Atlas of prostate cancer heritability in European and African-American men pinpoints tissue-specific regulation

Alexander Gusev et al.#

Although genome-wide association studies have identified over 100 risk loci that explain ~33% of familial risk for prostate cancer (PrCa), their functional effects on risk remain largely unknown. Here we use genotype data from 59,089 men of European and African American ancestries combined with cell-type-specific epigenetic data to build a genomic atlas of single-nucleotide polymorphism (SNP) heritability in PrCa. We find significant differences in heritability between variants in prostate-relevant epigenetic marks defined in normal versus tumour tissue as well as between tissue and cell lines. The majority of SNP heritability lies in regions marked by H3k27 acetylation in prostate adenocarcinoma cell line (LNCaP) or by DNaseI hypersensitive sites in cancer cell lines. We find a high degree of similarity between European and African American ancestries suggesting a similar genetic architecture from common variation underlying PrCa risk. Our findings showcase the power of integrating functional annotation with genetic data to understand the genetic basis of PrCa.

Correspondence and requests for materials should be addressed to A.G. (email: agusev@hsph.harvard.edu) or to B.P (email: pasaniuc@ucla.edu).

#A full list of authors and their affiliations appears at the end of the paper.
Family history is a well-established risk factor for prostate cancer (PrCa), which has an estimated heritability of 58%—one of the highest across common cancers.1 Genome-wide association studies (GWAS) have been particularly successful in identifying over 100 risk loci that capture ~33% of the estimated familial risk.2 Although most of the GWAS PrCa variants overlap prostate-specific regulatory elements (for example, androgen receptor-binding sites (ARBS))2-8, a quantification of the contribution of genetic variation from various chromatin marks to PrCa risk is currently lacking.

Recent work form the ENCODE/ROADMAP consortia has shown that a large fraction of the genome plays a role in at least one biochemical event, in at least one tissue. Although this functional atlas of the human genome has greatly enhanced our understanding of regulatory elements, such functional elements are often tissue specific10,11 making their interpretability in the context of PrCa risk challenging. Existing studies that have integrated PrCa GWAS findings with tissue-specific functional annotations have relied only on the GWAS significant variants (~100 in the most recent study) or single-nucleotide polymorphisms (SNPs) tagging them2,7, thus ignoring loci that do not reach genome-wide significance. Recent methodological advances have shown that the entire polygenic architecture of common traits can be interrogated using variance components across all assayed SNPs (typed and/or imputed) to increase power for detecting trait-specific functional annotations12. In addition to offering superior performance relative to methods that evaluate only GWAS SNPs, the variance components methods also allow for comparison of estimates across different studies and sample sizes. This is because variance components yield an unbiased estimate (under standard assumptions) of SNP heritability ($h^2$)—the variance in trait explained by SNPs that reside within elements of a given functional category12-15.

Here, we use targeted and genome-wide SNP array data from 59,089 male PrCa cases and controls of European ancestry to dissect the genetic risk of PrCa. We estimate the SNP heritability of previously implicated regulatory annotations1,12 and perform a broad analysis of 544 epigenetic marks from ENCODE/ROADMAP (ref. 9). Our approach interrogates the entire common polygenic architecture of PrCa while accounting for potential correlations between related functional categories. First, we find that SNPs near ARBS assayed in prostate tumour explain significantly more of the heritability of PrCa than ARBS SNPs assayed in prostate normal tissue. Second, we localize much of the heritability of PrCa to regions in the genome marked by three functional categories: (i) H3K27ac histone modifications in prostate adenocarcinoma cell lines (LNCaP; typically marking active enhancers17); (ii) androgen receptors in prostate tissue18; and (iii) DNase I hypersensitivity sites (DHS) in cancer cell lines. We replicate the LNCaP H3K27ac and DHS results across different ancestries and show that risk prediction from genome-wide SNP data is significantly improved with a predictor that incorporates the functional atlas as prior. Overall, our results suggest a similar genetic architecture from common variation of PrCa risk across men of European and African ancestry and highlight H3k27ac histone mark in LNCaP and ARBS in prostate tissue for follow-up studies of PrCa risk.

Results

Partitioning the genetic risk for prostate cancer. We analysed multiple functional annotations and quantified the fraction of variance in trait explained by SNPs that are localized within each functional class. Our approach models the phenotype (PrCa) of a set of individuals as being drawn from a multivariate normal distribution with variance components estimated based on genetic data (that is, SNPs) plus an environmental term (see Methods)13,14. For each functional category i, a genetic relationship matrix across all individuals is computed from all the SNPs residing in the given functional category to serve as a variance component. Multiple components are then jointly fitted using the restricted maximum likelihood (REML) as implemented in the GCTA software14 to estimate variance parameters ($\sigma^2_i$) for each component. The SNP heritability for component i is then estimated as $h^2_i = \sigma^2_i / \sum \sigma^2_i$, where the sum in the denominator is across all fitted components including the environmental term. Therefore, we view $h^2_i$ as an estimate of the variance in trait that can be explained by all the SNPs in the corresponding functional category with a linear model of the trait (that is, SNP heritability)1-2. We expect functional categories that are enriched with causal variants for PrCa to attain a higher estimated SNP heritability as compared with functional categories depleted of causal variants for PrCa. To focus our results on noncoding variation and account for potential confounders because of linkage disequilibrium (LD), we explicitly included coding and coding-proximal regulatory variation as ‘background’ components whenever we quantified the effect of each functional annotation tested (see Methods).

