

## Y chromosome haplogroups and prostate cancer in populations of European and Ashkenazi Jewish ancestry

Zhaoming Wang · Hemang Parikh · Jinping Jia · Timothy Myers · Meredith Yeager · Kevin B. Jacobs · Amy Hutchinson · Laurie Burdett · Arpita Ghosh · Michael J. Thun · Susan M. Gapstur · W. Ryan Diver · Jarmo Virtamo · Demetrius Albanes · Geraldine Cancel-Tassin · Antoine Valeri · Olivier Cussenot · Kenneth Offit · Ed Giovannucci · Jing Ma · Meir J. Stampfer · J. Michael Gaziano · David J. Hunter · Ana Dutra-Clarke · Tomas Kirchhoff · Michael Alavanja · Laura B. Freeman · Stella Koutros · Robert Hoover · Sonja I. Berndt · Richard B. Hayes · Ilir Agalliu · Robert D. Burk · Sholom Wacholder · Gilles Thomas · Laufey Amundadottir

Received: 22 November 2011 / Accepted: 4 January 2012 / Published online: 24 January 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** Genetic variation on the Y chromosome has not been convincingly implicated in prostate cancer risk. To comprehensively analyze the role of inherited Y chromosome variation in prostate cancer risk in individuals of European ancestry, we genotyped 34 binary Y chromosome markers in 3,995 prostate cancer cases and 3,815 control subjects drawn from four studies. In this set, we identified nominally significant association between a rare haplogroup,

E1b1b1c, and prostate cancer in stage I ( $P = 0.012$ , OR = 0.51; 95% confidence interval 0.30–0.87). Population substructure of E1b1b1c carriers suggested Ashkenazi Jewish ancestry, prompting a replication phase in individuals of both European and Ashkenazi Jewish ancestry. The association was not significant for prostate cancer overall in studies of either Ashkenazi Jewish (1,686 cases and 1,597 control subjects) or European (686 cases and 734 control subjects) ancestry ( $P_{\text{meta}} = 0.078$ ), but a meta-analysis of stage I and II studies revealed a nominally significant association with prostate cancer risk ( $P_{\text{meta}} = 0.010$ , OR = 0.77; 95% confidence interval 0.62–0.94). Comparing haplogroup frequencies between studies, we noted strong similarities between those conducted in the US and France, in which the majority

Z. Wang and H. Parikh are co-first authors.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00439-012-1139-5) contains supplementary material, which is available to authorized users.

Z. Wang · H. Parikh · J. Jia · T. Myers · M. Yeager · K. B. Jacobs · A. Hutchinson · L. Burdett · A. Ghosh · D. Albanes · M. Alavanja · L. B. Freeman · S. Koutros · R. Hoover · S. I. Berndt · S. Wacholder · L. Amundadottir  
Division of Cancer Epidemiology and Genetics,  
National Cancer Institute, National Institutes of Health,  
Bethesda, MD 20892, USA

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

H. Parikh · J. Jia · T. Myers · L. Amundadottir  
Laboratory of Translational Genomics, Division of Cancer  
Epidemiology and Genetics, National Cancer Institute,  
National Institutes of Health, Bethesda, MD 20877, USA

M. J. Thun · S. M. Gapstur · W. Ryan Diver  
Epidemiology Research Program, American Cancer Society,  
Atlanta, GA 30303, USA

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

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

K. Offit · A. Dutra-Clarke · T. Kirchhoff  
Clinical Genetics Service, Department of Medicine,  
Memorial Sloan-Kettering Cancer Center, Box 192,  
1275 York Avenue, New York, NY 10065, USA

E. Giovannucci · J. Ma · M. J. Stampfer · J. Michael Gaziano  
Channing Laboratory, Division of Preventive Medicine,  
Department of Medicine, Brigham and Women's Hospital,  
Harvard Medical School, Boston, MA 02115, USA

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

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

of men carried R1 haplogroups, resembling Northwestern European populations. On the other hand, Finns had a remarkably different haplogroup distribution with a preponderance of N1c and I1 haplogroups. In summary, our results suggest that inherited Y chromosome variation plays a limited role in prostate cancer etiology in European populations but warrant follow-up in additional large and well characterized studies of multiple ethnic backgrounds.

## Introduction

Family and twin studies have shown that prostate cancer has a clear heritable component which may be among the highest of all cancer types (Amundadottir et al. 2004; Lichtenstein et al. 2000). Over the last few years, genome wide association studies (GWAS) have successfully identified germline variants conferring risks of prostate cancer at over 45 loci (Amundadottir et al. 2006; Chung and Chanock 2011; Eeles et al. 2008, 2009; Gudmundsson et al. 2007a, b, 2008, 2009; Haiman et al. 2007; Kote-Jarai et al. 2011; Schumacher et al. 2011; Takata et al. 2010; Thomas et al. 2008; Yeager et al. 2007, 2009). These studies have not implicated variants on the Y chromosome in the risk of prostate cancer, possibly due to the fact that very few Y chromosome SNPs have been included on most genotyping chips used to date. Several groups have specifically investigated the role of Y chromosome haplogroups in prostate cancer risk. Many of these studies are inconclusive due to the small number of samples and/or markers used. One of

the larger studies was conducted within the multi-ethnic cohort (MEC) using samples from prostate cancer cases and control subjects drawn from four ethnic groups. Of the 41 haplogroups observed, one was significantly associated with prostate cancer in Japanese men (Paracchini et al. 2003) but this association was not replicated in a separate study from Korea (Kim et al. 2007). No association was seen between Y haplogroups and prostate cancer in a large Swedish study (Lindstrom et al. 2008).

The Y chromosome contains the largest non-recombining region in the human genome, spanning almost the entire length of the chromosome. This region is called the non-recombining Y (NRY) or the male-specific Y (MSY) (Rozen et al. 2003). In the absence of recombination, the NRY passes mostly unchanged from father to son and observed mutations reflect the evolutionary history of the Y chromosome. Binary markers can be used to classify Y chromosomes into haplogroups organized by a phylogenetic tree. A first generation phylogeny of the tree was published in 2002 by the Y Chromosome Consortium (2002) and further revised in 2008 (Karafet et al. 2008). The Y chromosome tree now consists of over 300 haplogroups organized into 20 major groups or clades (Karafet et al. 2008).

Multiple lines of evidence support a possible role for genes on the Y chromosome in prostate cancer etiology. Loss of the Y chromosome is one of the most frequent cytogenetic change seen in prostate tumors and may be an early event in tumorigenesis (Brothman et al. 1999; Jordan et al. 2001). In support of the previous assertion, chromosome transfer studies indicate that the human Y chromosome suppresses tumorigenicity of human prostate cell lines in vivo implying that it may harbor gene(s) with tumor suppressor function (Vijayakumar et al. 2005). Based on the essential role of the Y chromosome in secondary sexual differentiation and its potential role in disease pathogenesis, particularly related to the secondary sex organs, we explored this genomic region to investigate whether germline variation on this chromosome plays a role in prostate cancer risk.

## Results

We analyzed 7,810 men from the Cancer Genetic Markers of Susceptibility (CGEMS) scan in stage I of this study. Of the 34 chromosome Y markers genotyped, 26 were observed in our sample (8 markers were monomorphic). With such a sample size, we were able to accurately characterize and estimate the Y chromosome frequency distribution in populations of European ancestry for 28 haplogroups including three combined groups (R1b1b + R1b\*, R1a + R1\* and I2b + I2c) as the leaf nodes of the NRY tree (Fig. 1a). Stage I had 41, 76 and 95% power to detect an association with an odds ratio of 1.3 and a MAF

