Germline variation at 8q24 and prostate cancer risk in men of European ancestry

Gerald Andriole
Aleksandra Klim
Bettina F. Drake

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs
Germline variation at 8q24 and prostate cancer risk in men of European ancestry

Marco Matejcic, Edward J. Saunders et al.†

Chromosome 8q24 is a susceptibility locus for multiple cancers, including prostate cancer. Here we combine genetic data across the 8q24 susceptibility region from 71,535 prostate cancer cases and 52,935 controls of European ancestry to define the overall contribution of germline variation at 8q24 to prostate cancer risk. We identify 12 independent risk signals for prostate cancer \( (p < 4.28 \times 10^{-15}) \), including three risk variants that have yet to be reported. From a polygenic risk score (PRS) model, derived to assess the cumulative effect of risk variants at 8q24, men in the top 1% of the PRS have a 4-fold (95%CI = 3.62–4.40) greater risk compared to the population average. These 12 variants account for ~25% of what can be currently explained of the familial risk of prostate cancer by known genetic risk factors. These findings highlight the overwhelming contribution of germline variation at 8q24 on prostate cancer risk which has implications for population risk stratification.

Correspondence and requests for materials should be addressed to C.A.H. (email: Christopher.Haiman@med.usc.edu). †A full list of authors and their affiliations appears at the end of the paper.
Prostate cancer (PCa) is the most common cancer among men in the US, with 161,360 new cases and 26,730 related deaths estimated in 2017. Familial and epidemiological studies have provided evidence of substantial heritability of PCa, and ~170 common risk loci have been identified through genome-wide association studies (GWAS). The susceptibility region on chromosome 8q24 has been shown to be a major contributor to PCa risk, with multiple variants clustered in five linkage disequilibrium (LD) blocks spanning ~600 Mb that are independently associated with risk. Many of these association signals reported at 8q24 have been replicated across racial/ethnic populations, pointing to common shared functional variants within 8q24. However, rare ancestry-specific variants have also been detected, which confer larger relative risks of PCa (odds ratios [ORs] >2.0) than common risk variants in the region and signify allelic heterogeneity in the contribution of germline variation at 8q24 to PCa risk across populations.

In the current study, we perform a comprehensive investigation of genetic variation across the 1.4 Mb cancer susceptibility region at 8q24 (127.6–129.0 Mb) in relation to PCa risk. We combine genotyped and imputed data from two large GWAS consortia (PRACTICAL/ELLIPSE OncoArray and iCOGS) including >124,000 individuals of European ancestry to search for novel risk variants, as well as to determine the overall contribution of genetic variation at 8q24 to PCa heritability. Our findings underscore the sizable impact of genetic variation in the 8q24 region in explaining inter-individual differences in PCa risk, with potential clinical utility for genetic risk prediction.

**Results**

**Marginal and conditional association analysis.** Genotype data from the Illumina OncoArray and iCOGS array and imputation to 1000 Genomes Project (1KGP) were generated among 71,535 PCa cases and 52,935 controls of European ancestry from 86 case-control studies (see Methods). Of the 5600 genotyped and imputed variants at 8q24 (127.6–129.0 Mb) with minor allele frequency (MAF) >0.1% retained for analysis (see Methods), we identified 12 variants with conditional p-values from the Wald test between $2.93 \times 10^{-37}$ and $4.28 \times 10^{-15}$ (Table 1). None of the other variants were statistically significant at $p < 5 \times 10^{-8}$ after adjustment for the 12 independent hits (Fig. 1). The 8q24 region is shown in Supplementary Fig. 1. Of these 12 stepwise signals, three had alleles with extreme risk allele frequencies (RAFs) that conveyed large effects (rs77541621, RAF = 2%, OR = 1.85, 95%CI = 1.76–1.94; rs183373024, RAF = 1%, OR = 2.67, 95%CI = 2.43–2.93; rs190257175, RAF = 99%, OR = 1.60, 95%CI = 1.42–1.80). The remaining variants had RAfs between 0.11 and 0.92 and conditional ORs that were more modest and ranged from 1.10 to 1.37 (Table 1). For 8 of the 12 variants, the allele found to be positively associated with PCa risk was the predominant allele (i.e., >50% in frequency). For two variants, rs78511380 and rs190257175, the marginal associations were not genome-wide significant and substantially weaker than those in the conditional model. For rs78511380, the marginal OR was slightly protective (OR = 0.97; $p = 0.027$), but reversed direction and was highly statistically significant when conditioning on the other 11 variants (OR = 1.19; $p = 3.5 \times 10^{-18}$; Table 1).

