DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Fine Mapping the \(KLK3\) Locus on Chromosome 19q13.33 Associated With Prostate Cancer Susceptibility and PSA Levels

(Article begins on next page)

ORIGINAL INVESTIGATION

Fine mapping the *KLK3* locus on chromosome 19q13.33 associated with prostate cancer susceptibility and PSA levels

Hemang Parikh · Zhaoming Wang · Kerry A. Pettigrew · Jinping Jia · Sarah Daugherty · Meredith Yeager · Kevin B. Jacobs · Amy Hutchinson · Laura Burdett · Michael Cullen · Liqun Qi · Joseph Boland · Irene Collins · Thomas J. Albert · Lars J. Vatten · Kristian Hveem · Inger Njølstad · Geraldine Cancel-Tassin · Olivier Cussenot · Antoine Valeri · Jarmo Virtamo · Michael J. Thun · Heather Spencer Feigelson · W. Ryan Diver · Nilanjan Chatterjee · Gilles Thomas · Demetrius Albanes · Stephen J. Chanock · David J. Hunter · Robert Hoover · Richard B. Hayes · Sonja I. Berndt · Joshua Sampson · Laufey Amundadottir

Abstract Measurements of serum prostate-specific antigen (PSA) protein levels form the basis for a widely used test to screen men for prostate cancer. Germline variants in the gene that encodes the PSA protein (*KLK3*) have been shown to be associated with both serum PSA levels and prostate cancer. Based on a resequencing analysis of a 56 kb region on chromosome 19q13.33, centered on the *KLK3* gene, we fine mapped this locus by genotyping tag SNPs in 3,522 prostate cancer cases and 3,338 controls from five case–control studies. We did not observe a strong association with the *KLK3* variant, reported in previous studies to confer risk for prostate cancer (rs2735839; $P = 0.20$) but did observe three highly correlated SNPs (rs17632542, rs62113212 and rs62113214) associated with prostate cancer [$P = 3.41 \times 10^{-4}$, per-allele trend odds ratio (OR) = 0.77, 95% CI = 0.67–0.89]. The signal was apparent only for nonaggressive prostate cancer cases with Gleason score <7 and disease stage <III ($P = 4.72 \times 10^{-5}$, per-allele trend OR = 0.68, 95% CI = 0.57–0.82) and not for advanced cases with Gleason score ≥8 or stage ≥III ($P = 0.31$, per-allele trend OR = 1.12, 95% CI = 0.90–1.40). One of the three highly correlated SNPs,

H. Parikh · K. A. Pettigrew · J. Jia · I. Collins ·
L. Amundadottir
Laboratory of Translational Genomics, Division of Cancer
Epidemiology and Genetics, National Cancer Institute,
National Institutes of Health, Bethesda, MD 20877, USA

H. Parikh · Z. Wang · J. Jia · S. Daugherty · M. Yeager ·
K. B. Jacobs · A. Hutchinson · L. Burdett · M. Cullen · L. Qi ·
J. Boland · I. Collins · N. Chatterjee ·
G. Thomas · D. Albanes · S. J. Chanock · R. Hoover ·
S. I. Berndt · J. Sampson · L. Amundadottir
Division of Cancer Epidemiology and Genetics,
National Cancer Institute, National Institutes of Health,
Bethesda, MD 20892, USA

Z. Wang · M. Yeager · K. B. Jacobs · A. Hutchinson ·
L. Burdett · M. Cullen · L. Qi · J. Boland
Core Genotyping Facility, SAIC-Frederick, Inc., NCI-Frederick,
Frederick, MD 21702, USA

K. A. Pettigrew
Centre for Cancer Research and Cell Biology,
Queens University Belfast, Belfast BT9 7BL, Ireland

T. J. Albert
Roche NimbleGen, Madison, WI 53711, USA

L. J. Vatten · K. Hveem
Department of Public Health and General Practice,
Norwegian University of Science and Technology,
7489 Trondheim, Norway

I. Njølstad
Department of Community Medicine, Faculty of Health
Sciences, University of Tromsø, 9037 Tromsø, Norway

G. Cancel-Tassin · O. Cussenot · A. Valeri
Centre de Recherche pour les Pathologies Prostatiques
(CeRePP), Hôpital Tenon, Assistance Publique-Hôpitaux de
Paris, 75020 Paris, France

J. Virtamo
Department of Chronic Disease Prevention, National Institute
for Health and Welfare, Helsinki 00300, Finland

M. J. Thun · H. S. Feigelson · W. R. Diver
Epidemiology Research Program, American Cancer Society,
Atlanta, GA 30329, USA

G. Thomas
Synergie-Lyon-Cancer, INSERM U590, Centre Leon Berard,
69373 Lyon Cedex 08, France

rs17632542, introduces a non-synonymous amino acid change in the KLK3 protein with a predicted benign or neutral functional impact. Baseline PSA levels were 43.7% higher in control subjects with no minor alleles (1.61 ng/ml, 95% CI = 1.49–1.72) than in those with one or more minor alleles at any one of the three SNPs (1.12 ng/ml, 95% CI = 0.96–1.28) ($P = 9.70 \times 10^{-5}$). Together our results suggest that germline *KLK3* variants could influence the diagnosis of nonaggressive prostate cancer by influencing the likelihood of biopsy.

Introduction

Prostate cancer is the most commonly diagnosed non-cutaneous cancer in men in the United States with over 200,000 new cases and 30,000 deaths estimated in 2010 (Jemal et al. 2010). Over 35 prostate cancer risk loci have been reported to date from a series of genome-wide association studies (GWAS). One locus maps to a region on chromosome 19q13.33 marked by SNP rs2735839, located ∼600 bp downstream of *KLK3*, the gene that encodes prostate-specific antigen (PSA).

This locus was initially reported in a GWAS of prostate cancer conducted in subjects from the UK and Australia (with a *P* value of 1.5×10^{-18}) (Eeles et al. 2008). Other prostate cancer GWAS efforts, including the Cancer Genetic Markers of Susceptibility (CGEMS) project in the US, did not report a GWAS significant association between SNPs on chromosome 19q13.33 and prostate cancer risk (Gudmundsson et al. 2009; Sun et al. 2009; Thomas et al. 2008). An important difference in study design could account for the divergent results in the GWAS: in the study that reported the prostate cancer association for the *KLK3* locus, controls were selected to have low PSA levels (<0.5 ng/ml) (Eeles et al. 2008) whereas the other studies did not select controls based on PSA levels. In a follow-up analysis of the UK GWAS, the signal was observed although not as strongly as in the discovery data set (Kote-Jarai et al. 2008). In a subsequent re-analysis of the

D. J. Hunter
Program in Molecular and Genetic Epidemiology,
Department of Epidemiology, Harvard School of Public Health,
Boston, MA 02115, USA

R. B. Hayes
Division of Epidemiology, Department of Environmental
Medicine, New York University School of Medicine, New York,
NY 10016, USA

L. Amundadottir (✉)
Laboratory of Translational Genomics, Division of Cancer
Epidemiology and Genetics, National Institutes of Health,
8717 Grovemont Circle, Gaithersburg, MD 20877, USA
e-mail: amundadottirl@mail.nih.gov

CGEMS GWAS (within the Prostate, Lung, Colon Ovarian screening trial, PLCO) an association between rs2735839 and prostate cancer risk was seen only when control individuals were restricted to those with very low PSA levels (Ahn et al. 2008).