---

I. Agalliu · R. D. Burk  
Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, NY 10461, USA

R. D. Burk  
Department of Pediatrics, Albert Einstein College of Medicine, Bronx, New York, NY 10461, USA

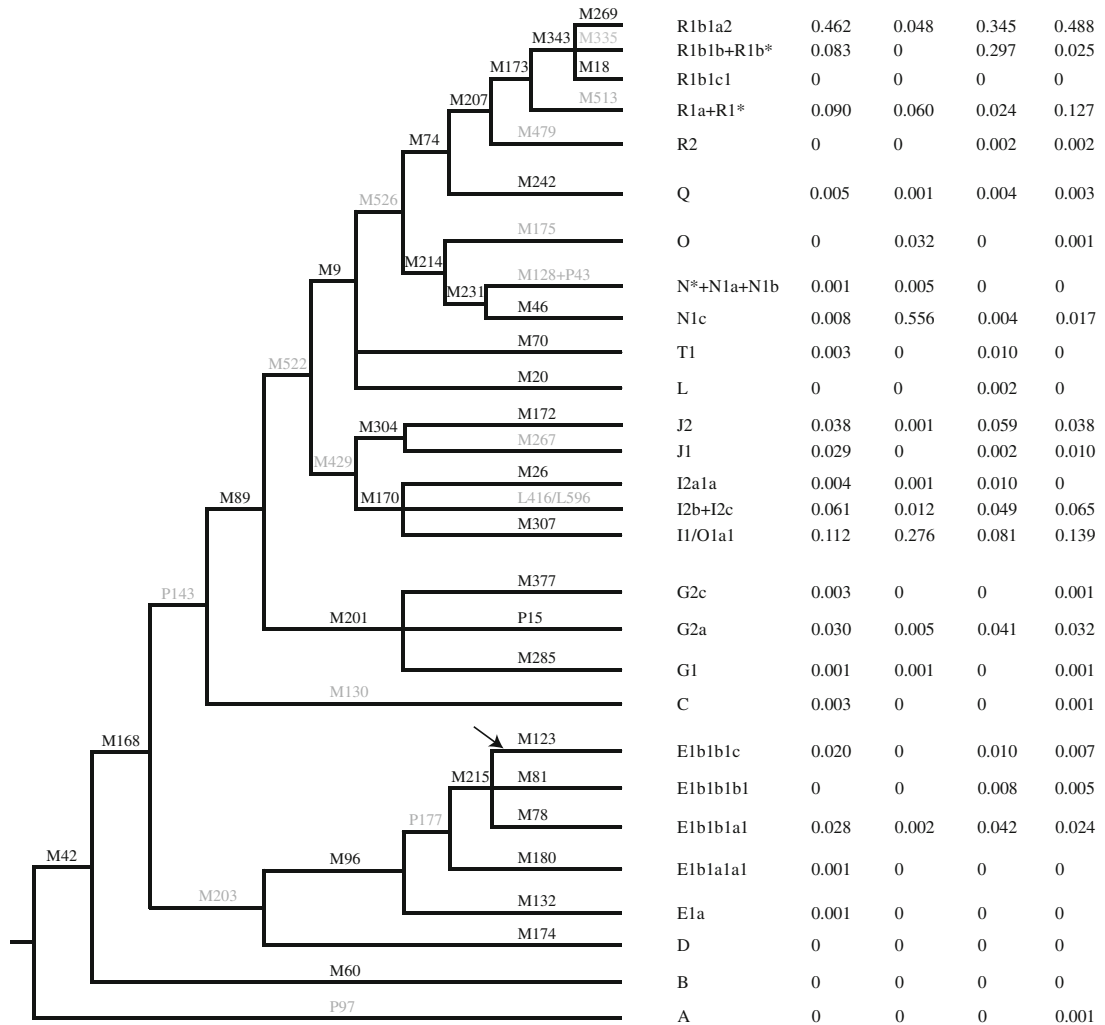
R. D. Burk  
Department of Microbiology & Immunology, Albert Einstein College of Medicine, Bronx, New York, NY 10461, USA

R. D. Burk  
Department of Obstetrics, Gynecology and Women's Health, Albert Einstein College of Medicine, Bronx, New York, NY 10461, USA

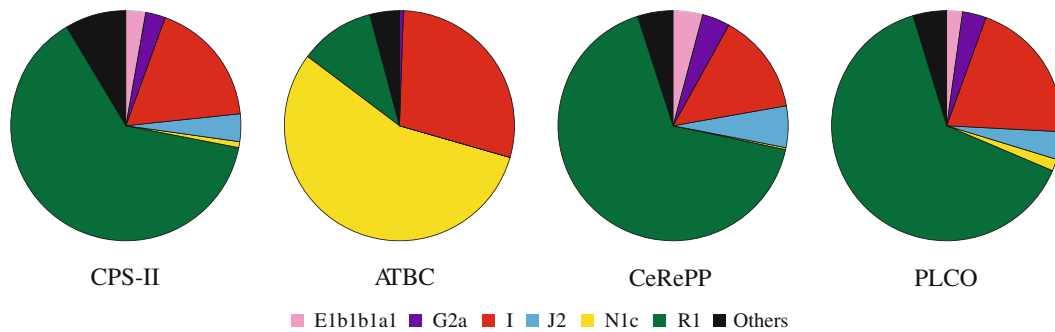
G. Thomas  
Synergie-Lyon-Cancer, Universite Lyon 1, Centre Leon Berard, 69373 Lyon Cedex 08, France

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

**A**



**B**



**Fig. 1** Chromosome Y haplogroup tree and frequency distribution in control subjects of European ancestry in Stage I. **a** Chromosome Y tree showing genotyped markers in *black* and those not genotyped in *light grey*. Haplogroup names are according to the International Society of Genetic Genealogy (ISOGG) 2011 update. The *arrow* points to the mutational event which gave rise to the E1b1b1c haplogroup. Stage I studies are the following: *CPS-II* American Cancer Society Cancer

Prevention Study II, *ATBC* Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study, *CeRePP* Centre de Recherche pour les Pathologies Prostatiques, and *PLCO* Prostate, Lung Colorectal and Ovarian Cancer Screening Trial. **b** The *circle plots* show frequencies for haplogroups with a derived frequency of 5% or higher in *different colors* for each Stage I cohort (remaining haplogroups are combined in one group shown in *black*)

of 0.02, 0.05 and 0.10, respectively (assuming prostate cancer prevalence of 1.5067% and alpha of 0.05) ([http://seer.cancer.gov/csr/1975\\_2007/](http://seer.cancer.gov/csr/1975_2007/)).

### Stage I association analysis

After genotyping quality control based on completion rates and concordance analysis, a total of 3,995 prostate cancer cases and 3,815 control subjects from four studies were used in the analysis (1994; Calle et al. 2002; Gohagan et al. 2000; Valeri et al. 2003). This included 1,531 men diagnosed with non-aggressive prostate cancer (Gleason score <7 and disease stage <III) and 2,142 men diagnosed with aggressive prostate cancer (Gleason score  $\geq$ 7 or stage  $\geq$ III).

Of the 26 haplogroup markers analyzed, one was significantly associated with overall prostate cancer at a nominal  $P$  value threshold of  $P \leq 0.05$  (Table 1). This was haplogroup E1b1b1c (locus M123, formerly named haplogroup E3b1c) ( $P = 0.012$ , allelic odds ratio (OR) 0.51; 95% confidence interval 0.30–0.87), a rare haplogroup with a 1.1% frequency in control subjects in our sample set. When the analysis was performed according to degree of differentiation and severity of prostate cancer, this haplogroup was significantly associated with non-aggressive prostate cancer ( $P = 0.017$ , allelic OR 0.33; 95% confidence interval 0.13–0.86) but not with aggressive prostate cancer ( $P = 0.091$ , allelic OR 0.59; 95% confidence interval 0.32–1.09). However, the difference between the two case groups was not significant ( $P = 0.48$ ).

Based on the Y chromosome haplogroup tree structure (see “Methods” section and Fig. 1a), we were able to test three additional haplogroups (J1, IJ, IJK) for which markers were not directly genotyped. These three haplogroups did not significantly associate with overall risk of prostate cancer (data not shown) (Table 1).

### Population substructure of E1b1b1c carriers

We assessed the population substructure for E1b1b1c haplogroup carriers using principal components analysis (PCA) from the initial CGEMS GWAS dataset (Thomas et al. 2008; Yeager et al. 2007, 2009). In addition we evaluated one common haplogroup, R1b1a2 (43.7% frequency in control subjects from stage I) as it had the second lowest  $P$  value in stage I ( $P = 0.054$ ). Carriers of the E1b1b1c haplogroup showed a distinct distribution of the first and second eigenvectors (EV1 and EV2) in this analysis that separates them from the majority of the European ancestry subjects, i.e. negative values for EV1 and positive values for EV2 (Fig. 2a). Conversely, the population substructure of R1b1a2 haplogroup carriers was similar to that of the majority of subjects in our study, implying Northwestern

European ancestry. We compared the population substructure pattern of E1b1b1c haplogroup carriers to the large number of individuals in the initial CGEMS prostate cancer scan (Thomas et al. 2008; Yeager et al. 2007, 2009) and a GWAS of breast cancer in families of Ashkenazi Jewish descent (Gold et al. 2008) and noted strong clustering with a group of individuals of self-reported Ashkenazi Jewish descent with similar values for EV1 and EV2 (Fig. 2b). The majority of E1b1b1c haplogroup carriers in our study (37 out of 62) were Ashkenazi Jewish alike by this comparison. Therefore, the frequency of E1b1b1c in inferred Ashkenazi Jewish individuals in our study was estimated to be approximately 15% (37/240). A similar number has been reported in men of Jewish ancestry (Hammer et al. 2009). These findings prompted a replication phase in sample sets of both European and Ashkenazi Jewish ancestry.

