



NIH PUBLIC ACCESS

Author Manuscript

Nat Genet. Author manuscript; available in PMC 2013 February 03.

Published in final edited form as:

Nat Genet. 2009 October ; 41(10): 1122–1126. doi:10.1038/ng.448.

Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility

Julius Gudmundsson^{1,*}, Patrick Sulem^{1,*}, Daniel F. Gudbjartsson¹, Thorarinn Blondal¹, Arnaldur Gylfason¹, Bjarni A. Agnarsson^{2,3}, Kristrun R. Benediktsdottir^{2,3}, Droplaug N. Magnusdottir¹, Gudbjorg Orlygsdottir¹, Margret Jakobsdottir¹, Simon N. Stacey¹, Asgeir Sigurdsson¹, Tiina Wahlfors⁴, Teuvo Tammela⁵, Joan P. Breyer⁶, Kate M. McReynolds⁶, Kevin M. Bradley⁶, Berta Saez^{7,8}, Javier Godino⁷, Sebastian Navarrete⁹, Fernando Fuertes⁹, Laura Murillo¹⁰, Eduardo Polo¹¹, Katja K. Aben^{12,13}, Inge M. van Oort¹⁴, Brian K. Suarez¹⁵, Brian T. Helfand¹⁶, Donghui Kan¹⁶, Carlo Zanon^{1,17}, Michael L. Frigge¹, Kristleifur Kristjansson¹, Jeffrey R. Gulcher¹, Gudmundur V. Einarsson¹⁸, Eirikur Jonsson¹⁸, William J. Catalona¹⁶, Jose I. Mayordomo^{7,8,19}, Lambertus A. Kiemeny^{12,13,14}, Jeffrey R. Smith^{6,20}, Johanna Schleutker⁴, Rosa B. Barkardottir², Augustine Kong¹, Unnur Thorsteinsdottir^{1,3}, Thorunn Rafnar¹, and Kari Stefansson^{1,3}

¹deCODE genetics, IS-101 Reykjavik, Iceland ²Department of Pathology, Landspítali-University Hospital, IS-101 Reykjavik, Iceland ³Faculty of Medicine, University of Iceland, IS-101 Reykjavik, Iceland ⁴Institute of Medical Technology, University of Tampere and Tampere University Hospital, FI-33014 Tampere, Finland ⁵Department of Urology, Tampere University Hospital, University of Tampere, FI-33014 Tampere, Finland ⁶Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA ⁷Health Science Institute of Aragon, 50009 Zaragoza, Spain ⁸Nanotechnology Institute of Aragon, 50009 Zaragoza, Spain ⁹Division of Radiation Oncology, Lozano Blesa University Hospital, University of Zaragoza, 50009 Zaragoza, Spain ¹⁰Division of Medical Oncology, Reina Sofia Hospital, 31500 Tudela, Spain ¹¹Division of Medical Oncology, Ernest Lluch Hospital, 50300 Calatayud, Spain ¹²Comprehensive Cancer Center East, 6501 BG Nijmegen, The Netherlands ¹³Department of Epidemiology, Biostatistics & HTA, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands ¹⁴Department of Urology, Radboud University Nijmegen Medical Center, 6500 HB Nijmegen, The Netherlands ¹⁵Department of Psychiatry, Washington University School of Medicine, St Louis, MO 63110, USA ¹⁶Department of Urology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA ¹⁷Laboratory of Molecular Genetics, Institute for

Correspondence and request for materials should be addressed to: K. S., (kstefans@decode.is) or J. G., (julius@decode.is).

*These authors contributed equally to this work.

AUTHOR CONTRIBUTION

The study was designed and results were interpreted by J.G., P.S., A.K., U.T., T.R., and K.S. Statistical analysis was carried out by P.S., D.F.G., J.G., M.L.F., and A.K. Subject recruitment, biological material collection and handling along with genotyping was supervised and carried out by J.G., B.A., K.R.B., D.N.M., G.O., M.J., S.N.S., A.S., T.W., T.T., J.P.B., K.M.Mc., K.M.B., B.S., J.Godino, S.N., F.F., L.M., E.P., K.K.A., I.M.vO., B.K.S., B.T.H., D.K., C.Z., K.K., J.R.G., G.V.E., E.J., W.J.C., J.I.M., L.A.K., J.R.S., J.S., R.B.B., U.T. and T.R. Authors J.G., P.S., D.F.G., T.R. and K.S. drafted the manuscript. All authors contributed to the final version of the paper. Principal investigators and corresponding authors for the respective replication study populations are: The Netherlands, L.A.K.; Spain, J.I.M.; Chicago, W.J.C.; Nashville, J.R.S.; Finland, J.S.

COMPETING INTERESTS STATEMENT

The authors from deCODE genetics declare competing financial interests.

URLS

Cancer Genetics Markers of Susceptibility (CGEMS) study (<http://cgems.cancer.gov/>) The Icelandic Cancer Registry (<http://www.krabbameinsskra.is/indexen.jsp?icd=C61>) The International HapMap Project (<http://hapmap.org/index.html>)

Cancer stat fact sheets at NORDCAN, The Association of the Nordic Cancer Registries (<http://www-dep.iarc.fr/NORDCAN/english/StatsFact.asp?cancer=241&country=352>)

Cancer Research and Treatment, University of Torino Medical School, I-10060 Candiolo, Italy ¹⁸Department of Urology, Landspítali-University Hospital, 101 Reykjavik, Iceland ¹⁹Division of Medical Oncology, University Hospital, University of Zaragoza, 50009 Zaragoza, Spain ²⁰Medical Research Service, VA Tennessee Valley Healthcare System, Nashville, Tennessee 37212, USA