In a previous study, we catalogued common SNPs and insertion/deletion polymorphisms (indels) in the *KLK3* region by a resequence analysis of a 56 kb region flanking rs2735839 (chr19:56,019,829–56,076,043 bps; NCBI Build 36.3) in 78 unrelated individuals of European ancestry (Parikh et al. 2010). Based on these results, we tagged the genomic region surrounding rs2735839 and genotyped 24 SNPs in five prostate cancer case–control studies from the United States, France, Norway and Finland (1994; Calle et al. 2002; Naess et al. 2008; Prorok et al. 2000; Valeri et al. 2003).

In our analysis, we did not observe a significant association between rs2735839 and prostate cancer risk, but found suggestive evidence for three highly correlated SNPs in the *KLK3* gene. Interestingly, the association with these markers was observed only for nonaggressive prostate cancer, defined by a Gleason score lower than 7 and disease stage <III, and not for cases with Gleason score 8 or greater or Stage III and above (also known as advanced disease). These three SNPs are also associated with baseline serum PSA levels. Our results suggest that the observed association between germline variation in the *KLK3* gene and nonaggressive prostate cancer, at least in part, may be mediated by the association between the germline variation and PSA levels.

Methods

Sample collections

This study included 3,522 prostate cancer cases and 3,338 control subjects from four case–control studies nested within cohorts and one hospital based case–control study (Supplemental Table 1), previously analyzed in the Cancer Genetics Markers of Susceptibility (CGEMS) study. The cohort studies were: the Prostate, Lung, Colon and Ovarian (PLCO) Cancer Screening Trial (Gohagan et al. 2000; Prorok et al. 2000), the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC) (1994), the American Cancer Society Prevention Study II Nutrition Cohort (CPS-II) (Calle et al. 2002) and the Cohort of Norway (CONOR) (Naess et al. 2008). The case–control study was the French Prostate Case–Control Study (CeRePP) (Valeri et al. 2003). Details of the PLCO, ATBC, CPS-II, CONOR and CeRePP have been previously described elsewhere in CGEMS (Yeager et al. 2009). All studies were approved by the appropriate institutional review boards.

SNP selection, pre-genotyping quality control (QC) and genotyping

Pairwise analysis of linkage disequilibrium identified a set of tagging single-nucleotide polymorphisms (SNPs) using a criteria of a linkage disequilibrium coefficient threshold of $r^2 \geq 0.8$ and minor allele frequency (MAF) ≥ 0.05 based on the resequencing data in the *KLK3* region on chr19:56,019,829–56,076,043 (NCBI Build 36.3) (Supplemental Table 2) (Parikh et al. 2010) in the program GLU (http://code.google.com/p/glu-genetics). Twelve tag SNPs, covering an 8 kb region surrounding rs2735839, were selected for genotyping. They capture 90% of common variation (MAF ≥ 0.05) across the 8 kb region surrounding rs2735839 ($r^2 \geq 0.8$) based on the 1,000 Genomes Project data (overall 40 SNPs in the Caucasians from Utah, USA (CEU) population samples, June 2010 release, http://www.1000genomes.org/). An additional 12 SNPs with $r^2 > 0.3$ with rs2735839 were also included. The program, Tagger, was used to estimate the proportion of common variation across the region captured by the tag SNPs based on a multi-marker tagging approach (de Bakker et al. 2005). The 24 SNPs (Fig. 1; Supplemental Table 3) extend across the entire *KLK3* gene and in addition cover 12,907 bps upstream and 3,424 bps downstream of the gene (Parikh et al. 2010).

After pre-genotyping quality control at the Core Genotyping Facility (CGF) of the National Cancer Institute (http://cgf.nci.nih.gov/operations/pregenotyping-qaqc.html), 24 SNPs were genotyped in 8,505 samples using TaqMan genotyping assays (ABI, Foster City, CA, USA). For quality control assessment, 5.2–6.1% of samples from each of the five studies were genotyped in duplicate (a total of 466 samples) with 98.1% genotype concordance overall. When 13 duplicate pairs with low completion rates (<80%) were excluded, the concordance for the remaining 453 duplicate sample increased to 99.6%. One SNP, rs2735839, had previously been genotyped as part of CGEMS for all studies except CONOR. Genotypes for this SNP were retrieved from CGEMS for subjects in PLCO, ATBC, CPS-II and CeRePP and combined with other genotypes before analysis.

Post-genotyping quality control

As per the standard CGF protocol, samples with <80% completion were excluded ($n = 612$), which increased the overall completion rate to >98% in the combined data set and the locus completion rate to >90%. An additional 90 subjects with <80% estimated CEU admixture were excluded, as well as 159 participants that did not meet eligibility requirements (Supplemental Table 4) based on prior GWAS and dense SNP genotyping (Yeager et al. 2007, 2009). Deviations from Hardy–Weinberg equilibrium (HWE) were tested using an exact test in the control group of each cohort. Significant deviation from HWE ($P < 1 \times 10^{-4}$) was observed for one SNP (CGF_34700) that was subsequently excluded in further analysis. A final genotype data set from 6,860 subjects for 23 SNPs was analyzed. A two group χ^2 test of equal proportions (Newcombe 1998) was performed to determine the differences in MAFs for each of the 23 SNPs analyzed in this study and the CEU population available in the 1,000 Genomes Project data (June 2010 release, http://www.1000genomes.org/).