### Limited evidence for association to prostate cancer in Stage II analysis

We attempted replication of the E1b1b1c and R1b1a2 haplogroups in three prostate cancer cohort studies of European ancestry from the continental USA, the Physicians’ Health Study (PHS) (Ma et al. 2008), the Health Professionals Follow-up Study (HPFS) (Chen et al. 2005), the Agricultural Health Study (AHS) (Alavanja et al. 1996); and in two case–control studies of Ashkenazi Jewish ancestry collected in the USA, from the Albert Einstein College of Medicine (Einstein) (Agalliu et al. 2009) and the Memorial Sloan Kettering Cancer Center (MSKCC) (Gallagher et al. 2010). The three European ancestry studies included a total of 1,272 prostate cancer cases and 1,932 control subjects; the two Ashkenazi Jewish ancestry studies included a total of 1,686 prostate cancer cases and 1,597 control subjects. Neither haplogroup was significantly associated with overall prostate cancer risk at a nominal  $P$  value in any study (Table 2) nor was a meta-analysis of the combined studies significant ( $P_{\text{meta}} = 0.078$  for E1b1b1c and  $P_{\text{meta}} = 0.36$  for R1b1a2). For non-aggressive prostate cancer, the E1b1b1c haplogroup was significantly associated at a nominal  $P$  value in the Einstein study only ( $P = 0.024$ , allelic OR 0.66; 95% confidence interval 0.45–0.95). However, a meta-analysis of non-aggressive cases in all stage II studies was nominally significant ( $P_{\text{meta}} = 0.025$ , allelic OR 0.68; 95% confidence interval 0.49–0.95). No other significant associations were noted.

A meta-analysis of stage I and II results for the E1b1b1c haplogroup revealed a nominally significant association with risk of prostate cancer overall ( $P_{\text{meta}} = 0.010$ ; allelic OR = 0.77; 95% confidence interval 0.62–0.94) and with risk of non-aggressive prostate cancer ( $P_{\text{meta}} = 0.0077$ ; allelic OR = 0.67; 95% confidence interval 0.50–0.90) but not

**Table 1** Association of chromosome Y variants with risk of prostate cancer (Stage I)

Haplogroup <sup>a</sup>	Locus <sup>b</sup>	rs number <sup>c</sup>	Location <sup>d</sup>	Alleles <sup>e</sup>	All cases				Nonaggressive cases				Aggressive cases			
					Subjects <sup>f</sup>	MAF <sup>g</sup>	P <sup>h</sup>	OR	Subjects <sup>i,j</sup>	MAF <sup>g</sup>	P <sup>h</sup>	OR	Subjects <sup>k</sup>	MAF <sup>g</sup>	P <sup>h</sup>	OR
E	M96	rs9306841	20238386	C/G	373913903	0.0420.043	0.638	1.06 (0.84–1.32)	323311476	0.0390.029	0.188	0.78 (0.54–1.13)	373921111	0.0420.056	0.242	1.16 (0.90–1.50)
E1b1b	M215	rs2032654	13977218	T/C	380713983	0.0420.042	0.746	1.04 (0.83–1.30)	329811527	0.0390.029	0.196	0.79 (0.55–1.13)	380721334	0.0420.055	0.303	1.14 (0.89–1.48)
E1b1b1a1	M78	na <sup>*</sup>	20352691	G/A	375313948	0.0230.029	0.112	1.26 (0.95–1.67)	324711512	0.0210.022	0.725	1.08 (0.70–1.67)	375321116	0.0230.036	0.082	1.33 (0.96–1.83)
E1b1b1b1	M81	rs2032640	20351960	G/A	324313372	0.0020.001	0.079	0.27 (0.06–1.29)	273811223	0.0010.000	NA	NA	324311945	0.0020.001	0.192	0.36 (0.07–1.77)
E1b1b1c	M123	na <sup>*</sup>	20223974	C/T	381413992	0.0110.005	0.012	0.51 (0.30–0.87)	330611531	0.0110.000	0.017	0.33 (0.13–0.86)	381421140	0.0110.007	0.091	0.59 (0.32–1.09)
F	M89	rs2032652	203276701	A/G	381313994	0.0430.044	0.641	1.05 (0.85–1.31)	330411531	0.0400.031	0.267	0.82 (0.58–1.16)	381321141	0.0430.057	0.255	1.16 (0.90–1.49)
G	M201	rs2032636	13536923	G/T	380413987	0.0290.026	0.528	0.92 (0.70–1.20)	329611529	0.0270.026	0.995	1.00 (0.68–1.48)	380421136	0.0290.029	0.388	0.87 (0.62–1.20)
G1	M285	rs13447378	21151128	G/C	380713990	0.0010.001	0.973	1.02 (0.26–4.11)	330011528	0.0010.002	0.440	1.86 (0.38–9.12)	380721140	0.0010.000	NA	NA
G2a	P15	na <sup>*</sup>	21653414	G/A	379113948	0.0260.023	0.328	0.87 (0.65–1.16)	330211529	0.0240.022	0.714	0.92 (0.61–1.41)	379121098	0.0260.025	0.310	0.84 (0.59–1.18)
G2c	M377	na <sup>*</sup>	13536827	A/G	375013950	0.0020.002	0.410	1.54 (0.55–4.35)	324111508	0.0020.002	0.843	1.16 (0.28–4.85)	375021120	0.0020.002	0.474	1.56 (0.46–5.29)
I	M170	rs2032597	13357186	A/C	381113990	0.2030.221	0.059	1.11 (1.00–1.24)	330211529	0.2130.247	0.042	1.17 (1.01–1.35)	381121139	0.2030.204	0.058	1.14 (1.00–1.31)
I1/O1a1	M307	rs13447354	21160339	G/A	380413988	0.11500.154	0.648	1.03 (0.91–1.17)	329811525	0.1610.176	0.737	1.03 (0.87–1.22)	380421141	0.1500.136	0.249	1.10 (0.94–1.29)
I2a1a	M26	rs2032629	20325209	C/T	371913875	0.0030.002	0.298	0.62 (0.25–1.53)	321511453	0.0020.003	0.537	1.51 (0.41–5.58)	371921107	0.0030.002	0.117	0.41 (0.13–1.30)
J	M304	rs13447352	21159241	A/C	374213922	0.0480.048	0.635	1.05 (0.85–1.31)	323411492	0.0460.035	0.302	0.84 (0.60–1.17)	374221116	0.0480.059	0.379	1.12 (0.87–1.43)
J2	M172	rs2032604	13479028	T/G	378013958	0.0330.033	0.657	1.06 (0.82–1.37)	327211516	0.0280.022	0.697	0.92 (0.61–1.39)	378021124	0.0330.043	0.493	1.11 (0.83–1.47)
K	M9	rs3900	20189645	C/G	379913975	0.3220.340	0.055	1.10 (1.00–1.21)	329411525	0.3240.338	0.299	1.07 (0.94–1.22)	379921129	0.3220.348	0.021	1.15 (1.02–1.29)
T1	M70	rs2032672	20353269	T/G	370813904	0.0030.003	0.922	1.04 (0.44–2.47)	320211492	0.0020.001	0.315	2.92 (0.33–25.65)	370821101	0.0030.004	0.882	0.93 (0.37–2.32)
N	M231	rs9341278	13979118	C/T	376913946	0.11390.136	0.664	0.96 (0.80–1.15)	326211505	0.1600.189	0.565	0.94 (0.76–1.17)	376921122	0.1390.062	0.076	0.78 (0.60–1.03)
N1c	M46/Tat	rs34442126	13431977	T/C	380913992	0.11370.134	0.604	0.95 (0.80–1.14)	330211530	0.1570.185	0.514	0.93 (0.75–1.16)	380921140	0.1370.062	0.058	0.77 (0.59–1.01)
NO	M214	rs2032674	13981319	A/G	380513977	0.11390.135	0.676	0.96 (0.80–1.15)	329711523	0.1590.186	0.578	0.94 (0.76–1.17)	380521135	0.1390.062	0.058	0.77 (0.59–1.01)
P	M74	rs2032635	20349155	C/T	354713766	0.4840.494	0.272	1.06 (0.96–1.18)	304711418	0.4910.455	0.959	1.00 (0.86–1.16)	354720041	0.4840.427	0.282	1.07 (0.95–1.21)
Q	M242	rs8179021	13527976	C/T	371913856	0.0040.004	0.886	1.05 (0.51–2.19)	322911487	0.0040.003	0.842	0.89 (0.27–2.88)	37192058	0.0040.005	0.353	1.47 (0.65–3.35)
R	M207	rs2032658	14091377	T/C	374513908	0.4860.499	0.088	1.09 (0.99–1.21)	323811497	0.4900.453	0.663	0.97 (0.84–1.12)	37452098	0.4860.436	0.094	1.11 (0.98–1.25)
R1	M173	rs2032624	13535818	A/C	369113876	0.4770.487	0.144	1.08 (0.97–1.19)	318911481	0.5000.467	0.783	0.98 (0.85–1.13)	36912076	0.4770.423	0.160	1.09 (0.97–1.23)
R1b	M343	rs9786184	2947824	C/A	380513976	0.4410.428	0.114	0.92 (0.84–1.02)	329711525	0.4100.371	0.673	0.97 (0.84–1.12)	380521131	0.4410.495	0.193	0.92 (0.82–1.04)
R1b1a2	M269	rs9786153	21148755	T/C	348913634	0.4370.421	0.054	0.90 (0.81–1.00)	302911386	0.4050.365	0.719	0.97 (0.84–1.13)	348911959	0.4370.491	0.093	0.90 (0.79–1.02)

