Carcinogenesis vol.34 no.6 pp.1260–1264, 2013
doi:10.1093/carcin/bgt055
Advance Access publication February 7, 2013

The G84E mutation of *HOXB13* is associated with increased risk for prostate cancer: results from the REDUCE trial

Zhuo Chen[1], Celia Greenwood[2,3], William B.Isaacs[4], William D.Foulkes[5], Jielin Sun[1,6], Sigun L.Zheng[1,6], Lynn D.Condreay[7] and Jianfeng Xu[1,6,8,9,*]

[1]Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA, [2]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec H3T 1E2, Canada, [3]Departments of Oncology and Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec H3A 1A2, Canada, [4]Department of Urology, John Hopkins Medical Institutions, Baltimore, MD 21205, USA, [5]Program in Cancer Genetics, McGill University, Montreal, Quebec H2W 1S6, Canada, [6]Department of Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston-Salem , NC 27157, USA, [7]Genetics, GlaxoSmithKline, Research Triangle Park, NC 27709, USA, [8]Department of Urology and [9]Department of Cancer Biology, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA

*To whom correspondence should be addressed. Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA. Tel: +1 336 713 7500; Fax: +1 336 713 7566; Email: jxu@wakehealth.edu

A novel rare mutation, homeobox B13 (*HOXB13*) G84E, was reported to co-segregate with prostate cancer (PCa) in hereditary PCa families and associate with PCa risk in unrelated cases and controls. In this study, we aim to compare the G84E mutation frequency among subjects of different races/ethnicities from various geographic regions in the world and to assess its risk for developing PCa, in the Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial. All the 3508 subjects had initial negative prostate biopsy and were biopsied at Year 2 and 4 for detection of PCa. The G84E mutation was detected only in Caucasians, with the highest carrier frequency in Northern Europe (1.06%), followed by Western Europe (0.60%) and North America (0.31%). No mutation carrier was observed in Southern Europe, Eastern Europe, Latin America, Australia and South Africa. In Caucasians, the G84E mutation frequency was 0.99% and 0.24% in positive and negative biopsy subjects, respectively ($P = 0.01$). In positive biopsy subjects, the frequency was significantly higher in subjects with a positive family history than those without (4.31% versus 0.34%, $P = 0.002$). In the 4 year follow-up, the PCa detection rate was 53.8% among the 13 mutation carriers and 22.0% among 3186 non-carriers, relative risk = 2.45 (95% confidence interval: 1.48–4.07). All mutation carriers shared a common haplotype, suggesting a founder effect. In Finland, the G84E mutation was estimated to occur in the year 1792 (95% credible interval: 1735–1831). In conclusion, the G84E mutation of *HOXB13*, a relatively recent mutation that likely occurred in Northern Europe, significantly increases risk for PCa.

Introduction

Prostate cancer (PCa) is one of the most prevalent cancers worldwide. The incidence of PCa is higher in African Americans and in the European descent than in Hispanics and Asians (1,2). It is widely hypothesized that genetic factors contribute to racial/ethnic differences in incidence of PCa. However, specific genes underlying the variation remain largely unknown.

By targeted exome sequencing in a PCa linkage region at 17q21–q22 among 94 probands of hereditary PCa families, Ewing *et al.* (3) recently identified a rare but recurrent non-synonymous mutation, G84E (rs138213197), in the transcription factor homeobox B13 (*HOXB13*) gene. The G84E mutation is located in the conserved functional domains of *HOXB13,* which mediates the binding of HOX13 paralogs (including *HOXB13*) to the MEIS homeodomain family of HOX cofactor proteins. The amino acid substitution from glycine to glutamic acid is predicted to be deleterious to HOXB13 protein function. In Ewing's study, the G84E mutation was found to co-segregate completely with PCa. In addition, they found that the mutation occurred in 1.4% of 5083 unrelated PCa cases and 0.1% of 1401 controls of Caucasian men. Carriers of the G84E mutation had 20.1 times the odds of PCa compared with the non-carriers (odds ratio [OR] = 20.1) (3). Subsequent studies by Breyer *et al.* (4) in 928 familial PCa probands and 930 controls of European descent and by Akbari *et al.* (5) in 1525 PCa sporadic cases and 1757 controls from Canada confirmed the association. However, the reported OR in the latter studies were 7.9 (95% confidence interval [CI]: 1.8–34.5) and 5.8 (95% CI: 1.3–26.5), respectively, which were considerably lower than the initial report (3).