Abstract

We report a genome-wide association follow up study on prostate cancer. We identify four variants associated with the disease in European populations: rs10934853-A (OR = 1.12, $P = 2.9 \times 10^{-10}$) on 3q21.3, two moderately correlated ($r^2 = 0.07$) variants on 8q24.21; rs16902094-G (OR = 1.21, $P = 6.2 \times 10^{-15}$) and rs445114-T (OR = 1.14, $P = 4.7 \times 10^{-10}$) and rs8102476-C (OR = 1.12, $P = 1.6 \times 10^{-11}$) on 19q13.2. We also refine a previous association signal on 11q13 with the SNP rs11228565-A (OR = 1.23, $P = 6.7 \times 10^{-12}$). In a multi-variant analysis, using 22 prostate cancer risk variants typed in the Icelandic population, we estimate that carriers belonging to the top 1.3% of the risk distribution have a risk of developing the disease that is more than 2.5 times greater than the population average risk estimates.

We and others have previously presented results from genome-wide association studies (GWAS) on prostate cancer reporting several common variants conferring risk of the disease¹⁻⁷. By scrutinizing both our Icelandic GWAS data and publicly available data, as well as through fine-mapping work of two previously published loci on 8q24.21 and 11q13, we identified four new variants conferring risk of prostate cancer and a variant refining the previously published association signal on 11q13.

In order to search for prostate cancer risk variants, we performed an analysis of the combined data from the Icelandic GWAS, generated using the Illumina 317K chip, and from the replication genotyping project, released (spring 2008) by the National Cancer Institute Cancer Genetic Markers of Susceptibility (CGEMS), using 25K SNPs genotyped on five study populations, including PLCO, CPS-II, HPFS, FPCC and ATBC^{5,7}. In the combined study, when excluding regions containing previously reported prostate cancer risk variants, only two SNPs have $P < 1 \times 10^{-5}$. These two new variants are: allele A of rs10934853 (rs10934853-A) located on 3q21.3 (a fully correlated marker used from the CGEMS data is rs4857841 with D' and $r^2 = 1$ according to CEU HapMap, Supplementary Table 1); allelic odds ratio (OR) of 1.14 and $P = 1.6 \times 10^{-7}$ and allele C of rs8102476 (rs8102476-C) located on 19q13.2; OR of 1.11 and $P = 2.6 \times 10^{-6}$. In the combined study of the Icelandic and CGEMS data, when analyzing variants located within 1 Mb regions of previously published prostate cancer risk loci, we found only the SNP rs445114, located on 8q24.21, to be associated with prostate cancer (OR = 1.14, $P = 3.1 \times 10^{-8}$) and not to be correlated with other signals at any of the previously reported prostate cancer loci.

As a part of ongoing fine-mapping project on the three previously published variants on 8q24.21 we re-sequenced a 527 kb candidate region on 8q24 using pools of either Icelandic case or control samples (see Online Methods). We prioritized the analysis of 7 SNPs producing suggestive association results for prostate cancer in the analysis of the pooled samples (Supplementary Table 1). These 7 SNPs were genotyped, using Centaurus single track assay, in our Icelandic set of up to 1,980 patients and 7,000 controls. Six of the SNPs were found to be correlated with one of the previously reported variants on 8q24, while rs16902094 was not found to be correlated with any of the previously reported prostate cancer variants on 8q24.21 (Supplementary Tables 1 and 2). Furthermore, rs16902094 was found to be associated with prostate cancer in Iceland, having an OR of 1.28 ($P = 3.5 \times 10^{-6}$; Table 1). rs16902094 is not present on the Illumina Hap 317K SNP chip and no data are

available for it in the HapMap project according to HapMap's official website. However, by genotyping SNPs located within the same LD-region as rs16902094 and that are present on the Illumina Hap 550 chip, we identified a highly correlated SNP rs16902104 ($D' = 0.98$, $r^2 = 0.96$ according to data from 2,633 Icelanders; Supplementary Table 1) which CGEMS has released data for. By combining the Icelandic data for rs16902094 and the CGEMS data for rs16902104 we obtained an OR of 1.21 ($P = 1.1 \times 10^{-9}$).

The two new SNPs on 8q24.21, rs16902094 and rs445114, are located in the same LD-region but the correlation between them is very low ($D' = 1$ and $r^2 = 0.07$ according to data from 5,450 Icelanders; Supplementary Table 1) and the results for both remain significant after adjustment for the other (Supplementary Table 3a). This suggests that a unique variant capturing the effect of both rs16902094 and rs445114 remains to be discovered or, alternatively, that the LD-region contains more than one variant that predisposes to prostate cancer. Of the previously published cancer variants on 8q24, only the breast cancer variant (rs13281615)⁸ is located within the same LD-region as the two new 8q24 SNPs and rs445114 is somewhat correlated with it ($D' = 0.76$, $r^2 = 0.44$; Supplementary Table 2), while rs16902094 is much less correlated with it ($D' = 0.61$, $r^2 = 0.06$; Supplementary Table 2). However, both rs16902094 and rs445114 show very little correlation with any of the previously published prostate^{1,2,7,9}, colon¹⁰⁻¹², or bladder cancer¹³ risk variants on 8q24 ($D' = 0.6$ and $r^2 = 0.13$; Supplementary Table 2 and Supplementary Fig. 1). The results in Iceland for rs16902094, rs445114 and the three previously published prostate cancer risk variants on 8q24, remain significant after being adjusted for each other (Supplementary Table 3b). Hence, rs16902094 and rs445114 should be included in the list of variants at 8q24 associated with prostate cancer risk. The five 8q24 variants conferring risk of prostate cancer are distributed between four LD-regions, spanning approximately 480 kb. One of the five prostate cancer risk variants, rs6983267, has been shown to also affect the risk of colorectal cancer. Other cancer risk variants on 8q24.21, conferring risk of breast or bladder cancer, have not been shown to predispose to prostate cancer^{13,14}.