The variance component model has previously been shown to yield robust estimates under the assumption that causal variants are typed and uniformly sampled from a given component13,20,21. Here, we perform additional simulations using the UK10K whole-genome sequence data to confirm the validity of this model for our data, and to assess how representative SNP estimates are of true underlying biology at common sequenced variants. The simulation framework uses real genotype data from the UK10K consortium to generate additive, polygenic phenotypes with a given heritability and then performs heritability estimation with the variance component model (see Methods). Although the UK10K data contains a much smaller set of individuals as the iCOGS data (3,047 versus 42,613 individuals, see Methods), it contains variation from whole-genome sequencing; this allows us to evaluate model performance by simulation when restricting to SNPs genotyped on the iCOGS platform. We focused on the LNCaP: H3k27ac annotation (which was most significant in our data, see below) to evaluate the multiple component models. Over thousands of simulations, we confirmed that the variance components approach correctly recovered the causal contribution to trait from a given functional category when causal variants were typed (Supplementary Table 1, see Methods). Under both null and enriched scenarios the estimates were unbiased and standard errors properly calibrated (Supplementary Table 1). For common sequenced variants not present on the iCOGS platform, relative estimates of noncoding enrichment/depletion were conservative, with the tagged effects distributed across the typed components (Supplementary Table 2). Deviations from the standard variance components model assumptions on the distribution of effect-sizes and ancestry-specific effects in African Americans yielded either well calibrated or conservative estimates of SNP heritability in the focal LNCaP: H3k27ac category (see Methods, Supplementary Tables 1-3).

Our primary functional analyses focus on the densely genotyped iCOGS sample (21,678 cases and 20,935 controls), whose large sample size allowed for highly accurate estimates of component-specific $h^2_i$. Although the iCOGS chip is custom built to oversample risk loci, it provides a broad coverage of the common variation genome wide1. To showcase the power of the variance components approach, we estimated the total SNP...
heritability of PrCa at 0.28 (s.e. 0.01) in the iCOGS data (not significantly different from the total SNP heritability estimate of 0.26 (s.e. 0.05) in the BPC3 data), a significant increase from the variance explained only by the known GWAS variants \( h_g^2 \) (s.e. 0.06; s.e. 0.001) (see Methods; Supplementary Table 4). Interestingly, the total SNP heritability in the African American sample, which was genotyped on a different platform than iCOGS (see Methods), was estimated at 0.32 (s.e. 0.06) indicating a similar aggregate contribution of common variation to PrCa risk across the two ethnicities despite higher overall risk in African Americans\(^\text{(22)}\) (Supplementary Table 4).

**Enrichment at androgen receptor-binding sites in tumours.** We first focused on SNPs localized in the ARBS; an epigenetic profile causally implicated in prostate tumorigenesis. In contrast to typical assays that focus on cell lines, the ARBS were defined by chromatin immunoprecipitation and high-throughput sequencing (ChiP-seq) directly in primary human tissue (seven normal and 13 tumour specimens)\(^\text{(18)}\). We observed that variants within 5 kb of tumour-specific ARBS explained 17.0% of the genome-wide \( h_g^2 \) (s.e. 1.7%; \( P = 2.6 \times 10^{-16} \) by Z-test), whereas the variants near normal-specific ARBS explained 0.0% of the \( h_g^2 \) (s.e. 0.9%; \( P = 0.11 \) by Z-test) (Fig. 1). The difference between these two groups was highly significant and demonstrates the importance of assaying functional marks in both normal and tumour tissues. We note that the 5 kb extension may also include other regulatory variants near the tumour/normal-specific ARBS (but not heritability from coding/untranslated region (UTR)/promoter variants, which were explicitly modelled, see Methods). Smaller flanking regions were also investigated but did not include enough markers for the variance components model to converge. We also quantified the proportion of SNP heritability explained directly by all ARBS variants (both normal and tumour without 5 kb flanks) at 10.7% of \( h_g^2 \) significantly different from the SNP heritability of ARBS variants assayed in prostate adenocarcinoma cancer cell line (LNCaP; 3.2% of \( h_g^2 \) \( P = 4.4 \times 10^{-7} \) for difference by Z-test) (Fig. 1). This difference is partially explained by the very low number of SNPs within cell line ARBS making their aggregate contribution small but not empowering us to place a strong bound on the enrichment. Overall, these findings highlight the increased complexity of ARBS in a sample of tissues as compared with the single LNCaP cell line.

**Identification of functional marks relevant to PrCa risk.** Next, we looked for marks that contribute to the heritability of PrCa across a broad spectrum of functional annotations without prior assumptions on relevance to disease. We investigated 544 epigenetic annotations spanning six major classes (DHs; H3k4me1; H3k4me3; H3k9ac; H3k27ac); and computationally predicted functional classes or ‘segmentations’\(^\text{(23,24)}\) averaging 101 cell types per class (see Methods). After accounting for multiple testing, we identified 82 annotations that exhibited statistically significant deviations in SNP heritability from what was expected based on the proportion of the genome covered by that particular annotation (see Fig. 2 and Supplementary Data).

We first focused on 17 functional marks measured in the prostate, of which 14 were statistically significant (Supplementary Table 5). The single most significant enrichment was observed for H3k27ac marks in LNCaP \( (P = 1 \times 10^{-32}) \) by Z-test), which localized 22% of the total \( h_g^2 \) to the 2.9% of genotyped SNPs within the annotation. This was followed by variants in DHS marks in LNCaP \( (P = 2 \times 10^{-18}) \) by Z-test; 16.7% of \( h_g^2 \) localized in 3.1% of genome). The DHS annotations allowed us to compare estimates across three major prostate cell lines: LNCaP; normal prostate epithelial (PrEC); and immortalized prostate epithelial (RWPE1) (overlapping by 25–50% with ARBS; Supplementary Fig. 1). We observed heritability explained by LNCaP DHS to be nominally significantly higher than PrEC \( (P = 0.01 \) by Z-test); and both LNCaP and PrEC to be significantly higher than RWPE1 \( (P = 1.5 \times 10^{-9}, P = 1.2 \times 10^{-5}) \) respectively, by Z-test) (Fig. 3). More broadly, 10 out of 16 DHS marks measured in cancer cell lines were observed as significant, with colorectal cancer as the next most significant cancer \( (P = 6.0 \times 10^{-10}) \) by Z-test; 9.4% of heritability localized in 2.0% of genome; Supplementary Data). H3k27ac in LNCaP remained the most significantly enriched mark across all 544 annotations (presented in detail in the Supplementary Data). The most depleted categories were repressed regions computationally predicted by Segway-chromHMM in HepG2 cells \( (P = 1.3 \times 10^{-19}) \) by Z-test; 51.9% of \( h_g^2 \) from 74.3% of SNPs; Supplementary Data), with similar levels of depletion in repressed regions from other cell types. These regions are typically associated with decreased gene expression and repressive histone marks\(^\text{(23–25)}\), further emphasizing the importance of active regulation.