To better understand the frequency of the G84E mutation in various races/ethnicities and geographic regions in the world and estimate its risk for PCa, we evaluated the mutation in the Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial, an international multicenter chemoprevention trial originally designed to evaluate the effect of dutasteride on PCa risk. The broad distribution of study subjects across the world and the prospective nature of the study design provide a unique opportunity to address these two questions.

Materials and methods

Study subjects

The study was performed in a total of 3508 subjects (3239 Caucasians, 45 African descendants, 199 Hispanics and 25 Asians from 31 countries) from the REDUCE trial who consented for genetic studies. The REDUCE trial is a 4 year, multicenter, randomized, double-blind, placebo-controlled, parallel group study. The design and rational of the REDUCE trial have been described with details in the previous reports (6,7). Among the 3508 subjects included in this study, 1794 subjects were randomized to the placebo arm, whereas 1714 subjects were in the dutasteride arm. The inclusion criteria for the REDUCE trial were as follows: (i) 50–75 years of age; (ii) prostate-specific antigen (PSA) level of 2.5–10 ng/ml for subjects <60 years old and 3–10 ng/ml for those ≥60 years; (iii) prostate volume ≤80 cm^3; (iv) negative initial prostate biopsy and (v) without high-grade prostatic intraepithelial neoplasia or atypical small acinar proliferation (6). During the 4 year follow-up period, PCa were detected in 770 of 3508 subjects (21.9%). Characteristics of the study subjects are presented in Supplementary Table 1, available at *Carcinogenesis* Online.

Genetic markers and genotyping

We genotyped the *HOXB13* G84E mutation and additional 14 variants within or flanking the *HOXB13* gene region (chromosome 17, bp 46, 719, 399-46, 827, 590, build 37). All variants were genotyped using the iPLEX MassARRAY system (Sequenom, San Diego, CA). Duplicates and negative (water) controls were included in each 96-well plates as quality control samples. All assays were performed in a blinded fashion. The overall genotyping call rate is 98.5%. Negative controls (water samples) did not give a genotype call. All of the variants were in Hardy–Weinberg Equilibrium in negative biopsy subjects (Supplementary Table 2, available at *Carcinogenesis* Online). Among the 15 genetic variants genotyped, 11 had a minor allele frequency over 0.1%. Therefore, the 11 variants were used for haplotype construction, as detailed in the previous study (8). Haplotypes were inferred based on the 11 genetic variants using the expectation–maximization algorithm implemented in the PLINK software package (9).

Abbreviations: *HOXB13*, homeobox B13; OR, odds ratio; PCa, prostate cancer; PSA, prostate-specific antigen; REDUCE, Reduction by Dutasteride of Prostate Cancer Events; RR, relative risk; SNP, single nucleotide polymorphism.

Statistical analyses

G84E frequencies were compared between positive prostate biopsy subjects and negative prostate biopsy subjects, using Fisher's exact test implemented in SAS software (SAS Institute, Cary, NC). Due to the prospective nature of this study, relative risk (RR, rate ratio) was used to measure the magnitude of association between G84E mutation and PCa among Caucasians. In this study, the RR is a ratio of the probability of PCa (positive prostate biopsy) occurring in the G84E mutation carriers versus the probability in the non-carriers. RR was estimated for all subjects and among the subsets of positive biopsy subjects versus all negative biopsy subjects. We also conducted the Breslow–Day test for homogeneity of stratified analysis according to family history, age at diagnosis and Gleason grade, respectively. RR estimation and the Breslow–Day test were conducted using SAS. Linear regression analysis was conducted using PLINK software (9) to assess the association between G84E mutation and baseline clinical variables, including total PSA, free PSA, prostate volume, testosterone, dihydrotestosterone and International Prostate Symptom Score, adjusting age as a covariate. All clinical variables were log transformed to approximate normal distributions. The interaction effect between G84E and treatment (dutasteride) was analyzed using logistic regression implemented in PLINK by adding an interaction term (9). All tests were two tailed.

