H3K27ac HiChIP in prostate cell lines identifies risk genes for prostate cancer susceptibility

Claudia Giambartolomei,[1,2,11] Ji-Heui Seo,[3,4,11] Tommer Schwarz,[5] Malika Kumar Freund,[6] Ruth Dolly Johnson,[7] Sandor Spisak,[3] Sylvan C. Baca,[3] Alexander Gusev,[3] Nicholas Mancuso,[8] Bogdan Pasaniuc,[2,5,6,9,10,*] and Matthew L. Freedman[3,4,*]

Summary

Genome-wide association studies (GWASs) have identified more than 200 prostate cancer (PrCa) risk regions, which provide potential insights into causal mechanisms. Multiple lines of evidence show that a significant proportion of PrCa risk can be explained by germline causal variants that dysregulate nearby target genes in prostate-relevant tissues, thus altering disease risk. The traditional approach to explore this hypothesis has been correlating GWAS variants with steady-state transcript levels, referred to as expression quantitative trait loci (eQTLs). In this work, we assess the utility of chromosome conformation capture (3C) coupled with immunoprecipitation (HiChIP) to identify target genes for PrCa GWAS risk loci. We find that interactome data confirm previously reported PrCa target genes identified through GWAS/eQTL overlap (e.g., *MLPH*). Interestingly, HiChIP identifies links between PrCa GWAS variants and genes well-known to play a role in prostate cancer biology (e.g., *AR*) that are not detected by eQTL-based methods. HiChIP predicted enhancer elements at the *AR* and *NKX3-1* prostate cancer risk loci, and both were experimentally confirmed to regulate expression of the corresponding genes through CRISPR interference (CRISPRi) perturbation in LNCaP cells. Our results demonstrate that looping data harbor additional information beyond eQTLs and expand the number of PrCa GWAS loci that can be linked to candidate susceptibility genes.

Introduction

Prostate cancer (PrCa [MIM: 176807]) is the most common cancer in men, is the second leading cause of cancer deaths worldwide, and has a strong familial component.[1–4] The large genetic heritability component of PrCa has been attributed to the inheritance of rare high-risk genetic variants as well as a polygenic inheritance of multiple lower-risk common variants.[5–8] Genome-wide association studies (GWASs) have identified 269 common single-nucleotide polymorphisms (SNPs) associated with increased risk of PrCa.[9] The largest publicly available PrCa GWAS to date reported 147 predominantly non-coding genomic loci estimated to collectively explain 28.4% of familial risk.[10] With the vast majority of risk variants discovered by PrCa GWASs located outside protein-coding regions, the molecular mechanisms driving pathogenesis have only been described for a handful of loci.[11–16] GWAS loci predominantly colocalize with tissue-specific regulatory elements, thus supporting the hypothesis that risk variants exert their effects on disease by influencing the transcriptional levels of their target genes. For example, a substantial proportion of PrCa heritability lies in regions marked by H3K27ac,[17] a histone modification marking active enhancers and promoters.[18–20]

A standard approach to link variants to genes is through expression quantitative trait locus (eQTL) analysis, which identifies genotypes that correlate with transcript levels across individuals in a target tissue of interest.[21–23] Overlapping eQTLs with variants identified by GWASs is a powerful tool to prioritize target genes for further functional investigation.[21,22,24–29] However, eQTLs have several limitations. They are primarily studied under steady-state conditions, and therefore eQTLs that are elicited under specific conditions are missed. For example, context-specific eQTLs have been identified after stimulation with interferon-γ and bacterial lipopolysaccharides in monocytes.[30] Second, statistical power for eQTL identification is dependent on sample and effect size, so many eQTLs are undetected (i.e., a false negative), particularly with hard-to-collect tissues and/or eQTLs in a rare cell type. Despite some PrCa GWAS loci mapping near well-known prostate cancer biology genes, such as *FOXA1* (MIM: 602294), *GATA2* (MIM: 137295), *AR* (MIM: 313700), and *MYC* (MIM: 190080), the absence of strong eQTL associations with these transcripts is notable. This

[1] Central RNA Lab, Istituto Italiano di Tecnologia, Genova 16163, Italy; [2] Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [3] Department of Medical Oncology, The Center for Functional Cancer Epigenetics, Dana Farber Cancer Institute, Boston, MA 02215, USA; [4] The Center for Cancer Genome Discovery, Dana Farber Cancer Institute, Boston, MA 02215, USA; [5] Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA 90095, USA; [6] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [7] Department of Computer Science, University of California, Los Angeles, Los Angeles, CA 90095, USA; [8] Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90032, USA; [9] Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [10] Johnson Comprehensive Cancer Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA

[11] These authors contributed equally

*Correspondence: mfreedman@partners.org (M.L.F.), pasaniuc@ucla.edu (B.P.)

https://doi.org/10.1016/j.ajhg.2021.11.007.

suggests that traditional eQTL/GWAS overlap may fail to detect important susceptibility genes for PrCa.

Chromosome conformation capture (3C)-based assays have recently emerged as powerful tools to evaluate physical interactions between regulatory elements and target genes in the context of GWASs, including prostate cancer risk,[31–33] and can be used as an orthogonal method to validate eQTL-based findings. Efficiencies of 3C-based techniques can be increased by enriching for interactions via chromatin immunoprecipitation against proteins, such as H3K27ac, a marker of active promoter/enhancer activity (HiChIP)[34–36] or through nucleic acid hybridization.[37]

Notably, data suggest that loops at enhancer-promoter contacts pre-exist at stimulation-responsive genes prior to the stimulus.[38–40] In other words, stimulation does not induce significant *de novo* looping even at genes that are transcriptionally responsive in differentiated cells. These observations raise the provocative hypothesis that looping (measured even in the steady state) can identify GWAS target genes beyond traditional GWAS/eQTL overlap; that is, looping could reveal eQTL-target gene relationships that are observable only under certain contexts (e.g., cell type) and/or capture eQTLs with small, but important, effects. Recent data suggest that the majority (>80%) of common loops identified in primary benign prostate tissue and prostate tumors are overlapping, further indicating that *de novo* looping is not a major driver of transcriptional differences in differentiated tissue.[41] While the current data suggest that loops are relatively static, more experiments across a wider variety of conditions are needed to understand the generalizability of these observations.

We utilized high-resolution H3K27ac-HiChIP data in the LNCaP prostate cancer cell line, the most widely used *in vitro* model of prostate cancer, to identify links between target genes and PrCa risk loci. We linked 98 out of 130 known susceptibility loci for PrCa with 665 genes and observed a significant overlap between eQTL-target gene pairs and loops. Notably, we identified looping overlapping candidate PrCa causal variants from GWASs and genes with established roles in PrCa biology at eQTL-negative loci. We used CRISPR interference (CRISPRi) to functionally validate two enhancer-promoter interactions for *NKX3-1* (MIM: 602041) and *AR*. Overall, our results confirm that 3C-based strategies can not only validate eQTLs but can also discover important target gene links not detected by current GWAS/eQTL strategies.

Material and methods

Cell culture and DHT treatment

Hormone-depleted LNCaP cells (ATCC CRL-1740) were grown in phenol red free RPMI (#11835030, GIBCO) with 10% charcoal stripped FBS (#100-119, GemBio) for 3 days and then were stimulated with 10 nM dihydrotestosterone (DHT) (5α-Androstan-17β-ol-3-one, A8380, Sigma) for either 4 h or 16 h. For the vehicle treatment group, the cells were treated with the same amount of 100% EtOH used to make 10 nM DHT for 16 h. Subsequently, cells were collected for further analysis accordingly. LNCaP cells were authenticated by sequencing and comparing short tandem repeats to parental LNCaP cells in ATCC database. Prior to experiments, cells have been tested for several strains of mycoplasma contamination with LookOut Mycoplasma PCR Detection Kit (Sigma-Aldrich #D9307).

H3K27ac ChIP-seq in LNCaP

H3K27ac ChIP in LNCaP was performed as previously described.[42] 10 million cells were fixed with 1% formaldehyde at room temperature for 10 min and quenched. Cells were collected in lysis buffer (1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS and protease inhibitor [#11873580001, Roche] in PBS).[43] Chromatin was sonicated to 300–800 bp with Covaris E220 sonicator (140PIP, 5% duty cycle, 200 cycle burst). H3K27ac antibody (C15410196, Diagenode, 1:600 ratio) was incubated with 40 μL of Dynabeads protein A/G (Invitrogen) for at least 6 h before immunoprecipitation with the sonicated chromatin overnight. Chromatin was washed with LiCl wash buffer (100 mM Tris [pH 7.5], 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate) six times for 10 min each time. Eluted sample DNA was prepared as the sequencing libraries with the ThruPLEX-FD Prep Kit (Rubicon Genomics). Libraries were sequenced with 150-base pair single reads on the Illumina platform (Illumina) at Novogene. For further analyses, we used the union of ChIP sequencing (ChIP-seq) narrow and broad peaks in regular media. This comprises 49,638 ChIP-seq peaks in LNCaP cells (length of peaks ranged from 146 to 129,126 bases). 43,335 out of 49,638 peaks overlap HiChIP anchors.

Peak calling

ChIP-seq was processed through the ChiLin pipeline.[44] Briefly, Illumina Casava1.7 software used for base calling and raw sequence quality and GC content was checked with FastQC (version 0.10.1). We used the Burrows–Wheeler Aligner (BWA, version 0.7.10) to align the reads to human genome hg19. Then, MACS2 (v. 2.1.0.20140616) was used for peak calling with a false discovery rate (FDR) q value threshold of 0.01. Bed files and Bigwig files were generated with bedGraphToBigWig for H3K27ac. The union of H3K27ac narrow and broad peaks was used in the downstream analyses. The following quality metrics were assessed for each sample: (1) percentage of uniquely mapped reads;, (2) PCR bottleneck coefficient to identify potential over amplification by PCR; (3) FRiP (fraction of non-mitochondrial reads in peak regions); (4) peak number; (5) number of peaks with 10-fold and 20-fold enrichment over background; (6) fragment size; (7) percentage of the merged peaks with promoter, enhancer, intron, or intergenic region; and (8) peak overlap with DNase I hypersensitivity sites. For datasets with replicates, ChiLin calculates the replicate consistency with two metrics: (1) Pearson correlation of ChIP-seq reads across the genome with UCSC software wigCorrelate after normalizing signal to reads per million and (2) percentage of overlapping peaks in the ChIP replicates.