As H3k27ac typically marks active enhancers, we further evaluated variants with respect to their enhancer or ‘super’-enhancer status (large clusters of enhancers that are enriched for genes involved in cell identity\(^\text{(26)}\)) (see Methods). We did not observe differences in average heritability explained by SNPs within the two marks across 49 cell lines (see Methods), with an average of 1.51 (1.47)-fold increase over random SNPs for

---

**Figure 1 | Functional partitioning for variants within ARBS for PrCa.** Bars graphs detailing %SNP heritability estimates from two models of PrCa relevant functional annotations. (a) Joint comparison of variants within 5 kb of tumour-only and normal-only regions in the ARBS in prostate tissue \( (P = 2.1 \times 10^{-19}) \) for difference by Z-test). (b) Estimates from ARBS in prostate tissue (no longer using a 5 kb flank) and ARBS in LNCaP cell lines \( (P = 4.4 \times 10^{-7}) \) for difference). The null \( (\% h_g^2 = \% \text{SNPs}) \) is labelled by the dashed lines. Error bars show analytical standard error of estimate.
enriched categories in the selected model, we found that SNPs localized to 8.6% of SNPs (Table 1 and Supplementary Tables 8 and 9). Estimates from imputed markers were more representative of underlying enrichment in our simulations (see Methods, Supplementary Table 2) but may include the effects of nearby markers\(^2\) and so we consider them as an upper bound. None of the estimates changed significantly after adjusting for known GWAS associations\(^2\) (79 of which were typed in this data), underscoring the polygenic nature of this effect.

Having inferred the selected model, we re-analysed each of the 82 marginally significant categories jointly with the selected model (see Methods). Only three marks remained significant: two H3k27ac annotations in the colon crypt and one H3k27ac annotation in pancreas (Supplementary Data). This implies that the marginal enrichment of the 82 annotations was primarily driven by the overlap with functional marks in the selected model. For example, the H3K4me1 mark in penis foreskin keratinocytes that was previously highly significant (24.6% \(h^2\), \(P = 3.0 \times 10^{-16}\) by Z-test, Fig. 1) was no longer enriched after conditioning on the selected model (7.1% \(h^2\), \(P = 0.29\) by Z-test, Supplementary Data). The reduction to a small number of categories in the selected model with limited loss in signal further emphasizes the extent to which the selected model has localized the functional sources of enrichment. Focusing on the two most enriched categories in the selected model, we found that SNPs present in both the prostate tissue ARBS and LNCaP H3k27ac marks yielded significantly higher average heritability per SNP than either mark individually (Supplementary Table 10).

**Figure 2 | Functional partitioning of heritability across six main epigenetic classes.** Each point corresponds to an estimate of % SNP heritability (y axis) from SNPs within a cell-type-specific functional annotation versus annotation size (%SNPs, x axis). Overall, 544 annotations were tested, and red points indicate significant deviations from the null of \(\% h^2\) equal to \%SNPs after accounting for all tests. The two most significant annotations in each class are shown with triangle/cross, respectively, and labelled in bottom right (see Supplementary Data for all annotations).
In contrast, the variants specific to ARBS or H3k27ac were comparable in SNP heritability.

Replication of genomic functional atlas across ancestries. We evaluated replication of our model using two separate genome-wide SNP data sets of PrCa, one of European ancestry (BPC3; 6,953 samples) and one of African ancestry (AAPC; 9,522 samples) for PrCa (see Methods). To account for the smaller sample size, we focused on the eight-component selected model, only retaining significant components and three coding-proximal classes (coding, UTR, promoter)\(^\text{12}\). Because of platform differences between the populations, we used post-QC imputed variants in each data set, which are most reflective of underlying differences between the populations, we used post-QC imputed variants within each mark were remarkably similar (\(r^2\) values computed by Z-test using \(h^2_g\) estimate and analytical standard error).

H3k27ac mark in LNCaP is specific to PrCa. As a negative control, we evaluated the selected model with imputed SNPs across 11 common non-cancer diseases from the Wellcome Trust Case Control Consortium (WTCCC) (see Methods, Supplementary Table 14) where we observed two main differences: the LNCaP H3k27ac annotation was no longer significantly enriched (1.1\% \(h^2_g\) with 2.6\% of SNPs); and the repressed regions were much less depleted from the null (28.1\%).
Table 1 | Partitioning of heritability across functional classes in prostate cancer.

