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.
 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 × 10^{-15} and 4.28 × 10^{-15} (Table 1). None of the other variants were statistically significant at p < 5 × 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 6 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 × 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 PCA 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 1 Marginal and conditional estimates for genetic markers at 8q24 independently associated with prostate cancer risk

<table>
<thead>
<tr>
<th>Variant IDa</th>
<th>Positionb</th>
<th>Allelec</th>
<th>RAFd</th>
<th>LD clustera</th>
<th>Conditional associationf</th>
<th>Marginal associationh</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) 1.09 (1.07–1.11)</td>
<td>p-value 3.07 × 10^{-21}</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) 1.16 (1.14–1.18)</td>
<td>p-value 2.97 × 10^{-24}</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) 1.83 (1.74–1.92)</td>
<td>p-value 4.33 × 10^{-13}</td>
</tr>
<tr>
<td>rs190257175</td>
<td>128103466</td>
<td>T/C</td>
<td>0.92</td>
<td>block 2</td>
<td>OR (95%CI) 1.60 (1.42–1.80)</td>
<td>p-value 6.90 × 10^{-10}</td>
</tr>
<tr>
<td>rs72725879</td>
<td>128103969</td>
<td>T/C</td>
<td>0.18</td>
<td>block 3</td>
<td>OR (95%CI) 1.31 (1.23–1.40)</td>
<td>p-value 9.36 × 10^{-9}</td>
</tr>
<tr>
<td>rs5013678</td>
<td>128103979</td>
<td>T/C</td>
<td>0.78</td>
<td>block 4</td>
<td>OR (95%CI) 1.10 (1.08–1.12)</td>
<td>p-value 4.44 × 10^{-18}</td>
</tr>
<tr>
<td>rs183373024</td>
<td>128104117</td>
<td>G/A</td>
<td>0.01</td>
<td>block 5</td>
<td>OR (95%CI) 2.67 (2.43–2.93)</td>
<td>p-value 4.00 × 10^{-8}</td>
</tr>
<tr>
<td>rs78511380</td>
<td>128114146</td>
<td>T/A</td>
<td>0.92</td>
<td>block 6</td>
<td>OR (95%CI) 1.19 (1.14–1.24)</td>
<td>p-value 5.60 × 10^{-9}</td>
</tr>
<tr>
<td>rs17464492</td>
<td>128342866</td>
<td>A/G</td>
<td>0.72</td>
<td>block 7</td>
<td>OR (95%CI) 1.16 (1.14–1.18)</td>
<td>p-value 9.07 × 10^{-8}</td>
</tr>
<tr>
<td>rs6982267</td>
<td>128413050</td>
<td>G/T</td>
<td>0.51</td>
<td>block 8</td>
<td>OR (95%CI) 1.18 (1.16–1.20)</td>
<td>p-value 9.05 × 10^{-8}</td>
</tr>
<tr>
<td>rs78512894</td>
<td>128520479</td>
<td>A/T</td>
<td>0.13</td>
<td>block 9</td>
<td>OR (95%CI) 1.57 (1.33–1.80)</td>
<td>p-value 3.15 × 10^{-7}</td>
</tr>
<tr>
<td>rs12549761</td>
<td>128540776</td>
<td>C/G</td>
<td>0.87</td>
<td>block 10</td>
<td>OR (95%CI) 1.21 (1.18–1.24)</td>
<td>p-value 1.25 × 10^{-7}</td>
</tr>
</tbody>
</table>

aVariants that remained genome-wide significantly associated with PCA risk (p < 5 × 10^{-8}) in the final stepwise model
bPosition in genomic coordinates
ccSNP
RAF
destimated based on human genome build 38
fRisk allele/reference allele
gHaplotype block
hPer-allele odds ratio and 95% confidence interval adjusted for country, 7(OncoArray)/8(iCOGS) principal components and all other variants in the table

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).
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, rs12549761) were conserved across the genomic regions.
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).4,7–10. 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 date4,11 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 PCa risk variants including 8q24 (see Methods). The 175 established PCa susceptibility loci identified to date4,11 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 PCa risk variants including 8q24 (see Methods).