**Haplotype analysis.** The haplotype analysis showed an additive effect of the 12 independent risk variants consistent with that predicted in the single variant test; co-occurrence of the 8q24 risk alleles on the same haplotype does not further increase the risk of PCa (Supplementary Table 1). The unique haplotype carrying the reference allele for rs190257175 (GCTTAT, 0.5% frequency) is also the sole haplotype associated with a reduced risk of PCa, suggesting that having the C allele confers a protective effect. The reference allele for rs78511380 (A, 8% frequency) occurs on a haplotype in block 2 together with the risk alleles for rs190257175, rs72725879 and rs5013678 (haplotype GTTTAA, 8%) which obscures the positive association with the T allele of rs78511380. Thus, the marginal protective effect associated with the risk allele for rs78511380 reflects an increased risk associated with the occurrence on a risk haplotype with other risk alleles (Supplementary Table 1).

<table>
<thead>
<tr>
<th>Variant ID</th>
<th>Position</th>
<th>Allele</th>
<th>RAF</th>
<th>LD cluster</th>
<th>Conditional association</th>
<th>Marginal association</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1914295</td>
<td>127910317</td>
<td>T/C</td>
<td>0.68</td>
<td>block 1</td>
<td>OR (95%CI): 1.10 (1.08–1.12)</td>
<td>OR (95%CI): 1.09 (1.07–1.11)</td>
</tr>
<tr>
<td>rs1487240</td>
<td>128021752</td>
<td>A/G</td>
<td>0.74</td>
<td>block 1</td>
<td>OR (95%CI): 1.20 (1.17–1.22)</td>
<td>OR (95%CI): 1.16 (1.14–1.18)</td>
</tr>
<tr>
<td>rs77541621</td>
<td>128077146</td>
<td>A/G</td>
<td>0.02</td>
<td>block 2</td>
<td>OR (95%CI): 1.85 (1.76–1.94)</td>
<td>OR (95%CI): 1.83 (1.74–1.92)</td>
</tr>
<tr>
<td>rs190257775</td>
<td>128103466</td>
<td>T/C</td>
<td>0.99</td>
<td>block 3</td>
<td>OR (95%CI): 1.68 (1.42–1.80)</td>
<td>OR (95%CI): 1.36 (1.22–1.53)</td>
</tr>
<tr>
<td>rs72725879</td>
<td>128103969</td>
<td>T/C</td>
<td>0.18</td>
<td>block 4</td>
<td>OR (95%CI): 1.31 (1.28–1.35)</td>
<td>OR (95%CI): 1.17 (1.14–1.19)</td>
</tr>
<tr>
<td>rs5013678</td>
<td>128103979</td>
<td>T/C</td>
<td>0.78</td>
<td>block 5</td>
<td>OR (95%CI): 1.10 (1.08–1.13)</td>
<td>OR (95%CI): 1.20 (1.17–1.22)</td>
</tr>
<tr>
<td>rs183373024</td>
<td>128104117</td>
<td>G/A</td>
<td>0.01</td>
<td>block 6</td>
<td>OR (95%CI): 2.67 (2.43–2.93)</td>
<td>OR (95%CI): 3.20 (2.92–3.50)</td>
</tr>
<tr>
<td>rs78511380</td>
<td>128114146</td>
<td>T/A</td>
<td>0.92</td>
<td>block 7</td>
<td>OR (95%CI): 1.19 (1.14–1.23)</td>
<td>OR (95%CI): 0.97 (0.94–1.00)</td>
</tr>
<tr>
<td>rs17464492</td>
<td>128342866</td>
<td>A/G</td>
<td>0.72</td>
<td>block 8</td>
<td>OR (95%CI): 1.16 (1.14–1.18)</td>
<td>OR (95%CI): 1.17 (1.15–1.19)</td>
</tr>
<tr>
<td>rs6982267</td>
<td>128413050</td>
<td>G/T</td>
<td>0.51</td>
<td>block 9</td>
<td>OR (95%CI): 1.18 (1.16–1.20)</td>
<td>OR (95%CI): 1.23 (1.21–1.25)</td>
</tr>
<tr>
<td>rs7752894</td>
<td>128520479</td>
<td>A/T</td>
<td>0.01</td>
<td>block 10</td>
<td>OR (95%CI): 1.27 (1.23–1.30)</td>
<td>OR (95%CI): 1.45 (1.41–1.49)</td>
</tr>
<tr>
<td>rs12549761</td>
<td>128540776</td>
<td>C/G</td>
<td>0.87</td>
<td>block 11</td>
<td>OR (95%CI): 1.21 (1.18–1.24)</td>
<td>OR (95%CI): 1.28 (1.25–1.31)</td>
</tr>
</tbody>
</table>

- **RAF** is the risk allele frequency.
- **LD cluster** is the linkage disequilibrium block.
- **Conditional association** is the OR (95% confidence interval) adjusted for country and 7(OncoArray)/8(iCOGS) principal components.
- **Marginal association** is the OR (95% confidence interval) estimated in 2772; see Methods).