| Functional category | %SNPs ICOGS genotyped | %SNPs ICOGS imputed | %SNPs BPC3 imputed | %SNPs AAPC imputed |
|---------------------|------------------------|----------------------|---------------------|---------------------|
|                     | $h_g^2$ s.e.m. | $h_g^2$ s.e.m. | $h_g^2$ s.e.m. | $h_g^2$ s.e.m. |
| Coding              | 1.8        | 3.0        | 1.3        | 0.9        | 2.9        | 0.2        | 10.1        | 3.3        | 11.1 |
| UTR                 | 1.9        | 1.6        | 1.4        | 3.0        | 3.1        | 21.0       | 11.3        | 5.9        | 11.2 |
| Promoter            | 3.4        | *7.8*      | 1.8        | 8.9        | 4.1        | 0.0        | 12.7        | 0.0        | 14.7 |
| LNCaP: H3k27ac      | 3.2        | **22.3**   | 2.1        | **27.0**   | 3.8        | **30.3**   | 12.1        | **28.9**   | 12.7 |
| ARBS                | 1.0        | *3.3*      | 1.1        | **9.1**    | 3.3        | 1.1        | 12.1        | 15.2       | 12.1 |
| LNCaP: FOXA1        | 1.5        | 1.5        | 1.3        | 16.3       | 19.6       | 3.5        | 24.8       | 12.3       | 24.8 |
| LNCaP: H3k4me1      | 2.0        | 1.3        | 1.4        | 7.3        | 4.1        | 2.4        | 11.3       | 7.6        | 11.3 |
| LNCaP: DHS          | 2.9        | 5.4        | 1.6        | 5.4        | 2.6        | 1.4        | 11.3       | 7.6        | 11.3 |
| DHS prostate        | 1.8        | 2.6        | 1.4        | 1.8        | 0.2        | 1.3        | 11.3       | 7.6        | 11.3 |
| DHS cancer          | 4.7        | *14.1*     | 2.3        | **49.6**   | 6.3        | **47.4**   | 21.4        | 46.6       | 22.4 |
| H3k4me1 (other)     | 16.3       | 19.6       | 3.5        | 7.3        | 4.1        | 2.4        | 11.3       | 7.6        | 11.3 |
| H3k27ac (other)     | 7.3        | 4.1        | 2.4        | 1.8        | 0.2        | 1.3        | 11.3       | 7.6        | 11.3 |
| DHS (other)         | 1.8        | 0.2        | 1.3        | 16.3       | 19.6       | 3.5        | 24.8       | 12.3       | 24.8 |
| repressed           | 48.7       | **11.0**   | 4.1        | **0.3**    | 7.0        | **0.0**    | 23.8        | **0.0**    | 24.5 |
| all other           | 1.7        | 0.7        | 1.2        | 2.0        | 0.7        | 1.2        | 2.0        | 0.7        | 1.2 |

ARBS, androgen receptor-binding sites; DHS, DNase I hypersensitivity sites; SNP, single-nucleotide polymorphism; UTR, untranslated region.

Genomic functional atlas improves polygenic risk prediction.

To validate our SNP heritability genomic atlas, we compared the accuracy of predicting case/control status from genetic data with or without the functional atlas. We evaluated three different prediction models in the iCOGS sample: (i) a genetic risk score (GRS) from the genome-wide significant SNPs; (ii) the single best linear unbiased predictor (BLUP) using a single variance component from all SNPs; and (iii) the weighted sum of individual BLUPs from each epigenetic category in the selected model (multi-BLUP; see Methods). Evaluated by cross-validation, the GRS yielded an $R^2 = 0.029$ with true phenotype, whereas the single BLUP yielded an $R^2 = 0.065$ and the multi-BLUP had an $R^2 = 0.071$ (Supplementary Table 15). In a joint model with all three predictors, the multi-BLUP was highly significant ($P = 5.3 \times 10^{-31}$ from multiple regression). When we constructed the GRS from SNPs recently discovered in a much larger PrCa GWAS (ref. 2), the resulting prediction $R^2$ increased to 0.084. However, including the single BLUP or the multi-BLUP as an additional predictor still increased the prediction $R^2$ to 0.096 (joint $P = 6.7 \times 10^{-4}$ from multiple regression) and 0.098 (joint $P = 1.3 \times 10^{-33}$ from multiple regression), respectively (Supplementary Table 15). The consistent statistical significance and increased prediction accuracy confirms the validity of the selected model in this data and in larger GWAS.

Discussion

Using large-scale genotype data from over 59,089 men of European and African American ancestries jointly with epigenetic annotations, we identified highly significant differences in SNP heritability ($h_g^2$) of PrCa across variants from different epigenetic classes, tissue types and cell lines. Focusing on marks measured in prostate, we observed significantly higher $h_g^2$ around tumour-specific ARBS; ARBS measured in primary tissue relative to cell line; and DHS measured in PrCa cell line relative to prostate epithelial cell line. The enrichment at tumour-specific ARBS was consistent with recent findings showing that these sites were enriched for nearby genes highly expressed in tumours.

These analyses are comprehensive and cover most commonly studied prostate cell lines except for vertebral cancer of the prostate, which were not well represented in the ENCODE/ROADMAP. A search across 544 diverse functional annotations restricted most of the $h_g^2$ to a small fraction of the genome marked by prostate regulatory elements. Consistent with previous findings in common disease, functionally repressed regions were significantly depleted in heritability, highlighting the role of active regulation in PrCa susceptibility. Subsequent model selection localized the enrichment from 82 individually significant annotations to six that remained significant in a joint model. In particular, the abundance of enrichment in H3k27ac marks (active enhancers) relative to H3k4me1/H3k4me3 (poised enhancers/promoters) underscores their role in PrCa, though further enrichment in super enhancers was not observed. The enrichment within LNCaP: H3K27ac and depletion at repressed regions was replicated across different ancestries and yielded significant improvements in polygenic risk prediction.