Differential gene expression analysis

RNA sequencing (RNA-seq) data were processed with the VIPER pipeline.[45] Reads were aligned to the hg19 human genome build with STAR.[46] Fragments per kilobase of transcript per million mapped reads (FPKM) values were calculated with Cufflinks[47] for 20,114 RefSeq genes included in the VIPER repository. We performed differential expression (DE) analyses with the DESeq2 R package[48] by using supervised analysis based on gene expression levels (counts from STAR) and cutoffs of FDR-adjusted p value (padj) < 0.05 and log2 fold-change > 1. DHT treatment of LNCaP for 4 or 16 h was compared to vehicle treatment (two replicates each). We called loops at FDR 1% and annotated genes to loops, following the procedure described below. The total numbers of loops called in 4 and 16 h were 98,960 and 183,958, respectively. The total numbers of genes annotated in 4 and 16 h from HiChIP loop data were 17,649 and 20,899, respectively. We considered only the genes in common between LNCaP and 4 h sample (Ngenes = 15,630) and between LNCaP and 16 h sample (Ngenes = 16,979).

H3K27ac HiChIP in LNCaP

HiChIP was performed mainly following an established procedure:[49] trypsinized 10 million LNCaP cells were fixed with 1% formaldehyde at room temperature for 10 min and quenched. Sample was lysed in HiChIP lysis buffer and digested with MboI (NEB) for 4 h. After 1 h of biotin incorporation with biotin dATP, the sample was ligated with T4 DNA ligase for 4 h and chipped with H3K27ac antibody (DiAGenode, C1541019) after chromatin. Reverse-crossed IP sample was pulled down with streptavidin C1 beads (Life Technologies) treated with Transposase (Illumina) and was amplified with reasonable cycle numbers based on the qPCR with 5-cycle pre-amplified library. Library was sequenced with 150-base pair end reads on the Illumina platform (Novogene). All of these libraries were generated with the 4-bp cutter MboI restriction fragment. We first trimmed the raw fastq files (paired-end data) to remove adaptor sequences by using Trim Galore.[50] We used HiC-Pro version 2.9.0[51] to align the reads to the hg19 human genome, assign reads to MboI restriction fragments, and remove duplicate reads. We used the following options: MIN_MAPQ = 20, BOWTIE2_GLOBAL_OPTIONS = –very-sensitive–end-to-end–reorder, BOWTIE2_LOCAL_OPTIONS = –very-sensitive–end-to-end–reorder, GENOME_FRAGMENT = MboI_resfrag_hg19.bed, LIGATION_SITE = GATCGATC, LIGATION_SITE = "GATCGATC," BIN_SIZE = "5000." All other default settings were used. The HiC-Pro pipeline selects only uniquely mapped *valid* read pairs involving two different restriction fragments to build the contact maps.

We applied FitHiChIP version 5.1[52] for bias-corrected peak calling and DNA loop calling. FitHiChIP models the genomic distance effect with a spline fit, normalizes for coverage differences with regression, and computes statistical significance estimates for each pair of loci. We used the FitHiChIP loop significance model to determine whether interactions are significantly stronger than the random background interaction frequency. We used 49,638 regions from H3K27ac LNCaP union of narrow and broad peaks as anchors to call loops. We used a 5 kb resolution and considered only interactions between 5 kb and 3 Mb. We used the peak to all for the foreground, meaning at least one anchor needed to be in the H3K27ac peak rather than both. The corresponding FitHiChiP options specification is "IntType=3." For the global background estimation of expected counts (and contact probabilities for each genomic distance), FitHiChIP can use either peak-to-peak (stringent) or peak-to-all (loose) loops for learning the background and spline fitting. We specified the suggested option to merge interactions close to each other to represent a single interaction when their originating bins are closer. The corresponding FitHiChiP options specifications are "UseP2PBackgrnd=0" and "MergeInt=1" (FitHiChIP(L + M)). We used the default FitHiChIP q value $<$ 0.01 to identify significant loops. For comparisons across replicates, we used the results not merged MergeInt = 0, as suggested by the authors. The length considered was between 5 kb and 3 Mb. We explored reproducibility across replicates in the following way. We have processed five biological replicates separately by using the FitHiChIP pipeline as well as all replicates together in a dataset called "merged" made up of the combined reads across the replicates. We used the q value < 0.01 cutoff to define high confidence loops and compared the level of accuracy achieved by one replicate and the merged data versus our high-confidence loops (i.e., the proportion of reference loops, reported by one replicate, that are captured at differing number of loop calls from other replicates, or from our combined library). The final number of significant loops (q value < 0.01) considered in these analyses are those using background 0 and merged FitHiChIP settings.

Mapping loops to enhancers and promoters

For loop annotations, we first extended loop anchors by 5 kb on either side. To identify potential gene targets, we defined promoter regions around the transcription start site (TSS) (± 500 bases) for 27,063 genes by using RefSeq hg19; 27 genes were removed because of ambiguous positions. We used the longest transcript to define TSS based on strand. To define enhancers, we used the subset of 49,638 regions from H3K27ac LNCaP in regular media (union of narrow and broad peaks). We then labeled the promoters and enhancer regions that overlap either right or left anchors and considered a loop as E-E if both the anchors overlap an enhancer region, a loop as P-P if both the anchors overlap a promoter region, a loop as E-P if only one anchor overlaps a promoter and the other an enhancer region, and a loop as E-O or P-O if one anchor overlaps a promoter or enhancer and the other overlaps a region without an H3K27ac or a TSS (± 500 bases).

Correlation of gene expression with looping

We considered only loops involving promoters (E-P, P-P, or P-O). In cases where there were duplicated loops (supporting the same two anchors but in opposite directions), we summed the paired-end tag (PET) read counts. For each gene, we considered two measures of gene connectivity: (1) the number of loops with one anchor overlapping the promoter (the opposing anchor can overlap a promoter, an enhancer, or neither ["other"]) and (2) the sum of PET counts across all loops overlapping the promoter. We compute a Spearman correlation between the expression across genes with both measures of connectivity. For the expression of every 13,274 genes that have looping and expression data, we averaged the FPKM value across two LNCaP RNA-seq replicates. FPKM was converted to transcripts per kilobase million (TPM) by dividing the each FPKM value by the sum of all FPKM values of the respective sample and multiplying by 1e6. We divided the genes into ten expression bins of equal size. We compare this to the log of the counts of loops (Figure S3A) and the log of the counts of PETs (Figure S3B). Spearman correlations were computed between average TPM and log of counts of loops or PETs. To estimate the amount that each loop/PET contributes to expression, we fitted a linear regression (with or without adjusting for H3K27ac): (1) expression (TPM) ~loops/PETs; (2) expression (TPM) ~loops/PETs + H3K27ac. The H3K27ac level per gene is extracted by overlapping H3K27ac broad peaks scores with each gene promoter. If more than one peak overlaps the region, we sum across scores. The coefficient in the model represents the per-loop/per-PET contribution to expression.

Comparison to the ABC model

We downloaded the positive enhancer–promoter pairs in LNCaP from Fulco et al.[53] The file contained 12,641 genes and 37,079 element-gene pairs with positive predictions of ABC model (ABC score $\geq$ 0.022), and only distal (non-promoter) elements are included. To compare our results with the ABC model, we only considered 35,749 E-P loops in our data, encompassing 13,787 promoters and 48,360 enhancer-promoter pairs (because each loop can connect more than one promoter), and there was a mean (median) of 4.8 (3) gene promoters contacted per enhancer. 9,333 genes were identified in both analyses as having a link with an element, and of these, 5,969 genes had overlapping enhancer elements.

We stress that there are differences in how an enhancer and a promoter are defined in the two methods. The enhancer anchors in this analysis are of length 15 kb, while the enhancer elements in the ABC model range from 500 to 3,507 bases (ABC elements are defined as ~500 bp regions centered on DNase hypersensitivity site [DHS] peaks). The promoter is defined with a region of 1 kb around the TSS in this paper, while the ABC model uses a padding of 500 bp and expression information to filter the data; lastly, the ABC model is designed for Hi-C datasets and uses information on activity additionally to the contact frequency.

Credible sets of causal variants from prostate cancer GWAS

We used 147 SNPs previously reported[10] to define regions for fine-mapping with GWAS summary statistics from Schumacher et al.[54] (N = 79,148 cases and 61,106 controls) across 20,370,946 SNPs. SNPs were compared by chromosome and position to the 1000 Genomes phase 3 for allele and rsIDs matching. After filtering by minor allele frequency (MAF) $\geq$ 0.001, 15,818,179 SNPs remained and 20,155 SNPs passed genome-wide significance ($p < 5e-8$). A ~150 kb window (adjusted manually) around the 147 index SNPs, which resulted in 137 regions after merging overlapping regions on chromosomes 4, 8, and 17, was considered. In six of the regions (chr18.51504059.51971031.rs8093601,
chr1.9945583.10662959.rs636291,
chr6.170305546.170805546.rs138004030,
chr7.1694537.2194537.rs527510716,
chr9.123429584.124429584.rs1571801,
chr9.21791998.22291998.rs17694493), a p value of genome-wide significance (p value $< 5e-8$) was not reached, so we do not consider these regions further.

The final number of PrCa risk loci considered was 130. We used PAINTOR,[55] a Bayesian statistical method, with no functional annotations and specifying a maximum of 1 causal SNP, to fine-map 129 regions (excluding *MYC*, which was previously fine-mapped). We then constructed a 95% credible set for the most likely causal variants by taking the cumulative sum of the posterior probability until a cumulative 95% posterior probability was reached. For the highly complex 8q24 region, since this region was deeply fine-mapped in previous efforts, we used 174 SNPs included in the 95% credible set from the JAM fine-mapping (Data S3 of Matejcic et al.[12]). As expected, our fine-mapping method assuming 1 causal variant identifies 1 SNP in the 95% credible set for each of the two regions considered: rs183373024 (posterior probability 1) and rs10090154 (posterior probability 0.97). Since it is very likely that multiple regions within 8q24 independently affect risk for prostate cancer, we preferred to use the available results from previous fine-mapping efforts not assuming one causal variant. The final probable causal set contained 3,243 (3,069 from PAINTOR, 174 from JAM) SNPs across 130 PrCa risk regions. We considered a "PrCa SNP" a SNP that is part of the 95% credible causal variants from fine-mapping across 130 PrCa regions by using the largest publicly available PrCa GWAS to date.[54]

We considered "genes with evidence for PrCa risk," any gene anchor in HiChIP data with either anchors (the gene anchor or the opposite anchor) overlapping a PrCa SNP.