We proceeded to genotype all the four newly discovered SNPs (rs10934853 on 3q21.3, rs8102476 on 19q13.2 and rs16902094 and rs16902104 both on 8q24.21) in at least two out of five prostate cancer study groups (deCODE follow up groups) of European descent. These groups come from The Netherlands, Spain, Finland and the United States (US) (see Supplementary Note). When results for SNPs successfully genotyped in these groups were combined with the Icelandic and CGEMS data discussed above, they were significant for all loci, surpassing the threshold of genome-wide significance set by many at $P > 10^{-7}$. By examining the literature we found that Duggan et al.¹⁵ had published data (OR = 1.21, $P = 0.045$) for rs10934853, on 3q21.3, from a study on aggressive prostate cancer in the CAPS study population from Sweden. Thereby, adding further significance to the combined analysis of the 3q21.3 locus (see Table 1 for combined OR and P -values).

When inspected, a test of heterogeneity in the OR for all variants and all study groups showed a nominally significant heterogeneity ($P_{het} = 0.039$) for the 3q21.3 locus, no significant difference was observed for the other three loci ($P_{het} > 0.1$). For rs10934853 on 3q21.3 the estimated OR tended to be greater for the combined US study groups (OR = 1.21 and the 95% CI: 1.13, 1.28) than for the combined European study groups (OR = 1.08 and the 95% CI: 1.04, 1.13). This observation, while interesting, needs to be further confirmed.

A prostate cancer risk variant, with an allelic frequency of about 0.5 and located on 11q13, was reported previously by two groups independently; Thomas et al.⁵ reported rs10896449, and Eeles et al.⁴ reported rs7931342 (the two SNPs are highly correlated with $D' = 1$ and $r^2 = 0.97$ according to Utah CEPH (CEU) HapMap data). In the Icelandic GWAS data set, the strongest association with prostate cancer on 11q13 was observed for allele G of rs10896450

(OR = 1.13, $P = 2.5 \times 10^{-4}$; Table 2), a SNP highly correlated with the two previously reported SNPs ($D' = 1$; $r^2 > 0.97$ according to CEU HapMap data). In order to investigate further the 11q13 locus, we selected six SNPs not present on the Illumina Hap300 chip but moderately correlated with each of the three anchor SNPs: rs10896449, rs7931342, and rs10896450 ($D' = 0.70$ and $r^2 > 0.24$ according to CEU HapMap data; Supplementary Table 1). Furthermore, these SNPs are not strongly correlated with any SNP on the Illumina Hap300 chip ($r^2 < 0.8$). By doing this, we attempted to find variants with greater risk and lower allelic frequency. The additional six SNPs were genotyped in 1,809 and 783 Icelandic cases and controls, respectively. The control individuals were randomly selected from the Icelandic population and were not known to have prostate cancer according to the nationwide Icelandic Cancer Registry (see URL below). Allele A of the refinement SNP rs11228565 was found to confer greater risk of the disease than the anchor SNP (rs10896450), with an OR of 1.24 ($P = 0.0057$) and a control frequency of 0.177 in the Icelandic study group (Table 1 and Supplementary Table 4). We then tested the two SNPs, rs10896450 and rs11228565, in the five deCODE follow up groups. For the Finnish study group, instead of genotyping rs10896450, we used data available for the highly correlated SNP rs7931342 ($D' = 1$; $r^2 > 0.97$ according to CEU HapMap data), previously reported by Kote-Jarai et al¹⁶. Combination of the results from Iceland and the deCODE follow up groups gave an estimated OR of 1.15 ($P = 2.6 \times 10^{-8}$) for rs10896450-G whereas for rs11228565-A the OR was estimated to be 1.23 ($P = 6.7 \times 10^{-12}$) (Table 1 and Supplementary Table 5). The estimated frequency of the risk allele of rs11228565 in the combined data set of the groups used in this study is 0.20 compared to 0.50 for the G allele of rs10896450 (Supplementary Table 5). Hence, the frequency of the risk allele for the new SNP is lower than for the previously reported one. Also, the point estimate of the OR is greater (OR = 1.23 for rs11228565 vs. 1.15 for rs10896450) although the difference is not significant, ($P = 0.08$ using results for all study groups; Supplementary Table 5). The results for rs11228565 were strongest for the study groups from Nashville and Finland but weakest for the Spanish group. However, a test of heterogeneity in the OR of all study groups showed no significant difference for the 11q13 locus ($P = 0.26$). When adjusting the results for the refinement SNP (rs11228565-A) using data the anchor SNPs, (rs10896450-G or rs7931342-G), it remained significant (OR = 1.19, $P = 5.0 \times 10^{-7}$) while the anchor SNPs were not significant after being adjusted for the refinement SNP (OR = 1.03, $P = 0.15$; Supplementary Table 5).