Results from the unconditional logistic regression of the genotypes generated in a total of 3,995 individuals with prostate cancer and 3815 control subjects are shown. The analysis was adjusted for age in 10-year categories, study and one principal component of population stratification

na<sup>\*</sup> no rs number, OR odds ratio, CI 95% confidence interval

<sup>a</sup> Haplogroup being tested based on the Y Chromosome Consortium and ISOGG 2011 nomenclature

<sup>b</sup> ChrY locus/marker name

<sup>c</sup> NCBI dbSNP identifier

<sup>d</sup> NCBI human genome build 36 location

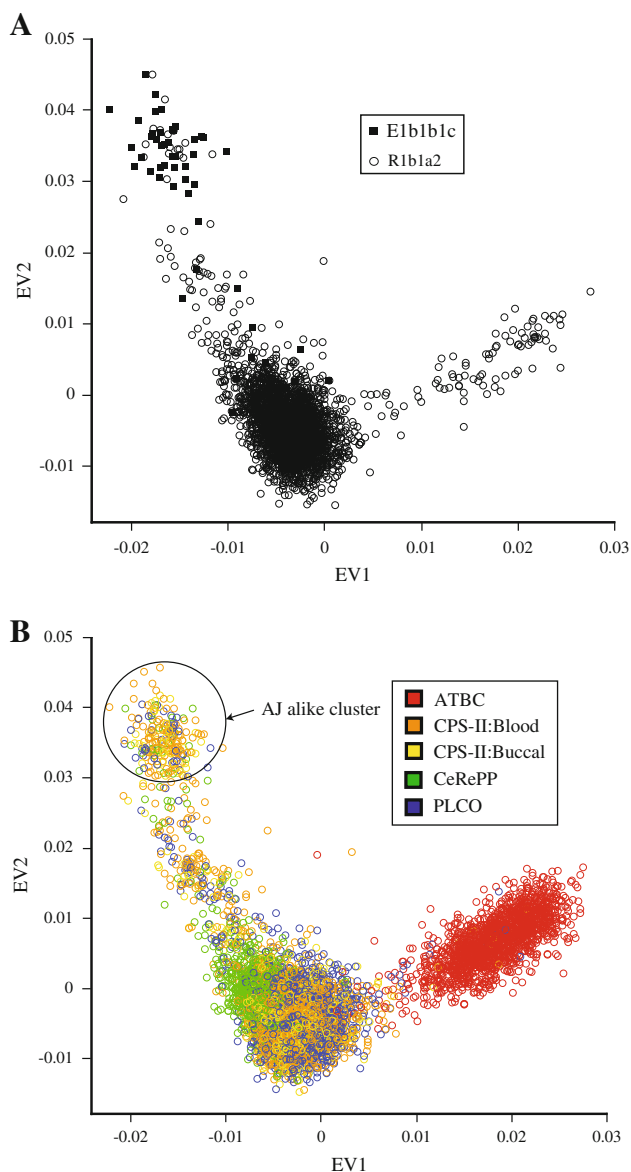
<sup>e</sup> Ancestral allele, derived allele

<sup>f</sup> Controls, cases

<sup>g</sup> Minor allele frequency in control and case participants

<sup>h</sup> Score test (1df)

<sup>i</sup> For the nonaggressive model, subjects in two categories (“FPCC” and “age <50”) could not be used because we there are no case subjects in either category. Therefore, the total number of controls is smaller than in the overall and aggressive models



**Fig. 2** Population substructure analysis by principal component analysis and comparison to CGEMS prostate cancer GWAS. **a** shows the distribution of the first two principal components, EV1 and EV2, for carriers of E1b1b1c (filled squares) and R1b1a2 (open circles) haplogroups in Stage I. Circles and squares denote eigenvalues for PCA analysis for each individual. The distribution of EV1 and EV2 for all Stage I subjects is shown in **b**. Studies are designed by different colors. CPS-II Blood blood derived DNA samples were used for genotyping, CPS-II Buccal buccal derived DNA samples were used for genotyping. DNA samples from ATBC, CeRePP and PLCO were all derived from blood. Individuals of inferred Ashkenazi Jewish ancestry are circled. PCA results were performed by EIGENSTRAT in CGEMS prostate cancer GWAS (Thomas et al. 2008; Yeager et al. 2007, 2009)

with risk of aggressive prostate cancer ( $P_{\text{meta}} = 0.28$ ). In Stage II, we had 60% power to detect a variant with 13% MAF and an OR of 0.8 in the Ashkenazi Jewish sample set, but only 25% power to detect a variant with 3% MAF and an OR of 0.7 in the European American sample set.

## Haplogroup frequency and population distribution

Y chromosome haplogroup frequency distribution in controls from each of the four study populations from phase I was summarized and compared in Fig. 1b. Two of the studies, namely CPS-II and PLCO, include subjects from continental USA. Their haplogroup frequencies are very similar with an average difference of 0.8% and a maximum difference of 5.8% for the combined category of haplogroups R1b1b + R1b\*. The CeRePP study, conducted in France, is relatively similar to the US studies with an average haplogroup frequency difference of 2.2%, and a maximum difference of 24.3% for the combined group of R1b1b + R1b\*. The greatest difference in frequency was seen for ATBC, a Finnish study, with an average haplogroup frequency difference of 5.2% and a maximum difference of 54.4%. This stems from a very high frequency of haplogroup N1c in this study (55.6%), while it is infrequent in the other three studies from the US and France (0.8% in CPS-II, 1.7% in PLCO and 0.4% in CeRePP). Second, R1b, the most frequent haplogroup overall, is seen in over 50% of subjects in PLCO, CPS-II and CeRePP but only in 4.8% of Finnish subjects. The third largest difference was noted for haplogroup I1 which was more common in Finns at 27.6%, as compared to 13.9% in PLCO, 11.2% in CPS-II and only 8.1% in CeRePP.

Haplogroup E1b was observed at low frequencies in all studies and its sub lineage E1b1b1c was seen in approximately 1–2% of subjects from the two US studies (PLCO and CPS-II) and the French study (CeRePP), whereas it was absent from the Finnish study (ATBC). Other haplogroups were absent or rare in the four studies.

## Discussion

In this study, we explored the role of germline Y chromosome variation in prostate cancer risk. Previous studies have not analyzed such a large sample size with as many markers in individuals of European ancestry. Because of the threshold for MAF chosen for this study ( $\geq 1\%$ ), we had limited capacity to detect risk variants with low to medium frequency and effect sizes. Prostate cancer GWAS to date have used arrays with limited coverage on the Y chromosome. As an example, in CGEMS, of the approximately 500,000 SNPs genotyped in stage I, only ten Y chromosome markers passed quality control assessment and were included in the primary analysis; this limited set of variants on the Y chromosome included only four that mark chromosome Y haplogroups (Thomas et al. 2008; Yeager et al. 2007, 2009). Other published prostate cancer GWAS studies have reported on a similar fraction of Y variants (Amundadottir et al. 2006; Chung and Chanock 2011; Eeles et al.