Fig. 1 **a** Genomic structure of the *KLK3* region indicating the location of the 24 SNPs genotyped. SNPs that associate with prostate cancer and/or PSA levels are colored in *black* (nonaggressive prostate cancer and PSA levels) and in *red* (only PSA levels). Locations of other SNPs are indicated by *arrows*. One SNP that was excluded from analysis based on deviation from Hardy–Weinberg equilibrium is shown with a *gray arrow*. The boundary of the *KLK3* gene is indicated with a *horizontal arrow*. Exons are indicated by *boxes* and introns are by *solid lines*. **b** Linkage disequilibrium (LD) map for the 24 genotyped SNPs in *KLK3*. Each *diamond* represents the correlation coefficient r^2, between every pair of SNPs. *Darker shades* represent stronger LD
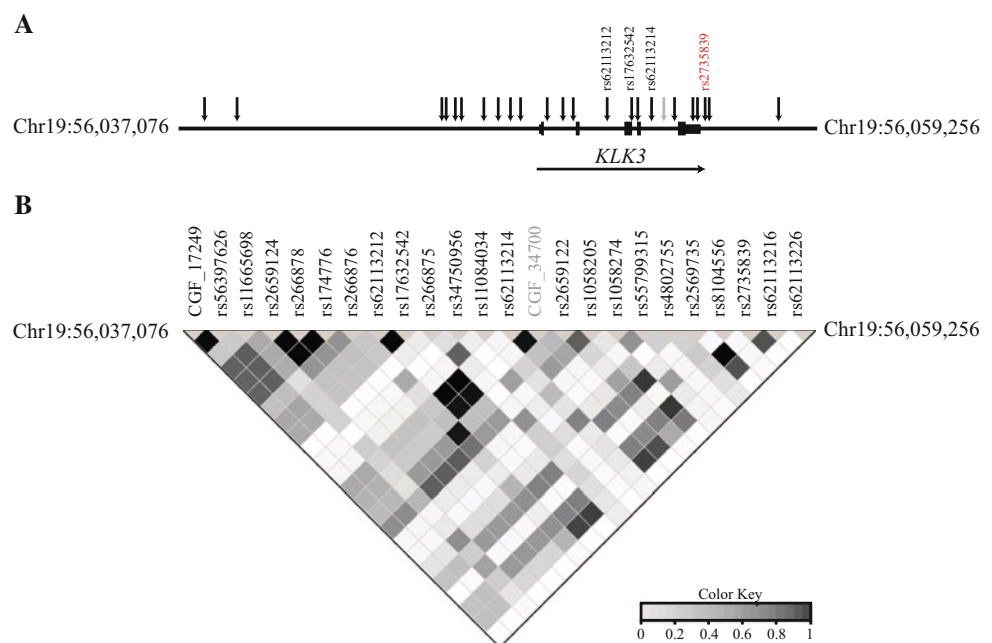
Power was estimated using Quanto under the log-additive model at $\alpha = 0.05$ [http://hydra.usc.edu/gxe/ (Gauderman 2002)]. This study had >80% power to detect an association with an odds ratio of 1.25 (assuming a MAF of 0.05, a prevalence of prostate cancer of 1.5067% and alpha of 0.05) (http://seer.cancer.gov/csr/1975_2007/).

Statistical analysis

Single locus case–control association analysis was performed using the programs, GLU and PLINK. In order to incorporate prostate cancer stage and grade at diagnosis, we discriminated between nonaggressive (Gleason score <7 AND disease stage <III, $n = 1,654$), advanced prostate cancer (Gleason score ≥8 OR disease stage ≥III, $n = 678$) or aggressive prostate cancer (Gleason score ≥7 OR disease stage ≥III, $n = 1,475$) as defined in CGEMS (Thomas et al. 2008; Yeager et al. 2007). Logistic regression models were adjusted for study (ATBC, CPS-II, CeRePP, PLCO and CONOR), arm (for CeRePP and PLCO), age (in 10-year categories) and two principal components of population structure as identified in the PLCO study (as these two principal components were significant predictors of the disease status identified under null logistic model). Both a two degree of freedom (df) genotype effect model and 1 df trend effect model were tested. We found no evidence of heterogeneity in genetic effects across study using the Q and I^2 statistics.

We analyzed the association between SNPs in *KLK3* and serum PSA levels in 702 control subjects and 461 incident prostate cancer cases from the PLCO cohort. PSA measurements at baseline (the first PSA test performed as men entered the PLCO Cancer Screening Trial) were available for this study (Thomas et al. 2008; Ahn et al. 2008). We excluded men diagnosed with prostate cancer in their first year of the PLCO trial ($n = 217$) as many present with extremely high PSA levels probably due to more advanced disease as compared to men diagnosed later in the trial (range: 0.32–617.80 ng/ml, average 19.11 ng/ml) and control subjects with missing PSA level information ($n = 20$). The average time between baseline PSA and diagnosis for the cases included in this study was 2.5 years (median 2.0 years). Average PSA levels were 1.53 ng/ml (median 1.11) in control subjects; 8.94 ng/ml (median 4.12) in all cases; and 4.23 ng/ml (median 3.43) in cases diagnosed after the first year of the PLCO trial. For each desired SNP, we defined a group of individuals without a minor allele (G = 0) and with at least one minor allele (G = 1) and estimated the mean baseline PSA in these two groups. To minimize the influence of the larger values, we also created a truncated PSA level for control subjects (at 4 ng/ml) and cases (at 8 ng/ml) and performed a similar analysis. For a formal statistical test, we compared PSA and genotype (G),

using a general linear model with a gamma distribution and a log link function. The results were essentially equivalent to those obtained by performing a linear model on the log-transformed PSA level. We adjusted for center, age group and population stratification. Only one principal component, estimated to adjust for population structure, was a significant predictor of PSA level under null linear model based on prior analyses in CGEMS (Yeager et al. 2007).

Imputation

We imputed genotypes for an approximately 56 kb region on chromosome 19q33.13 (chr19:56,019,829–56,076,043 bps; NCBI Build 36.3) for 2,303 samples (1,179 cases and 1,124 control subjects) from the PLCO Cancer Screening Trial using IMPUTEv2 (Howie et al. 2009). The PLCO subjects were genotyped as part of the CGEMS prostate cancer whole-genome scan (Phase 1A with Illumina Human-Hap300 and Phase 1B with Illumina HumanHap240 assays, Illumina Corp., San Diego, CA, USA) (Thomas et al. 2008; Yeager et al. 2007). Other studies that were part of this work did not have whole-genome SNP data and were not used for imputation. We used 120 estimated CEU haplotypes from the June 2010 release of the 1,000 Genomes Pilot 1 data (Durbin et al. 2010) as well as 298 estimated CEU and Tuscans in Italy (TSI) haplotypes from the HapMap3 project to carry out genotype imputation (Altshuler et al. 2010). Association analysis was conducted using SNPTESTv2 (Marchini et al. 2007) with logistic regression models (1 df trend test) adjusted for age (in 10-year categories) and two principal components of genetic structure as identified in the PLCO study (Yeager et al. 2007).