For the six study groups of European descent (excluding the five CGEMS groups and the study by Duggan et al¹⁵), where we had information about age at diagnosis, no age effect was seen for any of the five loci discussed above ($P = 0.3$). By computing the genotype specific ORs or inspecting the genotypic ORs from the public CGEMS data set for the variants in Table 1, we found that the multiplicative model provides an adequate fit for all five loci in the study groups analyzed (for the full- vs. the multiplicative model all $P > 0.1$; Supplementary Table 6 and Online Methods). Comparing the patients with a more aggressive phenotype (Gleason ≥ 7 and/or T3 or higher and/or node positive and/or metastatic disease) to the group with less aggressive tumors (Gleason < 7 and T2 or lower) showed no difference; an OR between 0.94 and 1.03 ($P > 0.1$) was observed for the five loci. The public CGEMS data also showed no difference between patients with more or less aggressive disease ($P > 0.40$) for the SNPs where data are available (3q21, 8q24, and 19q13.2). The study published by Duggan et al¹⁵ was on aggressive prostate cancer only and a comparison with less aggressive disease was therefore not possible.

The four novel loci reported here, add to the rapidly increasing number of prostate cancer susceptibility variants, identified through GWASs. In Table 2, we provide results from the Icelandic GWAS for risk variants reported by us and/or others (until February 2009) to confer risk of prostate cancer. For some of the SNP associations reported in other populations, the Icelandic results provide replication (in particular 7p15 and 10q11), while

we failed to replicate association for other SNPs in this population (in particular 9q33 and 22q13). In order to summarize the overall effect of the variants in Tables 1 and 2 we combined the effect of all variants affecting risk of prostate cancer in the Icelandic population. We performed a multi-variant analysis, using the multiplicative model for 22 risk variants. Testing the assumptions, no significant deviation from the multiplicative model was observed for any given variant as well as no interaction between variants (see Online Methods). Based on this analysis, the estimated risk is more than 2.5-fold greater for the top 1.3% of the risk distribution (including 0.3% of the population with a risk greater than 3), using the population average risk as a reference (Table 3). For these individuals this corresponds to a lifetime risk of over 30% of being diagnosed with prostate cancer, compared with a risk of 12%, on average in Iceland, of getting the disease before age 75, according to NORDCAN (see URL below) assuming no interaction between the effect of the variants and age at diagnosis. The combined risk estimates presented here are similar to those previously reported by Kote-Jarai et al. for 15 risk variants¹⁶. We note that the estimates provided here are based solely on our Icelandic study population, and that more accurate estimates could be obtained from large prospective studies. Also, given the fast pace of discoveries in the current era of GWASs, more variants associated with prostate cancer risk are likely to be discovered, which suggests the need for constant updating of such multivariate risk models.

Online Methods

Genotyping Methods

Illumina genotyping—1,968 and 35,382 Icelandic case- and control-samples respectively, were successfully assayed with the Infinium HumanHap300 SNP chip (Illumina, San Diego, CA, USA), containing 317,503 haplotype tagging SNPs derived from phase I of the International HapMap project. Of the SNPs assayed on the chip, 2,906 SNPs had a yield lower than 95%, 271 SNPs had a minor allele frequency, in the combined set of cases and controls, below 0.01 or were monomorphic. An additional 4,632 SNPs showed a significant distortion from Hardy-Weinberg equilibrium in the controls ($P < 1.0 \times 10^{-3}$). In total, 6,983 unique SNPs were removed from the study. Thus, the analysis reported in the main text utilizes 310,520 SNPs. Any samples with a call rate below 98% were excluded from the analysis.

Replication genotyping—Single SNP genotyping of the SNPs reported in the main text for the four case-control groups from Iceland, The Netherlands, Spain and Chicago was carried out by deCODE Genetics in Reykjavik, Iceland, applying the Centaurus²⁰ (Nanogen) platform. The quality of each Centaurus SNP assay was evaluated by genotyping each assay in the CEU and/or YRI HapMap samples and comparing the results with the HapMap publicly released data. Assays with >1.5% mismatch rate were not used and a linkage disequilibrium (LD) test was used for markers known to be in LD. We re-genotyped more than 10% of the samples and observed a mismatch rate lower than 0.5%. Genotyping of samples from Finland and Nashville was done using the same Centaurus assays as used in Iceland at the University of Tampere and Vanderbilt University, respectively, using standard protocols.

For each of the SNPs discussed in the main text, the yield was higher than 95% for those samples which genotyping was attempted for in every study group. The SNPs rs16902094 on 8q24 and rs11228565 on 11q13 are not present on the Human Hap300 chip. Therefore, using a single SNP assay for genotyping, an attempt was made to genotype 6,900 and 800 individuals, respectively, of the 35,382 Icelandic controls as well as 1,860 Icelandic cases and all available individuals from the replication study groups.

Discovery of new SNP on 8q24 by Solexa re-sequencing—In order to search for new SNPs on 8q24, a 527 kb region (128113108 – 128640337 bp, Build 36) was sequenced using the Solexa re-sequencing platform (Illumina Inc.). From our set of ~2,000 cases, 800 were selected randomly and split into two DNA-pools, each with 400 samples. Similarly, 800 control individuals, not known to have prostate cancer, were selected randomly and split into two DNA-pools. Dilutions were prepared in duplicates for the 4 pools and used for long-range PCR reactions (each amplicon consisted of about 10 kb). PCR fragments were run on 0.8% agarose gels and the DNA visualized with BlueView (Sigma Inc.) under normal light and their sizes estimated with HindIII digested lambda size marker (Fermentas Inc). Bands of correct sizes were excised out of the agarose gels and purified with Qiagen gel extraction kit (Qiagen Inc.). The PCR products were quantified by picogreen assay (Invitrogen Inc.) as described by the manufacturer. The preparation of the Solexa DNA libraries, the cluster generation and DNA sequencing was done as described by Bentley et al²¹. The SNP Analysis pipeline is composed of four components: Alignment, SNP calling, Filtering and Association analysis. Promising SNPs were selected for further study/confirmation using Centaurus single track SNP assays.