**Table 2** Replication analysis of notable chromosome Y signals in prostate cancer studies of Ashkenazi and European descent (Stage II)

Study	All cases				Nonaggressive cases				Aggressive cases			
	Subjects <sup>c</sup>	MAF <sup>d</sup>	P <sup>e</sup>	OR	Subjects <sup>c</sup>	MAF <sup>d</sup>	P <sup>e</sup>	OR	Subjects <sup>c</sup>	MAF <sup>d</sup>	P <sup>e</sup>	OR
<b>E1b1b1c<sup>a</sup> (M123<sup>b</sup>)</b>												
Einstein	1214928	0.1380.117	0.146	0.82 (0.63–1.07)	1214413	0.1380.097	0.024	0.66 (0.45–0.95)	1214456	0.1380.134	0.693	0.94 (0.68–1.29)
MSKCC	375750	0.1280.123	0.608	0.87 (0.52–1.46)	375212	0.1280.090	0.956	0.96 (0.47–1.94)	375362	0.1280.146	0.954	0.98 (0.53–1.83)
PHS	489471	0.0290.021	0.455	0.73 (0.32–1.67)	489194	0.0290.015	0.379	0.57 (0.16–2.02)	489167	0.0290.024	0.780	0.85 (0.28–2.63)
HPFS	244215	0.0330.023	0.540	0.70 (0.23–2.18)	244138	0.0330.036	0.848	1.12 (0.36–3.49)	244144	0.0330.000	0.210	0.00 (0.00–1.77)
<b>R1b1a<sup>2</sup> (M269<sup>b</sup>)</b>												
Einstein	1210912	0.0950.115	0.121	1.25 (0.94–1.67)	1210408	0.0950.120	0.147	1.30 (0.91–1.87)	1210446	0.0950.108	0.390	1.17 (0.82–1.69)
MSKCC	373739	0.1100.118	0.936	1.02 (0.61–1.71)	373206	0.1100.121	0.981	0.99 (0.51–1.93)	373358	0.1100.126	0.841	1.06 (0.59–1.92)
PHS	482463	0.5000.531	0.321	1.14 (0.88–1.47)	482190	0.5000.553	0.238	1.23 (0.87–1.72)	482164	0.5000.537	0.430	1.15 (0.81–1.65)
AHS	1159571	0.5890.578	0.662	0.96 (0.78–1.17)	1159326	0.5890.558	0.320	0.88 (0.66–1.09)	1159180	0.5890.612	0.703	1.09 (0.69–1.74)

Results from the unconditional logistic regression of the genotypes generated in Stage II are shown. The analysis was adjusted for age in 10-year categories

OR odds ratio, CI 95% confidence interval

<sup>a</sup> Haplogroup being tested based on the Y Chromosome Consortium and ISOGG 2011 nomenclature

<sup>b</sup> ChrY locus/marker name

<sup>c</sup> Controls, cases

<sup>d</sup> Minor allele frequency in control and case participants

<sup>e</sup> Score test (1df) for all scenarios except for aggressive cases from HPFS which is based on a Fisher exact test (1df)

2008, 2009; Gudmundsson et al. 2007a, b, 2008, 2009; Haiman et al. 2007; Kote-Jarai et al. 2011; Schumacher et al. 2011; Takata et al. 2010; Thomas et al. 2008; Yeager et al. 2007, 2009).

One haplogroup of interest was noted in phase I of our study; the E1b1b1c haplogroup was nominally significant in the overall prostate cancer and non-aggressive prostate cancer groups. The marker that denotes this haplogroup is located in the last intron of the taxilin gamma 2 pseudogene (*TXLNG2P*) on chromosome Yq11.222. This haplogroup was analyzed in a second phase using replication studies of European and Ashkenazi Jewish ancestry along with a more common haplogroup, R1b1a2. Neither haplogroup was significantly associated with overall prostate cancer risk in stage II. A meta-analysis of stage I and stage II results yielded a *P* value of 0.010 for the E1b1b1c haplogroup. Although nominally significant, this *P* value is unremarkable in comparison with the rigorous threshold required for significance in GWAS studies (Wellcome Trust Case Control Consortium 2007), suggesting that further studies are required to establish this association. Although our analysis does not provide strong evidence for a relationship between variation in the Y chromosome and prostate cancer, it can be argued that the appropriate statistical threshold to be applied to a study of approximately 30 markers should not be as stringent as a GWAS threshold. However, the probability of false-positive findings is high, even in a study of our size and power (Wacholder et al. 2004) especially in the first stage where E1b1b1c haplogroup frequency was very low. In addition, we cannot exclude a chance finding due to population stratification.

Our study represents the largest analysis to date of a possible association between Y chromosome variants and prostate cancer. The role of germline variation on the Y chromosome had been assessed previously, but with limited sample and/or marker sets. One of the most complete studies published was conducted within the MEC (Paracchini et al. 2003). Four ethnic groups with a total of 930 cases and 1,208 control subjects were included. One of the 41 haplogroups observed in the study was significantly associated with prostate cancer risk in Japanese men with a *P* value of 0.02 (Paracchini et al. 2003). Despite the large overall sample set in this study, each ethnic group only consisted of approximately 100–150 case–control pairs, limiting power considerably. No haplogroups were significantly associated with prostate cancer risk in a small Korean study that assessed 14 markers in approximately 106 cases and 110 control subjects, including the haplogroup reported in the MEC study (Kim et al. 2007). Lack of an association between Y haplogroups and prostate cancer was also reported in a Swedish study assessing five ChrY markers in 1,452 cases and 779 control subjects of N-European background (Lindstrom et al. 2008). Our results appear to con-

firm an overall lack of importance for germline variants on the Y chromosome and prostate cancer risk.

Frequencies of Y chromosome haplogroups vary considerably between different geographical regions and ethnic groups, and have turned out to be informative in studies of human evolution and migration. In Europe, marked differences in haplogroup frequencies are observed between countries in Northeastern, Northwestern, Southwestern, Southeast and Central Europe (Wiik 2008). In addition, the Ashkenazi Jewish community has a specific pattern that is reminiscent of non-Ashkenazi Jewish communities in the Near East (Behar et al. 2004). We observed a different distribution of major haplogroups in subjects of Northwestern European ancestry (represented by the majority of subjects from the US in PLCO and CPS-II), Northeastern European ancestry (represented by Finnish subjects in ATBC) and Western/Central European ancestry (represented by French subjects in CeRePP). Haplogroups in the US and French studies can mostly be accounted for by the R and I haplogroup clans with a combined frequency of 81–85%; R1b1a2 and I1 were the most common sub branches. The R1 haplogroup clan originated in Eurasia and migrated into Europe where it divided into two subgroups, R1a (common in Eastern Europe) and R1b (common in Western Europe) (Wiik 2008). R1b1a2 shows an East to West gradient in Europe and is very common in Spain, France, UK and Ireland (Balaresque et al. 2010). Haplogroup clan I1 appears to have originated in the Balkans and migrated north throughout Europe (Wiik 2008). It is most common in Scandinavia and Northwestern Europe and gradually decreases in Central and Southern Europe (Wiik 2008). Finnish subjects were strikingly different from the other three studies with a preponderance of N1c (56%) and I1 (28%) haplogroups and few R1b carriers. The N1c haplogroup is thought to have an Eastern or Central Asian origin and probably reached Eastern Europe via expansion through Siberia (Rootsi et al. 2007). The frequency of this haplogroup in Finland has been reported to be 58% (Wiik 2008).

Genotypes in stage II confirmed the scarcity of E1b1b1c in subjects of European ancestry (1–2%) and revealed a higher frequency in the two Ashkenazi Jewish studies (13–14%), in line with previous reports (Hammer et al. 2009) indicating similar Y chromosome haplogroup frequencies in men of Ashkenazi Jewish descent living in the US and those from Jewish communities in the Middle East. E1b1b1c may have arisen in Northeastern Africa, and migrated through the Levantine corridor to the Near East and Europe (Semino et al. 2004). In a similar manner, haplogroup R1b1a2 was seen in 50–59% of the subjects in different European American studies but only 10–11% in the two Ashkenazi Jewish studies.

In conclusion, we found limited evidence for an association between Y chromosome haplogroups and risk of



prostate cancer in populations of European and Ashkenazi Jewish ancestry using a large sample set close to 4,000 case–control pairs in Stage I and 2,300 case–control pairs in Stage II. Weak but consistent evidence for a protective effect for haplogroup E1b1b1c was seen in all studies with a nominally significant meta-analysis, thus, calling for additional replication efforts for this haplogroup in populations of Ashkenazi Jewish and European ancestry. The different frequencies seen in subjects from the four stage I studies may limit power to detect true associations for some branches of the Y haplogroup tree. Furthermore, correcting for population substructure based on autosomal SNPs may not be optimal, as Y chromosome inheritance only reflects male lineages that may have somewhat different characteristics throughout human history and population migration as compared to that of females. Although we cannot exclude a role for all chromosome Y haplogroups in prostate cancer etiology, our study has good power to detect common alleles with relatively large effects. Smaller or population specific effects for the haplogroups tested here, or for other haplogroups, could exist and should be studied by testing comprehensive sets of chromosome Y haplogroup markers in additional studies.

