eight missense cSNPs and found that four of them, all within the *PAI2* and *SERPINB10* genes, were associated with cancer risk among 676 Japanese subjects. To our knowledge, this is the first report describing susceptibility to prostate cancer conferred by *SERPINB*-family genes.

The SIFT program estimated potential changes in protein function that might be brought about by each variation we examined, and expressed the predictions in scores ranging from 0–1. A significantly low score (< 0.05) was given to only two SNPs, *PAI2*_S413C and *SERPINB8*_K158N, although the latter was not polymorphic in our test population. Nevertheless, we considered that the relatively low scores given to *SERPINB10*_R246C (score = 0.06) and *PAI2*_N120D (score = 0.16) were comparable to those reported elsewhere for variations in important susceptibility genes such as *ELAC2* (for S217L, score = 0.56; for A540T, score = 0.17) and *SRD5A2* (for A49T, score = 0.18; for V89L, score = 0.12) (Tavtigian et al. 2001; Hsing et al. 2001). Since our aim was to identify susceptibility variations whose effects on cancer risk might be only slight, we assumed that the recommended significance level of the SIFT program might be too conservative for our purposes because it had been designed with a stringency possibly required for identifying affective mutations in monogenic diseases. Moreover, because SIFT scores are, after all, products of mathematical prediction, we extended the possible range for susceptibility variations to scores of less than 0.2. As a result, we detected multiple cSNPs, including *PAI2*_N120D and *SERPINB10*_R246C, that showed significant association with prostate cancer in spite of having only relatively low SIFT scores.

Among them, the most significant association was detected for *PAI2*_N120D ($P = 5.0 \times 10^{-5}$), suggesting that the minor D-allele of *PAI2*_N120D might suppress emergence or progression of prostate cancer. However, because analysis of haplotypes within the *PAI2* and *SERPINB10* loci indicated that any of five missense cSNPs could be responsible for the cancer risk, several possibilities must be considered. Firstly, the *PAI2*_N120D polymorphism alone might have a direct effect on PAI2 function and the other missense SNPs have none. More likely, however, is that several of the missense SNPs we examined would have synergistic or additive effects on the function of both proteins; moreover, a contribution by other, as yet unidentified, functional polymorphisms in LD with the tested SNPs cannot be ruled out at present. To our knowledge, no

Table 5 Haplotype analysis of the LD block in *PAI2* and *SERPINB10* genes

Haplotype number	SNP names						
	PAI2[a]			*SERPINB10*		Carrier number/(Percentage)	
	N120D	N404K	S413C	I41M	R246C	Pca patients[b]	Control
Haplotype1	0	0	0	0	0	239 (69.5%)	381 (58.8%)
Haplotype2	1	1	1	1	1	92 (26.7%)	262 (40.4%)
Haplotype3	1	1	1	0	0	4 (1.2%)	1 (0.15%)
Others	–	–	–	–	–	9 (2.6%)	4 (0.61%)
					Total	344	648

[a] *0* Major allele of each SNP; *1* minor allele of each SNP
[b] *Pca* Prostate cancer

previous report has identified an effect of *SERPINB10* on occurrence of cancer. Thus, the functional significance of variations should be evaluated for both *PAI2* and *SERPINB10*, to clarify what the associations we observed actually mean.

Biochemical analysis should clarify whether the polymorphisms examined here do in fact affect protein function. In addition, the mechanism(s) that suppress prostate tumors could be clarified by means of biological assays. In vitro examination of potential cellular invasion and proliferation could be appropriately designed to exploit a known function of PAI2, inhibition of the urokinase-type plasminogen activator (uPA; Kruithof et al. 1995), and to examine additional specific aspects of that function including inhibition of apoptosis in cultured cell lines (Kumar and Baglioni 1991).