104 PrCa loci overlapped 1,953 LNCaP loop anchors and 665 genes, connecting 2,016 fine-mapped SNPs.

Alternative fine-mapping algorithm

To explore the sensitivity of our pipeline to the fine-mapping method used to identify the PrCa SNPs, we applied sum of single effects (SuSiE)[56] to 137 associated regions and applied our pipeline from the start by using the 95% credible set SNPs from this analysis. To run SuSiE, we used the LD information of only SNPs (bi-allelic-only strict) from 503 EUR individuals in the 1000G data (see web resources).

To calculate SuSiE credible sets, we used the following command in R CRAN package susieR:

susieR::susie_rss, c(list(z = gwas_subset$Z, R = R, z_ld_weight = 1/500, max_iter = 500), L = 5, estimate_prior_variance = TRUE, track_fit = TRUE, min_abs_corr = 0.1, tol = 1e-3).

We note that for several regions the algorithm does not converge with these specifications ("IBSS algorithm did not converge in *n* iterations") possibly because of the limited number of iterations and limited number of individuals for the LD used, so we limited our observations to the regions that converged. A thorough comparison of the fine-mapping methods PAINTOR[55] and SuSiE[56] was out of the scope of this paper.

The SuSiE algorithm converged for 101 out of the 130 PrCa loci. We extracted the SuSiE 95% credible sets, and we have re-annotated our HiChIP data in these regions by using the same procedure for the SuSiE fine-mapped.

1,818 SNPs with PAINTOR and 1,020 SNPs with SuSiE were in the 95% set across the 101 regions. After crossing with the HiChIP data, SuSiE identified 391 while PAINTOR identified 488 genes whose promoter overlap a 95% set SNP across 77 regions. 333 genes are prioritized in both methods. SuSiE identifies 58 genes not prioritized by PAINTOR.

MAGMA score for prioritized HiChIP genes

Hi-C coupled multimarker analysis of genomic annotation MAGMA (H-MAGMA) is a gene-based analysis tool that assigns genes to traits and incorporates information on chromatin interactions.[57] We used H-MAGMA v.1.08. As input, H-MAGMA required GWAS summary statistics (from Schumacher et al.[54]) and reference data with similar ancestry (1000G European samples was used) to estimate LD between SNPs. For the SNP-gene annotations, we used the 665 genes identified with a SNP from our fine-mapping analysis linking to a promoter in the HiChIP data. H-MAGMA p values could be computed for 628 genes out of the 665 genes. As expected, the p value from MAGMA is correlated with the minimum p value obtained in the fine-mapped region ("minPVAL_gwasregion" and "P.magma" in Table S8, Spearman 0.75, p value $< 2.2e-16$).

SNP-heritability enrichment for loop types

To estimate enrichment of PrCa risk heritability across functional categories, we ran stratified linkage disequilibrium score (LDSC) regression by using PrCa GWAS summary statistics from Schumacher et al.[54] First, we created custom bed tracks by using H3K27ac peaks called from ChIP-seq in LNCaP. We used the custom bed tracks to annotate SNPs genome wide, indicating their overlap. We computed annotation-specific LDSCs by using the above annotations with the baseline model containing 53 functional annotation[58] and estimated functional enrichments of heritability by using sLDSC.[59]

eQTL in prostate tissues

eQTL results on two prostate-specific datasets were used: Thibodeau eQTLs in prostate normal tissue[60] and TCGA eQTLs in prostate tumor tissue.[61] The Thibodeau dataset contains gene expression from tumor-adjacent normal prostate tissue on 471 individuals[60] on autosomal and X chromosomes. The TCGA dataset contains gene expression on 378 prostate cancer samples only on autosomal chromosomes.[61] We used MatrixEQTL[62] and RNA-seq and genotype data to conduct the gene expression linear regression association for each study datasets. We computed the *cis*-eQTLs by using a window of 3 Mb from the TSS around the RefSeq genes to match the HiChIP loop analysis. For the Thibodeau dataset, we tested 51,232,032 SNP-eQTL pairs and 15,673 genes, and we used a Bonferroni threshold of 0.05/51,232,032 to define "eQTLs" and "eGenes." After the Bonferroni cut-off, there were 85,435 significant eQTLs and 4,747 eGenes. For the TCGA dataset, we tested 219,256,418 SNP-eQTL pairs and 15,723 genes, and we used a Bonferroni threshold of 0.05/219,256,418 to define "eQTLs" and "eGenes." There were 86,555 significant eQTLs and 1,118 eGenes. Across both the datasets, we found 4,871 eGenes. To obtain the list of eQTL-eGene pairs, for each gene we selected the SNP with the most significant p value.

Genes with somatically acquired mutations in PrCa

Genes somatically mutated in prostate cancer were extracted from three publications: an exome sequencing study looking at both localized and advanced prostate cancer[63] (Table S4 "Known in prostate cancer and Recurrently altered in cancer" and Table S6 "cancer_pathways_mutation"); a study looking at recurrent alterations in primary (localized) prostate cancer[64] (Table S1C "SigMutated" and Table S1D "genes in wide peak"—we only took the gene highlighted in bold—and Table S1E "Fusions"); and an analysis describing recurrent alterations in metastatic prostate cancer[65] (Table S5 "SigMutated"). Together, these sources identified 122 unique genes, of which 119 were in RefSeq and also considered in our analyses (*MRE11A*, *FL1*, and *WHSC1L1* were missing). 43 genes were within 3 Mb of a credible risk SNP for PrCa.

Gene expression data

Genes lists associated with gene expression were retrieved from the sources listed in the web resources and manipulated as follows:

(1) 908 genes generally expressed in prostate tissue from GTEx[66] (TPM > 100 TPM). We downloaded GTEx median gene-level TPM by tissue from the GTEx website and selected genes in RefSeq with a TPM > 100.

(2) 803 genes from tissue-specific expression (top 10% t.stat from Finucane et al.[67]). The file contains t statistics comparing each gene in the particular tissue to all other tissues with 24,842 genes in GTEx. We used only the 18,026 RefSeq genes. We ordered the genes on the basis of absolute value of the t statistics in prostate and took the top 10%.

(3) 2,384 genes from differential gene expression tumor/normal.[68] Differential gene regulation in prostate tissues was detected with GEPIA, which uses the TCGA and GTEx projects databases to compare gene expression between tumor and normal tissues under Limma, both under- and overexpressed. We used the default thresholds of logFC of 1 and q value cut-off of 0.01.

Over-representation of eQTL-eGene links in HiChIP loops

eGenes that overlapped HiChIP were considered: n = 3,837 for Thibodeau and n = 870 for TCGA. For every gene, we chose a random SNP from the same window that eQTL was computed from (i.e., 3 Mb for Thibodeau and TCGA data) and computed how many times this overlaps a HiChIP anchor. We constructed 100 random control loop datasets by randomly flipping anchor1 (in 50% of the loops) or anchor2 (in the other 50%) and quantified eQTL-eGenes that are supported by HiChIP loops in real data versus the random data. We find the empirical p value by comparing the proportion of eQTL-eGenes supported by HiChIP loops in the actual versus the 100 artificial loop datasets. We also tested enrichment of loops containing an eQTL on one anchor but with the eGene that is not overlapping the other anchor and computed a fold change compared to random loops as described above.

GWAS/eQTL overlap

We tested for colocalization of GWAS and eQTL by using COLOC with default parameters.[69] COLOC TCGA tested 7,001 genes, and 11 genes (9 of the 130 regions) had posterior probability of a shared causal variant between eQTL and GWAS data (PP4) $\geq$ 0.75. COLOC Thibodeau tested 7,148 genes, and 42 genes (33 of the 130 regions) had PP4 $\geq$ 0.75. In total, 46 unique genes had evidence of colocalization. 32 unique genes (three in TCGA, 22 in Thibodeau, seven in common across the two datasets) were also eGenes under a strict Bonferroni threshold (see material and methods above). We used the multi-tissue transcriptome-wide association (TWAS) for PrCa from Mancuso et al.[21] This includes 892 significant TWAS associations (TWAS.p < 0.05/109170) in 45 tissues covering 217 genes (N = 4,458), including normal and tumor prostate. Restricting to only the genes in RefSeq, this includes 651 TWAS associations in 170 genes. We restrict to prostate-specific results, specifically the results in "TCGA.PRAD_SP.TUMOR," "TCGA.PRAD.TUMOR," and "GTEx.Prostate," which includes 190 signals and 74 unique genes, and 42 were also eGenes under a strict Bonferroni threshold. In total, 101 unique genes had evidence from COLOC or TWAS, and 74 were also eGenes. 29 genes (COLOC) and 40 genes (TWAS) were also HiChIP genes (n = 17,690), and 24 genes (COLOC) and 26 genes (TWAS) were also HiChIP genes looping to a 95% credible SNP (n = 665) (Table 1).