Bioinformatic analysis of a KLK3 coding variant

Four computational tools were used to assess the predictive value of a non-synonymous substitution caused by rs17632542 on the KLK3 protein. Polyphen-2 (Polymorphism Phenotyping) (Ramensky et al. 2002), SIFT (Sorting Intolerant From Tolerant) (Ng and Henikoff 2001), SNPs&GO (SNPs and Gene Ontology) (Calabrese et al. 2009) and Prophyler (Protein Phylogeny and Evolutionary Rates) (Binkley et al. 2010) predict functionally relevant disease-related mutations based on one or more of the following: sequence conservation, steric and biochemical properties of amino acid residues and protein functional annotation (SNPs&GO only).

Results

A total of 3,522 prostate cancer cases and 3,338 control subjects from five case–control studies were analyzed in

the current study (Supplemental Table 1), including 1,654 men diagnosed with nonaggressive prostate cancer (Gleason score <7 and disease stage <III) and 678 men with advanced prostate cancer (Gleason score ≥8 or stage ≥III). We also analyzed men with aggressive prostate cancer ($n = 1,475$) by including cases with a Gleason score of 7 or higher (Gleason score ≥7 or stage ≥III).

SNP selection and genotyping

Twenty-four SNPs in a 22 kb region surrounding rs2735839, the initial marker found to be associated with prostate cancer risk, were genotyped (physical locations shown in Fig. 1a). Twelve of these SNPs were selected to tag the genomic region of *KLK3* at a linkage disequilibrium (LD) coefficient threshold of $r^2 \geq 0.8$ and a minor allele frequency (MAF) ≥0.05. They capture 90% of common variation ($r^2 \geq 0.8$ and MAF ≥0.05) across the 8 kb region surrounding rs2735839 based on the 1,000 Genomes Project data. An additional 12 SNPs with an $r^2 > 0.3$ with rs2735839 were also genotyped to capture markers in both modest and high LD with this SNP. One SNP was excluded from further analysis because of substantial deviation from Hardy–Weinberg equilibrium. No significant differences were observed between minor allele frequencies of SNPs in control subjects of the current study and those of the CEU population from the 1,000 Genomes Project (Supplemental Table 3). The linkage disequilibrium structure across the *KLK3* locus in control subjects (Fig. 1b) reflects the relatively low degree of LD seen in our resequence analysis (Parikh et al. 2010).

Association of SNPs in *KLK3* with prostate cancer

Four SNPs were associated with overall prostate cancer at a significance level of $P < 1 \times 10^{-3}$ (Table 1). The most significant association was with rs17632542 ($P = 3.41 \times 10^{-4}$, per-allele trend OR 0.77, 95% CI 0.67–0.89; unconstrained heterozygote OR (OR_{het}) 0.80, 95% CI 0.69–0.93 and homozygote OR (OR_{hom}) 0.37, 95% CI 0.16–0.86). Two additional SNPs, both highly correlated to rs17632542 ($r^2 \geq 0.97$), showed a similar level of significance: rs62113214 ($P = 3.57 \times 10^{-4}$) and rs62113212 ($P = 1.17 \times 10^{-3}$). No other SNPs were nominally significant after adjusting for any of the three SNPs (data not shown). The marker initially reported to be associated with prostate cancer risk in a GWAS (Eeles et al. 2008) was not significant in our study (rs2735839, $P = 0.197$).

When the analysis was stratified by disease severity, the association was observed only with nonaggressive prostate cancers (Gleason score <7 and disease stage <III) for the three highly correlated SNPs (Table 2). SNP rs62113214 had the lowest P value ($P = 4.72 \times 10^{-5}$, per-allele trend

OR 0.68, 95% CI 0.57–0.82; unconstrained heterozygote OR (OR_{het}) 0.70, 95% CI 0.57–0.85 and homozygote OR (OR_{hom}) 0.34, 95% CI 0.11–1.04). The two highly correlated SNPs rs17632542 ($P = 5.49 \times 10^{-5}$) and rs62113212 ($P = 5.74 \times 10^{-5}$) showed similar results. SNP rs2735839 was marginally associated with nonaggressive prostate cancer ($P = 0.031$). No SNPs were associated with advanced prostate cancer ($P > 0.05$) as defined by a Gleason score ≥8 or disease stage ≥III (Table 3) or with aggressive prostate cancer ($P > 0.05$) as defined by a Gleason score ≥7 or stage ≥III when compared to control subjects (Supplemental Table 5). When nonaggressive cases were compared to advanced cases a significant difference was seen for the same three SNPs (rs62113212: $P = 3.23 \times 10^{-5}$, per-allele trend OR 0.59, 95% CI 0.46–0.76) indicating that this signal could be related to the less aggressive and often latent forms of prostate cancer, which are diagnosed partly as a consequence of PSA screening (Supplemental Table 6).

KLK3 SNPs and serum PSA levels

The PSA test is commonly used to screen for prostate cancer and has been shown to increase detection of early, nonaggressive prostate cancer. We assessed baseline PSA levels according to different genotypes at the three highly correlated SNPs: rs17632542, rs62113212 and rs62113214, as well as at rs2735839. Because PSA information was only available in the PLCO cohort, we performed the analysis in this subset (702 control subjects and 461 cases). Baseline PSA levels differed significantly (Table 4) between control individuals without or with one or more minor alleles at any one of the three highly correlated SNPs. Mean PSA levels were 1.61 ng/ml (95% CI 1.49–1.72) in men without a minor allele, and 1.12 ng/ml (95% CI 0.96–1.28) in men with at least one minor allele at rs62113212. The significance of the differences was confirmed by a general linear model ($P = 9.7 \times 10^{-5}$). The box plot (Fig. 2a) shows that the entire distribution of PSA levels was shifted to lower values in the group with the minor allele (G = 1). Because of the high correlation with rs62113212, the results for rs17632542 and rs62113214 were nearly identical (Table 4). After truncating PSA at 4 ng/ml in control subjects, to limit the influence of larger values, we still found the mean PSA levels for the two groups to be significantly different [1.49 ng/ml (95% CI 1.41–1.58) vs. 1.11 ng/ml (95% CI 0.96–1.26), $P = 1.5 \times 10^{-5}$ for rs62113212].