## Materials and methods

### Study population

Stage I of this study included 3,995 men diagnosed with adenocarcinoma of the prostate and 3,815 control subjects from three case–control studies nested within cohorts and one hospital based case–control study, previously analyzed in stages I and II of the Cancer Genetics Markers of Susceptibility study (CGEMS). Study details have been published previously (Thomas et al. 2008; Yeager et al. 2007, 2009). The cohort studies were: the Prostate, Lung Colorectal and Ovarian Cancer Screening Trial (PLCO, subjects from continental USA) (Gohagan et al. 2000); the American Cancer Society Cancer Prevention Study II (CPS-II, from continental USA) (Calle et al. 2002) and the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC, from Finland) (1994). The case–control study was the French Prostate Case–Control Study (CeRePP, Centre de Recherche pour les Pathologies Prostatiques, from France) (Valeri et al. 2003). The number of subjects included from each study is shown in Supplemental Table 1a. We incorporated prostate cancer stage and grade at diagnosis to distinguish between non-aggressive (Gleason score <7 and disease stage <III,  $n = 1,531$ ) and aggressive prostate cancer (Gleason score  $\geq 7$  and/or disease stage  $\geq$ III,  $n = 2,141$ ) as defined in CGEMS (Thomas et al. 2008).

Stage II included 471 prostate cancer cases and 490 control subjects of European descent from the Physicians' Health Study (PHS, from continental USA) (Ma et al. 2008); 215 prostate cancer cases and 244 control subjects of European descent from the Health Professionals Follow-up Study (HPFS, from continental USA) (Chen et al. 2005); 586 prostate cancer cases and 1198 control subjects of European descent from the Agricultural Health Study (AHS, from NC and IA) (Alavanja et al. 1996); 933 prostate cancer cases and 1,221 control subjects of Ashkenazic descent collected by the Albert Einstein College of Medicine (Einstein, majority recruited from NY, FL, CA or NJ, USA) (Agalliu et al. 2009); and 753 prostate cancer cases and 376 male control subjects of Ashkenazic descent collected at the Memorial Sloan Kettering Cancer Center (MSKCC, from Northeast USA) (Gallagher et al. 2010). Prostate cancer stage and grade at diagnosis were included to distinguish between non-aggressive (Gleason score <7 AND disease stage <III,  $n = 194$  for PHS;  $n = 172$  for HPFS,  $n = 338$  for AHS;  $n = 416$  for Einstein and  $n = 212$  for MSKCC) and aggressive prostate cancer (Gleason score  $\geq 7$  OR disease stage  $\geq$ III,  $n = 167$  for PHS;  $n = 80$  for HPFS,  $n = 85$  for AHS;  $n = 457$  for Einstein and  $n = 364$  for MSKCC).

The study protocols for each study were approved by the Institutional Review Board of each corresponding institution, and written informed consent was obtained from all study participants.

### Marker selection and genotyping

Markers were selected to detect chromosome Y haplogroups with minor allele frequencies (MAF)  $\geq 1\%$  in populations of European descent, using data from the International Society of Genetic Genealogy (ISOGG) ([http://www.isogg.org/tree/ISOGG\\_YDNA\\_SNP\\_Index.html](http://www.isogg.org/tree/ISOGG_YDNA_SNP_Index.html)) 2011 update, the Y Chromosome Consortium (<http://ycc.biosci.arizona.edu/>) (Karafet et al. 2008; Underhill et al. 2001) and from HapMap (<http://hapmap.ncbi.nlm.nih.gov/>). TaqMan custom genotyping assays (ABI, Foster City, CA, USA) were designed and optimized for 34 biallelic chromosome Y markers (32 SNPs and 2 insertion/deletion polymorphisms) based on the Y Chromosome Consortium, ISOGG and HapMap databases.

For stage I, DNA was extracted from blood samples for all studies except a subset of CPS-II where buccal cells were used for a subset of subjects ( $n = 939$ ). After pre-genotyping quality control at the Core Genotyping Facility (CGF) of the National Cancer Institute of the National Institutes of Health, Gaithersburg, MD, USA (<http://cgf.nci.nih.gov/operations/pregenotyping-qaqc.html>), 34 SNPs were genotyped on 9,501 samples in stage I using TaqMan genotyping assays (ABI, Foster City, CA, USA). The average concordance for 146 duplicate samples was

99.75%. Samples were excluded based on a completion rate <80% or  $\geq 2$  heterozygous genotypes. After genotype quality control (Supplemental Table 1a), 8,157 samples remained (including 8,011 subjects of which 7,810 men (3,995 cases and 3,815 controls) had all covariates used in the association analysis). Eight markers were monomorphic in our data set (Supplemental Table 2), thus leaving 26 polymorphic markers for analysis.

For stage II, DNA was isolated from blood for all studies except for the samples from Einstein where DNA was obtained from mouthwash. A subset of DNA samples from the AHS study ( $n = 1,858$ ) were whole genome amplified prior to genotyping using the GenomiPhi™ version 2 kit (GE Healthcare) at the Core Genotyping Facility (CGF) of the National Cancer Institute of the National Institutes of Health, Gaithersburg, MD, USA (<http://cgf.nci.nih.gov/operations/wga.html>). Two SNPs (M123 for haplogroup E1b1b1c and M269 for haplogroup R1b1a2) were genotyped in stage II on 6,695 samples using TaqMan genotyping assays (ABI, Foster City, CA, USA) at CGF and on 1,213 samples at MSKCC (for MSKCC samples). This included 6,487 subjects (2,958 case and 3,529 control subjects). The E1b1b1c haplogroup was genotyped in samples from Einstein, MSKCC, PHS and HPFS; the R1b1a2 haplogroup was genotyped in samples from Einstein, MSKCC, PHS and AHS. Genotype quality control was performed in a similar manner as for stage I studies (detailed in Supplemental Table 1b). Concordance rates for duplicate samples ( $n = 91$ ) were 99.9%.

#### Statistical analysis

The association between haplogroups of the Y chromosome and prostate cancer risk was examined using a logistic regression model adjusted for age, study center and first principle component previously constructed based on CGEMS genotype data (Thomas et al. 2008; Yeager et al. 2007, 2009) as it was significant in the base model, to correct for population stratification if available. All subjects were of self-described European ancestry.

The variance weighted fixed-effect meta-analysis was performed to assess the overall statistical significance of stage II studies as well as combination of stage I together with II studies. Results were not corrected for multiple testing because of the strong dependence among the markers on this chromosome. Because all the Y haplogroups map to a haplogroup evolutionary tree, each branch in the tree can be cut and thus creating a bipartition of all individuals. Individuals under each cut will have inherited the mutation incurred on that branch. The case/control imbalance could therefore be tested by comparing two groupings of subjects. This is exactly the same as testing the genotypes of individual markers. To make full use of the data, genotypes from

untyped branches were imputed if possible, based on its ancestor and sibling nodes in the tree. As an example, we could infer genotypes for J1 because both J and J2 were genotyped (Fig. 1a). We searched across all the branches in the tree and tested three additional untyped haplogroups, namely J1 (M267), IJ (M429) and IJK (M522). A branch was not analyzed if there was a directly genotyped derived/ancestor branch with a difference in frequency of <0.001. For example, M96 and M203 are almost the same because frequencies of haplogroup D in all study populations were close to 0. Thus, testing of the imputed marker M203 became redundant when the directly genotyped marker M96 was already analyzed.

#### Validation by sequencing

Genotypes for the two markers selected for replication (M123 for haplogroup E1b1b1c and M269 for haplogroup R1b1a2) were confirmed by sequencing in 94 subjects from the current study. They were chosen from the PLCO, CeRePP and CPS-II studies in stage I such that approximately one-third (E1b1b1c) or half (R1b1a2) carried each haplogroup. Primers were designed with the program Primer3 (<http://frodo.wi.mit.edu/primer3/>) and used for PCR amplification of the genomic regions containing the 2 markers (Supplemental Table 3). PCR amplifications were performed with 10 ng genomic DNA using the AmpliTaq Gold 360 master mix (ABI). The samples were cleaned using AMPure beads (Agencourt) on a Biomek FX (Beckman Coulter). After resuspending the beads in 50  $\mu$ l of water, PCR products were sequenced using primers for the two markers and an ABI PRISM Big Dye Terminator version 3.1 cycle sequencing kit (Applied BioSystems, Foster City, CA, USA). Sequencing was performed on an ABI 3730 capillary sequencer (Applied Biosystems). A 100% concordance was noted for E1b1b1c and 98.9% concordance for R1b1a2.

#### URLs

<http://cgf.nci.nih.gov/operations/pregenotyping-qaqc.html>  
<http://code.google.com/p/glu-genetics/>  
<http://ycc.biosci.arizona.edu/>  
[http://www.isogg.org/tree/ISOGG\\_YDNA\\_SNP\\_Index.html](http://www.isogg.org/tree/ISOGG_YDNA_SNP_Index.html)  
<http://hapmap.ncbi.nlm.nih.gov/>

**Acknowledgments** This study was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health (NIH). The authors acknowledge and thank both staff and participants in all studies for donating their time and making this study possible. The authors acknowledge support and constructive comments from Dr. Stephen J.