CRISPR-dCas9-mediated repression and gene expression analysis

Stable dCas9-KRAB expressing LNCaP cell line was created with lenti-KRAB-dCas9-blast lentiviral vector (Addgene, 89567). Antibiotic selection was performed (6 µg/mL blasticidin) for 2 weeks. gRNAs were designed against active epigenetic region containing

Table 1. **Genes where HiChIP looping data link their promoters to PrCa risk variants**

Genes for which looping links their promoters to prostate cancer (PrCa) risk variants with extra evidence from eQTL/GWAS overlap or somatically acquired in PrCa	
COLOC positive only	*CASP10* (MIM: 601762), *COL2A1* (MIM: 120140), *COMMD7* (MIM: 616703), *CTSK* (MIM: 601105), *DNMT3B* (MIM: 602900), *HNF1B* (MIM: 189907), *LARP4B* (MIM: 616513), *MSMB* (MIM: 157145), *PPFIBP2* (MIM: 603142), *RAD9A* (MIM: 603761), *SNRPC* (MIM: 603522)
TWAS positive only	*ACAT2* (MIM: 100678), *C10orf95*, *CNTROB* (MIM: 611425), *FAM84B* (MIM: 609483), *MLPH* (MIM: 606526), *NOL10* (MIM: 616197), *RGS17* (MIM: 607191), *SESN1* (MIM: 606103), *SPINT2* (MIM: 605124), *TSPO* (MIM: 109610), *USP20* (MIM: 615143), *USP39* (MIM: 611594), *ZGPAT* (MIM: 619577)
COLOC and TWAS positive	*C9orf78* (MIM: 619569), *CTBP2* (MIM: 602619), *FAM57A* (MIM: 611627), *GEMIN4* (MIM: 606969), *HAUS6* (MIM: 613433), *KRT8* (MIM: 148060), *MBNL1* (MIM: 606516), *MMP7* (MIM: 178990), *MYO9B* (MIM: 602129), *NCOA4* (MIM: 601984), *PPP1R14A* (MIM: 608153), *UHRF1BP1* (MIM: 619570), *VPS53* (MIM: 615850)
With somatically acquired mutations in PrCa	*AR* (MIM: 313700), *CCND1* (MIM: 168461), *CDKN1B* (MIM: 600778), *HD3* (MIM: 605166), *CIC* (MIM: 612082), *ERF* (MIM: 611888), *GATA2* (MIM: 137295), *KMT2D* (MIM: 602113), *MAP2K1** (MIM: 176872), *MED12* (MIM: 300188), *MYC* (MIM: 190080), *NKX3-1* (MIM: 602041), *PTEN* (MIM: 158350), *RNF43* (MIM: 612482), *RPRD2** (MIM: 614695), *SETDB1* (MIM: 604396), *TP53* (MIM: 191170), *ZMYM3* (MIM: 300061)

Out of the 665 genes linked by looping to PrCa risk variants (Figure 3), we list genes with evidence of eQTL/GWAS overlap (37) or somatically acquired mutations in PrCa (18); see Table S6 for all RefSeq genes. For 37 genes, looping links their promoter to PrCa GWAS variants, and they also show evidence of eQTL/GWAS overlap through colocalization (COLOC) and/or transcriptome-wide association (TWAS). 18 out of the 119 genes previously reported to have somatically acquired mutations in PrCa show loops linking their promoters to PrCa germline risk variants. The asterisks denote eGenes.

(rs1160267 [NKX3-1] and rs5964602 [AR]) genetic variants, and gRNA efficiency score was calculated and ranked.[70] Due to the PAM restriction (NGG) of the SpCas9 system, the design of gRNA properly targeting the genetic variants was challenging. Therefore, we designed gRNAs against the active epigenetic regions (four gRNAs per each peak) containing genetic variants to test their target gene regulatory potential. Non-human genome targeting negative control and *HPRT1* (MIM: 308000) promoter targeting positive control gRNAs were also selected. gRNA cassettes were synthesized (Integrated DNA Technologies) and cloned into lentiGuide-Puro (Addgene, 52963) vector. All gRNA sequences are listed in Table S10. LNCaP cells stably expressing KRAB-dCas9 were then subsequently infected with gRNA vectors and selected with 2 μg/mL puromycin for 5 days. For gene expression, qRT-PCR 500 ng total RNA (Macherey-Nagel) was reverse transcribed (High Capacity Reverse transcription kit, Life Technologies) and cDNA was diluted (20×). SYBR Green assay was performed on Light Cycler 480 instrument (2× Probe Master Mix, Roche). All primer sequences are listed in Table S10. Relative gene expression was calculated based on the ddCT (delta-delta-CT, dCT [sample] – dCT [control average]) method.[71] Each sample was measured by two biological and three technical replicates. We used *Beta-actin* (*ACTB* [MIM: 102630]) as a housekeeping gene to normalize gene expression among the samples.

CRISPRi data and analysis

The first loop we tested is contained in the fine-mapped region chr8: 23,283,623–23,783,623 (index SNP of the fine-mapped region is rs2928679) and is composed of an enhancer anchor positioned at chr8: 23,515,000–23,530,000 and the promoter anchor positioned at chr8: 23,530,000–23,545,000 overlapping the *NKX3-1* promoter. Specifically, the SNP rs1160267 (p value of association with PrCa 1.3e−58) is located in the enhancer anchor of the loop that connects the NKX3-1 promoter and overlaps an active epigenetic region. No other gene is linked through this loop. We note that we only validated the one enhancer-promoter loop, but *NKX3-1* has 13 other loops connecting the promoter (see Table S8, column “nLoops”). *NKX3-1* has weak evidence for eQTL association. The SNP rs4872175 is associated with prostate cancer in the GWAS data (p value 6.7e−58) and is also associated to gene expression (p value = 1.2e−5). Although it does not pass our threshold for calling this an eGene (material and methods), the evidence of a shared causal variant between the eQTL and the GWAS signal is high for this gene (the probability of colocalization between GWAS and eQTL data at this locus is 90%).

The second locus/loop we tested is contained in the fine-mapped region chrX: 66,608,321–67,108,321 (index SNP of the fine-mapped region is rs5919432) and connects an enhancer anchor positioned at chrX: 66,735,000–66,750,000 and the promoter anchor positioned at chrX: 66,755,000–66,770,000 overlapping the *AR* promoter. Specifically, SNP rs5964602 (p value of association with PrCa 4.4e–13) falls within the enhancer anchor and connects to the *AR* promoter. No other gene is overlapped in this loop. *AR* has no evidence of eQTL association.

We used the ddCT method[71] to determine relative gene expression alterations from cycle threshold (CT) values of the qPCR. Briefly, for each sample, we used the average of the housekeeping gene (*ACTB*) to calculate the dCT (delta-CT, CT [gene of interest] – CT [housekeeping gene]) values for each of the three technical replicates, and we computed the average dCT values for each sample. After this, using the control sample average, we computed the ddCT values for each replicate. This shows the relative deviation of each sample from the control condition. We then computed the expression values for each condition by using this formula: $\exp = 2^{-\mathrm{ddCT}}$. Finally, to compute the effects and the SEs, we combined the averages of the expression values for each sample across the two biological replicates by using a fixed effect meta-analysis. We then used a t test to compare each sample to the control.

Results

A high-resolution H3K27ac-HiChIP chromatin contact map for LNCaP

We performed H3K27ac HiChIP in LNCaP across five biological replicates and identified 126,280 loops (FitHiChIP, FDR < 0.01, material and methods and Figure 1). We called between 14,000 and 43,000 loops from a read depth ranging from 183,000 reads to 235,000 reads across individual replicates (Table S1). We observed that loop count and loop length depended on read depth. Replicates 1 and 5 had the highest number of high-quality uniquely mapped read pairs and final number of loops, and therefore the highest median loop length and PET counts (Figures S1A and S1B). As expected, the number of loops decreased with increasing loop distance (Figure S1B). The intersection of significantly called loops across the five replicates ranged from 20% to 60%. Once the individual replicates data was merged, each replicate shared 90% or more significantly called loops with the merged data (Table S2 and Figure S2). Pairwise correlations of PET counts supporting each loop were significant across all replicates ($\rho > 0.7$, p value < 0.001, Figure 2). Merging data across all replicates substantially increased the number of significant loops to 126,280. The mean (median) loop length was 173 kb (95 kb), and the mean (median) number of paired-end tags (PETs) per loop was 44.6 (19) (Table S1). We then overlapped the two anchors of HiChIP loops with gene promoters and enhancers (material and methods) and classified the 126,280 significant loops into different categories: enhancer-promoter loops (35,749), enhancer-enhancer loops (36,290), promoter-promoter loops (26,510), enhancer-other loops (12,814), and promoter-other loops (14,917). 17,690 genes (out of 27,063 RefSeq genes considered) had promoters overlapping at least one loop; each gene promoter overlaps a mean (median) of 7.9 (6) loops per gene, and there is a mean (median) of 351 (174) PETs per gene. Gene connectivity, defined as the number of loops per gene promoter (or as total number of PETs per gene), is moderately correlated with gene expression activity (as assayed by RNA-seq in the same cell line, material and methods) (Spearman $\rho = 0.489$; p value $< 2.2e{-}16$) (Figure S3). Since we were interested in linking a gene to its regulatory element in a PrCa risk region, in this paper we have focused on loops involving promoters (enhancer-promoter, promoter-promoter, promoter-other, see material and methods) and have summarized the data by gene. If instead we summarize the data by enhancer (defined with the H3K27ac LNCaP peaks), we have 32,469 enhancers; each enhancer overlaps a mean (median) of 5.3 (4) loops per enhancer, and there is a mean (median) of 206.8 (98) PETs per enhancer.

Prior studies have demonstrated that looping interactions do not qualitatively change in response to a perturbation.[38–40] To assess whether loop formation correlates with differential expression across a defined perturbation, we stimulated the LNCaP cell line with androgen (material and methods). We observed that genes that are destined to change under these conditions have pre-existing loops (e.g., upon androgen stimulation, genes with >1.5log2FC in transcript levels after 4 h have ten loops on average, whereas the average number of loops of any gene in LNCaP is nine, see Table S3; Table S4). This observation suggests that *de novo* loop formation is not a major mechanism underlying gene expression changes in this defined setting.

H3K27ac-HiChIP loops are enriched for eQTLs in prostate tissue

To evaluate whether chromatin interactions are enriched in prostate eQTLs, we assessed the overlap of loops with eQTLs measured in two prostate datasets: Thibodeau eQTLs in prostate normal tissue ($n = 471$)[60] and TCGA eQTLs in prostate tumor tissue ($n = 378$).[61] Local eQTLs (within 3 Mb of every gene) were called with standard approaches (material and methods). First, we found that the top eQTL variant for each eGene is 2 times more likely to fall within a HiChIP anchor than a random SNP in a 3 Mb region centered on the eGene (empirical p value $=$ 0.0099, Figure S4A). Second, we found that loops with eQTLs in one anchor were more likely to loop to the promoter of their eGene than expected by chance (1.2-fold change compared to random, empirical p value $=$ 0.0099, material and methods) across a range of distances from their target promoters (Figures S4B and S4C). If we only consider loops with an eQTL overlapping one anchor but with the eGene that is not overlapping the other anchor, the fold change is similar (1.2- and 1.3-fold change for TCGA and Thibideau, respectively, p value $= 0.0099$). This highlights that HiChip loops provide added information that is not captured by eQTL-eGene links.