With most GWAS associations falling outside coding regions, our analyses offer an important resource for prioritizing potential loci and focusing future studies on the most heritable genomic regions. The marginal analyses provide a ranking of 544 common functional assays, while the selected model localizes heritability to only those functional classes that are independently enriched. Emerging functional categories may further refine this signal or reveal other relevant epigenetic marks, though little enrichment beyond the selected model was observed in the comprehensive sampling of functional data analysed here. In general, the variance component model offers an opportunity to evaluate biological hypotheses in silico and without strictly relying on individually significant SNPs. However, as with any analysis of array-based data, the $h_g^2$ estimates will not include the contribution of SNPs that are untyped or poorly tagged, such
as rare variants or other contributors to the missing heritability. Future analyses of whole-genome sequencing, additional functional annotations, and larger sample sizes can yield important insights into functional mechanisms that are still not localized. Overall, our results suggest similar patterns of functional enrichment across men of European and African American ancestry and highlight ARBS, H3k27ac marks in LNCaP cell lines and DHS in cancer cell lines for follow-up studies of PrCa risk.

Methods

Epigenetic annotations. Sample collection and processing for functional annotations was made publically available by the ENCODE/ROADMAP consortia28. DHS, H3k4me1, H3k4me3, H3k9ac annotations and genome segmentations, enhancers and super enhancers26 and PrCa-specific annotations18 were assay and processed as detailed in the original studies. Tumour-only and normal-only ARBS were defined in seven normal and 13 tumour specimens in the original study18. All annotations curated for this paper (ENCODE/ROADMAP; Pomerantz et al. and Hazelett et al.) are available at https://data.broadinstitute.org/alkesgroup/ANNOTATIONS/PRCA/. The full list of individual annotations with web-links to the corresponding boundary definitions is provided in the Supplementary Data. Some functional marks are listed multiple times due to multiple independent assays or laboratory protocols.

ARS ChIP-seq in human tissue specimens. The ARBS assay was performed as described in REF (ref. 18) and summarized here. Fourteen subjects of European American ancestry were selected for ChIP analysis. Their chromatin was incubated overnight with 6μg antibody AR (N-20, Santa Cruz Biotechnology, Dallas, TX).

Figure 5 | Partitioning of heritability across functional classes in prostate cancer. Visual representation of heritability enrichment in three studies a,b: iCOGS; c: AAPC; d: BPC3 (shown numerically in Table 1). Each subplot corresponds to an analysis of the listed joint model, with coloured slices representing the functional annotations evaluated. Volume of each interior (light coloured) pie-chart slice represents the %SNP for the functional annotation, which is equal to the expected %h2 g under the null of no enrichment. Volume of each shaded pie-chart slice represents the actual %h2 g inferred by the model. Slices extending outside/inside the middle pie correspond to enrichment/depletion in SNP heritability, as indicated by the dotted lines. Colour coding is consistent across all subpanels. * (**) denotes significant deviation at P<0.05 (P<0.05/15) of fraction of SNP heritability (%h2 g from null model of %h2 g = %SNPs by Z-test; see Supplementary Table 6 for P values).
bound to protein A and protein G beads (Life Technologies, Carlsbad, CA). A fraction of the sample was not exposed to antibody to be used as control (input). The samples were de-crosslinked, treated with RNase and proteinase K, and DNA was extracted. The samples were then re-sheared to 100–300 base pairs using the ThruPLEX-FD Prep Kit (Rubicon Genomics, Ann Arbor, MI). DNA sequencing were performed using 50-base pair reads on the Illumina platform (Illumina, San Diego, CA) at Dana-Farber Cancer Institute. AR binding sites were generated using Model-Based Analysis of ChIP-seq 2 (MACS2), with a qvalue (false discovery rate, FDR) threshold of 0.01.

Diego, CA) at Dana-Farber Cancer Institute. AR binding sites were generated using Model-Based Analysis of ChIP-seq 2 (MACS2), with a qvalue (false discovery rate, FDR) threshold of 0.01.

Model-Based Analysis of ChIP-seq 2 (MACS2), with a qvalue (false discovery rate, FDR) threshold of 0.01.
We confirmed that estimates of $\hat{h}^2$ from a polygenic trait were accurate under the model where causal variants are typed (Supplementary Table S1). Under the null, the LNCaP H3K27ac component is expected to explain 3.22% of the SNP heritability, and the model estimated 3.50% (0.22%) and 3.68% (0.21%) under a low-frequency and high-frequency disease architecture, respectively (Supplementary Table S1). None of the estimates were significantly different from the true proportions of components tested. This is not a scenario where LNCaP H3K27ac explains 50% of the $\hat{h}^2$, the model predicted 51.13% (0.40%) and 46.98% (0.35%) under a low-frequency and high-frequency disease architecture, respectively (Supplementary Table S1). Although the high-frequency architecture (where common variants explain more variance in trait than rare variants) represented a series of well-calibrated misspecifications, this does not introduce substantial bias and is likely to slightly underestimate the SNP heritability at focal chromatin mark. In all cases, the empirical standard deviation over 500 simulations was similar to the average analytical s.e.m. computed by GCTA (REML algorithm), thus showing that the analytical standard error is well-calibrated (Supplementary Table S1). We note that the standard error is inversely related to the sample size{sup}, and is therefore much higher in these simulations than in the iCOGS data which is 14-fold larger.