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

<table>
<thead>
<tr>
<th>Risk category percentile</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>26,646</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 full stepwise model with allele dosage from OncoArray and iCOGS weighted by the per-allele conditionally adjusted odds ratios from the meta-analysis.

(a)Risk category groups were based on the percentile distribution of risk alleles in overall controls.

(b)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 7(OncoArray)/8(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>8q24a</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>HOX513b</td>
<td>1</td>
<td>1.91 (1.30–2.85)</td>
<td>5.2</td>
<td>0.004 (0.005)</td>
<td>3.0</td>
</tr>
<tr>
<td>All other variants &amp; 8q24</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>

*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 and iCOGS and 6 additional GWAS.

(c)Risk variants included from fine-mapping of PCa susceptibility loci in European ancestry populations.

### 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 date4,11 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 an individual susceptibility region to the FRR of PCa is the TERT region at chromosome 5p15, where 5 independent signals contributed 2.63% (95%CI = 2.34–3.00). No other individual GWAS
locus has been established as explaining >2% of the FRR, including the low frequency, non-synonymous, moderate penetrance HOXB13 variant (rs13821397) at chromosome 17q21 that is estimated to explain only 1.91% (95%CI = 1.20–2.85) of the FRR\(^\text{11}\).

**JAM analysis.** We explored our data with a second fine-mapping approach, JAM (Joint Analysis of Marginal summary statistics)\(^\text{12}\), 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 PCa FRR 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 mechanism(s) 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, rs7811380, rs6983267 and rs7812894) or through a ChromHMM enhancer annotation (rs17464492, rs6983267, rs7812894). Eight of the 12 stepwise hits (rs77541621, rs190257175, rs5013678, rs183373024, rs78511380, 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\(^\text{13}\). 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\(^\text{14,15}\). 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\(^\text{16}\). The rare variant with the largest effect on risk, rs183373024, shows high evidence of functionality based on overlap with multiple DNaseI 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\(^\text{17}\). 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\(^\text{11,18,19}\) and tissues\(^\text{15,20}\). Based on eQTL annotations in prostate adenocarcinoma cells, the independent signal rs1914295 and three correlated variants \((r^2 > 0.9);\) Supplementary Data 2) are associated with overexpression of FAM84B, a gene previously associated with progression and poor prognosis of PCa in animal studies\(^\text{21}\). Variants correlated at \(r^2 > 0.9\) with rs7812894 \((n = 9;\) Supplemental Table 4) are eQTLs for POU5F1B, a gene overexpressed in cancer cell lines and cancer tissues\(^\text{22,23}\), although its role in PCa development is unknown. Whilst we have successfully refined the 8q24 region and identified a subset of variants with putative biological function within our credible set, multi-ethnic comparisons may help refine the association signals even further and precisely identify the functional alleles and biological mechanisms that modify PCa risk.

Whereas the individual associations of the 8q24 variants with PCa risk are relatively modest (ORs < 2.0, except for rs183373024), their cumulative effects are substantial, with risk being 4-fold higher for men in the top 1% of the 8q24-only PRS. The contribution to the overall FRR of PCa is substantially greater for the 8q24 region \((9.42%)\) than for any other known GWAS locus, including the moderate penetrance non-synonymous variant in HOXB13 \((1.91%)\). The ability of these markers to explain ~25.4% of what can be currently explained by all known PCa risk variants is a clear indication of the important contribution of germline variation at 8q24 on PCa risk. Our study was predominantly powered to analyze variants with MAF > 1% as the imputed variants with MAF = 0.1-1% were most likely to fail quality control (QC); however, the high density of genotyped markers and haplotypes at 8q24 in the OncoArray and iCOGS studies provided a robust backbone for imputation and increased the chances to impute lower MAF variants with high imputation quality score. Understanding of the biology of these variants and the underlying genetic basis of PCa could provide new insights into the identification of reliable risk-prediction biomarkers for PCa, as well as enable the development of effective strategies for targeted screening and prevention.