*Table 1 Marginal and conditional estimates for genetic markers at 8q24 independently associated with prostate cancer risk*
Correlation with known risk loci. The 12 risk variants spanned across the five LD blocks previously reported to harbor risk variants for PCa at 8q24\(^4\), with block 2 harboring six signals, blocks 1 and 5 two signals each, and blocks 3 and 4 only one (Supplementary Fig. 2). Except for a weak correlation between rs72725879 and rs78511380 in block 2 (\(r^2 = 0.28\)), the risk variants were uncorrelated with each other (\(r^2 \leq 0.09\); Supplementary Data 1), which corroborates their independent association with PCa risk. Eight of the variants (rs1487240, rs77541621, rs72725879, rs5013678, rs183373024, rs17464492, rs6983267, and rs12549761) were significantly associated with PCa risk, as shown in the marginal and conditional association plots (Supplementary Fig. 2).
rs7812894) have been previously reported either directly (Supplementary Table 2) or are correlated ($r^2 \geq 0.42$) with known markers of PCa risk from studies in populations of European, African or Asian ancestry (Supplementary Data 1)\(^4\)-\(^{10}\). The marginal estimates for previously published PCa risk variants at 8q24 in the current study are shown in Supplementary Table 2. The variant rs1914295 in block 1 is only weakly correlated with the previously reported risk variants rs12543663 and rs10086908 ($r^2 = 0.17$ and 0.14, respectively), while rs7851380 is modestly correlated with the previously reported risk variant rs1016343 ($r^2 = 0.28$). The remaining two variants, rs190257175 and rs12549761, are not correlated ($r^2 < 0.027$) with any known PCa risk marker.

**Polygenic risk score and familial relative risk.** To estimate the cumulative effect of germline variation at 8q24 on PCa risk, a polygenic risk score (PRS) was calculated for the 12 independent risk alleles from the final model based on allele dosages weighted by the per-allele conditionally adjusted ORs (see Methods). Compared to the men at ‘average risk’ (i.e., the 25th–75th PRS range among controls), men in the top 10% of the PRS distribution had a 1.93-fold relative risk (95%CI = 1.86–2.01) (Table 2), with the risk being 3.99-fold higher (95%CI = 3.62–4.40) for men in the top 1%. Risk estimates by PRS category were not modified by family history (FamHist-yes: OR = 4.24, 95%CI = 2.85–6.31; FamHist-no: OR = 3.38, 95%CI = 2.88–3.97). To quantify the impact of germline variation at 8q24, we also estimated the proportion of familial relative risk (FRR) and heritability of PCa contributed by 8q24 and compared this to the proportions explained by all known PCa risk variants including 8q24 (see Methods). The 175 established PCa susceptibility loci identified to date\(^{3,11}\) are estimated to explain 37.08% (95%CI = 32.89–42.49) of the FRR of PCa, while the 12 independent signals at 8q24 alone capture 9.42% (95%CI = 8.22–10.88), which is 25.4% of the total FRR explained by known genetic risk factors for PCa (Table 3). This is similar to the proportion of heritability explained by 8q24 variants (22.2%) compared to the total explained heritability by the known risk variants (0.118). In comparison, the next highest contribution of rs1016343 ($r^2 < 0.027$) with any known PCa risk marker.

**Table 2 Relative risk of PCa for polygenic risk score (PRS) groups**

<table>
<thead>
<tr>
<th>Risk category percentile(^a)</th>
<th>No. of individuals</th>
<th>Risk estimates for PRS groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
<td>Cases</td>
</tr>
<tr>
<td>≤1%</td>
<td>530</td>
<td>339</td>
</tr>
<tr>
<td>1%-10%</td>
<td>4771</td>
<td>3636</td>
</tr>
<tr>
<td>10%-25%</td>
<td>7936</td>
<td>7359</td>
</tr>
<tr>
<td>25%-75%</td>
<td>26,464</td>
<td>32,743</td>
</tr>
<tr>
<td>75%-90%</td>
<td>7940</td>
<td>13,431</td>
</tr>
<tr>
<td>90%-99%</td>
<td>4766</td>
<td>11,451</td>
</tr>
<tr>
<td>&gt;99%</td>
<td>528</td>
<td>2576</td>
</tr>
</tbody>
</table>

Note: PRS were calculated for variants from the final stepwise model with allele dosage from OncoArray and iCOGS weighted by the per-allele conditionally adjusted odds ratios from the meta-analysis. Risk category groups were based on the percentile distribution of risk alleles in overall controls. Estimated effect of each PRS group relative to the interquartile range (25%-75%) in OncoArray and iCOGS datasets separately, and then meta-analyzed across the two studies; odds ratios were adjusted for country and % (OncoArray)/% (iCOGS) principal components.