To help make the results of this study more comparable with those of previous studies, we also estimated the OR, using PLINK software to measure the association of each single nucleotide polymorphism (SNP) with PCa risk under an additive genetic model in all subjects (9). For OR estimation, subjects with positive prostate biopsy were treated as cases, whereas subjects with negative prostate biopsy were treated as controls.

Phylogenetic tree generation

The phylogenetic tree of haplotypes, which appeared over 10 times in all subjects, was constructed using MEGA version 5.05 (10). The neighbor-joining method (11), a simplified version of the minimum evolution method (12), was used to assess the relationships among major haplotypes. In the neighbor-joining method, evolutionary distance measures are used to correct for multiple mutational events at the same nucleotide and the examination of different topologies is embedded in the algorithm (13) so that only one final tree is produced. The neighbor-joining method does not require the assumption of a constant rate of evolution.

Mutation age estimation

DMLE (Disease Mapping using Linkage disEquilibrium) software was used to estimate G84E mutation age in Finland, based on the observed linkage disequilibrium at multiple genetic markers (14). DMLE uses Markov Chain Monte Carlo methods to allow Bayesian estimation of the posterior probability density of the position of a disease mutation relative to a set of markers as well as the age of the mutation (14). The key parameters necessary for estimating the mutation age are the population growth rate and the proportion of disease chromosomes sampled. The population growth rate was calculated using the following formula: $p1 = p0 * r^g$ (15), where p0 is the population of Finland in 1750 AD, p1 is the Finnish population in 2003 when the REDUCE trial started, g is number of generations between these two population estimates and r is the growth rate per generation. According to Statistics Finland (http://www.stat.fi/meta/svt/index_en.html), the population was 421 500 in the year 1750 and 5 219 732 in 2003, respectively. Because the average age of primiparous women was 24.9 in 1975 in Finland (http://www.stat.fi/meta/svt/index_en.html) and that women gave birth at a relatively younger age in history than in recent years, we assumed 22 years per generation in the history in Finland. We assumed a constant growth rate over these 250 years as well. For the proportion of disease chromosomes sampled, we used the formula: pro = n/ (N*I*F), where pro is the desired proportion, n is the number of cases in the sample, N is the population of Finland in 2003, I is the lifetime risk of PCa and F is the estimated frequency of the G84E mutation in PCa cases (estimated from our sample). Other parameters of DMLE were left at their default values as follows: data as genotyped, 0; use fixed random seed, 0; run simulation, 0; mutation's low and high boundaries, −0.05 0.05; simultaneous runs, 1; staring value for recdist, −99; iterate ancestral states, mutation age, mutation location, allele frequency, 1100; flip all loci, 1; burn-in iterations, 1 000 000; iterations, 1 000 000; screen update and file update intervals, 10 010 010; number of histogram bars, 200; mutation age −99; mutation age boundaries, 0 5000; star genealogy, 0; loci for the root, −99; use sequence weights 0; weighs for exons, introns and non-genes, 0.

To test whether the mutation age estimation is sensitive to the population growth rate, estimations assuming 20 or 25 years per generation (population growth rate 1.22 or 1.28 per generation, respectively) were conducted.

Results

G84E carrier frequency distributions

The distribution for G84E carriers is presented in Table I. Briefly, the G84E mutation of *HOXB13* was observed in 13 of 3239 Caucasian men (0.4%) but was not detected in other races/ethnicities, including 45 African descendants, 199 Hispanics and 25 Asians. Among Caucasians, carrier frequency of the G84E mutation varied from different geographic regions; it was highest in Northern Europe (1.06%), followed by Western Europe (0.60%) and North America (0.31%). No mutation carrier was observed in Southern Europe ($N = 187$), Eastern Europe ($N = 419$), Latin America ($N = 233$), Australia ($N = 37$) and South Africa ($N = 31$).