**Methods**

**Study subjects, genotyping, and quality control.** We combined genotype data from the PRACTICAL/ELLIPSE OncoArray and iCOGS consortia\(^\text{24,25}\), which included 143,699 men of European ancestry from 86 case-control studies largely based in either the US or Europe. In each study, cases primarily included men with incident PCa while controls were men without a prior diagnosis of the disease.

Both of the OncoArray and iCOGS custom arrays were designed to provide high coverage of common alleles (minor allele frequency [MAF] ≥ 5%) across 8q24 \((127.6–129.0\) Mb) based on the 1000 Genomes Project (1KGP) Phase 3 for OncoArray, and the European ancestry (EUR) panel from HapMap Phase 2 for iCOGS. A total of 57,580 PCa cases and 37,927 controls of European ancestry were genotyped with the Illumina OncoArray, and 24,198 PCa cases and 23,994 controls of European ancestry were genotyped with the Illumina iCOGS array. For both studies, sample exclusion criteria included duplicate samples, first-degree relatives, samples with a call rate <95% or with extreme heterozygosity \((p < 0.05)\), and samples with an estimated proportion of European ancestry <0.83\(^\text{24}\). In total, genotype data for 53,449 PCa cases and 36,224 controls from OncoArray and 18,086 PCa cases and 16,711 controls from iCOGS were included in the analysis. Genetic variants with call rates <0.95, deviation from Hardy-Weinberg equilibrium \((p < 10^{-10})\), and genotype discrepancy in >2% of duplicate samples were excluded. Of the final 498,417 genotyped variants on the OncoArray and 201,598 on the iCOGS array that passed QC, 1581 and 1737 within the 8q24 region, respectively, were retained for imputation.

All studies complied with all relevant ethical regulations and were approved by the institutional review boards at each of the participating institutions. Informed consent was obtained from all study participants. Additional details of each study are provided in the Supplementary Note 1.

**Imputation analysis.** Imputation of both OncoArray and iCOGS genotype data was performed using SHAPEIT\(^\text{25}\) and IMPUTE2\(^\text{26}\) 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 (termed FAME) was performed on the remaining variants individually and with PCa risk from the meta results ($p < 0.05, n = 2772$). 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 $p < 5 	imes 10^{-8}$ after adjustment for the other variants in the model. Consistent associations 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 Oncoarray 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 $\geq 0.5\%$ 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 PRS estimates, with the risk for each category 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 that it had replicated at genome-wide significance in the meta-analysis; this identified a total of 175 variants independently associated with PCa, and the FRR and heritability calculations. For these analyses, we used estimated conditional estimates 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 weighted correction2, which incorporates the uncertainty in the effect estimate into the final estimates of the bias-corrected ORs, 95% CIs and the corresponding calculations of percent FRR explained.

**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 $r^2 = 0.9$, 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 most significant combinations across the 4 JAM runs, and subsequently analyzed in 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 $r^2 > 0.9$ 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 (CADDv1.3) and overlap with prospective regulatory elements in prostate-specific datasets (DNasel hypersensitivity sites, Enhancer Atlas (TCGA PRAD; https://gdc-portal.nci.nih.gov) that passed QC3), the eQTL analysis was performed using FastQTL with 1000 permutations for each gene with a 1Mb 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 article and the supplementary information file. 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 ELucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research [U19 CA 148537 for ELucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIRD) under contract number HHSN2682012000081]. Additional analytic support was provided by NIH NCI U01 CA188392 (Ph. Schumacher). The PRACTICAL consortium ([http://practical.icr.ac.uk/](http://practical.icr.ac.uk/)) was supported by Cancer Research UK Grants CS/07/77357, C12677/A16653, C5047/A3354, C5047/A110692, 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 wish to thank all GWAS study groups contributing to the data set from which this study was conducted: OncoArray; iCOGS; The PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium; and The GAME-ON/ELLIPSE Consortium. Detailed acknowledgements and funding information for all GWAS study groups and from all the individual studies involved in the PRACTICAL Consortium are included in Supplementary Note 1. We would also like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK (now Prostate Action), The Orchid Cancer Appeal, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIH funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust.