**Table 3 Proportion of familial relative risk (FRR) and heritability ($h^2$) of PCa explained by known risk variants**

<table>
<thead>
<tr>
<th>Source</th>
<th>No. of variants</th>
<th>Proportion of FRR (95%CI)</th>
<th>% of total FRR</th>
<th>$h^2$ (SE)</th>
<th>% of total $h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8q24(^a)</td>
<td>12</td>
<td>9.42 (8.22–10.88)</td>
<td>25.4</td>
<td>0.027 (0.011)</td>
<td>22.2</td>
</tr>
<tr>
<td>HOXB13(^b)</td>
<td>1</td>
<td>1.91 (1.20–2.85)</td>
<td>5.2</td>
<td>0.004 (0.005)</td>
<td>3.0</td>
</tr>
<tr>
<td>All other variants(^b,c)</td>
<td>162</td>
<td>25.77 (22.94–29.36)</td>
<td>69.5</td>
<td>0.092 (0.010)</td>
<td>74.9</td>
</tr>
<tr>
<td>Total</td>
<td>175</td>
<td>37.08 (32.89–42.49)</td>
<td>100</td>
<td>0.118 (0.012)</td>
<td>100</td>
</tr>
</tbody>
</table>

\(^a\)Conditional estimates were derived by fitting a single model with all variants from OncoArray data.

\(^b\)Risk estimates and allele frequencies for regions with a single variant are from a meta-analysis of OncoArray, iCOGS and 6 additional GWAS\(^3\)

\(^c\)Risk variants included from fine-mapping of PCa susceptibility loci in European ancestry populations\(^3\)
locus has been established as explaining >2% of the FRR, including the low frequency, non-synonymous, moderate penetration HOXB13 variant (rs138213197) at chromosome 17q21 that is estimated to explain only 1.91% (95%CI = 1.20–2.85) of the FRR.

JAM analysis. We explored our data with a second fine-mapping approach, JAM (Joint Analysis of Marginal summary statistics), which uses GWAS summary statistics to identify credible sets of variants that define the independent association signals in susceptibility regions (see Methods). The 95% credible set for the JAM analysis confirmed all of the independent signals from stepwise analysis except rs190257175, for which evidence for an association was weak (variant-specific Bayes factor (BF) = 1.17). There were 50 total variants included in the 95% credible set, and 174 after including variants in high LD ($r^2 > 0.9$) with those in the credible set (Supplementary Data 2).

Discussion

In this large study of germline genetic variation across the 8q24 region, we identified 12 independent association signals among men of European ancestry, with three of the risk variants (rs1914295, rs190257175, and rs12549761) being weakly correlated ($r^2 \leq 0.17$) with known PCa risk markers. The combination of these 12 independent signals at 8q24 capture approximately one quarter of the total FCA PRS explained by known genetic risk factors, which is substantially greater than any other known PCa risk locus.

The 8q24 region is the major susceptibility region for PCa; however, the underlying biological mechanisms through which germline variation in this region influences PCa risk remains uncertain. For each of the 12 risk variants at 8q24, the 95% credible set defined noteworthy (i.e., putative functional) variants based on summary statistics while accounting for LD. To inform biological functionality, we overlaid epigenetic functional annotation using publicly available datasets (see Methods) with the location of the 12 independent signals (and corresponding 174 variants within their 95% credible sets; Supplementary Data 3).

Of the 12 independent lead variants, 6 are situated within putative transcriptional enhancers in prostate cell-lines; either through intersection with H3K27AC (rs72725879, rs5013678, rs78511380, rs6983267 and rs7812894) or through a ChromHMM enhancer annotation (rs17464492, rs6983267, rs7812894). Eight of the 12 stepwise hits (rs77541621, rs190257175, rs5013678, rs183373024, rs183373024, rs78511380, rs6983267, and rs7812894) also intersect transcription factor binding site peaks from multiple ChIP-seq datasets representing the AR, ERG, FOXA1, GABPA, GATA2, HOXB13, and NKX3.1 transcription factors, with all 8 intersecting a FOXA1 mark and half an AR binding site. These variants may therefore exert their effect through regulation of enhancer activity and long-range expression of genes important for cancer tumorigenesis and/or progression. The variant rs6983267 has also been shown to act in an allele-specific manner to regulate prostate enhancer activity and expression of the proto-oncogene MYC in vitro and in vivo. However, despite the close proximity to the MYC locus, no direct association has been detected between 8q24 risk alleles and MYC expression in normal and tumor human prostate tissues. The rare variant with the largest effect on risk, rs183373024, shows high evidence of functionality based on overlap with multiple DNase1 and transcription factor binding site peaks (for AR, FOXA1, HOXB13, and NKX3.1), which supports previous findings of an allele-dependent effect of this variant on the disruption of a FOXA1 binding motif. Seven independent signals (rs1914295, rs1487240, rs77541621, rs72725879, rs5013678, rs183373024, rs78511380) and variants correlated at $r^2 > 0.9$ with these signals (Supplementary Data 2) are located within or near a number of prostate cancer-associated long noncoding RNAs (lncRNAs), including PRNCR1, PCAT1, and CCAT2, previously reported to be upregulated in human PCa cells and tissues.