Statistical analysis

Association analysis—For SNPs that were in strong LD, whenever the genotype of one SNP was missing for an individual, the genotype of the correlated SNP was used to provide partial information through a likelihood approach as previously described⁹. This ensured that results presented in Supplementary Table 5 were based on the same number of individuals, allowing meaningful comparisons of results for correlated SNPs. A likelihood procedure described in a previous publication²² and implemented in the NEMO software was used for the association analyses.

We tested the association of an allele to prostate cancer using a standard likelihood ratio statistic that, if the subjects were unrelated, would have asymptotically a χ^2 distribution with one degree of freedom under the null hypothesis. Allelic frequencies rather than carrier frequencies are presented for the markers in the main text. Allele-specific ORs and associated *P* values were calculated assuming a multiplicative model for the two chromosomes of an individual²³. Results from multiple case-control groups were combined using a Mantel-Haenszel model²⁴ in which the groups were allowed to have different population frequencies for alleles, haplotypes and genotypes but were assumed to have common relative risks (see Gudmundsson *et al.*³ for a more detailed description of the association analysis).

The control groups from Iceland, The Netherlands, Spain, and Finland include both male and female controls. No significant difference between male and female controls was detected for SNPs presented in Table 1 for each of these four groups. Controls from other study groups include only males.

In order to assess the association for the SNP rs4962416 on 10q26, which is in the CEU section of the Hapmap database but absent from the Illumina Hap300 chip, we use a method based on haplotypes of two markers (rs7077275 and rs893856) present on the chip. We used a method we have previously employed²⁵, that is an extension of the two-marker haplotype tagging method²⁶ and is similar in spirit to two other proposed methods^{27,28}. We computed associations with a linear combination of the different haplotypes chosen to act as surrogates to HapMap markers in the regions. These calculations were based on 1,724 prostate cancer cases and 35,322 controls genotyped on chip.

Multivariate analysis—In a multi-variant analysis, we combined the effect of 22 variants affecting risk of prostate cancer using estimates based on data from only the Icelandic study

group. A multiplicative model was assumed at each variant and between all variants. For the 21 autosomal variants we tested for deviation from the multiplicative model comparing it to the full model of genotypic odds ratio (OR). No significant (given the number of tests performed) deviation from the multiplicative model was detected (all P-values > 0.0024, corresponding to 0.05 divided by 21). For the variant located on the X chromosome (rs5945572), there exist only two male genotypes (carrier and no-carrier). Also, we tested the pair-wise interactions between the 22 risk variants using logistic regressions including terms corresponding to the 231 possible pairs. No significant (given the number of tests performed) deviation from the multiplicative model was detected with a level of significance below 0.00022 (corresponding to $P = 0.05$ divided by 231). Similarly, the absence of interaction between variants has previously been reported by others^{5,16,29}. Odds ratios were calculated for all possible genotype combinations based on these 22 variants and expressed relative to the average general population risk. The combined OR estimates were then divided into OR-ranges and presented along with the percentage of the population within each OR-range (See Table 3 in the main text). The general population risk was determined using a frequency-weighted average risk for all possible genotypes.

The Icelandic samples were part of the initial discovery study populations for 10 of these 22 variants, therefore the estimates for these may be inflated due to winners curse. From Table 1 we are using the estimates for the following variants: rs10934853 on 3q21.3, rs11228565 on 11q13 and rs8102476 on 19q13.2. For the five variants (rs1447295, rs6983267, rs16901979, rs16902094 and rs445114) on 8q24.21 we are using the adjusted estimates as reported in Supplementary Table 3b. From Table 2, we use the estimates for all variants except rs1571801 on 9q33 since its effect was in the opposite direction compared to the original publication, and rs10896450 on 11q13 for which data for the refinement SNP (rs11228565) in Table 1 was used. Furthermore, for the two variants on 17q12, reported in Table 2, we are using the estimates adjusted for each other since the two markers are located ~25 kb to each other although not correlated ($D' = 0.03$, $r^2 = 0.0004$ according to CEU HapMap data); rs4430796 has an adjusted OR of 1.17 and rs11649743 has an adjusted OR of 1.06.

We apply the combined genetic relative risk to the lifetime risk by multiplying them together. Since the association with the disease for these variants has not been shown to depend on age at diagnosis^{3,16}, we assume that the individual estimates for each variant will have a similar effect at any given age. The lifetime risk of getting prostate cancer is estimated to be 12%, on average in Iceland, before the age of 75, according to NORDCAN (see URL).

Analysis of the CGEMS data—For the five individual study populations from the CGEMS study^{5,7} (ACS = American Cancer Society Prevention Study II (US); ATBC = Alpha-Tocopherol, Beta-Carotene Prevention Study (Finland); FPCC: CeRePP French Prostate Case-Control Study (France); HPFS = Health Professionals Follow-up Study (US); PLCO = Prostate, Lung, Colon, Ovarian Trial (US)), when assessing the allelic effect we used the pre-computed data (released in spring, 2008) corresponding to “All case versus control (dichotomous), genotype trend effect model, adjusted”. When assessing the genotypic effect at each loci for the CGEMS study we used the pre-computed “All case versus control (dichotomous), genotype-specific effect model, adjusted, ALL (ACS, HPFS, FPCC, ATBC, PLCO)”.

