

Sara Lindström · Fredrik Wiklund
Björn-Anders Jonsson · Hans-Olov Adami
Katarina Bälter · Anthony J. Brookes · Jianfeng Xu
S. Lilly Zheng · William B. Isaacs · Jan Adolfsson
Henrik Grönberg

Comprehensive genetic evaluation of common E-cadherin sequence variants and prostate cancer risk: strong confirmation of functional promoter SNP

Received: 1 June 2005 / Accepted: 17 August 2005 / Published online: 28 September 2005
© Springer-Verlag 2005

Abstract The E-cadherin gene (CDH1) has been proposed as a prostate cancer (PC) susceptibility gene in several studies. Aberrant protein expression has been related to prognosis and progression in PC. In addition, a functional promoter SNP (rs16260) has been found to associate with PC risk. We performed a comprehensive genetic analysis of CDH1 by using the method of haplotype tagged SNPs in a large Swedish population-based case-control study consisting of 801 controls and 1,636 cases. In addition, Swedish PC families comprising a total of 157 cases sampled for DNA were analyzed for selected SNPs. Seven SNPs, including the promoter SNP

rs16260, that captured over 96% of CDH1 haplotype variation were selected as haplotype tagging SNPs and analyzed for associated PC risk. We observed significant confirmation of rs16260 ($P=0.003$) for cases with a positive family history of PC (FH+) both in an independent case-control population and in PC families. In addition, a common haplotype (HapB, 25%) including the variant allele of rs16260 was associated ($P=0.004$) with PC risk among FH+ cases. The promoter SNP rs16260 as well as HapB were significantly transmitted to affected offspring in PC families. We report strong confirmation of the association between PC risk in FH+ cases and a functional CDH1 promoter SNP in an independent population. In conjunction with the biological importance of CDH1 our findings encourage further evaluation of genetic variation in CDH1 in relation to PC etiology. Due to the difficulties in replication of genetic association studies, this finding is unusual and novel.

S. Lindström (✉) · F. Wiklund · B. A. Jonsson · H. Grönberg
Department of Radiation Sciences, Oncology,
University of Umeå, 901 87 Umeå, Sweden
E-mail: sara.lindstrom@oc.umu.se
Tel.: +46-90-7854734
Fax: +46-90-127464

H. O. Adami · K. Bälter
Department of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm, Sweden

A. J. Brookes
Center for Genomics and Bioinformatics, Karolinska Institutet,
Stockholm, Sweden

A. J. Brookes
Department of Genetics, University of Leicester, Leicester, UK

J. Xu · S. L. Zheng
Center for Human Genomics,
Wake Forest University School of Medicine,
Winston-Salem, NC, USA

W. B. Isaacs
Department of Urology, Johns Hopkins Medical Institutions,
Baltimore, MD, USA

J. Adolfsson
Oncologic Center, Karolinska Institutet, CLINTECH,
Stockholm, Sweden

Introduction

Evidence from epidemiological and genetic studies provides increasing support for the existence of inherited susceptibility to prostate cancer (PC) (Schaid 2004). Despite large efforts made during the last two decades, however, no major PC genes have been identified. Instead, genetic epidemiological studies suggest several low-penetrance genes, each contributing to a small risk.

The E-cadherin gene (CDH1, OMIM 192090) located on chromosome 16q22.1 encodes a transmembrane glycoprotein that mediates cell–cell adhesion and—in conjunction with cytoplasmic catenin proteins—also cell signaling. The cadherin/catenin complex plays a crucial part in cell polarity, preservation of normal tissue

morphology and cellular differentiation (Grunwald 1993). Since CDH1 also acts as a suppressor of invasion, aberrations in CDH1 expression is related to PC progression, metastasis and poor prognosis among patients with PC (Umbas et al. 1994; Dunsmuir et al. 2000; Chunthapong et al. 2004). Frequent loss of heterozygosity (Suzuki et al. 1996; Latil et al. 1997) and suggestive linkage (Suarez et al. 2000; Goddard et al. 2001) in close proximity to CDH1 on the 16q arm provides further support for CDH1 as a PC susceptibility gene. A promoter SNP (rs16260) of CDH1 has been shown to decrease gene transcription (Li et al. 2000; Nakamura et al. 2002; Lei et al. 2002) and the association with PC susceptibility has been reported (Verhage et al. 2002). Recently, we reported a significant association between hereditary PC risk and the promoter SNP in a large Swedish population-based study (Jonsson et al. 2004).

The aim of the present study was to confirm the association between the CDH1 promoter SNP and PC risk in an independent population study. In addition, we performed a systematic evaluation of common CDH1 sequence variants in relation to PC risk using haplotype tagged SNPs (htSNPs). Selected SNPs were genotyped both in a large Swedish population-based case-control study and in Swedish PC families.

Subjects and methods

Cases and controls

Cancer Prostate in Sweden (CAPS) is a population-based case-control study described in detail elsewhere (Lindmark et al. 2004). All of the men living in the northern and central parts of Sweden, under the age of 80 years, and all of the men living in the Stockholm region and south-eastern part of Sweden, under the age of 65 years, at any time from March 2001 through September 2002, formed the study base. All men with a newly diagnosed adenocarcinoma of the prostate, cytologically or pathologically verified, in this population were eligible to participate in the study. Utilization of regional cancer registries, to which reporting of newly diagnosed cancer cases is compulsory, allowed complete case ascertainment. From the Swedish National Prostate Cancer Registry (<http://www.roc.se>) detailed clinical information such as Gleason score, PSA level at the time of diagnosis and tumor -node-metastasis stage was obtained.