Imputation analysis. Imputation of both OncoArray and iCOGS genotype data was performed using SHAPEIT and IMPUTE2 to the October 2014 (Phase 3)
release of the 1KGp reference panel. A total of 10,136 variants from OncoArray and 10,360 variants from iCOGS with MAF > 0.1% were imputed across the risk region at 8q24 (127.6-129.0 Mb). Variants with an imputation quality score >0.8 were retained for a total of 5600 overlapping variants between the two datasets.

Statistical analysis. Unconditional logistic regression was used to estimate per-allele odds ratios (ORs) and 95% confidence intervals (CIs) for the association between genetic variants (single nucleotide polymorphisms and insertion/deletion polymorphisms) and PCa risk adjusting for country and principal components (7 for OncoArray and 8 for iCOGS). Allele dosage effects were tested through a 1-degree of freedom two-tailed Wald trend test. The marginal risk estimates for the 5600 variants at 8q24 that passed QC were combined by a fixed effect meta-analysis with inverse variance weighting using METAL27. A modified forward and backward stepwise model selection with inclusion and exclusion criteria based on the p-value was performed on 10,136 variants marginally associated with PCa risk from the meta-analysis (p < 0.05, n = 2772). At each step, the effect estimates for the candidate variants from both studies (OncoArray and iCOGS) were meta-analyzed and each variant was incorporated into the model based on the strength of association. All remaining variants were included one-at-a-time into the logistic regression model conditioning on those already incorporated in the model. We applied a conservative threshold for independent associations, with variants kept in the model if their meta p-value from the Wald test was genome-wide significant at p ≤ 5 × 10−8 after adjustment for the other variants in the model. Conditional association variants in the final model and previously published PCa risk variants at 8q24 were estimated using the 1KGp Phase 3 EUR panel (Supplementary Data 1).

Haplotype analysis. Haplotypes were estimated in the Oncorray data only using variants from the final stepwise model selection (n = 12) and the EM algorithm28 within LD block regions inferred based on recombination hotspots using Haploview 4.2 (Broad Institute, Cambridge, MA, USA). Only haplotypes with an estimated frequency ≥ 0.5% were tested.

Polygenic risk score and familial relative risk. An 8q24-only polygenic risk score (PRS) was calculated for variants from the final model (n = 12) with allele dosage from OncoArray and iCOGS weighted by the per-allele conditionally adjusted ORs from the meta-analysis. Categorization of the PRS was based on the percentile distribution of controls, where the risk for each category was estimated relative to the interquartile range (25−75%) in OncoArray and iCOGS separately, and then meta-analyzed across the two studies. We estimated the contribution of 8q24 variants to the familial (first-degree) relative risk (FRR) of PCa (FRR = 2.5)30 under a multiplicative model, and compared this to the FRR explained by all known PCa risk variants including 8q24 (Supplementary Data 4). We also estimated heritability of PCa using the LMM approach as implemented in GCTA31. For regions which have been fine-mapped using the OncoArray meta-analysis data, we used the updated representative lead variants, otherwise the originally reported variant was included provided it had replicated at genome-wide significance in the meta-analysis; this identified a total of 175 (71 independently associated) PCa variants for the FRR and heritability calculations30,31. For these analyses, we used estimated conditions from fitting a single model with all variants in the OncoArray dataset for regions with multiple variants and the overall marginal meta-analysis results from Schumacher et al.3 for regions with a single variant. To correct for potential bias in effect estimates of newly discovered variants, we implemented a Bayesian framework for joint analysis of marginal SNP effects. Genet. Epidemiol. 40, 188–201 (2016).

JAM analysis. To confirm the stepwise results and identify candidate variants for potential functional follow-up, we used a second fine-mapping approach, JAM (Joint Analysis of Marginal summary statistics)32. JAM is a multivariate Bayesian variable selection framework that uses GWAS summary statistics to identify the most likely number of independent associations within a locus and define credible sets of variants driving those associations. JAM was applied to summary statistics from the meta-analysis results using LD estimated from imputed individual level data from 20,000 cases and 20,000 controls randomly selected from the OncoArray sub-study. LD pruning was performed using Priority Pruner (http://prioritypruner.sourceforge.net/) on the 2772 marginally associated variants at r2 = 0.9, resulting in 825 tag variants analyzed in four independent JAM runs with varying starting seeds. Credible sets were determined as the tag variants that were selected in the top 50% of the cumulative posterior probability for each of the independent JAM runs, plus their designated high LD proxy variants from the pruning step.