Lastly, we performed the real data partitioned analysis in subsets of individuals to evaluate bias and power to detect significant enrichment. We confirmed that no non-cancer variants were observed (Supplementary Table S2). The causal variant was the one expected (Supplementary Table S2). 83.6% of common UK10K SNPs lie within 100 kb of an iCOGS SNP, so some common variation is likely to be partially tagged by the chip. If the imputed and/or genotyped SNPs served as a good proxy for the common sequence variation, then we would expect their estimates of $\hat{h}^2$ to match the simulated fractions. When no functional category was enriched with causal variants, small but significant differences were observed for genotyped coding variants (4.75% $\hat{h}^2$ estimated as compared with simulated 0.67%) and imputed intergenic variants (56.09% $\hat{h}^2$ as compared with 50.52% simulated) but not the focal LNCaP H3K27ac category. Similar deviations were observed for the disease architecture where common variants explain more variance in trait than rare variants (Supplementary Table S2). The causal variant was the one expected (Supplementary Table S2). All predictions were carried out by cross-validation in the full iCOGS data, removing 1,000 individuals in each fold. Prediction $R^2$ was then computed from a regression of phenotype on the predictor score with 10 PCs included as covariates to account for ancestry, subsequently subtracting the $R^2 = 0.021$ from a model with no PCs. $P$ values were estimated for each of the coefficients in the multiple regression of phenotype $\sim$ GRS + single-BLUP + multi-BLUP + PCs. To ensure that prediction across data sets was independent, we carefully removed all iCOGS individuals with a GRM value of $>0.05$ to any individual in the BPC3 when computing BLUP coefficients. We separately analysed the predictor in 26,000 iCOGS samples that had age at diagnosis, but did not observe significant differences before/after including age as a covariate.

Causal variants not tagged on the iCOGS genotyping platform. We used the sequenced UK10k common variants to evaluate how well the iCOGS genotyped and imputed SNPs captured underlying heritability by simulating phenotypes using causal variants from sequencing and estimating heritability from the iCOGS SNPs (that is, ignoring variants that were not genotyped or imputed). Supplementary Table S2). 83.6% of common UK10K SNPs lie within 100 kb of an iCOGS SNP, so some common variation is likely to be partially tagged by the chip. If the imputed and/or genotyped SNPs served as a good proxy for the common sequence variation, then we would expect their estimates of $\hat{h}^2$ to match the simulated fractions. When no functional category was enriched with causal variants, small but significant differences were observed for genotyped coding variants (4.75% $\hat{h}^2$ estimated as compared with simulated 0.67%) and imputed intergenic variants (56.09% $\hat{h}^2$ as compared with 50.52% simulated) but not the focal LNCaP H3K27ac category. Similar deviations were observed for the disease architecture where common variants explain more variance in trait than rare variants (Supplementary Table S2). The causal variant was the one expected (Supplementary Table S2). All predictions were carried out by cross-validation in the full iCOGS data, removing 1,000 individuals in each fold. Prediction $R^2$ was then computed from a regression of phenotype on the predictor score with 10 PCs included as covariates to account for ancestry, subsequently subtracting the $R^2 = 0.021$ from a model with no PCs. $P$ values were estimated for each of the coefficients in the multiple regression of phenotype $\sim$ GRS + single-BLUP + multi-BLUP + PCs. To ensure that prediction across data sets was independent, we carefully removed all iCOGS individuals with a GRM value of $>0.05$ to any individual in the BPC3 when computing BLUP coefficients. We separately analysed the predictor in 26,000 iCOGS samples that had age at diagnosis, but did not observe significant differences before/after including age as a covariate.

References

1. Hjelmborg, J. B. et al. The heritability of prostate cancer in the Nordic twin study of cancer. Cancer Epidemiol. Biomarkers Prev. 23, 2303–2310 (2014).
2. Al Olama, A. A. et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat. Genet. 46, 1103–1109 (2014).
3. Castro, E. et al. Germline BRCA mutations are associated with higher risk of nodal involvement, distant metastasis, and poor survival outcomes in prostate cancer. J. Clin. Oncol. 31, 1748–1757 (2013).
4. Eeles, R. A. et al. Identification of 25 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat. Genet. 45, 385–391 (2013).
5. Saunders, E. J. et al. Fine-mapping the HOXB region detects common variants tagging a rare coding allele: evidence for synthetic association in prostate cancer. PloS Genet. 10, e1004129 (2014).
6. Ewing, C. M. et al. Germline mutations in HOXB13 and prostate-cancer risk. N. Engl. J. Med. 366, 131–141 (2012).
7. Hazlett, D. J. et al. Comprehensive functional annotation of 77 prostate cancer risk loci. PloS Genet. 10, e1004102 (2014).
8. Hazlett, D. J., Coetzee, S. G. & Coetzee, G. A. A rare variant, which destroys a FOXA1 site at 8q24, is associated with prostate cancer risk. Cell Cycle 12, 379–380 (2013).
9. ENCODE Project Consortium et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
10. Stamatoyannopoulos, J. A. What does our genome encode? Genome Res. 22, 1602–1611 (2012).
11. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190–1195 (2012).
12. Gusé, A. et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am. J. Hum. Genet. 95, 535–552 (2014).
13. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42, 565–569 (2010).
14. Yang, J. et al. Heritability estimation with genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82 (2011).
15. Yang, J. et al. Genome partitioning of genetic variation for complex traits using common SNPs. N. Engl. J. Med. 368, 1509–1515 (2013).
16. Schumacher, F. R. et al. Genome-wide association study identifies new prostate cancer susceptibility loci. Hum. Mol. Genet. 20, 3867–3875 (2011).
17. Haiman, C. A. et al. Characterizing genetic risk at known prostate cancer susceptibility loci in African Americans. PloS Genet. 7, e1001387 (2011).
18. Pemmaraju, M. et al. The androgen receptor cis-limited is extensively reprogrammed in human prostate tumorigenesis. Nat. Genet. 47, 1346–1351 (2015).
19. Shlyueva, D., Stemple, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. 15, 272–286 (2014).
20. Gusé, A. et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am. J. Hum. Genet. 95, 535–552 (2014).
21. Lee, S. H. et al. Estimation of SNP heritability from dense genotype data. Am. J. Hum. Genet. 93, 1151–1155 (2013).
22. Cancer Facts & Figures for African Americans 2009–2010. Accessed: December 2015.
27. Hoffman, M. M. et al. Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res. 41, 827–841 (2013).