H3K27ac-HiChIP loops link prostate cancer risk variants to target genes

Next, we quantified the enrichment of GWAS signals at the anchor regions of the loops by using stratified linkage disequilibrium score (LDSC) regression analysis[58] integrating the H3K27ac-HiChIP loops with the largest PrCa GWAS to date.[10] Variants residing in loop anchors show high enrichment of prostate cancer heritability as compared to random variants in the genome overall (s-LDSC enrichment of 2.61, $p = 1.06e{-}9$).

We observed that enrichment is even greater when considering only variants within H3K27ac peaks (s-LDSC enrichment of 14.06, $p = 2.54e{-}5$ for all loop types, Table S5; our calling allows for one anchor not to be acetylated thus yielding many loops with one non-acetylated anchor, see material and methods). Conversely, restricting analysis to variants residing in a loop anchor that is not additionally supported by H3K27ac peaks, we found lower, but statistically similar, enrichment levels as those estimated from all loop residing variants (s-LDSC enrichment of

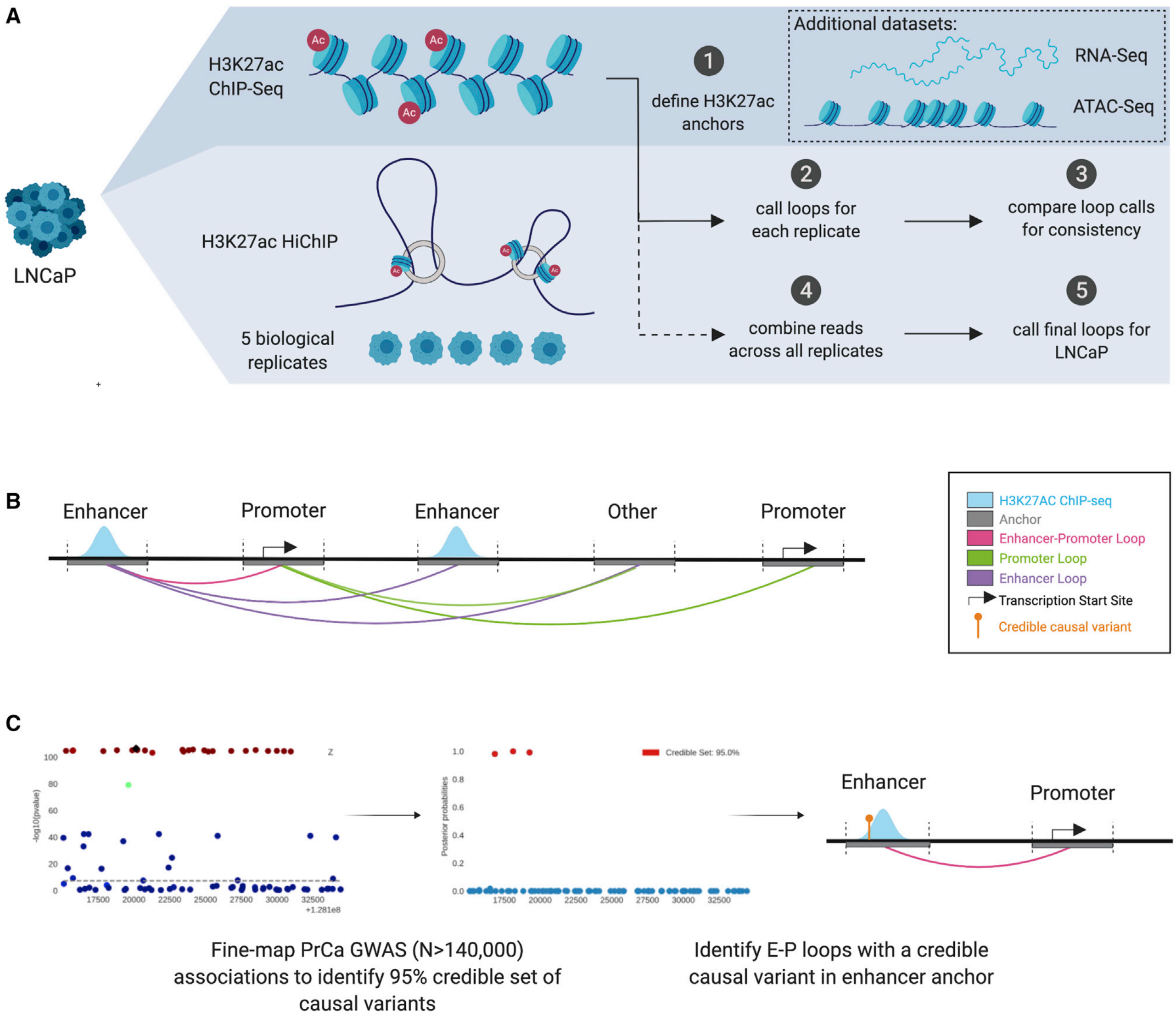

Figure 1. Experimental design
(A) Description of experimental datasets and workflow of HiChIP analysis to call total LNCaP loops.
(B) Anchor and loop type definitions.
(C) Possible regulatory mechanism describing a causal variant in an enhancer anchor interacting with a promoter of a target gene.

2.49, p = 5.22e−7, "All minus H3K27" Table S5). Interestingly, we observed a greater enrichment when restricting loops to specific categories (e.g., 18.72 for enhancer-promoter, p = 8.93e−6, 14.25 for enhancer-enhancer, p = 6.26e−5, Table S5). Together, our results confirm that H3K27ac-HiChIP looping localizes relevant PrCa GWAS signal.

Next, we performed probabilistic fine-mapping of the prostate cancer GWAS at 130 of the 147 previously reported risk loci except 8q24 for which exhaustive fine-mapping is available[12] (Table S6). We integrated 1,953 HiChIP loops with at least one anchor overlapping a PrCa credible causal variant in 104 risk loci regions (Table S7). Overall, 2,016 PrCa credible causal variants linked to 665 genes across these 104 loci (Figure 3, Table S8). 48 (out of 104) regions have three or fewer HiChiP genes overlapping credible GWAS SNPs (Table S6). Of the 665 genes, 37 are eGenes in prostate tissue and have evidence of GWAS/eQTL overlap with PrCa risk variants either through TWAS and/or colocalization (Figure 3, Table 1, Table S6). One example is *MLPH* (MIM: 606526) (Figure 4A), which is an eGene in prostate tissue and has been previously reported in normal prostate tissue from GTEx.[66] By contrast, four genes (*CEACAM21* [MIM: 618191], *MOB2* [MIM: 611969], *ASCL2* [MIM: 601886], and *GDF7* [MIM: 604651]) are eGenes and show evidence of TWAS/colocalization but are not supported by any HiChiP loop (Table S6).

To learn whether our results are robust to the fine-mapping algorithm, we repeated analyses by using results from another fine-mapping method called SuSiE (Wang et al.,[56] material and methods) and have found the results are largely concordant (Figure S6).

We then looked at evidence in support of the genes and found that 276 out of the 665 genes we have prioritized

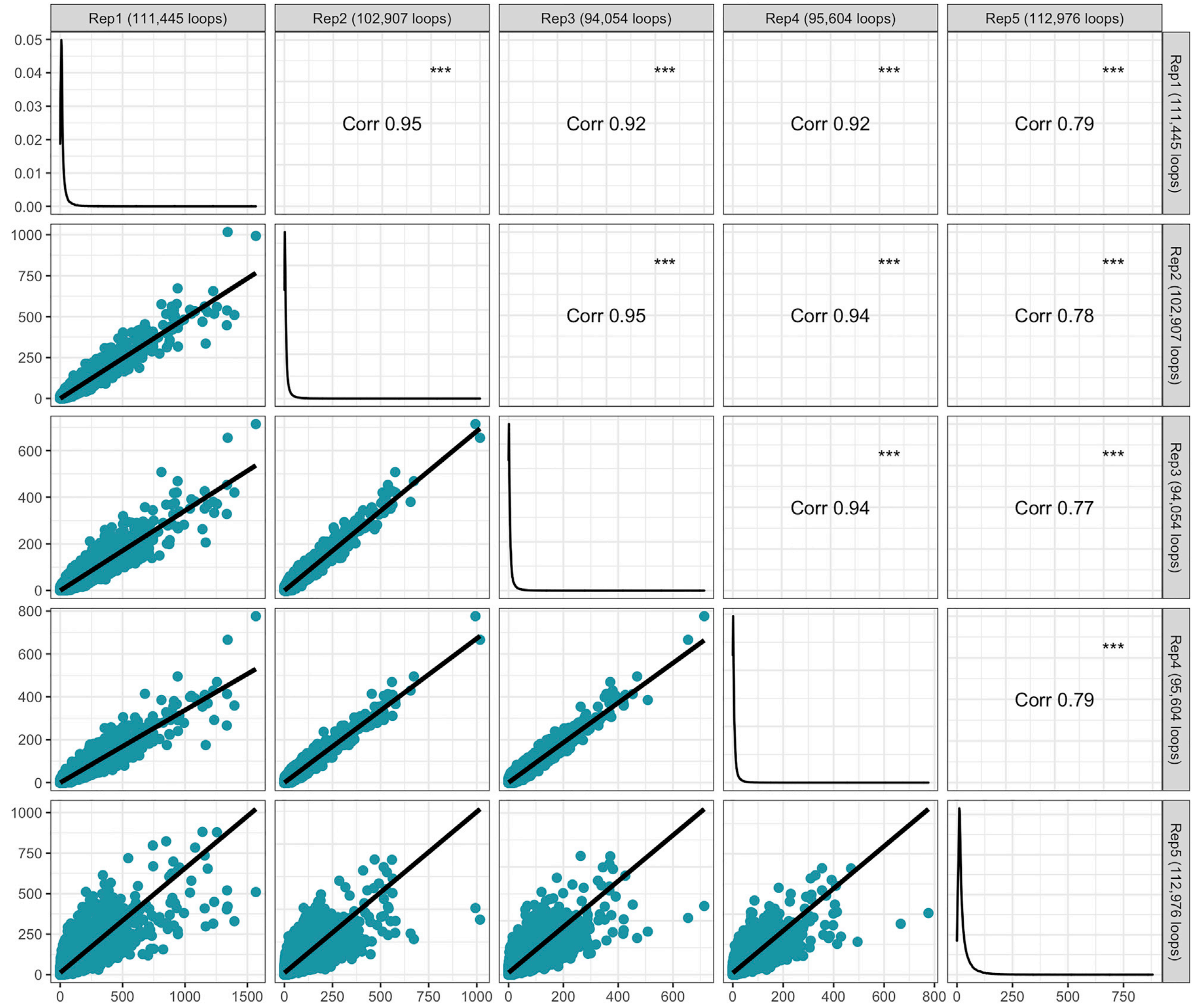

Figure 2. Pairwise correlations of paired-end tag (PET) counts at loops called significant in at least one of the LNCaP-HiChIP replicates