Functional annotation. Variants in the 95% credible set (n = 50) plus variants correlated at r2 > 0.9 with those in the credible set (n = 174) were annotated for putative evidence of biological functionality using publicly available datasets as described by Dadaev et al. Briefly, variants were annotated for proximity to gene (GENCODEv19), miRNA transcripts (miRBase release 20), evolutionary constraint (according to GERP++, SiPhy and PhastCons algorithms), likelihood of pathogenicity (CADDv13.3) and overlap with prospective regulatory elements within prostate-specific datasets (Dnasel hypersensitivity sites, H3K27Ac, H3K27me3 and H3K4me3 histone modifications, and for AR, CTCF, ERG, FOXA1, GATA2, HOXB13, and NKX3.1 transcription factor binding sites) in a mixture of LNCaP, PC-3, PrEC, RWPE1, and VCaP cell lines and human prostate tumor tissues downloaded from the Cistrome Data Browser (http://cistrome.org/db/). The chromatin state in which each variant resides was assessed using ChromHMM annotations from two prostate cell lines (PrEC and PC3). Cis-gene regulation was evaluated using 359 prostate adenoma cases from The Cancer Genome Atlas (TCGA PRAD; https://gdac.tcc.nci.nih.gov) that passed QC33. The eQTL analysis was performed using FastQTL with 1000 permutations for each gene with a false discovery rate window. We then used the method by Nica et al.34 that integrates eQTLs and GWAS results in order to reveal the subset of association signals that are due to cis-eQTLs. For each significant eQTL, we added the candidate variant to the linear regression model to assess if the inclusion better explains the change in expression of the gene. We retrieved the p-value of the model, assigning p-value of 1 if the eQTL and variant are the same. Then we ranked the p-values in decreasing order for each eQTL and finally calculated the colocalization score for each pair of eQTL and variants. In general, if an eQTL and candidate variant represent the same signal, this will be reflected by the variant having a high p-value, a low rank and consequently a high colocalization score.

Data availability

The authors declare that data supporting the findings of this study are available within the paper and the supplementary information files. However, some of the data used to generate the results of this study are available from the first author and the FRAC-TICAL Consortium upon request.

Received: 7 February 2018 Accepted: 1 October 2018
Published online: 05 November 2018

References


**Acknowledgements**

Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) [U19 CA 148537 for ELAcidating Loci Involved in Prostate Cancer Susceptibility (ELLIPE) project and R01HG07492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN268201000081]. Additional analytic support was provided by NIH NCI U01 CA188392 (PI: Schumacher). The PRACTICAL consortium (http://practicalacr.ac.uk/) was supported by Cancer Research UK Grants C5047/A10692, C16913/A6135, C5047/A3354, C5047/A16536, C5047/A11038, C1287/A16536, C5047/A11038, C1287/A16536, C5047/A3354, C5047/A10692, C16913/A6135, European Commission’s Seventh Framework Programme grant agreement n° 223175 (HEALTH-F2-2009-223175), and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON initiative). We