Chanock, LTG, DCEG, National Cancer Institute. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

**Conflict of interest** All authors report no financial interests or potential conflicts of interests.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Agalliu I, Gern R, Leanza S, Burk RD (2009) Associations of high-grade prostate cancer with BRCA1 and BRCA2 founder mutations. *Clin Cancer Res* 15:1112–1120. doi:[10.1158/1078-0432.CCR-08-1822](https://doi.org/10.1158/1078-0432.CCR-08-1822)
- Alavanja MC, Sandler DP, McMaster SB, Zahm SH, McDonnell CJ, Lynch CF, Pennybacker M, Rothman N, Dosemeci M, Bond AE, Blair A (1996) The Agricultural Health Study. *Environ Health Perspect* 104:362–369
- Amundadottir LT, Thorvaldsson S, Gudbjartsson DF, Sulem P, Kristjansson K, Arnason S, Gulcher JR, Bjornsson J, Kong A, Thorsteinsdottir U, Stefansson K (2004) Cancer as a complex phenotype: pattern of cancer distribution within and beyond the nuclear family. *PLoS Med* 1:e65. doi:[10.1371/journal.pmed.0010065](https://doi.org/10.1371/journal.pmed.0010065)
- Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, Sigurdsson A, Benediktsdottir KR, Cazier JB, Sainz J, Jakobsdottir M, Kostic J, Magnusdottir DN, Ghosh S, Agnarsson K, Birgisdottir B, Le Roux L, Olafsdottir A, Blondal T, Andresdottir M, Gretarsdottir OS, Bergthorsson JT, Gudbjartsson D, Gylfason A, Thorgeirsson G, Manolescu A, Kristjansson K, Geirsson G, Isaksson H, Douglas J, Johannsson JE, Balter K, Wiklund F, Montie JE, Yu X, Suarez BK, Ober C, Cooney KA, Gronberg H, Catalona WJ, Einarsson GV, Barkardottir RB, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K (2006) A common variant associated with prostate cancer in European and African populations. *Nat Genet* 38:652–658. doi:[10.1038/ng1808](https://doi.org/10.1038/ng1808)
- Balaresque P, Bowden GR, Adams SM, Leung HY, King TE, Rosser ZH, Goodwin J, Moisan JP, Richard C, Millward A, Demaine AG, Barbujani G, Previdere C, Wilson IJ, Tyler-Smith C, Jobling MA (2010) A predominantly neolithic origin for European paternal lineages. *PLoS Biol* 8:e1000285. doi:[10.1371/journal.pbio.1000285](https://doi.org/10.1371/journal.pbio.1000285)
- Behar DM, Garrigan D, Kaplan ME, Mobasher Z, Rosengarten D, Karafet TM, Quintana-Murci L, Ostrer H, Skorecki K, Hammer MF (2004) Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. *Hum Genet* 114:354–365. doi:[10.1007/s00439-003-1073-7](https://doi.org/10.1007/s00439-003-1073-7)
- Brothman AR, Maxwell TM, Cui J, Deubler DA, Zhu XL (1999) Chromosomal clues to the development of prostate tumors. *Prostate* 38:303–312. doi:[10.1002/\(SICI\)1097-0045\(19990301\)38:4<303:AID-PROS6>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-0045(19990301)38:4<303:AID-PROS6>3.0.CO;2-E)
- 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. doi:[10.1002/ncr.101970](https://doi.org/10.1002/ncr.101970)
- Chen YC, Giovannucci E, Lazarus R, Kraft P, Ketkar S, Hunter DJ (2005) Sequence variants of toll-like receptor 4 and susceptibility to prostate cancer. *Cancer Res* 65:11771–11778. doi:[10.1158/0008-5472.CAN-05-2078](https://doi.org/10.1158/0008-5472.CAN-05-2078)
- Chung CC, Chanock SJ (2011) Current status of genome-wide association studies in cancer. *Hum Genet* 130:59–78. doi:[10.1007/s00439-011-1030-9](https://doi.org/10.1007/s00439-011-1030-9)
- 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, Arden-Jones AT, Hall AL, O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon 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. doi:[10.1038/ng.90](https://doi.org/10.1038/ng.90)
- Eeles RA, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, Muir K, Hopper JL, Henderson BE, Haiman CA, Schleutker J, Hamdy FC, Neal DE, Donovan JL, Stanford JL, Ostrander EA, Ingles SA, John EM, Thibodeau SN, Schaid D, Park JY, Spurdle A, Clements J, Dickinson JL, Maier C, Vogel W, Dork T, Rebbeck TR, Cooney KA, Cannon-Albright L, Chappuis PO, Hutter P, Zeegers M, Kaneva R, Zhang HW, Lu YJ, Foulkes WD, English DR, Leongamornlert DA, Tymrakiewicz M, Morrison J, Arden-Jones AT, Hall AL, O'Brien LT, Wilkinson RA, Saunders EJ, Page EC, Sawyer EJ, Edwards SM, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As N, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Cooper CS, Southey MC, Lophatananon A, Liu JF, Kolonel LN, Le Marchand L, Wahlfors T, Tammela TL, Auvinen A, Lewis SJ, Cox A, FitzGerald LM, Koopmeiners JS, Karyadi DM, Kwon EM, Stern MC, Corral R, Joshi AD, Shahabi A, McDonnell SK, Sellers TA, Pow-Sang J, Chambers S, Aitken J, Gardiner RA, Batra J, Kedda MA, Lose F, Polanowski A, Patterson B, Serth J, Meyer A, Luedeke M, Stefllova K, Ray AM, Lange EM, Farnham J, Khan H, Slavov C, Mitkova A, Cao G et al (2009) Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet* 41:1116–1121. doi:[10.1038/ng.450](https://doi.org/10.1038/ng.450)
- Gallagher DJ, Vijai J, Cronin AM, Bhatia J, Vickers AJ, Gaudet MM, Fine S, Reuter V, Scher HI, Hallden C, Dutra-Clarke A, Klein RJ, Scardino PT, Eastham JA, Lilja H, Kirchoff T, Offit K (2010) Susceptibility loci associated with prostate cancer progression and mortality. *Clin Cancer Res* 16:2819–2832. doi:[10.1158/1078-0432.CCR-10-0028](https://doi.org/10.1158/1078-0432.CCR-10-0028)
- 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
- Gold B, Kirchoff T, Stefanov S, Lautenberger J, Viale A, Garber J, Friedman E, Narod S, Olshen AB, Gregersen P, Kosarin K, Olsh A, Bergeron J, Ellis NA, Klein RJ, Clark AG, Norton L, Dean M, Boyd J, Offit K (2008) Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci USA* 105:4340–4345. doi:[10.1073/pnas.0800441105](https://doi.org/10.1073/pnas.0800441105)
- Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, Rafnar T, Bergthorsson JT, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Xu J, Blondal T, Kostic J, Sun J, Ghosh S, Stacey SN, Mouy M, Saemundsdottir J, Backman VM, Kristjansson K, Tres A, Partin AW, Albers-Akkers MT, Godino-Ivan Marcos J, Walsh PC, Swinkels DW, Navarrete S, Isaacs SD, Aben KK, Graif T, Cashy J, Ruiz-Echarri M, Wiley KE, Suarez BK, Witjes JA, Frigge M, Ober C, Jonsson E, Einarsson GV, Mayordomo JI, Kiemeny LA, Isaacs WB, Catalona WJ, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K (2007a) Genome-wide