Every dot represents a HiChIP loop called at FDR < 0.01 in any of the replicates; the union of across replicates N = 114,142 loops. We consider all the loops called at no FDR cut-off. Out of these, we then take the loops called at FDR 0.01 in any of the replicates (i.e., the union of loops across replicates, N = 114,142 loops). Within the bottom left panels, the axis represents the number of PETs observed in the two compared replicates. The diagonal panel shows the distribution of each replicate's PET counts. Pearson correlation coefficients are shown on the top right panels. If the PET count is missing for a loop in one replicate but it is present in the other, it is treated as 0 in the correlation estimation. See Figures S1 and S2 for a comprehensive comparison across replicates.

also have support from the ABC model from Fulco et al.[53] We report the gene-enhancer pair with the maximum ABC score for the 276 genes in Table S8. We also compared whether the 665 genes are highly ranked by using a gene-based analysis tool that assigns genes to traits after incorporating chromatin interactions called H-MAGMA.[57] Out of the 665 genes, H-MAGMA found 628 genes containing valid SNPs in genotype data. The gene p values for the 628 genes range from 4.5e−5 to 1e−50, further confirming with an orthogonal method the importance of these genes to prostate cancer. For example, *NKX3-1* has a gene p value in H-MAGMA of 1.3e−12, *MYC* has a gene p value in H-MAGMA of 1.7e−13, and *AR* has a gene p value in H-MAGMA of 7.7e−9.

Looping maps PrCa GWAS variants to known PrCa biology genes

We noted that many genes previously implicated in PrCa biology, such as *GATA2*, *AR*, *MYC*, and *NKX3-1*, are nearby PrCa GWAS risk loci. Despite the clear role of these genes in PrCa, compelling evidence linking risk variants to these genes through expression-based methods is currently lacking. Interestingly, looping provides links from promoters of these genes to GWAS risk regions. We highlight three such genes: *MYC*, *AR*, and *NKX3-1*.

MYC has an uncontested role in cancer biology and has been associated with numerous cancer types through GWASs.[72–74] A study from Matejcic et al.[12] fine-mapped the prostate cancer susceptibility region at 8q24 where

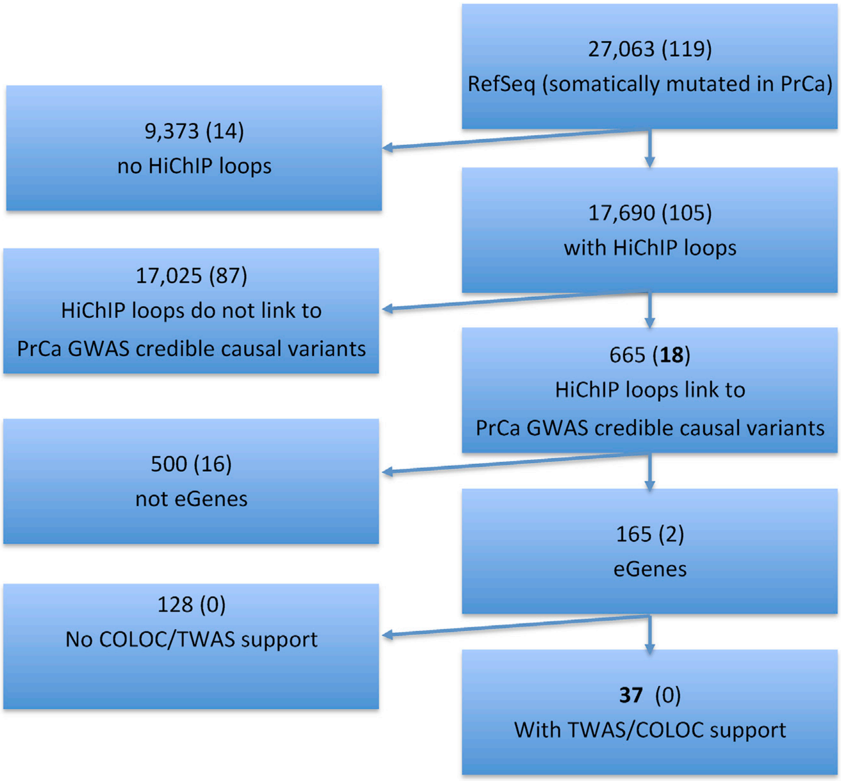

Figure 3. Breakdown of genes with various forms of evidence for linked to PrCa risk variants
Out of 27,063 genes in RefSeq, 17,690 show a HiChIP loop overlaying their promoter. 665 genes at 104 PrCa GWAS regions have a loop linking their promoter to a PrCa credible causal variant. 165 (out of 665) are also eGenes in tissues relevant to PrCa and 37 also show evidence of colocalization and/or transcriptome-wide association. The numbers in parentheses showcase the breakdown of the 119 genes with evidence of somatically acquired mutations in prostate cancer (material and methods).

MYC is located and observed 174 variants in the 95% credible set. 169 of these candidate causal variants overlap one of 225 HiChIP loops, and the majority (152/169) link to the *MYC* promoter (Figure 4B, Table S9). *MYC* has never been reported as an eGene. In the same fine-mapped region where *MYC* is located (chr8: 127,600,000–129,000,000), the genes *FAM84B* (MIM: 609483) and *POU5F1B* (MIM: 615739) have evidence from both eQTL data as well as from HiChIP data (Tables S6–S8). In particular, the SNP rs7839958 (chr8: 127,575,595) is an eQTL for the gene *FAM84B* (p value of association with gene expression in prostate tissue[60] is 7.90e−11, p value of association with PrCa[54] is 0.377). This gene has previously been reported to be a TWAS gene.[21,54] The SNP rs75555058 (chr8: 127,818,446) is an eQTL for the gene *POU5F1B* (p value of association with gene expression in prostate tissue[60] is 2.20e−33, p value of association with PrCa in Schumacher et al.[54] is 7.86e−6, p value of association with PrCa in previously reported Matejcic et al.[12] [Table S9] is 9.94e−7).

AR is critical for prostate cancer because PrCa is dependent on the actions of androgens and therefore on the function of the *AR* gene. More than half of primary tumors and almost all tumor metastases are associated with overexpressed or deregulated *AR*,[75,76] and mutations in the *AR* gene have been associated particularly with tumor progression.[42,77,78] We identified H3K27Ac-HiChIP loops that link PrCa credible SNPs located in the X chromosome with the *AR* promoter, potentially pointing to enhancer regions important for PrCa (Figure 4D). In our analysis, *AR* contains PrCa SNPs in both the promoter as well as the enhancer anchor of the HiChIP loops, increasing its evidence of this gene's involvement in PrCa biology. The PrCa SNPs in this region have not been implicated as eQTLs.

NKX3-1 is another interesting example because it is one of the most androgen-responsive genes in the LNCaP cell line, it is involved in prostate development, and is recurrently mutated in advanced PrCa.[65,79–81] In our data, *NKX3-1* has weak evidence for eQTL association. Although it does not pass our threshold for calling this an eGene (material and methods), the evidence of a shared causal variant between the eQTL and the GWAS signal is high for this gene (the probability of colocalization between GWAS and eQTL data at this locus is 90%). We measured looping before and after androgen stimulation in the LNCaP cell line and we observed that the number of *NKX3-1* promoter loops is largely unchanged (material and methods, Table S4 and Figure 4C). For example, the expression changes by ∼2-fold after 4 (16) h of androgen stimulation (logFC $=$ 2.23 (1.88) p value $<$ 1e−50), while the total number of loops change minimally (from 14 loops to 11 and to 13, respectively, for the two time points, Table S4). These data indicate that transcriptionally dynamic genes, which may represent context-dependent eQTL targets are discoverable through looping.

Looping identifies germline-somatic interactions

We next investigated germline-somatic interactions by evaluating whether genes known to be somatically mutated in PrCa oncogenesis also show evidence of looping to germline PrCa GWAS. We identified a set of 119 prostate cancer genes curated from large-scale PrCa studies that show evidence of somatically acquired mutations[63–65] (material and methods). Interestingly these genes are on average closer to a PrCa credible causal variant when compared to other genes within a 3 Mb region, providing additional evidence of their importance to PrCa (Figure S5, 30% of PrCa genes compared to 8% of all