In total, 1,961 PC patients were identified and invited to the study. Of them, 1,444 (74%) agreed to participate and from those a blood sample and a questionnaire concerning risk factors and family history were collected. For cases that reported at least one close relative being diagnosed with PC, a detailed family history was obtained by a second questionnaire through a telephone interview. Whenever possible, blood samples were collected from the affected relatives in these families.

Control subjects were randomly selected from the Swedish Population Registry and frequency matched according to gender, the expected age distribution of cases (groups of 5-year interval) and geographical region (two regions, representing north and south of Sweden including Stockholm). Control subjects were recruited concurrent with case subjects. A total of 1,697 controls were invited to the study and 866 (51%) agreed to participate. At the time of this study, DNA samples were available for 1,247 sporadic cases (SPC), 232 family positive cases [at least two relatives with a PC diagnosis within a nuclear family (FH+)] and 801 controls.

PC families

Recruitment of families with multiple PC cases has been conducted at the Department of Oncology, Umeå University, since 1995. Ascertainment of families has been mainly based on referrals by urologists and oncologists throughout Sweden, described in detail elsewhere (Grönberg et al. 1999). Families were initially taken for linkage analysis and almost all the families fulfilled the Carter criteria (Carter et al. 1993) for hereditary PC. The Carter criteria includes a cluster of three or more relatives affected with PC in any nuclear family, PC in three successive generations in either of the probands' paternal or maternal lineages or two close relatives affected with PC at 55 years of age or younger. Blood samples have been collected from as many family members as possible. At the time of this study, a total of 81 families, including 157 FH+ cases sampled for DNA, were available for analysis.

Total study population

The total study population comprises 1,636 PC patients and 801 unaffected controls. Of the 1,636 case patients, 1,247 (76%) are classified as sporadic (SPC) and 389 (24%) as FH+. The mean age at diagnosis of PC was 66.7 (47.3–80.4) for SPC cases and 65.5 (43.0–81.0) for FH+ cases. The mean age of inclusion for the controls was 67.8 years with a range of (45.5–80.0). Written informed consent was obtained from each subject. The ethics committee at the Karolinska Institutet and Umeå University approved the study.