Author contributions


Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-06863-1.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
The PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium

Chongqing, China. The People’s Hospital of Liaoning Province and The People’s Hospital of China Medical University, 110001 Shenyang, China.

Department of Urology, Zhongshan Hospital, Fudan University Medical College, 200032 Shanghai, China. Precision Medicine, School and Clinical Sciences at Monash Health, Monash University, Clayton, VIC 3168, Australia. Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS 7000, Australia. Faculty of Medical Science, John Radcliffe Hospital, University of Oxford, Oxford OX1 2JD, UK.

Department of Radiation Oncology, Complexo Hospitalario Universitario de Santiago, SERGAS, 15706 Santiago de Compostela, Spain. Division of Family Medicine, Department of Neurobiology, Care Science and Society, Karolinska Institutet, Huddinge, SE-171 77 Stockholm, Sweden. Scandinavian Development Services, 182 33 Danderyd, Sweden. University of Cantabria-IDEVAL, 39005 Santander, Spain. School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA 70112, USA. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA. Department of Urology, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, 2200 Copenhagen, Denmark.

Copenhagen Prostate Cancer Center, Department of Urology, Rigshospitalet, Copenhagen University Hospital, DK-2730 Herlev, Denmark. Saarland Cancer Registry, 66119 Saarbrücken, Germany. Department of Urology, University Hospital Ulm, 89075 Ulm, Germany. Cancer Prevention Institute of California, Fremont, CA 94538, USA. Department of Health Research and Policy (Epidemiology) and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305-5101, USA. Department of Radiotherapy, Ghent University Hospital, B-9000 Gent, Belgium. Cancer Research Malaysia (CRM), Outpatient Centre, Subang Jaya Medical Centre, 47500 Subang Jaya, Selangor, Malaysia. Urogenital Unit, Division of Medical Oncology, Department of Oncology, University Hospital Centre Zagreb, Šalata 2, 10000 Zagreb, Croatia. Department of Urology, University Hospital Center Zagreb, University of Zagreb School of Medicine, Šalata 2, 10000 Zagreb, Croatia. Department of Urology and Alexandrovska University Hospital, Medical University of Sofia, 1431 Sofia, Bulgaria.

Department of Urology, University Hospitals Leuven, BE-3000 Leuven, Belgium. Southampton General Hospital, The University of Southampton, Southampton SO16 6YD, UK. Genetic Oncology Unit, CHU Vilobos Hospital, Instituto de Investigación Biomédica Galicia Sur (IISGS), Complexo Hospitalario Universitario de Vigo, 36204 Vigo (Pontevedra), Spain. Moores Cancer Center, Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA 92093-0012, USA. Department of Clinical Chemistry, Erasmus University Medical Center, 3015 CE Rotterdam, The Netherlands. Program for Personalized Cancer Care, NorthShore University HealthSystem, Evanston, IL 60201, USA.

Department of Epidemiology, University of Washington, Seattle, WA 98195, USA. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London SW7 2AZ, UK. Genomics England, Queen Mary University of London, Dawson Hall, Charterhouse Square, London EC1M 6BQ, UK. Genomic Epidemiology Unit, German Cancer Research Center (DKFZ), D-69120 Heidelberg, Germany.

Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI 96813, USA. Dana-Farber Cancer Institute, Boston, MA 02215, USA. Paris-Sud University, UMR 1018, Cedex 94807 Villejuif, France. Deceased: Brian E. Henderson.