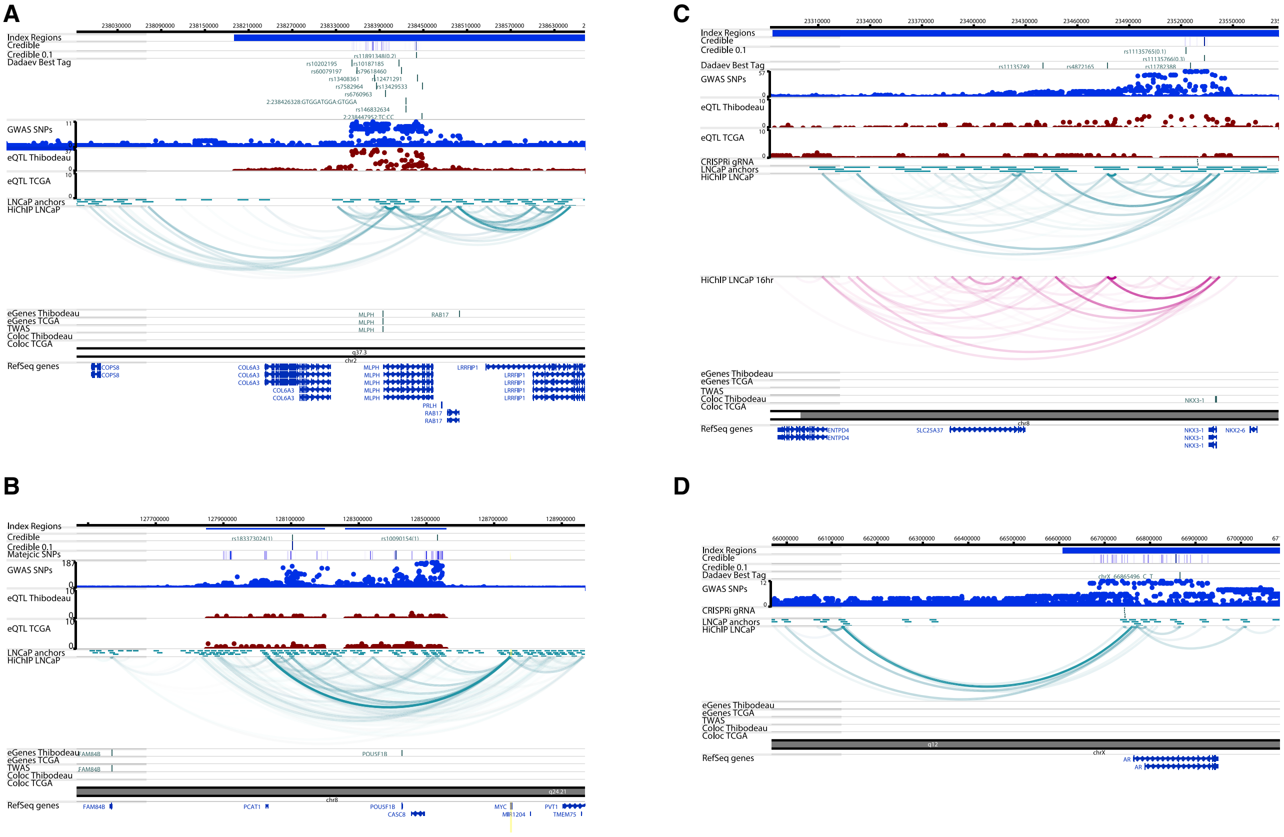

Figure 4. Examples of genes supported by HiChIP loops and GWAS

(A–D) *MLPH*, with evidence of eQTLs in prostate tissue using both prostate tissue datasets (Thibodeau and TCGA) (A); *MYC* (B) and *NKX3-1* (C), with weak evidence of eQTLs in prostate tissue. *AR*, with no evidence of eQTLs (D).

The tracks shown are listed below.

Index regions: the PrCa GWAS fine-mapped region around previously reported index SNP from Schumacher et al.[10] (see material and methods).

Credible: position of the SNPs included in the 95% credible SNP set (see material and methods); the color of the track is deeper for SNPs with higher probability of being causal from 0 to 1.

Credible 0.1: position and name of SNPs that reach a posterior probability of being causal of 0.1.

Dadaev best tag: position and name of SNPs within 341 "best tag" SNPs fine-mapped in Dadaey et al.[2]

GWAS SNPs: each hollow circle represents a SNP; y axis is the position and x axis is the −log10(p value) genome-wide association with PrCa risk. PrCa GWAS summary statistics is from Schumacher et al.[10] (N = 79,148 cases and 61,106 controls) across 20,370,946 SNPs.

LNCaP anchors: regions of the genome containing HiChIP anchors.

HiChIP LNCaP: loops from merged data of five replicates.

eGenes Thibodeau: highlighted genes that are eGenes in Thibodeau dataset.[60]

eGenes TCGA: highlighted genes that are eGenes in TCGA dataset.[61]

eQTL Thibodeau: each hollow circle represents a SNP; y axis is the position and x axis is the associations between SNP and gene expression (−log10(p value)) for SNPs within 3 Mb window around a gene promoter. Associations were run with gene expression and genotype data from 471 samples from normal prostate tissue.[60]

eQTL TCGA: each hollow circle represents a SNP; y axis is the position and x axis is the associations between SNP and gene expression (−log10(p value)) for SNPs within 3 Mb window around a gene promoter. Associations were run with gene expression and genotype data from 378 samples in prostate tumor tissue.[61]

TWAS: highlighted genes that have a significant transcriptome-wide association (TWAS) signal in prostate cancer.

Coloc Thibodeau: highlighted genes that have a significant colocalization (COLOC) signal in prostate cancer via Thibodeau eQTL data.

Coloc TCGA: highlighted genes that have a significant COLOC signal in prostate cancer via TCGA eQTL data.

RefSeq genes: known human protein-coding and non-protein-coding genes taken from the 2019 RefSeq release.

The following tracks are specific to the particular locus.

CRISPRi gRNA: gRNA positions targeted for the CRISPRi experiment in loci *AR* and *NKX3-1* (C and D).

Matejcic SNPs: position of fine-mapped SNPs from Matejcic et al. for locus *MYC* (B).[12]

HiChIP LNCaP 16 h: HiChIP loops measured after 16 h from androgen stimulation for locus *NKX3-1* (C).

genes are within 100 kb of PrCa credible causal variant). 18 (out of 104) genes have HiChIP loops linking a credible causal variant to the promoter of the gene (Table 1). Strikingly, none show significant colocalization GWAS/eQTL, and only two genes are eGenes in existing prostate transcriptomic data (Figure 3).

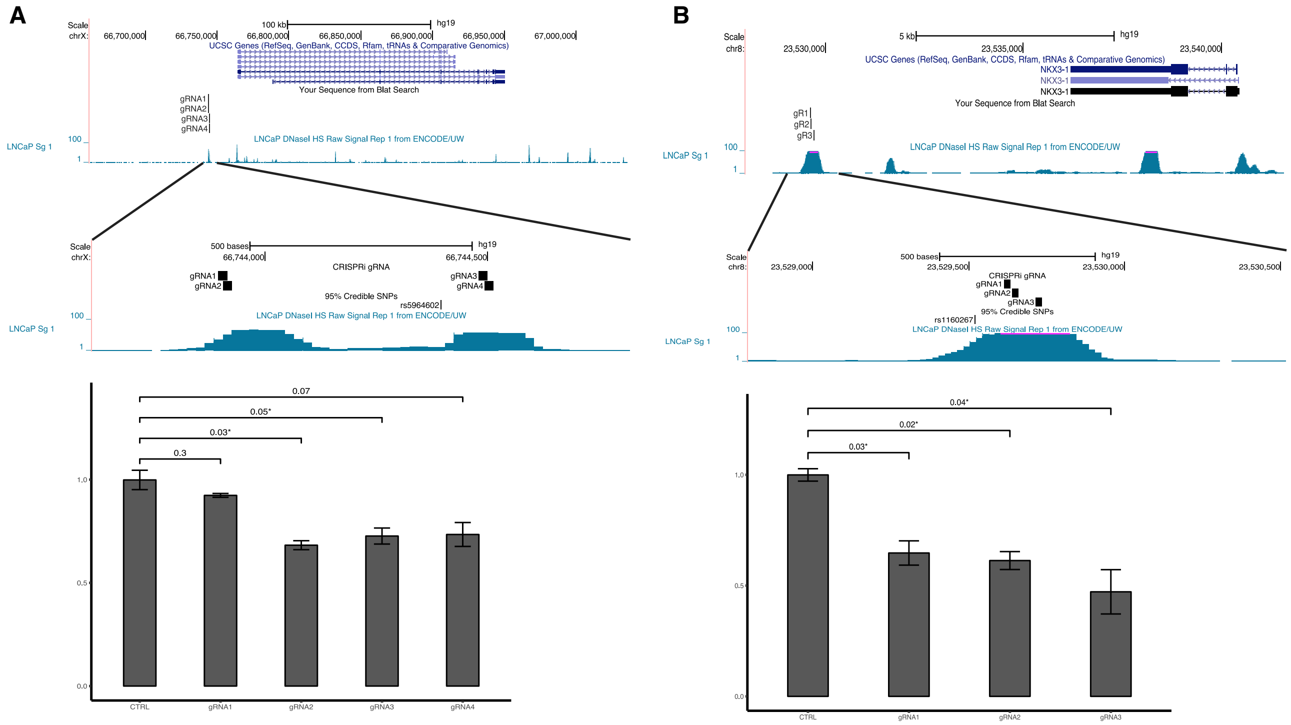

Figure 5. CRISPRi functional validation of the AR and NKX3-1 loci

Functionally relevant enhancers were identified by integrating epigenetic datasets (DHS peaks and HiChIP anchors in LNCaP cell line) in PrCa risk loci.

(A) AR gene genomic region, LNCaP DHS signals, 95% credible SNP set, and gRNA positions targeting DHS peaks for the CRISPRi experiment. Bottom panel shows the suppression effect on the AR gene with four different gRNAs compared to the non-human targeting negative control. Three out of the four tested gRNAs showed ~1-fold significant suppression on the AR gene expression. Columns demonstrate averages of two biological replicates, and error bars represent standard errors. In the bottom panel, the bar heights represent the fixed-effect estimates and the error bars represent standard errors from a meta-analysis across two biological replicates (averages of three replicates each). p values were obtained from a two-sample t test (equal variances) comparing each sample to the control. A single asterisk indicates significant p value (below 0.05).

(B) NKX3-1 genomic region, LNCaP DHS signals, 95% credible SNP set, and gRNA positions targeting DHS peaks for the CRISPRi experiment. Bottom panel shows the suppression effect on the NKX3-1 gene with three different gRNAs compared to the non-human targeting negative control. All three tested gRNAs showed ~1-fold significant suppression on the NKX3-1 gene expression. The bottom panel bars are defined above.