Effects of G84E on PCa

In the end of the 4 year follow-up of the REDUCE trial, we found that 7 out of 13 (53.8%) G84E mutation carriers developed PCa, whereas the number in non-carriers was 701 of 3186 (22.0%). The G84E mutation carriers were at 2.45-fold increased risk of developing PCa, compared with non-carriers (RR = 2.45, 95% CI: 1.48–4.07) (Table II). The G84E carrier frequency was 0.99% (7/708) in positive biopsy subjects compared with 0.24% (6/2491) in negative biopsy subjects (Table II).

In addition, we compared carrier frequency of G84E mutation in the positive biopsy subjects stratified by family history, age at diagnosis and Gleason grade (Table II). Among 708 positive biopsy subjects, we found 5 G84E mutation carriers in 116 subjects with a family history of PCa and 2 G84E mutation carriers in 592 subjects without a family history, which was significantly different (4.31% versus 0.34%, $P = 0.002$). With regard to age at diagnosis, we did not detect significant difference of G84E carrier frequency between positive biopsy subjects diagnosed before or after 65 years old (1.05% [3/285] versus 0.95% [4/423], $P = 0.99$]. However, we noticed that all of the three positive biopsy subjects carrying G84E mutation diagnosed before 65 years old had a family history of PCa. Although carrier frequency of the G84E mutation was not significantly different in positive biopsy subjects with high (≥7) or low (<7) Gleason grade (1.71% [4/234) versus 0.63% (3/471), $P = 0.23$], G84E mutation carriers were at 4.73-fold increased risk of developing high Gleason grade PCa (RR = 4.72, 95% CI: 2.19–10.19) compared with all negative biopsy subjects.

No significant association was detected between G84E mutation and PCa-related baseline clinical variables, including total PSA, free PSA, prostate volume, testosterone, dihydrotestosterone and International Prostate Symptom Score (Table III). In addition, no significant interaction between G84E mutation and treatment (dutasteride) was found ($P = 0.57$).

To make help our results more comparable with previous studies in which OR was used as a measurement for association, we also estimated OR to measure the association of each SNP with PCa risk under an additive genetic model among all subjects. Carriers of the G84E mutation had 4.14 times the odds of PCa compared with the non-carriers (OR = 4.14, 95% CI: 1.38–12.28, $P = 0.006$). None of the other SNPs were significantly associated with PCa risk (Supplementary Table 2, available at *Carcinogenesis* Online).

Haplotype inference

A total of 91 haplotypes were inferred. All 13 G84E mutation carriers shared a unique haplotype G-T-A-G-A-C-C-***T***-A-C-T with the G84E mutation at the 10th locus, suggesting a founder effect. To investigate the haplotype relationships and the potential evolutionary events in this region, we performed phylogenetic analysis of 40 haplotypes, which appeared over 10 times in 3239 Caucasians. In the phylogenetic tree, the unique haplotype shared by G84E carriers (0.50% and 0.12% in positive and negative biopsy subjects, respectively, $P = 4E\text{-}4$) was located in a subbranch with another haplotype G-T-A-G-A-C-C-***C***-A-C-T (0.50% and 0.72% in positive and negative biopsy subjects, respectively, $P = 0.46$). The two haplotypes differed only at the eighth locus where the G84E mutation is situated, indicating that the G84E

Table I. Geographic distributions of *HOXB13* G84E carrier frequency in Caucasians