PSA levels were also lower in men carrying one or more minor alleles at rs2735839 (1.28 ng/ml for carriers of at least one minor allele compared to 1.64 ng/ml) as shown in Table 4 and Fig. 2b. Although not statistically significant, cases with at least one minor allele at any of the four

Table 1 Association results for all prostate cancer cases as compared to control subjects for the 23 SNPs in the *KLK3* region

dbSNP rs number	Position	Alleles[a]	MAF[b]		Subjects		χ^2	P^c	Allelic OR (95% CI)	Genotype OR$_{Het}$ (95% CI)	Genotype OR$_{Hom}$ (95% CI)
			Control	Case	Control	Case					
CGF_17249	56037076	T/C	0.094	0.079	3,321	3,503	11.49	6.98×10^{-04}	0.81 (0.71–0.91)	0.81 (0.70–0.92)	0.66 (0.38–1.16)
rs56397626	56038334	T/C	0.093	0.078	3,322	3,509	12.14	4.95×10^{-04}	0.80 (0.71–0.91)	0.80 (0.70–0.92)	0.64 (0.37–1.14)
rs11665698	56046223	C/A	0.305	0.281	3,310	3,496	10.15	1.44×10^{-03}	0.88 (0.82–0.95)	0.91 (0.82–1.01)	0.75 (0.63–0.90)
rs2659124	56046409	T/A	0.128	0.116	3,325	3,516	5.41	0.020	0.88 (0.79–0.98)	0.88 (0.78–0.99)	0.80 (0.54–1.19)
rs266878	56050926	C/G	0.128	0.115	3,327	3,510	5.80	0.016	0.88 (0.79–0.98)	0.87 (0.77–0.98)	0.81 (0.54–1.22)
rs174776	56051664	C/T	0.128	0.116	3,317	3,510	5.00	0.025	0.89 (0.80–0.99)	0.88 (0.78–0.99)	0.84 (0.56–1.25)
rs266876	56052630	T/C	0.256	0.246	3,318	3,497	2.59	0.108	0.94 (0.86–1.01)	1.00 (0.90–1.11)	0.78 (0.64–0.96)
rs62113212	56052652	C/T	0.072	0.060	3,325	3,508	10.53	1.17×10^{-03}	0.79 (0.69–0.91)	0.82 (0.70–0.95)	0.41 (0.18–0.92)
rs17632542	56053569	T/C	0.072	0.059	3,325	3,504	12.83	3.41×10^{-04}	0.77 (0.67–0.89)	0.80 (0.69–0.93)	0.37 (0.16–0.86)
rs266875	56053749	A/G	0.465	0.478	3,312	3,499	2.48	0.115	1.06 (0.99–1.13)	1.00 (0.89–1.13)	1.12 (0.98–1.29)
rs34750956	56053808	C/T	0.107	0.106	3,327	3,512	0.08	0.783	0.98 (0.88–1.10)	1.03 (0.91–1.17)	0.74 (0.49–1.14)
rs11084034	56054116	C/T	0.354	0.366	3,246	3,429	2.87	0.090	1.07 (0.99–1.15)	1.03 (0.93–1.15)	1.16 (0.99–1.36)
rs62113214	56054349	T/G	0.072	0.060	3,321	3,504	12.74	3.57×10^{-04}	0.77 (0.67–0.89)	0.80 (0.69–0.93)	0.37 (0.16–0.87)
rs2659122	56054838	T/C	0.242	0.234	3,302	3,486	1.69	0.193	0.95 (0.87–1.03)	1.00 (0.90–1.11)	0.81 (0.65–1.00)
rs1058205	56055210	T/C	0.168	0.153	3,329	3,515	6.32	0.012	0.89 (0.80–0.97)	0.91 (0.81–1.01)	0.70 (0.51–0.96)
rs1058274	56055260	A/G	0.332	0.331	3,323	3,517	0.08	0.781	0.99 (0.92–1.06)	1.01 (0.91–1.13)	0.96 (0.81–1.13)
rs55799315	56056035	C/G	0.172	0.177	3,316	3,504	0.22	0.637	1.02 (0.93–1.12)	1.03 (0.92–1.15)	1.01 (0.77–1.33)
rs4802755	56056052	C/T	0.469	0.454	3,321	3,501	2.91	0.088	0.94 (0.88–1.01)	0.90 (0.80–1.01)	0.89 (0.77–1.02)
rs2569735	56056081	G/A	0.135	0.127	3,320	3,505	2.02	0.156	0.93 (0.84–1.03)	0.94 (0.84–1.06)	0.77 (0.52–1.15)
rs8104556	56056192	C/T	0.158	0.153	3,320	3,510	0.46	0.496	0.97 (0.88–1.06)	0.93 (0.83–1.05)	1.06 (0.79–1.42)
rs2735839	56056435	G/A	0.136	0.130	2,980	3,150	1.66	0.197	0.93 (0.83–1.04)	0.94 (0.83–1.07)	0.79 (0.52–1.21)
rs62113216	56056615	T/A	0.096	0.087	3,332	3,517	4.59	0.032	0.88 (0.78–0.99)	0.91 (0.79–1.03)	0.59 (0.35–0.99)
rs62113226	56059256	T/C	0.069	0.071	3,258	3,427	0.44	0.508	1.05 (0.91–1.20)	1.02 (0.88–1.19)	1.38 (0.72–2.66)

OR odds ratio, *Het* heterozygous, *Hom* homozygous for minor allele, *CI* 95% confidence interval. Coordinates are based on NCBI genome build 36.3. CGF denotes novel SNPs

[a] Major allele/minor allele

[b] Minor allele frequency

[c] 1 *df* score test

Table 2 Association results for nonaggressive prostate cancer (Gleason score <7 and disease stage <III) as compared to control subjects for the three highly correlated SNPs and rs2735839

dbSNP rs number	Position	Alleles[a]	MAF[b]		Subjects		χ^2	P^c	Allelic OR (95% CI)	Genotype OR$_{Het}$ (95% CI)	Genotype OR$_{Hom}$ (95% CI)
			Control	Case	Control	Case					
rs62113212	56052652	C/T	0.072	0.053	3,325	1,652	16.19	5.74×10^{-05}	0.69 (0.57–0.82)	0.70 (0.58–0.85)	0.33 (0.11–0.99)
rs17632542	56053569	T/C	0.072	0.054	3,325	1,652	16.27	5.49×10^{-05}	0.69 (0.57–0.82)	0.70 (0.57–0.85)	0.35 (0.11–1.05)
rs62113214	56054349	T/G	0.072	0.054	3,321	1,644	16.56	4.72×10^{-05}	0.68 (0.57–0.82)	0.70 (0.57–0.85)	0.34 (0.11–1.04)
rs2735839	56056435	G/A	0.136	0.122	2,980	1,422	4.65	0.031	0.85 (0.74–0.99)	0.85 (0.72–1.00)	0.76 (0.44–1.29)

OR odds ratio, *Het* heterozygous, *Hom* homozygous for minor allele, *CI* 95% confidence interval. Coordinates are based on NCBI genome build 36.3