- association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39:631–637. doi:[ng1999](https://doi.org/10.1038/ng1999)[pii][10.1038/ng1999](https://doi.org/10.1038/ng1999)
- Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediksdottir KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B, Ghosh S, Thorlacius S, Magnusdottir D, Stefansdottir G, Kristjansson K, Bagger Y, Wilensky RL, Reilly MP, Morris AD, Kimber CH, Adeyemo A, Chen Y, Zhou J, So WY, Tong PC, Ng MC, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Tres A, Fuertes F, Ruiz-Echarri M, Asin L, Saez B, van Boven E, Klaver S, Swinkels DW, Aben KK, Graif T, Cashy J, Suarez BK, van Vierssen Trip O, Frigge ML, Ober C, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Palmer CN, Rotimi C, Chan JC, Pedersen O, Sigurdsson G, Benediktsson R, Jonsson E, Einarsson GV, Mayordomo JI, Catalona WJ, Kiemeny LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K (2007b) Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 39: 977–983. doi:[10.1038/ng2062](https://doi.org/10.1038/ng2062)
- Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Gudbjartsson D, Agnarsson BA, Sigurdsson A, Benediksdottir KR, Blondal T, Jakobsdottir M, Stacey SN, Kostic J, Kristinsson KT, Birgisdottir B, Ghosh S, Magnusdottir DN, Thorlacius S, Thorleifsson G, Zheng SL, Sun J, Chang BL, Elmore JB, Breyer JP, McReynolds KM, Bradley KM, Yaspan BL, Wiklund F, Stattin P, Lindstrom S, Adami HO, McDonnell SK, Schaid DJ, Cunningham JM, Wang L, Cerhan JR, St Sauver JL, Isaacs SD, Wiley KE, Partin AW, Walsh PC, Polo S, Ruiz-Echarri M, Navarrete S, Fuertes F, Saez B, Godino J, Weijerman PC, Swinkels DW, Aben KK, Witjes JA, Suarez BK, Helfand BT, Frigge ML, Kristjansson K, Ober C, Jonsson E, Einarsson GV, Xu J, Gronberg H, Smith JR, Thibodeau SN, Isaacs WB, Catalona WJ, Mayordomo JI, Kiemeny LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K (2008) Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* 40:281–283. doi:[10.1038/ng.89](https://doi.org/10.1038/ng.89)
- Gudmundsson J, Sulem P, Gudbjartsson DF, Blondal T, Gylfason A, Agnarsson BA, Benediksdottir KR, Magnusdottir DN, Orlygsdottir 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, Kiemeny 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. doi:[10.1038/ng.448](https://doi.org/10.1038/ng.448)
- Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, Le Marchand L, Kolonel LN, Frasco M, Wong D, Pooler LC, Ardlie K, Oakley-Girvan I, Whittemore AS, Cooney KA, John EM, Ingles SA, Altshuler D, Henderson BE, Reich D (2007) Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39:638–644. doi:[10.1038/ng2015](https://doi.org/10.1038/ng2015)
- Hammer MF, Behar DM, Karafet TM, Mendez FL, Hallmark B, Erez T, Zhivotovsky LA, Rosset S, Skorecki K (2009) Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. *Hum Genet* 126:707–717. doi:[10.1007/s00439-009-0727-5](https://doi.org/10.1007/s00439-009-0727-5)
- Jordan JJ, Hanlon AL, Al-Saleem TI, Greenberg RE, Tricoli JV (2001) Loss of the short arm of the Y chromosome in human prostate carcinoma. *Cancer Genet Cytogenet* 124:122–126. pii:S0165-4608(00)00340-X
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830–838. doi:[10.1101/gr.7172008](https://doi.org/10.1101/gr.7172008)
- Kim W, Yoo TK, Kim SJ, Shin DJ, Tyler-Smith C, Jin HJ, Kwak KD, Kim ET, Bae YS (2007) Lack of association between Y-chromosomal haplogroups and prostate cancer in the Korean population. *PLoS One* 2:e172. doi:[10.1371/journal.pone.0000172](https://doi.org/10.1371/journal.pone.0000172)
- Kote-Jarai Z, Olama AA, Giles GG, Severi G, Schleutker J, Weischer M, Campa D, Riboli E, Key T, Gronberg H, Hunter DJ, Kraft P, Thun MJ, Ingles S, Chanock S, Albanes D, Hayes RB, Neal DE, Hamdy FC, Donovan JL, Pharoah P, Schumacher F, Henderson BE, Stanford JL, Ostrander EA, Sorensen KD, Dork T, Andriole G, Dickinson JL, Cybulski C, Lubinski J, Spurdle A, Clements JA, Chambers S, Aitken J, Gardiner RA, Thibodeau SN, Schaid D, John EM, Maier C, Vogel W, Cooney KA, Park JY, Cannon-Albright L, Brenner H, Habuchi T, Zhang HW, Lu YJ, Kaneva R, Muir K, Benlloch S, Leongamornlert DA, Saunders EJ, Tymrakiewicz M, Mahmud N, Guy M, O'Brien LT, Wilkinson RA, Hall AL, Sawyer EJ, Dadaev T, Morrison J, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As N, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Cooper CS, Lophatonanon A, Southey MC, Hopper JL, English DR, Wahlfors T, Tammela TL, Klarskov P, Nordestgaard BG, Roder MA, Tybjærg-Hansen A, Bojesen SE, Travis R, Canzian F, Kaaks R, Wiklund F, Aly M, Lindstrom S, Diver WR, Gapstur S, Stern MC, Corral R, Virtamo J, Cox A, Haiman CA, Le Marchand L, Fitzgerald L, Kolb S et al (2011) Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat Genet* 43:785–791. doi:[10.1038/ng.882](https://doi.org/10.1038/ng.882)
- 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* 343:78–85. doi:[10.1056/NEJM200007133430201](https://doi.org/10.1056/NEJM200007133430201)
- Lindstrom S, Adami HO, Adolfsson J, Wiklund F (2008) Y chromosome haplotypes and prostate cancer in Sweden. *Clin Cancer Res* 14:6712–6716. doi:[10.1158/1078-0432.CCR-08-0658](https://doi.org/10.1158/1078-0432.CCR-08-0658)
- Ma J, Li H, Giovannucci E, Mucci L, Qiu W, Nguyen PL, Gaziano JM, Pollak M, Stampfer MJ (2008) Prediagnostic body-mass index, plasma C-peptide concentration, and prostate cancer-specific mortality in men with prostate cancer: a long-term survival analysis. *Lancet Oncol* 9:1039–1047. doi:[10.1016/S1470-2045\(08\)70235-3](https://doi.org/10.1016/S1470-2045(08)70235-3)
- Paracchini S, Pearce CL, Kolonel LN, Altshuler D, Henderson BE, Tyler-Smith C (2003) A Y chromosomal influence on prostate cancer risk: the multi-ethnic cohort study. *J Med Genet* 40:815–819
- Rootsi S, Zhivotovsky LA, Baldovic M, Kayser M, Kutuev IA, Khusainova R, Bermisheva MA, Gubina M, Fedorova SA, Ilumae AM, Khusnutdinova EK, Voevoda MI, Osipova LP, Stoneking M, Lin AA, Ferak V, Parik J, Kivisild T, Underhill PA, Villems R (2007) A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur J Hum Genet* 15:204–211. doi:[10.1038/sj.ejhg.5201748](https://doi.org/10.1038/sj.ejhg.5201748)
- Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423:873–876. doi:[10.1038/nature01723](https://doi.org/10.1038/nature01723)
- Schumacher FR, Berndt SI, Siddiq A, Jacobs KB, Wang Z, Lindstrom S, Stevens VL, Chen C, Mondul AM, Travis RC, Stram DO, Eeles RA, Easton DF, Giles G, Hopper JL, Neal DE, Hamdy FC, Donovan JL, Muir K, Al Olama AA, Kote-Jarai Z, Guy M, Severi G, Gronberg H, Isaacs WB, Karlsson R, Wiklund F, Xu J, Allen NE, Andriole GL, Barricarte A, Boeing H, Bas Bueno-de-Mesquita H,

- Crawford ED, Diver WR, Gonzalez CA, Gaziano JM, Giovannucci EL, Johansson M, Le Marchand L, Ma J, Sieri S, Stattin P, Stampfer MJ, Tjonneland A, Vineis P, Virtamo J, Vogel U, Weinstein SJ, Yeager M, Thun MJ, Kolonel LN, Henderson BE, Albanes D, Hayes RB, Spencer Feigelson H, Riboli E, Hunter DJ, Chanock SJ, Haiman CA, Kraft P (2011) Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum Mol Genet* 20:3867–3875. doi:[10.1093/hmg/ddr295](https://doi.org/10.1093/hmg/ddr295)
- Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* 74:1023–1034. doi:[10.1086/386295](https://doi.org/10.1086/386295)
- Takata R, Akamatsu S, Kubo M, Takahashi A, Hosono N, Kawaguchi T, Tsunoda T, Inazawa J, Kamatani N, Ogawa O, Fujioka T, Nakamura Y, Nakagawa H (2010) Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet* 42:751–754. doi:[10.1038/ng.635](https://doi.org/10.1038/ng.635)
- The ATBC Cancer Prevention Study Group (1994) The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. *The ATBC Cancer Prevention Study Group. 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, Giovannucci 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. doi:[10.1038/ng.91](https://doi.org/10.1038/ng.91)
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65:43–62. pii:S0003480001008582
- 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. pii:022
- Vijayakumar S, Garcia D, Hensel CH, Banerjee M, Bracht T, Xiang R, Kagan J, Naylor SL (2005) The human Y chromosome suppresses the tumorigenicity of PC-3, a human prostate cancer cell line, in athymic nude mice. *Genes Chromosomes Cancer* 44:365–372. doi:[10.1002/gcc.20250](https://doi.org/10.1002/gcc.20250)
- 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
- 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. doi:[10.1038/nature05911](https://doi.org/10.1038/nature05911)
- Wiik K (2008) Where did European men come from? *J Gen Genealogy* 4:35–85
- Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Michiello 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. doi:[10.1038/ng2022](https://doi.org/10.1038/ng2022)
- Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–48. doi:[10.1101/gr.217602](https://doi.org/10.1101/gr.217602)
- 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. doi:[10.1038/ng.444](https://doi.org/10.1038/ng.444)