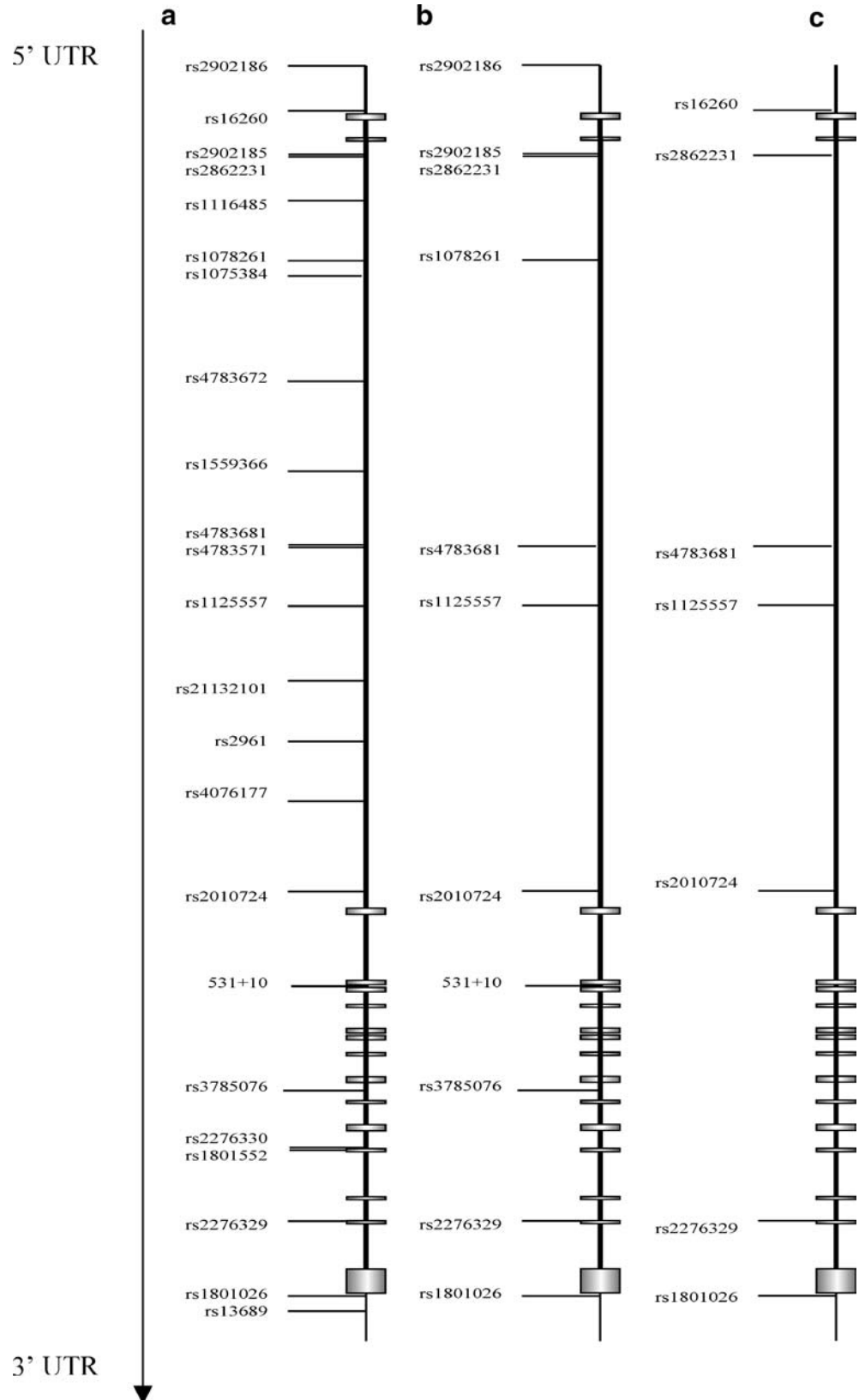
Selection of SNPs

The CDH1 gene covers 98 kb and has 16 exons. The target region for the selection of SNPs included 10 kb of the promoter region, all exons, introns and 5 kb of the 3'UTR. Using the public database SNPper (<http://SNPper.chip.org/>) we identified altogether 82 SNPs. A subset of 23 SNPs were selected, with the criteria of one SNP per 3 kb, reported minor allele frequency (MAF) of

at least 5% and as low as possible proportion of repetitive sequences. We prioritized notably validated SNPs and additional attention was paid to SNPs in the promoter region. The selected SNPs were distributed as

two SNPs in the promoter region, 14 in intron two, one in intron four, one in intron 10, one in intron 12, one in exon 13, one in intron 14 and two in exon 16 (Fig. 1a).

Fig. 1 **a** The original selection of SNPs. **b** SNPs succeeded in genotyping. **c** Selected htSNPs genotyped in the whole material



We genotyped these 23 SNPs in a randomly selected subset of 94 controls from the CAPS study. As a result, we eliminated four SNPs because of repetitive sequences, three due to failed assay, three due to unreliable results and two due to monomorphic results. From the remaining 11 SNPs (Fig. 1b), haplotypes were inferred using the software PHASE (<http://www.stat.washington.edu/stephens/software.html>). By haplotype deviation analysis using the htSNP2 software (<http://www-gene.cimr.cam.ac.uk/clayton/software/stata/htSNP2/>), six SNPs that explained over 96% of the haplotype diversity were selected as htSNPs (Fig. 1c). Together with the promoter SNP-160C/A (rs16260), a total of seven SNPs were genotyped in the whole study population.

Genotyping

DNA samples were extracted from leukocytes using standard methods. Each DNA plate contained two CEPH controls, a water bank and blinded internal replicates.

We used dynamic allele-specific hybridization (DASH) as previously described (Howell et al. 1999; Prince et al. 2001; Jobs et al. 2003). For this, two PCR primers and one DASH probe per target mutation/SNP were designed by means of custom software (Fredman et al. 2004) provided by DynaMetrix Ltd (UK). Oligonucleotides were provided and HPLC purified, by Biomers GmbH (Germany). The DASH PCRs entailed amplifying short genomic fragments spanning the variant of interest, with one of the primers carrying a 5'-biotin label. Amplifications were performed in 5 μ l volume, containing 1–2 ng genomic DNA, 0.38 μ M biotinylated primer, 0.75 μ M non-biotinylated primer, 0.03 U AmpliTaq Gold (PE Biosystems, CA, USA), 10% dimethylsulphoxide, 1 \times AmpliTaq Gold Buffer including 1.5 mM of MgCl₂ (PE Biosystems) and 0.2 mM each dNTP. Thermal cycling was conducted on an MBS 384 device (Thermo-Electron, USA) as follows; 1 \times (10 min at 94°C), 35 \times (15 s at 94°C, 30 s at annealing temperature).

To verify a successful amplification, 0.5 μ l of several randomly chosen samples were examined on a 3.0% low-melt agarose gel. DASH analysis of the PCR product was conducted on membrane macroarrays, using the DASH-2 protocol (Jobs et al. 2003). Briefly, this entailed transferring samples to the membrane array by centrifugation (Jobs et al. 2002) or robotic gridding. The resulting individual arrays with up to 9,600 distinct samples/features were rinsed in 0.1 M NaOH to denature the PCR products, and then exposed to 2 ml HE buffer (0.1 M HEPES, 10 mM EDTA, pH 7.9) containing 4 nmol of suitable probe, itself end-labelled with ROX. After heating to 85°C and air-cooling to room temperature, the membrane was briefly rinsed in HE buffer. The array was then

soaked in 40 ml HE-buffer containing SYBR GreenI dye at 1:20,000 dilution for up to 3 h. Using a DASH-2 device (DynaMetrix Ltd, UK), the membrane was taken through a DASH heating ramp (heating at 3°C/min from room temperature to 75°C) whilst fluorescence from the ROX acceptor dye on the probe was monitored. Data was collected at an interval of 0.5°C. Fluorescence changes with temperature (DNA-melting profiles) were used to distinguish different alleles, and this was done by means of generic melt-curve analysis software (DynaScore; DynaMetrix Ltd, UK). This software uses negative derivatives of fluorescence against temperature to reveal peak of denaturation rate (target-probe melting temperatures; T_m) and thereby automatically assigns DNA samples into genotype groups.