Correction for relatedness—Some individuals in the Icelandic case-control groups were related to each other, causing the aforementioned χ^2 test statistic to have a mean >1. We estimated the inflation factor by using a previously described procedure³⁰ in which we simulated genotypes through the genealogy of the 37,350 Icelanders analyzed in the present

study (number of simulations = 100,000). The inflation factor was estimated to be 1.10. Results from the Icelandic samples presented in the main text are based on adjusting the χ^2 statistics by dividing each of them by 1.10.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the individuals who participated in the study and whose contribution made this work possible. This project was funded in part by contract number 202059 (PROMARK) from the 7th Framework Program of the European Union to deCODE genetics (T.R., and L.A.K.), in part by FP7-MC-IAPP Grant agreement no.: 218071 (CancerGene) to deCODE genetics, in part by a V Foundation award and US Department of Veterans Affairs grants to J.R.S, and in part by Academy of Finland, Sigrid Juselius Foundation, Finnish Cancer Organisations and the Competitive Research Funding of the Pirkanmaa Hospital District, Tampere University Hospital to J.S.

References

1. Gudmundsson J, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet.* 2007; 39:631–7. [PubMed: 17401366]
2. Haiman CA, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet.* 2007; 39:638–44. [PubMed: 17401364]
3. Gudmundsson J, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet.* 2007; 39:977–83. [PubMed: 17603485]
4. Eeles RA, et al. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet.* 2008; 40:316–21. [PubMed: 18264097]
5. Thomas G, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet.* 2008; 40:310–5. [PubMed: 18264096]
6. Gudmundsson J, et al. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet.* 2008; 40:281–3. [PubMed: 18264098]
7. Yeager M, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet.* 2007; 39:645–9. [PubMed: 17401363]
8. Easton DF, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* 2007; 447:1087–93. [PubMed: 17529967]
9. Amundadottir LT, et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet.* 2006; 38:652–8. [PubMed: 16682969]
10. Tomlinson I, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet.* 2007; 39:984–8. [PubMed: 17618284]
11. Zanke BW, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet.* 2007; 39:989–94. [PubMed: 17618283]
12. Haiman CA, et al. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet.* 2007; 39:954–6. [PubMed: 17618282]
13. Kiemeny LA, et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet.* 2008; 40:1307–12. [PubMed: 18794855]
14. Ghousaini M, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. *J Natl Cancer Inst.* 2008; 100:962–6. [PubMed: 18577746]
15. Duggan D, et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst.* 2007; 99:1836–44. [PubMed: 18073375]
16. Kote-Jarai Z, et al. Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:2052–61. [PubMed: 18708398]

17. Rafnar T, et al. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat Genet.* 2009; 41:221–7. [PubMed: 19151717]
18. Sun J, et al. Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. *Nat Genet.* 2008; 40:1153–5. [PubMed: 18758462]
19. Sun J, et al. Sequence variants at 22q13 are associated with prostate cancer risk. *Cancer Res.* 2009; 69:10–5. [PubMed: 19117981]
20. Kutuyavin IV, et al. A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Research.* 2006; 34:e128. [PubMed: 17012270]
21. Bentley DR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 2008; 456:53–9. [PubMed: 18987734]
22. Gretarsdottir S, et al. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet.* 2003; 35:131–8. [PubMed: 14517540]
23. Falk CT, Rubinstein P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet.* 1987; 51(Pt 3):227–33. [PubMed: 3500674]
24. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959; 22:719–48. [PubMed: 13655060]
25. Styrkarsdottir U, et al. Multiple genetic loci for bone mineral density and fractures. *N Engl J Med.* 2008; 358:2355–65. [PubMed: 18445777]
26. Pe'er I, et al. Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet.* 2006; 38:663–7. [PubMed: 16715096]
27. Nicolae DL. Testing untyped alleles (TUNA)-applications to genome-wide association studies. *Genet Epidemiol.* 2006; 30:718–27. [PubMed: 16986160]
28. Zaitlen N, Kang HM, Eskin E, Halperin E. Leveraging the HapMap correlation structure in association studies. *Am J Hum Genet.* 2007; 80:683–91. [PubMed: 17357074]
29. Zheng SL, et al. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med.* 2008; 358:910–9. [PubMed: 18199855]
30. Stefansson H, et al. A common inversion under selection in Europeans. *Nat Genet.* 2005; 37:129–37. [PubMed: 15654335]

Table 1
Summary association results for the SNPs on 3q21.3, 8q24, 19q13.2 and 11q13

Study population	Cases (N)	Controls (N)	Frequency		OR (95% CI)	P-value
			Cases	Controls		
A. Results for rs10934853 [A] or rs4857841 [A] on 3q21.3						
Iceland ^d	1,968	35,227	0.295	0.269	1.14 (1.06, 1.22)	3.2E-04
Chicago, Illinois	1,077	1,003	0.313	0.273	1.21 (1.06, 1.39)	4.4E-03
Finland	2,638	1,716	0.330	0.319	1.05 (0.96, 1.15)	0.27
The Netherlands	1,084	1,827	0.306	0.286	1.10 (0.98, 1.24)	0.10
Nashville, Tennessee	596	687	0.283	0.270	1.07 (0.90, 1.27)	0.47
Spain	811	1,605	0.306	0.314	0.96 (0.84, 1.09)	0.54
ACS ^b	1,758	1,775	0.300	0.258	1.25 (1.12, 1.39)	4.3E-05
ATBC ^b	928	921	0.309	0.319	0.96 (0.84, 1.10)	0.59
FPCC ^b	654	657	0.291	0.272	1.09 (0.92, 1.29)	0.34
HPFS ^b	595	609	0.313	0.278	1.18 (0.99, 1.40)	0.070
PLCO ^b	1,167	1,093	0.308	0.266	1.23 (1.08, 1.41)	2.5E-03
CAPS ^c	498	494	0.329	0.288	1.21 (1.00, 1.46)	0.045
All combined^d	13,774	47,614	-	0.284	1.12 (1.08, 1.16)	2.9E-10
P_{het}					0.039	
I²					46.4	
B. Results for rs16902094 [G] or rs16902104 [T] on 8q24						
Iceland ^d	1,858	6,853	0.168	0.136	1.28 (1.15, 1.41)	3.5E-06
Chicago, Illinois	797	758	0.166	0.147	1.16 (0.95, 1.41)	0.14
Finland	2,197	1,725	0.248	0.222	1.15 (1.03, 1.28)	9.9E-03
The Netherlands	831	837	0.161	0.138	1.20 (0.99, 1.44)	0.066
Nashville, Tennessee	669	733	0.170	0.130	1.37 (1.11, 1.69)	3.0E-03
Spain	643	952	0.162	0.137	1.21 (1.00, 1.48)	0.055
ACS ^b	1,759	1,774	0.156	0.132	1.22 (1.06, 1.39)	4.3E-03