The etiology of cancer, as in many other polygenic diseases, involves participation and interaction of multiple environmental and genetic factors. However, in prostate cancers, the contribution of genetic factors is thought to be relatively high, and thus linkage analyses in familial cases have identified several reliable susceptibility genes including *RNASEL* and *ELAC2* (Tavtigian et al. 2001; Carpten et al. 2002). Also, a lesser number of CAG repeats in the polymorphic microsatellite within the *AR* gene reproducibly associate with sporadic prostate cancer (Irvine et al. 1995). In our study, we detected a strong association of missense variations in the *PAI2* and *SERPINB10* genes, although reproducibility could be a problem in view of the relatively small size of our test population. Prospective analyses of large cohorts are in progress as part of a clinical study being undertaken in a consortium of university hospitals in Japan, to investigate the real meaning of the correlations we have reported here. It would also be of value to investigate further the possibility of interactive effects of these polymorphisms. For example, the suggested protective effects conferred by minor variants of *PAI2* SNPs should be examined in combination with the rare variations in *RNASEL* or *ELAC2*, or with the CAG-repeat polymorphism of the *AR* gene. Molecular genetic and epidemiological investigations that integrate known risk factors including age, androgen status, and possible environmental exposure to carcinogens would help to clarify the complex etiological basis of prostate cancer.

In summary, we have detected an association of multiple missense cSNPs in the *PAI2* and *SERPINB10* genes with prostate-cancer risk, by genotyping 292 Japanese cancer patients and 384 healthy controls. Although the functional significance of these variations is still obscure, and more confirmative replication studies are necessary, any novel candidates added to the catalogue of susceptibility-associated genetic variations will increase our understanding of prostate cancer. A robust genome-wide strategy, recently undertaken by several research institutes (Wang et al. 2005), is expected to identify many reliable candidates for multi-factorial common diseases, in spite of technical limitations on truly genome-wide approaches. Alternatively, large-scale association analyses focusing on intolerant missense variations could be a fruitful approach. When multiple susceptibility genes for prostate cancer have been identified with certainty, epidemiological investigations that incorporate multiple risk factors should lead to identification of prognosticators, preventive approaches, and novel therapies for prostate cancer.

Acknowledgements We thank Mitsuko Kajita, Mina Kodaira, and Miho Kawagoe for their expert technical assistance. This work was supported by a grant for Strategic Research from the Ministry of Education, Science, Sports, and Culture of Japan; by a Research Grant for Research from the Ministry of Health and Welfare of Japan; and by a Research for the Future Program Grant of the Japan Society for the Promotion of Science.

References

Berthon P, Valeri A, Cohen-Akenine A, Drelon E, Paiss T, Wohr G, Latil A, Millasseau P, Mellah I, Cohen N, Blanche H, Bellane-Chantelot C, Demenais F, Teillac P, Le Duc A, de Petriconi R, Hautmann R, Chumakov I, Bachner L, Maitland NJ, Lidereau R, Vogel W, Fournier G, Mangin P, Cussenot O et al. (1998) Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2-43. Am J Hum Genet 62:1416–1424

Carpten J, Nupponen N, Isaacs S, Sood R, Robbins C, Xu J, Faruque M, Moses T, Ewing C, Gillanders E, Hu P, Bujnovszky P, Makalowska I, Baffoe-Bonnie A, Faith D, Smith J, Stephan D, Wiley K, Brownstein M, Gildea D, Kelly B, Jenkins R, Hostetter G, Matikainen M, Schleutker J, Klinger K, Connors T, Xiang Y, Wang Z, De Marzo A, Papadopoulos N, Kallioniemi OP, Burk R, Meyers D, Gronberg H, Meltzer P, Silverman R, Bailey-Wilson J, Walsh P, Isaacs W, Trent J (2002) Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. Nat Genet 30:181–184

Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC (1992) Mendelian inheritance of familial prostate cancer. Proc Natl Acad Sci USA 89:3367–3371

Gibbs M, Stanford JL, McIndoe RA, Jarvic GP, Kolb S, Goode EL, Chakrabarti L, Schster EF, Buckley VA, Miller EL, Brandzel S, Li S, Hood L, Ostrander EA (1999) Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36. Am J Hum Genet 64:776–787

Haga H, Yamada R, Ohnishi Y, Nakamura Y, Tanaka T (2002) Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190,562 genetic variations in the human genome. J Hum Genet 47:80–87