Statistical methods

Hardy–Weinberg equilibrium test for each SNP as well as pair-wise linkage disequilibrium estimations were performed using a replication method as implemented in the GENETICS package (<http://lib.stat.cmu.edu/R/CRAN/index.html>) for the R programming language. For each test, 10,000 permutations were run.

The association between each SNP and PC risk was evaluated by a likelihood ratio test based on an unconditional logistic regression model, adjusted for age (in 5-year intervals) and two geographical regions representing the north and south of Sweden including Stockholm as implemented in the STATA software. Dependence among genotypes within the families was taken into account by adjusting the confidence intervals for the odds ratios (ORs) using the Huber/White/sandwich robust technique (Wooldridge 2002).

The HAPLO.STATS package (Schaid et al. 2002) in the software language R was used for haplotype analysis. This method provides both global and haplotype-specific tests. Only unrelated individuals were included in the haplotype analyses and all computations were adjusted for age and geographical region as described above. Haplotypes with a frequency of <0.005 were pooled together. Empirical P values were computed by randomly permuting the trait and covariates (Besag and Clifford 1991). Precision criteria for the P values were set to a sample standard error of one fourth of the estimated P value but at least 1,000 permutations were run for each simulation.

The association between the CDH1 markers and PC risk was also evaluated by a family-based transmission test as described by Clayton (1999) using the software TRANSMIT (<http://www-gene.cimr.cam.ac.uk/clayton/software/>). With this method, transmission of multi-locus haplotypes is possible to evaluate, even if the phase is unknown, and parental genotypes may be unknown. All reported P values are based on two-sided tests.

Table 1 Genotype frequencies and ORs for the promoter SNP rs16260 in unaffected controls and FH+ cases

	Study population	Genotype	Number of subjects (%)		<i>P</i> value ^a	OR ^b (95% CI)
			Controls	FH+ cases		
Case subjects utilized in Jonsson et al. (2004)		C/C	129 (52)	66 (39)	0.03	1.00 (Ref.)
		C/A	96 (39)	80 (48)		1.45 (0.91–2.33)
		A/A	21 (9)	22 (13)		2.08 (1.02–4.23)
Independent population		C/C	268 (53)	87 (44)	0.003	1.00 (Ref.)
		C/A	209 (41)	88 (44)		1.38 (0.95–2.01)
		A/A	29 (6)	24 (12)		2.78 (1.45–5.34)
Pooled sample		C/C	397 (53)	153 (42)	0.0001	1.00 (Ref.)
		C/A	305 (40)	168 (46)		1.47 (1.10–1.96)
		A/A	50 (7)	46 (12)		2.61 (1.57–4.31)

^aScore test of a covariate equal to number of variant alleles based on a logistic regression model

^bORs are adjusted for age and geographical region

Results

Based on 115 duplicated samples, an estimated genotyping error rate of 0.3% was observed. SNPs were in pair-wise LD with each other (Range: 0.15–0.99) and they fulfilled HWE ($P=0.12$ –0.90).

In order to perform an independent test of association between the CDH1 promoter SNP (rs16260) and PC risk, all individuals included in our initial analysis (Jonsson et al. 2004) were excluded. This revealed an independent replication study population comprising 540 control subjects, 612 SPC cases, and 211 FH+ cases. Comparing controls and family positive cases within this replication population we found strong evidence of association between the promoter SNP and risk of PC ($P=0.003$, assuming an additive genetic model; Table 1). Genotype-specific risk estimates were essentially the same as observed among individuals included in the initial analysis. In the total study population, the OR for CA-heterozygous carriers was 1.5 (1.1–2.0) and for AA-homozygous carriers 2.6 (1.6–4.3) compared to CC-homozygous carriers. No significant difference in genotype frequency of the promoter SNP was observed between the SPC cases and controls ($P=0.37$; Table 2).

To comprehensively evaluate common sequence variants in CDH1 with respect to PC risk we genotyped an additional six SNPs in the gene, thereby capturing the majority of the haplotype diversity. Among cases with a positive family history the rs4783681 SNP was significantly associated with PC risk ($P=0.005$; Table 2). Among SPC the prevalence of two SNPs was significantly different compared to controls; rs2010724 ($P=0.02$) and rs1801026 ($P=0.04$).

We inferred nine CDH1 haplotypes with a frequency over 1% among the controls (Table 3). A significant difference in haplotype distribution was observed between cases with a positive family history and controls ($P=0.05$ global test). The second most common haplotype (HapB, control frequency = 0.25) was associated with positive family history of PC ($P=0.004$; Table 3). Interestingly, HapB was the only haplotype with a frequency over 1% carrying the variant A-allele for promoter SNP rs16260. In addition, haplotype frequency for HapD was significantly

higher among SPC cases compared to controls ($P=0.02$; Table 3).