Study population	Cases (N)	Controls (N)	Frequency		OR (95% CI)	P-value
			Cases	Controls		
ATBC ^b	929	920	0.255	0.193	1.43 (1.22, 1.67)	1.0E-05
FPCC ^b	656	657	0.152	0.145	1.06 (0.85, 1.31)	0.61
HPFS ^b	596	611	0.127	0.133	0.93 (0.74, 1.18)	0.57
PLCO ^b	1,167	1,093	0.145	0.137	1.09 (0.92, 1.30)	0.31
All combined^d	12,102	16,913	-	0.150	1.21 (1.15, 1.26)	6.2E-15
P_{het}					0.14	
I²					32.9	
C. Results for rs445114 [T] on 8q24						
Iceland ^d	1,727	35,382	0.710	0.672	1.20 (1.11, 1.29)	5.0E-06
The Netherlands	910	1,832	0.676	0.650	1.13 (1.00, 1.27)	0.048
Spain	490	1,387	0.660	0.624	1.17 (1.01, 1.36)	0.041
ACS	1,757	1,768	0.651	0.618	1.15 (1.05, 1.27)	4.0E-03
ATBC	925	919	0.702	0.661	1.22 (1.06, 1.40)	6.6E-03
FPCC	655	655	0.647	0.635	1.05 (0.90, 1.24)	0.52
HPFS	595	608	0.613	0.633	0.91 (0.77, 1.07)	0.26
PLCO	1,175	1,100	0.641	0.618	1.13 (1.00, 1.28)	5.7E-02
All combined^d	8,234	43,651	-	0.639	1.14 (1.10, 1.19)	4.7E-10
P_{het}					0.13	
I²					37.0	
D. Results for rs8102476 [C] on 19q13.2						
Iceland ^d	1,941	35,330	0.517	0.495	1.09 (1.03, 1.17)	6.4E-03
Chicago, Illinois	1,086	1,172	0.612	0.579	1.15 (1.02, 1.29)	0.024
Finland	2,629	1,739	0.481	0.435	1.21 (1.11, 1.31)	2.1E-05
The Netherlands	1,086	1,830	0.567	0.528	1.17 (1.05, 1.30)	4.2E-03
Nashville, Tennessee	596	689	0.565	0.553	1.05 (0.90, 1.23)	0.55
Spain	728	1,389	0.641	0.619	1.10 (0.96, 1.25)	0.16

Study population	Cases (N)	Controls (N)	Frequency		OR (95% CI)	P-value
			Cases	Controls		
ACS	1,755	1,766	0.574	0.551	1.10 (1.00, 1.21)	0.043
ATBC	926	919	0.473	0.461	1.05 (0.93, 1.20)	0.43
FPCC	656	655	0.607	0.563	1.19 (1.02, 1.40)	0.027
HPFS	595	609	0.574	0.57	1.03 (0.88, 1.21)	0.74
PLCO	1,175	1,100	0.571	0.545	1.10 (0.98, 1.24)	0.11
All combined^d	13,173	47,198	-	0.536	1.12 (1.08, 1.15)	1.6E-11
P_{het}					0.63	
I²					0.0	
E. Results for rs11228565 [A] on 11q13.						
Iceland ^a	1809	783	0.210	0.177	1.24 (1.06, 1.43)	5.7E-03
Chicago, Illinois	755	878	0.235	0.210	1.16 (0.98, 1.37)	8.0E-02
Finland	2643	1689	0.210	0.169	1.30 (1.16, 1.45)	3.2E-06
The Netherlands	992	1781	0.229	0.202	1.17 (1.02, 1.34)	0.021
Nashville, Tennessee	592	685	0.291	0.223	1.43 (1.20, 1.71)	8.5E-05
Spain	394	1399	0.240	0.224	1.09 (0.91, 1.30)	0.34
All combined^d	7185	7215	-	0.201	1.23 (1.16, 1.31)	6.7E-12
P_{het}					0.26	
I²					23.3	

All P values shown are two-sided. Shown are the corresponding numbers of cases and controls (N), allelic frequencies of variants in affected and control individuals, the allelic odds-ratio (OR) with 95% confidence interval (95% CI) and P value. Also shown, are the P-values for the heterogeneity of the ORs (P_{het}) for all study groups as well as I² which lies between 0% and 100% and describes the proportion of total variation in study estimates that is due to heterogeneity.

^aResults presented for Iceland were adjusted for relatedness (see Online Methods).

^bThe results for the five CGEMS groups on 3q21.3 and 8q24 are for the SNPs rs4857841 [A] and rs16902104[T], which are highly correlated with rs10934853[A] and rs169020948[G], respectively (D' and r² > 0.96 according to Icelandic and CEU HapMap data).