28. Haiman, C. A. et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat. Genet. 39, 638–644 (2007).

29. Hoffman, M. M. et al. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. Nat. Methods 9, 473–476 (2012).

30. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. Cell 155, 934–947 (2013).

31. Farh, K. K. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518, 337–343 (2014).

32. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).

33. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am. J. Hum. Genet. 94, 559–573 (2014).

34. Williams, A. L., Patterson, N., Glessner, J., Hakonarson, H. & Reich, D. Phasing of many thousands of genotyped samples. Am. J. Hum. Genet. 91, 238–251 (2012).

35. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. G3 (Bethesda) 1, 457–470 (2011).

36. Schumacher, F. R. et al. Genome-wide association study identifies new prostate cancer susceptibility loci. Nature Genet. 43, 3867–3875 (2011).

37. Haiman, C. A. et al. Characterizing genetic risk at known prostate cancer susceptibility loci in African Americans. PLoS Genet. 7, e1001387 (2011).

38. Kolonel, L. N. et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am. J. Epidemiol. 151, 346–357 (2000).

39. Alexander Gusev1,2, Huwenbo Shi3, Gleb Kichaev3, Mark Pomerantz4, Fugen Li5,6, Henry W. Long4,5, Andrew A. Adjei15,16, Evelyn T. Taylor15,16, Ann Truelove17, Shelley Niwa17, Anand P. Chokkalingam18, Wei Zheng13, Curtis A. Pettaway14, Edward D. Yeboah15,16, Yao Tettey15,16, Richard B. Biritwum15,16, Esther M. John19,20, Adam B. Murphy21, Lisa B. Signorello1,22, John Carpten23, M. Cristina Leske11, Suh-Yuh Wu11, Anslern J.M Hennis11,24, Christine Neslund-Dudas10, Ann W. Hsing19,20, Lisa Chiu19,20, Phyllis J. Goodman25, Eric A. Klein26, John S. Witte27,28, Graham Casey7, Sam Kaggwa29, Michael B. Cook30, Daniel O. Stram7, Wei Zheng13, Curtis A. Pettaway14, Edward D. Yeboah15,16, Yao Tettey15,16, Richard B. Biritwum15,16, Esther M. John19,20, Adam B. Murphy21, Lisa B. Signorello1,22, John Carpten23, M. Cristina Leske11, Suh-Yuh Wu11, Anslern J.M Hennis11,24, Christine Neslund-Dudas10, Ann W. Hsing19,20, Lisa Chiu19,20, Phyllis J. Goodman25, Eric A. Klein26, John S. Witte27,28, Graham Casey7, Sam Kaggwa29, Michael B. Cook30, Daniel O. Stram7, and X51003); Award Number P30CA042014 from the National Cancer Institute. This work was supported by NIH fellowship F32 GM106584 (AG), NIH grants R01 MH101244(A.G.), R01 CA184537-01 (the GAME-ON initiative); Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007 and C5047/A10902), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537-01 (the GAME-ON initiative); Cancer Research UK (C1287/A10118, C1287/A10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007 and C5047/A10902), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148605 and 1U19 CA114812—the GAME-ON initiative), the Department of Defense (W81XWH-10-1-0341), A Linneus Centre (Contract ID 70867902), Swedish Research Council (grant no K2010-70X-20430-04-3), the Swedish Cancer Foundation (grant no 09-0677), grants ROICCA056678, ROICCA082664 and ROICCA092579 from the US National Cancer Institute, National Institutes of Health; US National Cancer Institute (R01CA72818); support from the National Health and Medical Research Council, Australia (126402, 209057, 251533, 396414, 450104, 504700, 504702, 504715, 623204, 409304 and 614296); NIH grants CA63464, CA54281 and CA098758; US National Cancer Institute (RO1CA128831, PI: J.Y. Park); Bulgarian National Science Fund, Ministry of Education and Science (contract DOO-119/2009; DUNKO1/2-2009; DNPB-01/28(2012); Cancer Research UK grants [C1897/A10123] and [C1897/A10865]; grant code G0500966/75466; NIH Health Technology Assessment Programme (projects 96/20/96 and 96/20/99); Cancer Research UK grant number CS2/A8649, Medical Research Council of England grant number G0500966, ID 75466 and The NCRI, UK; The US Dept of Defense award W81XWH-04-1-0280; Australia Project Grant [390130, 109948] and Enabling Grant [614296 to APCR]; the Prostate Cancer Foundation of Australia (Project Grant [PG7] and Research infrastructure grant [to APCR]); NIH grant R01 CA092447; Vanderbilt-Ingram Cancer Center (P30 CA68485); Cancer Research UK [C490/A10124] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge; Competitive Research Funding of the Tampere University Hospital (NNO09 and X51003); Award Number P30CA042014 from the National Cancer Institute.

Author contributions
A.G., L.P., M.L.F., C.A.H. and B.P. planned the study and wrote the paper. A.G., H.S., G.K. and B.P. performed primary analysis. All authors contributed to study design, data collection and analysis of individual data and annotations.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/ncermuscommunications

Competing financial interests: The authors declare no competing financial interests.