For family-based transmission tests 123 informative PC families, originated from both CAPS and collected PC families, were available for analysis. In total, these families included 340 PC cases and 464 unaffected relatives. DNA was available for an average of 2.1 affected cases within each family. The promoter SNP rs16260 were transmitted in a greater extent than expected to affected offspring compared to unaffected ($P=0.02$; Table 4). In addition, three more SNPs (rs4783681, rs1125557 and rs2276329) were significantly over-transmitted to an affected offspring (Table 4). A global test revealed a significant difference in haplotype transmission between affected and unaffected offsprings ($P=0.01$). Haplotype-specific tests revealed that HapB was transmitted in a greater extent than expected ($P=0.02$). No other haplotypes were significantly over-transmitted to affected offspring.

Discussion

We report a strong confirmation of the association between PC risk in FH+ cases and the functional CDH1 promoter SNP rs16260. Previously we identified an elevated risk for carriers of the A-allele in cases with a positive family history of PC and suggested an additive inheritance model. Our findings in an independent study population strongly ($P=0.003$) support this hypothesis.

This replication is unique in the context of the lack of positive replication studies in genetic epidemiology. Considering the high false-positive rate of reported associations between complex diseases and genetic variants, estimated to be as high as 95% (Colhoun et al. 2003), emphasizes the importance of replication studies in independent populations.

One source of confounding contribution to false-positive association results is the effect of population stratification. However, since a family-based association test revealed a significant ($P=0.02$) over-transmission of rs16260 to affected offspring, we are confident that this is not an issue in our study. Besides the significant replication result, the impact of rs16260 in PC etiology

Table 2 Allele frequencies for the E-cadherin gene (CDH1) SNPs and corresponding *P* values among control subjects and PC patients with sporadic (SPC) and cases with a positive family history of PC (FH+)

SNP	Relative position	Allele	Allele frequencies (%)			<i>P</i> value ^a	
			Controls (n = 801)	SPC (n = 1,247)	FH+ (n = 389)	SPC vs. controls	FH+ vs. controls
rs16260	-285	C	73	75	65	0.37	0.004
		A	27	25	35		
rs2862231	+2,890	G	81	80	85	0.49	0.19
		A	19	20	15		
rs4783681	+35,190	G	73	75	67	0.54	0.005
		T	27	25	33		
rs1125557	+38,281	A	57	59	53	0.06	0.25
		G	43	41	47		
rs2010724	+61,096	A	67	70	62	0.02	0.11
		G	33	30	38		
rs2276329	+91,996	T	94	93	96	0.36	0.09
		C	6	7	4		
rs1801026	+96,138	C	88	85	87	0.04	0.58
		T	12	15	13		

^aGeneral two-degree of freedom likelihood ratio test

Table 3 Haplotype frequencies in CDH1 among controls, SPC and FH+ with corresponding *P* values

Haplotype id	SNP id ^a 1 2 3 4 5 6 7	Haplotype frequency (%)			<i>P</i> value	
		Controls (N = 801)	SPC (N = 1,247)	FH+ (N = 259)	SPC vs. controls	FH+ vs. controls
A	CGGAATC	32	30	32	0.47	0.86
B	AGTGGTC	25	24	31	0.13	0.004
C	CAGGATC	10	11	8	0.58	0.26
D	CGGAATT	9	11	11	0.02	0.87
E	CAGAATC	7	7	5	0.58	0.12
F	CGGAACC	4	5	3	0.38	0.08
G	CGGAGTC	3	2	3	0.15	0.74
H	CGGGGTC	2	2	2	0.08	0.27
I	CAGGATT	1	1	1	0.95	0.69
Global test for maximum value of score statistics					0.28	0.05

All haplotypes listed have an estimated frequency above 1% in controls

^aSNP id (1) rs16260 (2) rs2862231 (3) rs4783681 (4) rs1125557 (5) rs2010724 (6) rs2276329 (7) rs1801026

is further supported by the haplotype analyses. The only haplotype carrying the variant allele of rs16260 was also the only haplotype that was significantly associated with PC risk among family-positive cases. This haplotype has a high prevalence (25%) in the Swedish population and includes four SNPs with variant alleles (rs16260, rs4783681, rs1125557, rs2010724) that were all positively correlated with PC (data not shown).

In addition, to replicate our previous finding for the promoter SNP, we also performed a comprehensive evaluation of the CDH1 gene. By means of htSNPs, we captured over 96% of the haplotype variation and found association between several CDH1 sequence variants and PC risk. Apart from rs16260, rs4783681 was strongly associated with family-positive PC. However, these two SNPs are in strong LD with each other ($D' = 0.95$). We observed no association between the functional promoter SNP and SPC risk. This suggests a joint effect from several genetic variants necessary for disease development, with rs16260 as one polymorphism. The absence of LD between rs16260 and the other variants will result in independent inheritance, implying a low proportion of the general population carrying all risk alleles. However, within PC families a higher proportion of risk allele carriers are expected.