Country/region	Total count	G84E carriers		G84E non-carriers	
		Count	Frequency	Count	Frequency
Northern Europe	374	4	1.06%	370	98.94%
Denmark	2	0	0	2	1
Finland	101	2	1.96%	99	98.04%
Norway	56	0	0	56	1
Sweden	215	2	0.92%	213	99.08%
Southern Europe	183	0	0	183	1
Greece	12	0	0	12	1
Italy	6	0	0	6	1
Portugal	165	0	0	165	1
Western Europe	981	6	0.60%	975	99.40%
Austria	31	0	0	31	1
Belgium	26	1	3.85%	25	96.15%
Germany	602	3	0.49%	599	99.51%
Ireland	2	0	0	2	1
Netherlands	165	1	0.60%	164	99.40%
Switzerland	24	0	0	24	1
UK	131	1	0.75%	130	99.25%
Eastern Europe	413	0	0	413	1
Belarus	2	0	0	2	1
Bulgaria	31	0	0	35	1
Estonia	11	0	0	11	1
Hungary	14	0	0	14	1
Latvia	86	0	0	86	1
Lithuania	67	0	0	67	1
Poland	20	0	0	20	1
Romania	12	0	0	12	1
Russia	100	0	0	100	1
Slovakia	66	0	0	66	1
North America	947	3	0.31%	946	99.69%
Canada	315	1	0.32%	314	99.68%
USA	634	2	0.31%	632	99.69%
Latin America	231	0	0	231	1
Argentina	230	0	0	230	1
Chile	1	0	0	1	1
Others	68	0	0	68	1
South Africa	31	0	0	31	1
Australia	37	0	0	37	1
Total	3199	13	0.41%	3186	99.59%

mutation is likely to be a recent mutational event at the eighth locus on the background of the G-T-A-G-A-C-C-***C***-A-C-T haplotype.

Mutation age estimation

We estimated age of mutation for G84E in subjects from Finland, where the highest G84E carrier frequency (1.96%) was observed, except in Belgium where only 29 subjects were included and may not be representative. Assuming 22 years per generation (population growth rate 1.24 per generation), the G84E mutation was estimated to originate approximately 9.6 (95% credible interval: 7.8–12.2) generations ago in Finland. In term of year, the G84E mutation was estimated to originate around the year 1792 (95% credible interval: 1735–1831). When assuming 20 or 25 years per generation in Finland, the G84E mutation may originate in the year 1807 (95% credible interval: 1753–1843) or 1763 (95% credible interval: 1703–1808), respectively. Overall, the *HOXB13* G84E is most likely to originate around the turn of the 19th century.

Discussion

In this study, we confirmed the rare mutation G84E in *HOXB13* gene as susceptibility loci for PCa in Caucasians from the REDUCE trial in a prospective manner. In addition, in positive biopsy subjects, the G84E frequency was significantly higher in subjects with a positive family history than those without. Moreover, we found the G84E mutation was likely to be a founder mutation, suggesting this mutation affects risk of hereditary PCa. Given the G84E frequency in Northern Europe (1.06%), Western Europe (0.60%) and North America (0.31%), we speculated that G84E might spread from Finland to other geographic regions during historic population migration, as nearly 0.7 million Finnish emigrated to Sweden, the USA and Canada during 1866–1970 (16).

We compared the effect of the G84E mutation on PCa risk in our study with previous reports (3–5). Only OR estimates were available from the previous reports due to the case-control study design adopted. Therefore, we also estimated OR in our study. The OR estimate in our study (4.14) was smaller than that in the previous reports by Ewing *et al.* (3) (OR = 20.1), Breyer *et al.* (4) (OR = 7.9) and Akbari *et al.*(5) (OR = 5.8). Two possible reasons may account for the variations of the point estimates. The first possible reason may be due to the different sampling strategy used in the studies. For case selection, PCa patients in the original study by Ewing *et al.*,(3) were either hereditary PCa patients or diagnosed at a young age (<55 years). Similarly, study by Breyer *et al.* (4) was also conducted in familial cases, whereas only 15.8% patients in our study and 17.7% patients in Akbari's study had a family history of PCa (5). In addition to case selection, sampling strategy of control may also influence the point estimation of OR. For example, controls in Ewing's study were selected based on low PSA levels and digital rectal examination, whereas controls in our study had a PSA ranging from 2.5–10 ng/ml with negative prostate biopsy. Therefore, the variations in OR estimates may due to different sampling strategies used among studies. Second, potential population stratification may also influence OR estimation. Although

Table II. *HOXB13* G84E carrier frequency in positive and negative biopsy subjects