Hsing AW, Chen C, Chokkalingam AP, Gao YT, Dightman DA, Nguyen HT, Deng J, Cheng J, Sesterhenn IA, Mostofi FK, Stanczyk FZ, Reichardt JK (2001) Polymorphic markers in the SRD5A2 gene and prostate cancer risk: a population-based case-control study. Cancer Epidemiol Biomarkers Prev 10:1077–1082

Iida A, Saito S, Sekine A, Mishima C, Kitamura Y, Kondo K, Harigae S, Osawa S, Nakamura Y (2002) Catalog of 77 single-nucleotide polymorphism (SNPs) in the carbohydrate sulfotransferase 1 (CHST1) and carbohydrate sulfotransferase 3 (CHST3) genes. J Hum Genet 47:14–19

Irvine RA, Yu MC, Ross RK, Coetzee GA (1995) The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. Cancer Res 55:1937–1940

Iwasaki H, Emi M, Ezura Y, Ishida R, Kajita M, Kodaira M, Yoshida H, Suzuki T, Hosoi T, Inoue S, Shiraki M, Swensen J, Orimo H (2003) Association of a Trp16Ser variation in the gonadotropin releasing hormone signal peptide with bone mineral density, revealed by SNP-dependent PCR typing. Bone 32:185–190

Kruithof EK, Baker MS, Bunn CL (1995) Biological and clinical aspects of plasminogen activator inhibitor type 2. Blood 86:4007–4024

Kumar S, Baglioni C (1991) Protection from tumor necrosis factor-mediated cytolysis by overexpression of plasminogen activator inhibitor type-2. J Biol Chem 266:20960–20964

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K (2000) Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 13:78–85

Livak KJ (1999) Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet Anal 14:143–149

Miller PT, Sarçñina IB, Saccone NL, Putzel J, Laitinen T, Cao A, Kere J, Pilia G, Rice JP, Kwok PY (2000) Juxtaposed regions of extensive minimal linkage disequilibrium in human Xq25 and Xq28. Nat Genet 25:324–328

Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. Genome Res 11:863–874

Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y (2001) A high-throughput SNP typing system for genome-wide association studies. J Hum Genet 46:471–477

Ostrander EA, Stanford JL (2000) Genetics of prostate cancer: too many loci, too few genes. Am J Hum Genet 67:1367–1375

Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin PB, Gettins PG, Irving JA, Lomas DA, Luke CJ, Moyer RW, Pemberton PA, Remold-O'Donnell E, Salvesen GS, Travis J, Whisstock JC (2001) The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. J Biol Chem 276:33293–33296

Smith JR, Freije D, Carpten JD, Gronberg H, Xu J, Isaacs SD, Brownstein MJ, Bova GS, Guo H, Bujnovszky P, Nusskern DR, Damber JE, Bergh A, Emanuelsson M, Kallioniemi OP, Wlsh PC, Collins FS, Trent JM, Isaacs WB (1996) Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 274:1371–1374

Tavtigian SV, Simard J, Teng DH, Abtin V, Baumgard M, Beck A, Camp NJ, Carillo AR, Chen Y, Dayananth P, Desrochers M, Dumont M, Farnham JM, Frank D, Frye C, Ghaffari S, Gupte JS, Hu R, Iliev D, Janecki T, Kort EN, Laity KE, Leavitt A, Leblanc G, McArthur-Morrison J, Pederson A, Penn B, Peterson KT, Reid JE, Richards S, Schroeder M, Smith R, Snyder SC, Swedlund B, Swensen J, Thomas A, Tranchant M, Woodland AM, Labrie F, Skolnick MH, Neuhausen S, Rommens J, Cannon-Albright LA (2001) A candidate prostate cancer susceptibility gene at chromosome 17p. Nat Genet 27:172–180

Wang WY, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109–118

Xu J, Meyers D, Freije D, Isaacs S, Wiley K, Nusskern D, Ewing C, Wilkens E, Bujnovszky P, Bova GS, Walsh P, Isaacs W, Schleutker J, Matikainen M, Tammela T, Visakorpi T, Kallioniemi OP, Berry R, Schaid D, French A, McDonnell S, Schroeder J, Blute M, Thibodeau S, Trent J et al. (1998) Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 20:175–179