We also identified a marginally significant association between two SNPs (rs2010724, rs1801026) and sporadic PC risk. To investigate the possibility of false-positive findings due to multiple testing, we performed a data simulation by randomly permuting case-control status and then re-evaluated association for each SNP. *P* values

Table 4 Family based transmission results for SNPs in Swedish PC families

SNP	Allele	Allele frequency (%)	<i>P</i> value
rs16260	C	69	0.02 (A)
	A	31	
rs2862231	G	83	0.91
	A	17	
rs4783681	G	70	0.03 (T)
	T	30	
rs1125557	A	56	0.003 (G)
	G	44	
rs2010724	A	65	0.10
	G	35	
rs2276329	T	94	0.03 (T)
	C	6	
rs1801026	C	86	0.14
	T	14	

Allele in brackets implies overtransmitted allele

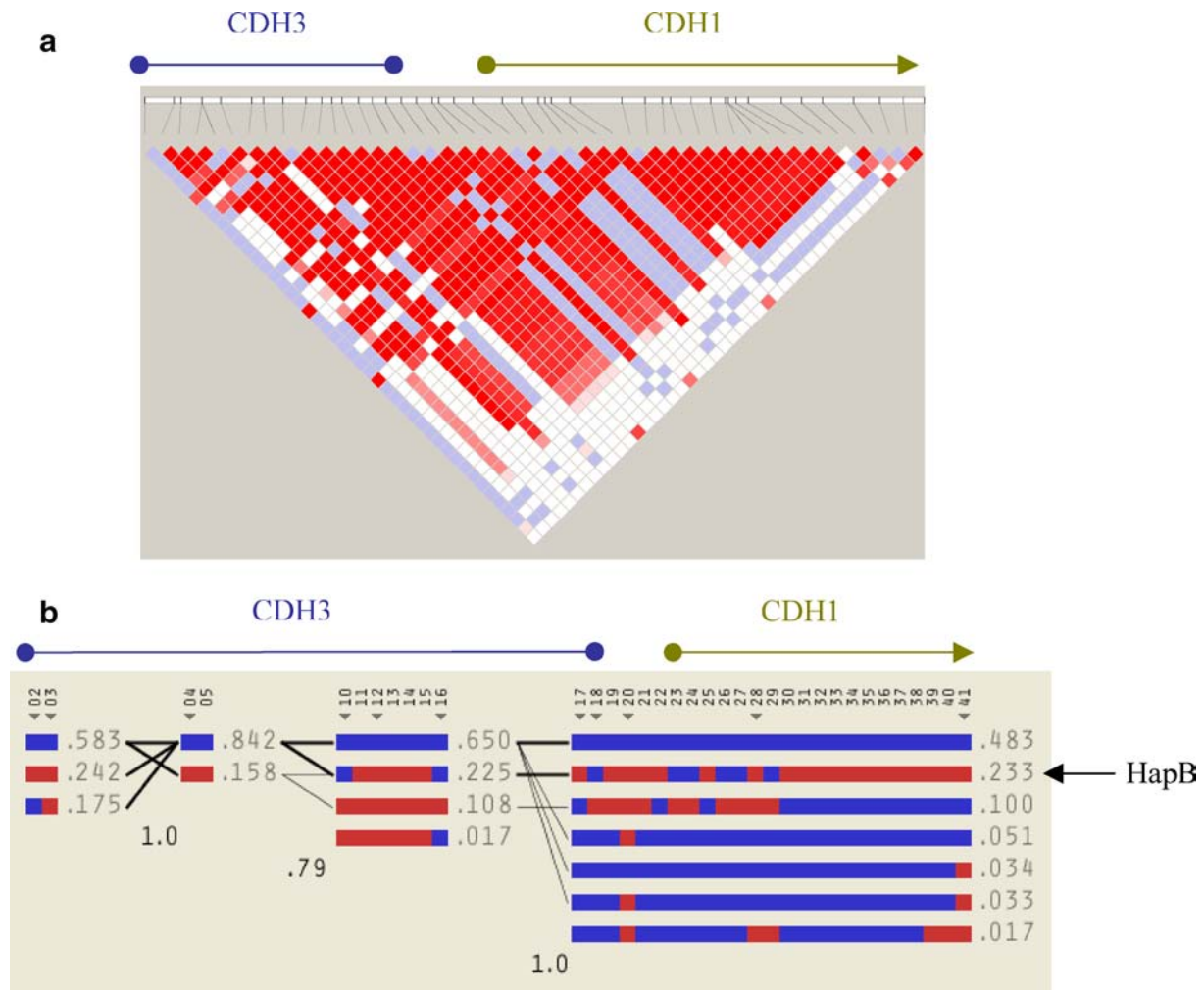


Fig. 2 **a** Pair wise LD-plot for the region of chromosome 16q22.1 spanning over 197 kb. *Red color* indicates strong pair wise LD. **b** Haplotype blocks defined by the method of LD confidence interval (Gabriel et al. 2002). The picture is constructed in Haploview and data is taken from the HapMap project. For each

SNP *blue color* represents the normal allele and *red* represents the variant allele. For a detailed description of the picture the reader is referred to (<http://www.broad.mit.edu/personal/jcbarret/haplo/documentation.php>)

adjusted for multiple testing were then computed based on the empirical distribution of the maximum of the seven test statistics. In addition, we estimated the probability of observing at least two significant associations (at the 5% level) out of seven tests under the null hypothesis of no associations. Based on 10,000 permutations, all adjusted *P* values were non-significant; however, the probability for observing at least two significant associations was estimated to only 7%.