	G84E carriers	G84E non-carriers	G84E carrier frequency	RR (95% CI)[a]	*P*	Homogeneity *P*[b]
Negative biopsy subjects	6	2485	0.24%	1.00	NA	
Positive biopsy subjects	7	701	0.99%	2.45 (1.48–4.07)	0.01	
Family history						
Positive family history	5	111	4.31%	10.63 (5.43–20.83)	6.0E-5	0.004
Negative family history	2	590	0.34%	1.31 (0.39–4.34)	0.65	
Age at diagnosis						
<65 y	3	282	1.05%	3.28 (1.29–8.31)	0.06	0.91
≥65 y	4	419	0.95%	2.78 (1.29–5.97)	0.04	
Gleason grade						
High grade (≥7)	4	230	1.71%	4.73 (2.19–25.71)	0.01	0.29
Low grade (<7)	3	471	0.63%	2.10 (0.83–5.30)	0.16	

[a]RR was estimated for overall or subsets of positive biopsy subjects versus all negative biopsy subjects.
[b]Homogeneity *P* of Breslow–Day test was estimated for homogeneity of stratified analysis according to family history, age at diagnosis and Gleason grade, respectively.

Table III. Associations between *HOXB13* G84E mutation and baseline clinical variables

Clinical variables	Mean[a]		*P*[a]
	G84E carriers	G84E non-carriers	
PSA (ng/ml)	5.43	5.58	0.78
Free PSA (ng/ml)	0.70	0.87	0.07
Prostate volume (cm^3)	37.26	43.38	0.16
Testosterone (nmol/l)	16.98	14.38	0.14
Dihydrotestosterone (nmol/l)	1.35	1.20	0.49
International Prostate Symptom Score	6.14	7.14	0.47

[a]Tests were based on variables by linear regression analysis, adjusted for age; mean were back-transformed to original scale.

all the subjects in the previous studies (3–5) were Caucasians, they may represent subjects with different originations from Europe. As we showed in our study, the G84E mutation frequency varied among different geographic regions in Europe. Thus, OR might be overestimated, if a higher percentage of subjects were originated from geographic regions with higher G84E frequency in cases than in controls, and *vice versa*.

We recognize there are several limitations in this study. First, the subjects included in the REDUCE trial may not be representative of the general population, due to the inclusion criteria used for recruitment. Specifically, only subjects with a relatively increased PSA level of 2.5–10 ng/ml (for subjects <60 years old) or 3–10 ng/ml (for subjects ≥60 years old) were recruited in the REDUCE trial. However, only subjects with a negative prostate biopsy prior to the recruitment were included in the study and men who developed PCa were excluded. Taken together, participants in the REDUCE trial is highly selected and whether they are at increased risk of PCa remains undetermined and further studies are warranted. Second, G84E frequency in men from certain ethnicities may not be representative because of the small sample size. For example, we did not observe any G84E mutation in men of African descents in our study. However, there are only 45 subjects of African descents in total, which are not sufficient to get a precise estimate of the G84E frequency. The same issue also applied in the frequency estimates in subpopulations in Austria, Belgium and Switzerland, where 31, 27 and 26 subjects were included in this study, respectively. This was also the reason why age of the G84E mutation was not estimated in Belgium, where the G84E frequency was higher in Finland. However, the strategy we used to pool subjects with similar genetic background may compensate to some extent. Future studies with larger sample size are needed to get a precise estimation of G84E frequency in certain subgroups. Third, the statistical power may be insufficient for the association analyses of G84E with baseline clinical variables. Studies with larger sample size are needed in the future. In addition, population stratification may confound the results of the study analyses.

Overall, this study provided further characteristics of *HOXB13* G84E mutation distribution among subjects of different races/ethnicities from various geographic regions in the world. In addition, we found G84E mutation carriers were at a 2.45-fold increased risk for developing PCa compared with non-carriers in the REDUCE trial, presenting larger effect in PCa risk, compared with common SNPs identified by genome-wide association studies with ORs ranging from 1.1 to 1.6 (17–24). In addition, we found *HOXB13* G84E is likely to be a recent founder mutation, which originated in Finland around the turn of the 19th century. This recent origin is likely to contribute to its low frequency in most populations studied.

Funding

National Cancer Institute RC2 grant (CA148463 to J.X.); and a research contract by GlaxoSmithKline to J.X.