NATURE COMMUNICATIONS | DOI: 10.1038/s41467-019-06863-1

© The Author(s) 2018

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018

1Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA 90033, USA. 2The Institute of Cancer Research, London SW7 3RP, UK. 3Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Strangeways Research Laboratory, University of Cambridge, Cambridge CB1 8RN, UK. 4Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0QQ, UK. 5Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH 44106-7219, USA. 6Seidman Cancer Center, University Hospitals, Cleveland, OH 44106, USA. 7Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD 20892, USA. 8Institute of Population Health, University of Manchester, Manchester M13 9PL, UK. 9Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK. 10Epidemiology Research Program, American Cancer Society, 250 Williams Street, Atlanta, GA 30303, USA. 11SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. 12Australian Prostate Cancer Research Centre–Qld, Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, QLD 4059, Australia. 13Translational Research Institute, Brisbane, QLD 4102, Australia. 14Department of Medical Epidemiology and Biostatistics, Karolinska Institute, SE-171 77 Stockholm, Sweden. 15Centre for Cancer Genetic Epidemiology, Department of Oncology, Strangeways Research Laboratory, University of Cambridge, Cambridge CB1 8RN, UK. 16Department of Applied Health Research, University College London, London WC1E 7HB, UK. 17Department of Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, FI-20014 Turku, Finland. 18Tyyk Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital, 20521 Turku, Finland. 19BioMediTech, University of Tampere, 33520 Tampere, Finland. 20Division of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, SE-171 77 Stockholm, Sweden. 21Division of Cancer Sciences, Manchester Academic Health Science Centre, Radiotherapy Related Research, Manchester NIHR Biomedical Research Centre, The Christie Hospital NHS Foundation Trust, University of Manchester, Manchester M13 9PL, UK. 22Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA. 23Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. 24GRC N°5 ONCOTYPE-URO, UPnC Univ Paris 06, Tenon Hospital, F-75020 Paris, France. 25CeRePP, Tenon Hospital, F-75020 Paris, France. 26Department of Molecular Medicine, Aarhus University Hospital, 8200 Aarhus N, Denmark. 27Department of Clinical Epidemiology, Aarhus University, 8200 Aarhus N, Denmark. 28Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway. 29Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. 30Department of Oncology, Addenbrooke’s Hospital, University of Cambridge, Cambridge CB2 QOQ, UK. 31Research Cancer UK Cambridge Cambridge Research Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK. 32Nuffield Department of Surgical Sciences, University of Oxford, Oxford OX1 2JD, UK. 33School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK. 34Cancer Epidemiology, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK. 35Department of Surgical Oncology, Princess Mary Margaret Cancer Centre, Toronto, ON MSG 2M9, Canada. 36Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. 37Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029-5674, USA. 38Centre for Molecular Oncology, John Vane Science Centre, Barts Cancer Institute, Queen Mary University of London, London E1CM 6BQ, UK. 39Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, VIC 3004, Australia. 40Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC 3010, Australia. 41Department of Urologic Surgery, Bringham and Womens Hospital, Boston, MA 02115, USA. 42Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica, CIBERER, IDIS, 15706 Santiago de Compostela, Spain. 43Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 29208, USA. 44Centre for Research on Environmental Epidemiology (CREAL), Barcelona Institute for Global Health (ISGlobal), 08003 Barcelona, Spain. 45CIBER Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain. 46IMIM (Hospital del Mar Research Institute), 08003 Barcelona, Spain. 47Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain. 48Channing Division of Network Medicine, Department of Medicine, Brigham and Woman's Hospital/ Harvard Medical School, Boston, MA 02118, USA. 49Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL 33612, USA. 50Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109-1024, USA. 51Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA 98195, USA. 52International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, 70-115 Szczecin, Poland. 53Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark. 54Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, 2200 Copenhagen, Denmark. 55Division of Clinical Cancer Epidemiology and Aging Research, German Cancer Research Center (DKFZ), D-69120 Heidelberg, Germany. 56German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), D-69120 Heidelberg, Germany. 57Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), 69120 Heidelberg, Germany. 58Institute for Human Genetics, University Hospital Ulm, 89075 Ulm, Germany. 59Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. 60Department of Genetics, Portuguese Oncology Institute of Porto, 4200-072 Porto, Portugal. 61Biomedical Sciences Institute (ICBAS), University of Porto, 4050-313 Porto, Portugal. 62Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, CA 91010, USA. 63Ghent University, Faculty of Medicine and Health Sciences, Basic Medical Sciences, B-9000 Gent, Belgium. 64Department of Medical Genetics, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia. 65Department of Urology, University of Washington, Seattle, WA 98195, USA. 66Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, D-20246 Hamburg, Germany. 67Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University of Sofia, 1437 Sofia, Bulgaria. 68Department of Oncology, Cross Cancer Institute, University of Alberta, Edmonton, AB T6G 2C2, Canada. 69Division of Radiation Oncology, Cross Cancer Institute, University of Alberta, Edmonton, AB T6G 2Z2, Canada. 70Molecular Endocriology Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, BE-3000 Leuven, Belgium. 71Manchester Cancer Research Centre, Faculty of Biology Medicine and Health, Manchester Academic Health Science Centre, NIHR Manchester Biomedical Research Centre, Health Innovation Manchester, University of Manchester, Manchester M13 9WL, UK. 72Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigacion Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saúde, SERGAS.
The PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium


83Molecular Cancer Epidemiology Laboratory, QIMR Berghofer Institute of Medical Research, Herston, QLD 4006, Australia. 84School of Medicine, University of Queensland, Herston, QLD 4006, Australia. 85Royal Brisbane and Women’s Hospital, Herston, QLD 4029, Australia. 86The Kinghorn Cancer Centre (TKCC), Victoria, NSW 2010, Australia. 87Prostate Cancer Research Group, South Australian Health and Medical Research Institute, Adelaide, SA 5000, Australia. 88Department of Physiology, Biomedical Research Institute, Cancer Program, Monash University, Melbourne, VIC 3800, Australia. 89University of Adelaide, North Terrace, SA 5005, Australia. 90Finnlab Laboratories, Tampere University Hospital, FI-33520 Tampere, Finland. 91Finnish Cancer Registry, FI-00130 Helsinki, Finland. 92Department of Urology, Helsinki University Central Hospital and University of Helsinki, FI-00014 Helsinki, Finland. 93Division of Biostatistics and Bioinformatics, University of Maryland Greenbaum Cancer Center, and Department of Public Health and Public Health, University of Maryland School of Medicine, Baltimore, MD 21201, USA. 94Cancer Genomics Research Laboratory (CGR), Division of Cancer Epidemiology and Genetics, FNLCR Leidos Biomedical Research, National Cancer Institute, Frederick, MD 21701, USA. 95DNA Extraction and Staging Laboratory (DESL), Cancer Genomics Research Laboratory (CGR), Division of Cancer Epidemiology and Genetics, FNLCR Leidos Biomedical Research, National Cancer Institute, Frederick, MD 21701, USA. 96Sheffield Institute for Nucleic Acids, University of Sheffield, Sheffield S10 2TN, UK. 97Cambridge Cancer Trials Centre, Cambridge Clinical Trials Unit-Cancer Theme, Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, UK. 98Department of Medical Imaging, University Health Network, Toronto, ON M5G 2C4, Canada. 99Department of Pathology, University Health Network, Toronto, ON M5G 2C4, Canada. 100Advanced Radiation Biology Research Program, Research Center for Charged Particle Therapy, National Institute of Radiological Sciences, Chiba 263-8555, Japan. 101Department of Urology, National Hospital Organization Tokyo Medical Center, Tokyo 152-8902, Japan. 102Department of Urology, Nanfang Hospital, Southern Medical University, 510515 Guangzhou, China. 103Department of Pathology, Changzheng Hospital, The Second Military Medical University, 200433 Shanghai, China. 104Department of Urology, First Affiliated Hospital, Medical College, Zhengzhou University, 450003 Zhengzhou, China. 105Department of Urology, North Sichuan Medical College, 637000 Nanchong, China. 106Department of Nutrition Science, Shenyang Medical College, 110034 Shenyang, China. 107Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, 08036 Barcelona, Spain. 108Centre de Recerca Biomèdica CELLEX, 08036 Barcelona, Spain. 109Department of Urology, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, 08036 Barcelona, Spain. 110Department of Urology, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, 08036 Barcelona, Spain. 111Department of Urology, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, 08036 Barcelona, Spain. 112Department of Urology, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, 08036 Barcelona, Spain. 113Department of Urology, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, 08036 Barcelona, Spain. 114Department of Urology, Klinikum rechts der Isar der Technischen Universität Muenchen, 81675 Munich, Germany. 115Department of General and Clinical Pathology, Alexandrovskaya University Hospital, Medical University, 1431 Sofia, Bulgaria. 116Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah 21252 3270, Saudi Arabia. 117Grupo de Medicina Xénomica, CIBERER, CIMUS, Universidad de Santiago de Compostela, Avenida de Barcelona, 15782 Santiago de Compostela, Spain. 118Héritage Cancer Registry, Montpellier cedex 5, 34298 Montpellier, France. 119Urology Department, Clinique Beaune Soleil, 34070 Montpellier, France. 120INSERM U1147, 75013 Paris, France. 121Department of Clinical Science, Intervention and Technology, Karolinska Institutet, SE-171 77 Stockholm, Sweden. 122Division of Health Technology Assessment and Assessment of Social Services, SE-102 30 Stockholm, Sweden. 123Swedish Agency for Health Technology Assessment and Assessment of Social Services, SE-102 30 Stockholm, Sweden. 124Department of Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, SE-901 85 Umeå, Sweden. 125Department of Urology, Uppsala University, SE-751 85 Uppsala, Sweden. 126Department of Urology, Faculty of Medicine and Health, Orebro University, SE-701 82 Orebro, Sweden. 127Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK. 128National Institute for Health Research (NIHR) Biomedical Research Centre, University of Bristol, Bristol BS8 1TH, UK. 129Department of Urology, Cancer Therapy and Research Centre, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA. 130Department of Medical Physics, Complexo Hospitalario Universitario de Santiago, SERGAS, 15706 Santiago de Compostela, Spain. 131Urology Department, Hospital Germans Trias i Pujol, 08916 Barcelona, Spain. 132Laboratory and Department of Urology, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, 08036 Barcelona, Spain. 133Department of Urology, Hospital Clinic, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, 08036 Barcelona, Spain. 134Dame Roma Mitchell Lifehouse (COBLH), Camperdown, Sydney, NSW 2010, Australia. 135Garvan Institute of Medical Research, Sydney, NSW 2010, Australia. 136Research Program, Cancer Research Division, Peter MacCallum Cancer Centre, Melbourne, VIC 3000, Australia. 137Department of Pathology and Pediatrics, Albert Einstein College of Medicine and Surgery, Karolinska Institutet, and Department of Urology, Karolinska University Hospital, 171 76 Stockholm, Sweden. 138Department of Clinical Sciences at Danderyd Hospital, Karolinska Institutet, 182 88 Stockholm, Sweden. 139Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. 140Department of Urology, Tampere University Hospital, University of Tampere, Kalevantie 4, FI-33014 Tampere, Finland. 141Faculty of Medicine and Life Sciences, University of Tampere, FI-33014 Tampere, Finland. 142Department of Urology, The First Affiliated Hospital, Chongqing Medical University, 200032 Shanghai, China. 143Department of Urology, University of Rochester Medical Center, Rochester, NY 14620, USA. 144Washington University School of Medicine, StLouis, MO 63110, USA. 145Department of Urology, Aarhus University Hospital, 8200 Aarhus N, Denmark. 146Department of Radiation Oncology, University of Rochester Medical Center, Rochester, NY 14620, USA. 147Department of Pathology and Pediatrics, Albert Einstein College of Medicine, Bronx, NY 10461, USA. 148Second Military Medical University, 200433 Shanghai, China. 149Wuxi Second Hospital, Nanjing Medical University, 214003 Wuxi, Jiangzhu, China. 150Department of Urology, The First Affiliated Hospital, Chongqing Medical University, 200032 Shanghai, China. 151Department of Urology, James Cook University, Queensland, Australia.