^cResults for the Swedish CAPS study group are for rs10934853[A] published by Duggan et al.¹⁵

^dFor the combined study populations, the reported control frequency was the average, unweighted control frequency of the individual populations, while the OR and the P value were estimated using the Mantel-Haenszel model.

Table 2
Association results from a GWAS in Iceland for variants reported to confer risk of prostate cancer identified through GWAS

Marker, [risk allele] and (correlated marker(s)) ^a	Locus	Cases (N)	Controls (N)	Frequency		OR (95% CI)	P-value
				Cases	Controls		
rs2710646 [A], (rs721048) ⁶	2p15	1,882	35,145	0.224	0.203	1.14 (1.05, 1.24)	2.5×10 ⁻³
rs2660753 [T] ⁴	3p12	1,725	35,362	0.110	0.100	1.11 (0.99, 1.25)	0.075
rs401681 [C] ¹⁷	5p15	1,962	35,400	0.562	0.547	1.07 (1.00, 1.14)	0.066
rs9364554 [T] ⁴	6q25	1,725	35,399	0.322	0.309	1.06 (0.99, 1.15)	0.11
rs10486567 [G] ⁵	7p15	1,725	35,392	0.787	0.765	1.13 (1.04, 1.24)	4.4×10 ⁻³
rs6465657 [C] ⁴	7q21	1,724	35,358	0.432	0.421	1.04 (0.97, 1.12)	0.26
rs1447295 [A] ⁹	8q24 (1)	1,821	35,470	0.165	0.111	1.58 (1.43, 1.74)	2.2×10 ⁻¹⁹
rs16901979 [A] ¹	8q24 (2)	1,726	35,403	0.073	0.042	1.80 (1.55, 2.09)	2.5×10 ⁻¹⁴
rs6983267 [G] ⁷	8q24 (3)	1,724	35,367	0.581	0.551	1.13 (1.05, 1.22)	7.5×10 ⁻⁴
rs1571801 [A] ¹⁵	9q33 ^b	1,721	35,303	0.261	0.276	0.93 (0.85, 1.01)	0.068
rs10993994 [T] ^{4,5}	10q11	1,727	35,397	0.410	0.384	1.11 (1.04, 1.20)	3.7×10 ⁻³
rs4962416 [C] ⁵	10q26 ^c	1,724	35,322	0.223	0.221	1.02 (0.94, 1.11)	0.68
rs10896450 [G], (rs10896449) ⁵ , rs793 1342 ⁴	11q13	1,951	35,394	0.501	0.469	1.13 (1.06, 1.21)	2.5×10 ⁻⁴
rs4430796 [A] ³	17q12	1,726	35,397	0.559	0.517	1.19 (1.10, 1.28)	8.3×10 ⁻⁶
rs11649743 [G] ¹⁸	17q12	1,747	35,405	0.812	0.799	1.09 (0.99, 1.19)	0.066
rs1859962 [G] ³	17q24.3	1,746	35,124	0.493	0.455	1.16 (1.08, 1.25)	3.7×10 ⁻⁵
rs2735839 [G] ⁴	19q13.33	1,726	35,376	0.879	0.865	1.14 (1.02, 1.27)	0.021
rs9623117 [C] ⁹	22q13 ^b	1,724	35,389	0.208	0.208	1.00 (0.91, 1.10)	0.99
rs5945572 [A] ⁶ , (rs5945 619) ⁶	Xp11	1,899	35,384	0.416	0.369	1.22 (1.11, 1.34)	6.1×10 ⁻⁵

^aShown in the table are GWAS from Iceland for variants that have been identified through GWAS results (published up to February 2009) and the original publication(s). Highly correlated markers are shown in parenthesis as well as the study reporting them. All P values are two-sided. Shown are the corresponding numbers of cases and controls (N), allelic frequencies of variants in affected and control individuals, the allelic odds-ratio (OR) with 95% confidence interval (95% CI) and P value adjusted for relatedness.

^bThe original results published for the loci on 9q33¹⁵ and 22q13¹⁹ were from a study on cases with aggressive prostate cancer. Results for these two loci in Icelandic cases (N = 693) with more aggressive prostate cancer (Gleason score >6 and/or T3 or higher and/or node positive and/or metastatic disease), using the same set of controls, were not significant (rs1571801; OR_{aggr} = 0.90 and P = 0.080; rs9623117; OR_{aggr} = 1.00 and P = 0.94).

^cThe SNP marker, rs4962416, at the 10q26 locus is not on the Illumina Hap300 chip, results shown for it are based on a weighted combination of two marker haplotypes generated from rs7077275 and rs893856 that are present on the chip and tag the SNP (rs4962416).

Table 3
Population distribution in Iceland of ORs for 22 prostate cancer susceptibility variants

OR-range	Population percentage
< 0.5	9.5%
0.5–0.75	25.2%
0.75–1	24.7%
1–1.5	27.6%
1.5–2	9.1%
2–2.5	2.7%
> 2.5	1.3%

Results from a multi-variant risk model analysis for prostate cancer in Iceland based on susceptibility variants in tables 1 and 2. Results from Iceland were used for all variants in table 1 and 2, except rs1571801 on 9q33 since its effect was in the opposite direction, and rs10896450 on 11q13 for which data for the refinement SNP in table 1 was used. Odds ratios (OR) were calculated for all possible genotype combinations based on 22 variants and expressed relative to the average general population risk, assuming the multiplicative model between variants. The combined OR estimates were then divided into OR-ranges and presented along with the percentage of the population within each OR-range. The general population risk was determined using a frequency-weighted average risk for all possible genotypes.