[a] Major allele/minor allele

[b] Minor allele frequency

[c] 1 *df* score test

Table 3 Association results for advanced prostate cancer (Gleason score ≥8 or disease stage ≥III) as compared to control subjects for the three highly correlated SNPs and rs2735839

dbSNP rs number	Position	Alleles[a]	MAF[b]		Subjects		χ^2	P^c	Allelic OR (95% CI)	Genotype OR$_{Het}$ (95% CI)	Genotype OR$_{Hom}$ (95% CI)
			Control	Case	Control	Case					
rs62113212	56052652	C/T	0.072	0.085	3,325	674	1.02	0.312	1.12 (0.90–1.40)	1.17 (0.92–1.49)	0.72 (0.20–2.56)
rs17632542	56053569	T/C	0.072	0.081	3,325	671	0.31	0.578	1.07 (0.85–1.34)	1.13 (0.89–1.43)	0.48 (0.11–2.17)
rs62113214	56054349	T/G	0.072	0.084	3,321	676	0.69	0.407	1.10 (0.88–1.37)	1.14 (0.90–1.45)	0.73 (0.21–2.60)
rs2735839	56056435	G/A	0.136	0.154	2,980	603	2.10	0.148	1.15 (0.95–1.38)	1.16 (0.94–1.43)	1.22 (0.61–2.44)

OR odds ratio, *Het* heterozygous, *Hom* homozygous for minor allele, *CI* 95% confidence interval. Coordinates are based on NCBI genome build 36.3. Results for aggressive prostate cancer as determined by a Gleason score ≥7 or disease stage of ≥III are shown in supplemental Table 5

[a] Major allele/minor allele

[b] Minor allele frequency

[c] 1 *df* score test

SNPs also had lower PSA levels than their respective counterparts.

Genotype imputation

Imputation was performed in the PLCO GWAS data set genotyped within the Cancer Genetic Markers of Susceptibility (CGEMS) study (Thomas et al. 2008; Yeager et al. 2007) utilizing the publicly available HapMap and 1,000 genomes pilot project data sets (June 2010 release). This added 239 imputed SNPs to the 27 SNPs already genotyped as part of the CGEMS study (Thomas et al. 2008; Yeager et al. 2007) in the 56 kb region (chr19: 56,020,000–56,076,000) for a combined data set of 266 SNPs in 1,179 cases and 1,124 control subjects. We did not observe a stronger association ($P < 1 \times 10^{-4}$) for either prostate cancer overall or specifically for nonaggressive prostate cancer using imputed genotype data (Supplemental Tables 7 and 8). However, the imputation was performed in a subset of the cohorts (approximately 20% of subjects).

Prediction of functional effects of rs17632542 on KLK3

The three SNPs that most significantly associated with prostate cancer risk in our study: rs62113212, rs62113214 and rs17632542 are all located in the *KLK3* gene (Fig. 1a). The first two are located in introns 2 and 4, respectively, whereas rs17632542 lies within exon 4. The minor allele (C) of rs17632542 causes a non-synonymous amino acid change from isoleucine (hydrophobic) to threonine (polar) at position 179 (Ile179Thr) in KLK3. This amino acid is conserved in humans, chimpanzee and rhesus but not in other mammals or vertebrates. The amino acid change seems to have deleterious effects based on Sorting Intolerant from Tolerant (SIFT) analysis with a SIFT score of 0.03. A benign impact is predicted by Polyphen-2 and a moderate impact is predicted by ProPhylER (multivariate analysis of protein polymorphism (MAPP) *P* value of 0.0058). Finally, when taking functional annotation into account (SNPs&Go) a neutral substitution is predicted (reliability index 5 out of 10).

Table 4 Association between the three highly correlated SNPs as well as rs2735839 and baseline PSA levels in the PLCO cohort

dbSNP rs number	Position	Alleles[b]	Means[a]				General linear model (adjusted)		
			G = 0[c]		G = 1[d]				
			n	Mean (95% CI)	n	Mean (95% CI)	β	Reduction[e]	P
Controls									
rs62113212	56052652	C/T	584	1.61 (1.49–1.72)	108	1.12 (0.96–1.28)	−0.36	0.30	9.70×10^{-05}
rs17632542	56053569	T/C	582	1.61 (1.49–1.72)	107	1.13 (0.97–1.29)	−0.35	0.30	1.50×10^{-04}
rs62113214	56054349	T/G	583	1.60 (1.49–1.72)	107	1.14 (0.98–1.30)	−0.34	0.29	2.40×10^{-04}
rs2735839	56056435	G/A	491	1.64 (1.51–1.77)	199	1.28 (1.12–1.44)	−0.25	0.22	1.00×10^{-03}
Cases									
rs62113212	56052652	C/T	387	4.35 (3.83–4.86)	51	3.46 (3.01–3.91)	−0.19	0.17	0.12
rs17632542	56053569	T/C	386	4.35 (3.84–4.86)	48	3.57 (3.13–4.01)	−0.16	0.15	0.20
rs62113214	56054349	T/G	387	4.34 (3.83–4.85)	50	3.48 (3.03–3.94)	−0.19	0.17	0.13
rs2735839	56056435	G/A	325	4.24 (3.73–4.76)	112	3.92 (3.19–4.64)	−0.10	0.10	0.23

Coordinates are based on NCBI genome build 36.3

β β-coefficient, *CI* 95% confidence interval

[a] Baseline PSA levels are shown in ng/ml

[b] Major allele/minor allele

[c] G = 0, individuals without a minor allele

[d] G = 1, individuals with at least one minor allele

[e] Reduction, % reduction in G = 1 as compared to G = 0

Fig. 2 *Box plots* showing PSA levels (ng/ml) at baseline in control subjects from the PLCO study for carriers (G = 1) and noncarriers (G = 0) of minor alleles at rs62113212 (**a**) and rs2735839 (**b**). Median values are indicated by a *horizontal line* and 25 and 75% are at the *bottom* and *top* of the *boxes*
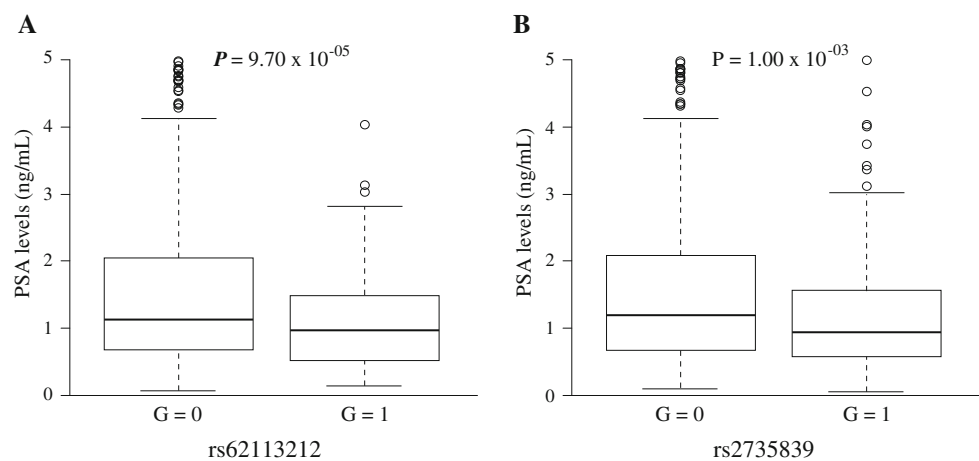
Discussion