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

How to cite this article: Gusev, A. et al. Atlas of prostate cancer heritability in European and African-American men pinpoints tissue-specific regulation. Nat. Commun. 7:10979 doi: 10.1038/ncomms10979 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/
The PRACTICAL consortium

Margaret Cook33, Michelle Guy31, Kovela Govindasami31, Daniel Leongamornlert31, Emma J. Sawyer31, Rosemary Wilkinson31, Edward J. Saunders31, Malgorzata Tymrakiewicz31, Tokhir Dadaev31, Angela Morgan31, Cyril Fisher31, Steve Hazle31, Naomi Livni31, Artitaya Lophatananon34,35, John Pedersen103, John L. Hopper37, Jan Adolfsen39, Paer Stattin39, Jan-Erik Johansson39, Carin Cavalli-Bjerkman39, Ami Karlsson39, Michael Brooks39, Anssi Auvinen104, Paula Kujala105, Liisa Maeattaena106, Teemu Murtola107,108, Kimmo Taari109, Maren Weischer45, Peter Kaarskul410, Andreas Roder111, Peter Iversen111, Hans Wallinder112, Sven Gustafsson112, Angela Cox113, Paul Brown114, Anne George115, Gemma Marsden116, Athene Lane117, Michael Davis118, Wei Zheng119, Lisa B. Signorello115, William J. Blot116,117, Lori Tillmans118, Shaun Riska58, Liang Wang58, Antje Rinckleb60, Jan Lubisiski61, Christa Stegmaier118, Julio Pio-Sang68, Hyun Park68, Selina Radleb68, Maria Rincon68, James Haley68, Babu Zachariah68, Darina Kachakova71, Elenko Popov70, Atanaska Mitkova71, Aleksandra Vlahova76, Tihomir Dikov76, Svetlana Christova76, Peter Heathcote72, Glenn Wood72, Greg Malone72, Pamela Saunders72, Allison Eckert72, Trina Yeadon72, Kris Kerr72, Angus Collins72, Megan Turner72, Srilakshmi Srinivasan72, Mary-Anne Kedda72, Kimberly Alexander72, Tracy Omara72, Huihai Wu119, Rui Henrique74, Pedro Pinto74, Joana Santos74& Joao Barros-Silva74

103 Tissupath Pty Ltd., Melbourne, Victoria 3122, Australia. 104 Department of Epidemiology, School of Health Sciences, University of Tampere, Tampere 33014, Finland. 105 Fimlab Laboratories, Tampere University Hospital, Tampere, Finland. 106 Finnish Cancer Registry, Helsinki, Finland. 107 School of Medicine, University of Tampere, Tampere, Finland. 108 Department of Urology, Tampere University Hospital, Tampere, Finland. 109 Department of Urology, Helsinki University Central Hospital and University of Helsinki, Helsinki 00100, Finland. 110 Department of Urology, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, Herlev DK-230, Denmark. 111 Copenhagen Prostate Cancer Center, Department of Urology, Rigshospitalet, Copenhagen University Hospital, Tagensvej 20, 7251, Copenhagen DK-2200, Denmark. 112 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London SW7 2AZ, UK. 113 CR-UK/YCR Sheffield Cancer Research Centre, University of Sheffield, Sheffield S10 2TN, UK. 114 Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 800, Nashville, Tennessee 37232, USA. 115 National Cancer Institute, NIH, 9609 Medical Center Drive, Suite 2W-172, MSC 9712, Bethesda, MD 20892-9712 (mail), Rockville, Maryland 20850 (FedEx/Courier), USA. 116 International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. 117 Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA. 118 George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, Utah 84132, USA. 119 Division of Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany. 120 German Cancer Consortium (DKTK), Heidelberg 69120, Germany. 121 Department of Cancer Epidemiology, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, Florida 33612, USA. 122 Biostatistics Program, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, Florida 33612, USA. 123 Department of Urology and Alexandrovsky University Hospital, Medical University, Sofia 1431, Bulgaria. 124 Department of Medical Chemistry and Biochemistry, Molecular Medicine Center, Medical University, Sofia, 2 Zdrave Str., Sofia 1431, Bulgaria. 125 Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, Queensland 4000, Australia. 126 Molecular Cancer Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland 4000, Australia. 127 Department of Genetics, Portuguese Oncology Institute, Porto 4200, Portugal. 128 Biomedical Sciences Institute (ICBAS), University of Porto, Porto 4200, Portugal. 129 The University of Surrey, Guildford, Surrey GU2 7XH, UK. 130 Department of Preventive Medicine, Norris Cancer Center, University of Southern California Keck School of Medicine, Los Angeles, California 90033, USA. 131 Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institute of Health, Bethesda, Maryland 20892, USA. 132 Department of Radiation Oncology and Molecular Radiation Sciences, Johns Hopkins Medicine, Baltimore, Maryland 21287, USA. 133 Genomic Epidemiology Group, German Cancer Research Center (DKFZ), 69121 Heidelberg, Germany. 134 Urologic Oncology, University of Colorado at Denver Health Sciences Center, Denver, Colorado 80220, USA. 135 Epidemiology Research Program, American Cancer Society, Atlanta, Georgia 30303, USA. 136 Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. 137 Division of Aging, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. 138 Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. 139 Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii 96813, USA. 140 Department of Epidemiology, Regional Health Authority, Murcia 30009, Spain. 141 CIBER Epidemiologia y Salud Publica (CIBERESP), Barcelona 28029, Spain. 142 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London SW7 2AZ, UK. 143 Department of Genomics of Common Disease, School of Public Health, Imperial College London, London SW7 2AZ, UK. 144 Bureau of Epidemiologic Research, Academy of Athens, Athens 106 79, Greece. 145 Hellenic Health Foundation, Athens 106 79, Greece. 146 HuGeF Foundation, Turin 10126, Italy. 147 School of Public Health, Imperial College London, London SW7 2AZ, UK. 148 Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. 149 Welcome Trust Sanger Institute, Welcome Trust Genome Campus, Cambridge CB10 1SA, UK. 150 Institute of Inflammation and Repair, University of Manchester, Manchester M13 9PT, UK. 151 Departments of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, California 90095, USA. 152 Department of Human Genetics, University of California Los Angeles, Los Angeles, California 90095, USA. 153 A full list of consortium members appears at the end of the paper.