The E-cadherin gene is a major cell adhesion molecule and determinant of cell polarity. It is a component of the cadherin/catenin complex which is important for cellular polarity, normal tissue morphology and cellular differentiation (Grunwald 1993). The gene is a part of the adherens junction at which it provides cell-cell adhesion through Ca^{2+} -dependent homophilic binding between molecules on adjacent epithelial cells. In most malignancies originated from epithelial tissue, CDH1-mediated cell-cell adhesion is lost when the tumor becomes more malignant. Somatic loss of E-Cadherin is

considered to be a defining feature in invasive lobular breast cancers and diffuse gastric cancers (Berx et al. 1998), and germline mutations, since 1998 have been reported to predispose to hereditary diffuse gastric cancer (Oliveira et al. 2003). Aberrations in CDH1 expression have been shown to associate with PC progression, metastasis and poor prognosis (Umbas et al. 1994; Dunsmuir et al. 2000; Chunthapong et al. 2004). The variant A-allele of promoter SNP rs16260 has been associated with a reduction of 68% in gene transcriptional activity (Li et al. 2000). Genetic epidemiological studies on PC have found suggestive linkage as well as LOH on chromosome 16q in close proximity to the gene (Suzuki et al. 1996; Latil et al. 1997; Suarez et al. 2000; Goddard et al. 2001).

Three other studies concerning rs16260 and PC risk have been published; Verhage et al. (2002) analyzed 82 cases and 188 controls and found a significant increased risk for carriers of the A-allele. The study population utilized by Verhage included both sporadic and

hereditary PC cases and all subjects were Caucasians. Hajdinjak and Toplak (2004) identified a non-significant excess risk between 183 sporadic PC patients and 198 control subjects. All participants were Caucasians. In contrast to these studies, Tsukino et al. (2004) found no association among 219 cases and 219 controls in a Japanese population.

There are some strong points worth mentioning in this study. CAPS is a large population-based study with a well-characterized phenotype with all PC diagnosis histologically or cytologically confirmed. Using the method developed by Wacholder et al. (2004), the false-positive report probability (FPRP) of the positive findings for rs16260 is in the range of 0.25–0.06 assuming prior probabilities of 0.01–0.05 and a true OR of 1.5. The Swedish population is genetically homogenous which minimizes possible confounding by population stratification, notably because we matched on geographical region. Furthermore, several of the observed associations in the case-control population were confirmed in family-based association tests.

It should be stressed that even though we captured the majority of genetic variation in the gene, there may still exist causal rare alterations that were not identified in this study.

To date, no major PC gene has been identified, indicating the genetic susceptibility of PC to be complex, depending on both individual genes and their interplay. The cadherin/catenin complex—including CDH1, α -catenin, β -catenin and p120—are strong candidate genes interacting with each other. In the human genome, the close neighborhood of CDH1 reveals a strong LD pattern with its neighbor gene P-cadherin (CDH3), (Fig. 2a). Utilizing the HapMap project (<http://www.hapmap.org>) and the Haploview software (<http://www.broad.mit.edu/personal/jcbarret/haplo>), we inferred haplotype block structures using the method of LD confidence interval (Gabriel et al. 2002). This revealed an LD-block including the promoter region of CDH1 extending up-stream into the CDH3 gene (Fig. 2b). As CDH1, CDH3 is a cell adhesion molecule and this block extension makes CDH3 another interesting candidate gene for PC susceptibility.

In conclusion, we report a strong confirmation of the association between family-positive PC risk and a functional CDH1 promoter SNP in an independent population. This association was confirmed by both haplotype- and family-based association analyses. Due to the intrinsic difficulties in replication of genetic association studies, particularly in PC epidemiology, our findings are important and novel. These results together with the biological importance of CDH1 encourage further evaluation of genetic variation in cell adhesion molecules and their relation to PC risk.

Acknowledgements The authors thank all the participants in the CAPS study and in PC families, Ulrika Undén for skilfully coordinating the study center at Karolinska Institute, all urologists including their patients in the CAPS study and all urologists for

providing clinical data to the national registry of prostate cancer. We also thank Karin Andersson, Susan Lindh, Gabriella Thorén and Margareta Åswärd at the Regional Cancer Registries. In addition, we thank the CAPS steering committee including Pär Stattin, Jan-Erik Johansson and Eberhart Varenhorst. Furthermore, we thank Sören Holmgren and the personnel at the Medical Biobank in Umeå for skilfully handling the blood samples.