Here we describe a comprehensive fine mapping effort in the *KLK3* gene based on resequence data across a region of chr19q13.33, surrounding rs2735839, a SNP previously associated with prostate cancer (Eeles et al. 2008; Kote-Jarai et al. 2008) and/or serum prostate-specific antigen (PSA) levels (Ahn et al. 2008). Based on a comprehensive catalog of common variation in a 56 kb region surrounding *KLK3* by deep resequencing (Parikh et al. 2010) we now report the follow-up genotyping of tag SNPs in close to 7,000 subjects drawn from five prostate cancer studies. We did not observe significant association between rs2735839 and overall prostate cancer risk. However, three highly correlated, neighboring SNPs in *KLK3* were associated with overall prostate cancer risk. This signal was detected only in association with nonaggressive prostate cancer and was not present in men with advanced prostate cancer. The significance level for these three SNPs is modest and did not reach the threshold for genome-wide significance in GWAS (2007). Nevertheless, our findings suggest that these three SNPs are associated with either developing or being diagnosed with nonaggressive prostate cancer, potentially due to differential case identification related to PSA level. Two of the SNPs are intronic but the third causes a non-synonymous amino acid substitution in the

KLK3 protein albeit with minimal predicted shift in function.

A recent report did not find a significant association between SNPs in *KLK* genes and prostate cancer risk in a Swedish population-based case–control study comprising 1,419 prostate cancer cases and 736 controls. This study analyzed SNPs in coding and promoter regions of all 15 kallikrein genes and included both rs17632542 and rs2735839 (but not rs62113212 or rs62113214) (Klein et al. 2010). We did not observe additional signals in *KLK3* or in the neighboring *KLK2* and *KLK15* genes by imputation.

We also observed that baseline PSA levels were lower in control subjects carrying one or more minor alleles at the three highly correlated SNPs. This may indicate that one or more of these alleles directly cause a reduction in serum PSA levels, possibly through regulatory effects (on transcription of the gene), through altered protein stability or reduced detection of serum PSA. However, since many factors could play a role in regulating serum PSA levels this question still remains open and requires validation and further investigations. Reduced PSA levels were also noted in carriers of minor alleles at rs2735839 although the effect was less pronounced.

The implications of our study directly pertain to the relationship between germline *KLK3* variants, serum PSA levels and detection of prostate cancer through PSA screening. Since the introduction of the PSA test as a screening tool in the US in the late 1980s, reported prostate cancer incidence rates have risen sharply (Welch and Albertsen 2009), especially for early stage disease and younger individuals (Shao et al. 2009). A significant number of men are likely to be 'over-diagnosed' as a result of PSA screening, where patients with clinically insignificant cancers that are unlikely to progress within their lifetime are diagnosed and treated, often with grave adverse effects on quality of life (Welch and Albertsen 2009). Two large randomized screening trials are now ongoing to assess the benefits of the PSA test for prostate cancer screening and reduction of mortality. Although interim results have been published, it is not clear if the test reduces mortality from prostate cancer (Andriole et al. 2009; Schroder et al. 2009).

We have fine mapped a challenging association signal in the *KLK3* gene to three highly correlated SNPs centromeric to the original signal reported by Eeles et al. (2008). Since we observed that the same three SNPs were also significantly associated with baseline PSA levels in the prospective PLCO cohort, we interpret our results to suggest that the effect may perhaps be a consequence of PSA screening as well. It is possible that the *KLK3* locus also contributes to prostate cancer risk. Further studies are needed to dissect the contribution of genetic variation in *KLK3* to PSA levels and prostate risk separately in an effort to elucidate the possibility of pleotropic effects of the *KLK3* locus.

URLs

References

Ahn J, Berndt SI, Wacholder S, Kraft P, Kibel AS, Yeager M, Albanes D, Giovannucci E, Stampfer MJ, Virtamo J, Thun MJ, Feigelson HS, Cancel-Tassin G, Cussenot O, Thomas G, Hunter DJ, Fraumeni JF Jr, Hoover RN, Chanock SJ, Hayes RB (2008) Variation in KLK genes, prostate-specific antigen and risk of prostate cancer. Nat Genet 40:1032–1034 author reply 1035–1036

Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Bonnen PE, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Gonzaga-Jauregui C, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Zhang Q, Ghori MJ, McGinnis R, McLaren W, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467:52–58

Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR, Fouad MN, Gelmann EP, Kvale PA, Reding DJ, Weissfeld JL, Yokochi LA, O'Brien B, Clapp JD, Rathmell JM, Riley TL,

Hayes RB, Kramer BS, Izmirlian G, Miller AB, Pinsky PF, Prorok PC, Gohagan JK, Berg CD (2009) Mortality results from a randomized prostate-cancer screening trial. N Engl J Med 360:1310–1319

Binkley J, Karra K, Kirby A, Hosobuchi M, Stone EA, Sidow A (2010) ProPhylER: a curated online resource for protein function and structure based on evolutionary constraint analyses. Genome Res 20:142–154

Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 30:1237–1244

Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML, Feigelson HS, Thun MJ (2002) The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer 94:2490–2501

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D (2005) Efficiency and power in genetic association studies. Nat Genet 37:1217–1223

Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurles ME, McVean GA (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073

Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, Ardern-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lopha-tananon A, Bryant SL, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Fisher C, Jamieson C, Cooper CS, English DR, Hopper JL, Neal DE, Easton DF (2008) Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 40:316–321

Gauderman WJ (2002) Sample size requirements for matched case–control studies of gene–environment interaction. Stat Med 21:35–50

Gohagan JK, Prorok PC, Hayes RB, Kramer BS (2000) The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. Control Clin Trials 21:251S–272S