Acknowledgements

We thank the patients enrolled in the REDUCE trial (sponsored by GlaxoSmithKline) who provided consent and genetic samples, which enabled this study, and the clinicians who contributed their expertise in recruiting study patients for the REDUCE clinical study. Dave Pulford, Jennifer Aponte, Jon Charnecki and Mary Ellyn Volk participated in consent reconciliation and sample management to enable genetic sample selection for inclusion and genotype determination. Karen King provided data management support for this project. We appreciate the assistance of Lauren Marmor in coordinating the support of the Avodart Collaborative Research Team.

Conflict of Interest Statement: L.D.C. was a GlaxoSmithKline employee during this study and holds stock in GlaxoSmithKline. J.X. certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (e.g. employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received or pending) have been disclosed.

References

1. Siegel,R. *et al.* (2011) Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA. Cancer J. Clin.*, **61**, 212–236.
2. Crawford,E.D. (2003) Epidemiology of prostate cancer. *Urology*, **62**(6 suppl. 1), 3–12.
3. Ewing,C.M. *et al.* (2012) Germline mutations in HOXB13 and prostate-cancer risk. *N. Engl. J. Med.*, **366**, 141–149.
4. Breyer,J.P. *et al.* (2012) Confirmation of the HOXB13 G84E germline mutation in familial prostate cancer. *Cancer Epidemiol. Biomarkers Prev.*, **21**, 1348–1353.

5. Akbari,M.R. *et al.* (2012) Association between germline HOXB13 G84E mutation and risk of prostate cancer. *J. Natl. Cancer Inst*., **104**, 1260–1262.
6. Andriole,G. *et al.*; REDUCE Study Group. (2004) Chemoprevention of prostate cancer in men at high risk: rationale and design of the Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial. *J. Urol*., **172**(4 Pt 1), 1314–1317.
7. Andriole,G.L. *et al.*; REDUCE Study Group. (2010) Effect of dutasteride on the risk of prostate cancer. *N. Engl. J. Med*., **362**, 1192–1202.
8. Karlsson,R. *et al.* (2012) A population-based assessment of germline HOXB13 G84E mutation and prostate cancer risk. *Eur. Urol*. 2012 Jul 20. [Epub ahead of print].
9. Purcell,S. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet*., **81**, 559–575.
10. Kumar,S. *et al.* (1994) MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput. Appl. Biosci*., **10**, 189–191.
11. Saitou,N. *et al.* (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*., **4**, 406–425.
12. Rzhetsky,A. *et al.* (1993) Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol. Biol. Evol*., **10**, 1073–1095.
13. Nei,M. *et al.* (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York, NY.
14. Reeve,J.P. *et al.* (2002) DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics*, **18**, 894–895.
15. Greenwood,C.M. *et al.* (2010) How old is this mutation? - a study of three Ashkenazi Jewish founder mutations. *BMC Genet*., **11**, 39.
16. Kero,R. (1980) Migration from Finland 1866–1970. In *Les Migrations Internationales de la Fin Du XVIII* Éditions du Centre national de la recherche scientifique, Paris, France, pp. 392–400.
17. Eeles,R.A. *et al.*; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators; PRACTICAL Consortium. (2009) Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat. Genet*., **41**, 1116–1121.
18. Eeles,R.A. *et al.*; UK Genetic Prostate Cancer Study Collaborators; British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators. (2008) Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet*., **40**, 316–321.
19. Gudmundsson,J. *et al.* (2009) Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat. Genet*., **41**, 1122–1126.
20. Gudmundsson,J. *et al.* (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet*., **39**, 631–637.
21. Kote-Jarai,Z. *et al.*; UK Genetic Prostate Cancer Study Collaborators/ British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators, The Australian Prostate Cancer BioResource; PRACTICAL Consortium. (2011) Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat. Genet*., **43**, 785–791.
22. Sun,J. *et al.* (2009) Sequence variants at 22q13 are associated with prostate cancer risk. *Cancer Res*., **69**, 10–15.
23. Thomas,G. *et al.* (2008) Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet*., **40**, 310–315.
24. Yeager,M. *et al.* (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet*., **39**, 645–649.

Received September 13, 2012; revised January 30, 2013; accepted February 3, 2013