References

- Berx G, Becker KF, Höfler H, van Roy F (1998) Mutations of the human E-cadherin (CDH1) gene. *Hum Mutat* 12:226–237
- Besag J, Clifford P (1991) Sequential Monte Carlo P values. *Biometrika* 78:301–304
- Carter BS, Bova GS, Beaty TH, Steinberg GD, Childs B, Isaacs WB, Walsh PC (1993) Hereditary prostate cancer: epidemiology and clinical features. *J Urol* 150:797–802
- Chunthapong J, Seftor EA, Khalkhali-Ellis Z, Seftor RE, Amir S, Lubaroff DM, Heidger PM, Hendrix MJ (2004) Dual roles of E-cadherin in prostate cancer invasion. *J Cell Biochem* 91:649–661
- Clayton D (1999) A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* 65:1170–1177
- Colhoun HM, McKeigue PM, Davey Smith G (2003) Problems of reporting genetic associations with complex outcomes. *Lancet* 361:865–872
- Dunsmuir WD, Gillett CE, Meyer LC, Young MP, Corbishley C, Eeles RA, Kirby RS (2000) Molecular markers for predicting prostate cancer stage and survival. *BJU Int* 86:869–878
- Fredman D, Jobs M, Strömquist L, Brookes AJ (2004) DFold: PCR design that minimizes secondary structure and optimizes downstream genotyping applications. *Hum Mutat* 24:1–8
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229
- Goddard KA, Witte JS, Suarez BK, Catalona WJ, Olson JM (2001) Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4. *Am J Hum Genet* 68:1197–1206
- Grönberg H, Smith J, Emanuelsson M, Jonsson BA, Bergh A, Carpten J, Isaacs W, Xu J, Meyers D, Trent J, Damber JE (1999) In Swedish families with hereditary prostate cancer, linkage to the HPC1 locus on chromosome 1q24–25 is restricted to families with early-onset prostate cancer. *Am J Hum Genet* 65:134–140
- Grunwald GB (1993) The structural and functional analysis of cadherin calcium-dependent cell adhesion molecules. *Curr Opin Cell Biol* 5:797–805
- Hajdinjak T, Toplak N (2004) E-cadherin polymorphism–160 C/A and prostate cancer. *Int J Cancer* 109:480–481
- Howell WM, Jobs M, Gyllensten U, Brookes AJ (1999) Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nat Biotechnol* 17:87–88
- Jobs M, Howell WM, Brookes AJ (2002) Creating arrays by centrifugation. *Biotechniques* 32:1322–1324, 1326, 1329
- Jobs M, Howell WM, Stromqvist L, Mayr T, Brookes AJ (2003) DASH-2: flexible, low-cost, and high-throughput SNP genotyping by dynamic allele-specific hybridization on membrane arrays. *Genome Res* 13:916–924
- Jonsson BA, Adami HO, Häggqvist M, Bergh A, Göransson I, Stattin P, Wiklund F, Grönberg H (2004) –160C/A polymorphism in the E-cadherin gene promoter and risk of hereditary, familial and sporadic prostate cancer. *Int J Cancer* 109:348–352
- Latil A, Cussenot O, Fournier G, Driouch K, Lidereau R (1997) Loss of heterozygosity at chromosome 16q in prostate adenocarcinoma: identification of three independent regions. *Cancer Res* 57:1058–1062

- Lei H, Sjöberg-Margolin S, Salahshor S, Werelius B, Jandáková E, Hemminki K, Lindblom A, Vorechovský I (2002) CDH1 mutations are present in both ductal and lobular breast cancer, but promoter allelic variants show no detectable breast cancer risk. *Int J Cancer* 98:199–204
- Li LC, Chui RM, Sasaki M, Nakajima K, Perinchery G, Au HC, Nojima D, Carroll P, Dahiya R (2000) A single nucleotide polymorphism in the E-cadherin gene promoter alters transcriptional activities. *Cancer Res* 60:873–876
- Lindmark F, Zheng SL, Wiklund F, Bensen J, Bälter KA, Chang B, Hedelin M, Clark J, Stattin P, Meyers DA, Adami HO, Isaacs W, Grönberg H, Xu J (2004) H6D polymorphism in macrophage-inhibitory cytokine-1 gene associated with prostate cancer. *J Natl Cancer Inst* 96:1248–1254
- Nakamura A, Shimazaki T, Kaneko K, Shibata M, Matsumura T, Nagai M, Makino R, Mitamura K (2002) Characterization of DNA polymorphisms in the E-cadherin gene (CDH1) promoter region. *Mutat Res* 502:19–24
- Oliveira C, Seruca R, Caldas C (2003) Genetic screening for hereditary diffuse gastric cancer. *Expert Rev Mol Diagn* 3:201–215
- Prince JA, Feuk L, Howell WM, Jobs M, Emahazion T, Blennow K, Brookes AJ (2001) Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res* 11:152–162
- Schaid DJ (2004) The complex genetic epidemiology of prostate cancer. *Hum Mol Genet* 13(Spec No 1):R103–R121
- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 70:425–434
- Suarez BK, Lin J, Burmester JK, Broman KW, Weber JL, Banerjee TK, Goddard KA, Witte JS, Elston RC, Catalona WJ (2000) A genome screen of multiplex sibships with prostate cancer. *Am J Hum Genet* 66:933–944
- Suzuki H, Komiya A, Emi M, Kuramochi H, Shiraishi T, Yatani R, Shimazaki J (1996) Three distinct commonly deleted regions of chromosome arm 16q in human primary and metastatic prostate cancers. *Genes Chromosomes Cancer* 17:225–233
- Tsukino H, Kuroda Y, Imai H, Nakao H, Qiu D, Komiya Y, Inatomi H, Hamasaki T, Kohshi K, Osada Y, Katoh T (2004) Lack of evidence for the association of E-cadherin gene polymorphism with increased risk or progression of prostate cancer. *Urol Int* 72:203–207
- Umbas R, Isaacs WB, Bringuier PP, Schaafsma HE, Karthaus HF, Oosterhof GO, Debruyne FM, Schalken JA (1994) Decreased E-cadherin expression is associated with poor prognosis in patients with prostate cancer. *Cancer Res* 54:3929–3933
- Verhage BA, van Houwelingen K, Ruijter TE, Kiemeny LA, Schalken JA (2002) Single-nucleotide polymorphism in the E-cadherin gene promoter modifies the risk of prostate cancer. *Int J Cancer* 100:683–685
- Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 96:434–442
- Wooldridge J (2002) *Econometric analysis of cross section and panel data*. Massachusetts Institute of Technology, Cambridge