Gudmundsson J, Sulem P, Gudbjartsson DF, Blondal T, Gylfason A, Agnarsson BA, Benediktsdottir KR, Magnusdottir DN, Orlygs-dottir G, Jakobsdottir M, Stacey SN, Sigurdsson A, Wahlfors T, Tammela T, Breyer JP, McReynolds KM, Bradley KM, Saez B, Godino J, Navarrete S, Fuertes F, Murillo L, Polo E, Aben KK, van Oort IM, Suarez BK, Helfand BT, Kan D, Zanon C, Frigge ML, Kristjansson K, Gulcher JR, Einarsson GV, Jonsson E, Catalona WJ, Mayordomo JI, Kiemeney LA, Smith JR, Schleutker J, Barkardottir RB, Kong A, Thorsteinsdottir U, Rafnar T, Stefansson K (2009) Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat Genet 41:1122–1126

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5:e1000529

Jemal A, Siegel R, Xu J, Ward E (2010) Cancer statistics, 2010. CA Cancer J Clin 60:277–300

Klein RJ, Hallden C, Cronin AM, Ploner A, Wiklund F, Bjartell AS, Stattin P, Xu J, Scardino PT, Offit K, Vickers AJ, Gronberg H, Lilja H (2010) Blood biomarker levels to aid discovery of cancer-related single-nucleotide polymorphisms: kallikreins and prostate cancer. Cancer Prev Res (Phila) 3:611–619

Kote-Jarai Z, Easton DF, Stanford JL, Ostrander EA, Schleutker J, Ingles SA, Schaid D, Thibodeau S, Dork T, Neal D, Donovan J, Hamdy F, Cox A, Maier C, Vogel W, Guy M, Muir K,

Lophatananon A, Kedda MA, Spurdle A, Steginga S, John EM, Giles G, Hopper J, Chappuis PO, Hutter P, Foulkes WD, Hamel N, Salinas CA, Koopmeiners JS, Karyadi DM, Johanneson B, Wahlfors T, Tammela TL, Stern MC, Corral R, McDonnell SK, Schurmann P, Meyer A, Kuefer R, Leongamornlert DA, Tymrakiewicz M, Liu JF, O'Mara T, Gardiner RA, Aitken J, Joshi AD, Severi G, English DR, Southey M, Edwards SM, Al Olama AA, Eeles RA (2008) Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium. Cancer Epidemiol Biomarkers Prev 17:2052–2061

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

Naess O, Sogaard AJ, Arnesen E, Beckstrom AC, Bjertness E, Engeland A, Hjort PF, Holmen J, Magnus P, Njolstad I, Tell GS, Vatten L, Vollset SE, Aamodt G (2008) Cohort profile: cohort of Norway (CONOR). Int J Epidemiol 37:481–485

Newcombe RG (1998) Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med 17:873–890

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

Parikh H, Deng Z, Yeager M, Boland J, Matthews C, Jia J, Collins I, White A, Burdett L, Hutchinson A, Qi L, Bacior JA, Lonsberry V, Rodesch MJ, Jeddeloh JA, Albert TJ, Halvensleben HA, Harkins TT, Ahn J, Berndt SI, Chatterjee N, Hoover R, Thomas G, Hunter DJ, Hayes RB, Chanock SJ, Amundadottir L (2010) A comprehensive resequence analysis of the KLK15-KLK3-KLK2 locus on chromosome 19q13.33. Hum Genet 127:91–99

Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, Fogel R, Gelmann EP, Gilbert F, Hasson MA, Hayes RB, Johnson CC, Mandel JS, Oberman A, O'Brien B, Oken MM, Rafla S, Reding D, Rutt W, Weissfeld JL, Yokochi L, Gohagan JK (2000) Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Control Clin Trials 21:273S–309S

Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30:3894–3900

Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, Kwiatkowski M, Lujan M, Lilja H, Zappa M, Denis LJ, Recker F, Berenguer A, Maattanen L, Bangma CH, Aus G, Villers A, Rebillard X, van der Kwast T, Blijenberg BG, Moss SM, de Koning HJ, Auvinen A (2009) Screening and prostate-cancer mortality in a randomized European study. N Engl J Med 360:1320–1328

Shao YH, Demissie K, Shih W, Mehta AR, Stein MN, Roberts CB, Dipaola RS, Lu-Yao GL (2009) Contemporary risk profile of prostate cancer in the United States. J Natl Cancer Inst 101:1280–1283

Sun J, Zheng SL, Wiklund F, Isaacs SD, Li G, Wiley KE, Kim ST, Zhu Y, Zhang Z, Hsu FC, Turner AR, Stattin P, Liu W, Kim JW, Duggan D, Carpten J, Isaacs W, Gronberg H, Xu J, Chang BL (2009) Sequence variants at 22q13 are associated with prostate cancer risk. Cancer Res 69:10–15

The ATBC Cancer Prevention Group (1994) The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. Ann Epidemiol 4:1–10

Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A, Crenshaw A, Cancel-Tassin G, Staats BJ, Wang Z, Gonzalez-Bosquet J, Fang J, Deng X, Berndt SI, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Gio-vannucci E, Willett WC, Cussenot O, Valeri A, Andriole GL, Crawford ED, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover

R, Hayes RB, Hunter DJ, Chanock SJ (2008) Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 40:310–315

Valeri A, Briollais L, Azzouzi R, Fournier G, Mangin P, Berthon P, Cussenot O, Demenais F (2003) Segregation analysis of prostate cancer in France: evidence for autosomal dominant inheritance and residual brother-brother dependence. Ann Hum Genet 67:125–137

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

Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678

Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N, Wang Z, Welch R, Staats BJ, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Gelmann EP, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover R, Hunter DJ, Chanock SJ, Thomas G (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 39:645–649

Yeager M, Chatterjee N, Ciampa J, Jacobs KB, Gonzalez-Bosquet J, Hayes RB, Kraft P, Wacholder S, Orr N, Berndt S, Yu K, Hutchinson A, Wang Z, Amundadottir L, Feigelson HS, Thun MJ, Diver WR, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Cancel-Tassin G, Cussenot O, Valeri A, Andriole GL, Crawford ED, Haiman CA, Henderson B, Kolonel L, Le Marchand L, Siddiq A, Riboli E, Key TJ, Kaaks R, Isaacs W, Isaacs S, Wiley KE, Gronberg H, Wiklund F, Stattin P, Xu J, Zheng SL, Sun J, Vatten LJ, Hveem K, Kumle M, Tucker M, Gerhard DS, Hoover RN, Fraumeni JF Jr, Hunter DJ, Thomas G, Chanock SJ (2009) Identification of a new prostate cancer susceptibility locus on chromosome 8q24. Nat Genet 41:1055–1057