CRISPRi validates HiChIP-predicted enhancer-target gene relationships

Next, we validated two enhancer-target gene relationships, demonstrating altered target gene expression with epigenetic CRISPRi silencing of the enhancers. We first selected two loops where anchors containing a candidate causal variant linked genes that play clear roles in PrCa biology, *NKX3-1* and *AR* (Figure 5, material and methods). As described above, these two genes have not been strongly implicated through eQTL- or COLOC/TWAS-based analyses. Second, putative regulatory enhancer regions overlapping DNase hypersensitivity sites (DHSs) were identified and targeted. To evaluate the impact of suppressing the regulatory element on the expression level of the target gene, guide RNAs (gRNAs) were designed against DHS peaks falling within the designated putative enhancer (Figure 5, Table S10). We note that, although the gRNA does not directly overlap a PrCa candidate causal variant (due to gRNA design constraints), the DHS peak is within the anchor. Notably, the anchors at these two risk loci only looped to a single target gene (*NKX3-1* and *AR*). Targeted epigenetic suppression of these enhancer regions significantly reduced RNA levels in the target genes predicted by the HiChIP loops (Figure 5, Table S11).

Discussion

A central issue driving post-GWAS studies is a mechanistic understanding of non-protein-coding risk loci, which account for over 90% of GWAS variants. In this work, we outlined a systematic approach, based on chromosome conformation capture technology, to link regulatory element(s) to possible candidate target genes in GWAS PrCa risk regions.

We used H3K27Ac-HiChIP methodology as a means to assay genome-wide chromatin interactions. 3C-based methods measure physical interactions and thus complement other approaches, such as eQTLs, which are based on association between genotypes and transcript levels. eQTL studies can be confounded by LD and are dependent on sample size whereby interactome maps do not suffer from these limitations.

An additional limitation of large-scale expression-based studies is that they are based on steady-state transcript levels. By contrast, studies have demonstrated that looping is less dynamic in response to defined perturbations.[38,40,82] Stated another way, looping identifies the *potential* of an enhancer-promoter interaction to be active and is less informative as a quantitative readout of transcriptional levels of genes. This observation raises the provocative notion that looping can identify latent stimulus- and context-dependent eQTLs and highlight important candidate genes without requiring experiments that directly measure these other conditions. This rationale is consistent with recent reports showing that steady-state eQTLs are insufficient to explain the majority of disease heritability.[83] Indeed, our results showed that important PrCa biology genes interact with risk loci that previously escaped detection through expression-based methods.

Other techniques can be used to assay enhancer-promoter interactions. Traditional Hi-C can assay all loops; however, Hi-C can lack resolution in mapping the enhancer-promoter link if not sequenced to extremely high depths. Through enrichment based on immunoprecipitation of a target protein, H3K27Ac HiChIP is an efficient way to detect enhancer-promoter loops and has been shown to identify a similar number of loops with 10-fold less sequencing compared to Hi-C.[49] ChIA-PET is another whole-genome 3C-based assay able to detect protein-centric long range contacts, however this method still requires hundreds of millions of cells per experiment and is less efficient than HiChIP.[49]

When we compared the genes identified in enhancer-promoter loops to a previously reported genome-wide map of enhancer-gene connections in LNCaP (Fulco et al.[53]), 74% of the genes identified in the ABC model (9,333/12,641 genes) were genes with an E-P loop in our HiChIP data, and 64% of these (5,969/9,333 genes) had overlapping element-gene pairs (Figure S7). The two methods HiChiP and ABC seem to offer different insights into gene prioritization in this data, and we leave a thorough comparison of the two methodologies as future work.

We validated two enhancer-target gene relationships, demonstrating altered target gene expression with epigenetic CRISPRi silencing of the enhancer. We note that this experiment does not prove variant causality. This type of work requires more intensive and precise genome editing strategies as we have previously shown and represents a logical next step.[15] Additionally, we note that HiChIP loop does not link one gene to one regulatory region and there is rarely a one-to-one relationship. Instead, we demonstrate a complex relationship where the same loop can connect multiple genes to multiple regions of the genome. While this complicates the detection of one causal gene, the shared regulatory regions could be an important aspect of gene regulation and important for future understanding of the genetics of the PrCa risk loci considered herein.

As with any method, there are limitations of HiChIP. First, all 3C-based methods have limited power to confidently detect nearby genes. Second, the computational pipelines to analyze HiChIP data are still in their infancy and future developments could affect results. HiChIP loop calling is dependent on H3K27ac ChIP-seq peak calling used for loop anchors. We considered only 27,063 cataloged RefSeq genes (GRCh37/hg19). Any other gene outside of this list will be missed from this analysis. We defined promoters on the basis of the longest transcript from TSS from RefSeq, which reports the most representative initiation site across different cell types. Using this definition, novel genes with as-yet unannotated start sites are missed. Furthermore, a 5-kb resolution of the HiChIP data analysis limits our definition of anchors, and there is the risk that the additional 5-kb padding added to anchors on either side (resulting in 15 kb anchors) could result in decreased resolution of enhancer-promoter links. In this work, we focused on H3K27ac histone modification. In future studies, assays using other modifications (e.g., H3K4me1, H3K4me2) could be used in combination to better refine enhancer and promoter regions.[84]

The limitations of this work include having performed HiChIP in a single cell line. Further studies should explore additional cell lines, models, and primary tissues to systematically annotate the interactome. We prioritized genes on the basis of HiChIP looping and PrCa association. Although these results are specific for LNCaP cell line, we propose a process for post-GWAS functional follow-up, which can be re-applied to run on different cell lines. Our pipeline makes use of results from fine-mapping methods to link genes that were not previously linked to PrCa. Fine-mapping methods that use summary-level data are dependent on the GWAS summary statistics and information about the LD structure in the region of interest to calculate posterior probabilities of being causal for each variant. Furthermore, imputation errors can affect the relative probability of SNPs being determined as causal in any statistical fine-mapping strategy. There is therefore a need to follow up these analyses with further experiments. Moreover, this study focuses on common variants studied by GWAS and eQTL studies. Rare, large effect variation with regulatory function would not be observed.

We demonstrate the benefit of using HiChIP to both validate eQTLs as well as prioritize genes for PrCa that are missed by eQTL-based methodologies. It is possible that eQTL fails to detect potentially causal cancer genes. Alternatively, eQTL and HiChIP genes could be part of the same causal pathways. We leave this to future analyses. Moving forward, we propose to utilize the complementary techniques of eQTL-based methodologies and HiChIP to prioritize genes at GWAS loci. This work also creates an opportunity to create a unified model combining expression and interaction information to extract the strengths of both methods.

Data and code availability

We provide HiChIP interactome maps integrated with GWAS and eQTL information generated as a resource to the research community to investigate PrCa GWAS mechanisms. Processed HiChIP data and intermediate files used in this study can be downloaded and visualized interactively via the WashU Epigenome Browser link.[85] The code used for annotation of peaks and other R scripts are available online.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.11.007.

Acknowledgments

This work was funded in part by the National Institutes of Health under award R01CA251555. C.G. has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 754490 – MINDED project. A.G. is supported by NIH R01CA227237, NIH R01CA244569, the Louis B. Mayer Foundation, the Doris Duke Charitable Foundation, and the Phi Beta Psi Sorority. N.M. received funding from NIH R01CA258808. M.L.F. is supported by the Claudia Adams Barr Program for Innovative Cancer Research, the Department of Defense (W81XWH-19-1-0565 and W81XWH-21-1-0234), the H.L. Snyder Medical Research Foundation, the Donahue Family Fund, the Mayer Foundation, and the Cutler Family Fund for Prevention and Early Detection. We would like to thank Daofeng Li for his help with the Epigenome Browser.

Declaration of interests

The authors declare no competing interests.

Received: April 2, 2021
Accepted: November 2, 2021
Published: November 24, 2021

Web resources

1000 Genomes data, ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/integrated_call_samples_v3.20130502.ALL.panel

ABC results, https://files.osf.io/v1/resources/uhnb4/providers/osfstorage/5d1cd944e745a4001798b19c?direct&version=1

Epigenome Browser, https://epigenomegateway.wustl.edu/legacy/?genome=hg19&datahub=https://wangftp.wustl.edu/~dli/Claudia/LNCAP_freedman_ALL.json

GEPIA, http://gepia2.cancer-pku.cn/#degenes

GTEx project, https://gtexportal.org/home/

LD score regression of Specifically Expressed Genes "LDSC SEG," https://alkesgroup.broadinstitute.org/LDSCORE/LDSC_SEG_ldscores/tstats/GTEx.tstat.tsv

Prostate Cancer GWAS, http://practical.icr.ac.uk/blog/?page_id=8164

Source code and R scripts, https://github.com/bogdanlab/hichip

UCSC Genome Browser RefSeq hg19, http://genome.ucsc.edu

References

1. Benafif, S., Kote-Jarai, Z., Eeles, R.A.; and PRACTICAL Consortium (2018). A Review of Prostate Cancer Genome-Wide Association Studies (GWAS). Cancer Epidemiol. Biomarkers Prev. *27*, 845–857.
2. Dadaev, T., Saunders, E.J., Newcombe, P.J., Anokian, E., Leongamornlert, D.A., Brook, M.N., Cieza-Borrella, C., Mijuskovic, M., Wakerell, S., Olama, A.A.A., et al. (2018). Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. Nat. Commun. *9*, 2256.
3. Hjelmborg, J.B., Scheike, T., Holst, K., Skytthe, A., Penney, K.L., Graff, R.E., Pukkala, E., Christensen, K., Adami, H.-O., Holm, N.V., et al. (2014). The heritability of prostate cancer in the Nordic Twin Study of Cancer. Cancer Epidemiol. Biomarkers Prev. *23*, 2303–2310.
4. Mucci, L.A., Hjelmborg, J.B., Harris, J.R., Czene, K., Havelick, D.J., Scheike, T., Graff, R.E., Holst, K., Möller, S., Unger, R.H., et al. (2016). Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. JAMA *315*, 68–76.
5. Eeles, R. (2016). Prostate cancer genome-wide association study from 89,000 men using the OncoArray chip to identify novel prostate cancer susceptibility loci. J. Clin. Oncol. *34*, 1525.
6. Amin Al Olama, A., Dadaev, T., Hazelett, D.J., Li, Q., Leongamornlert, D., Saunders, E.J., Stephens, S., Cieza-Borrella, C., Whitmore, I., Benlloch Garcia, S., et al. (2015). Multiple novel prostate cancer susceptibility signals identified by fine-mapping of known risk loci among Europeans. Hum. Mol. Genet. *24*, 5589–5602.
7. Eeles, R., Al Olama, A.A., Berndt, S., Wiklund, F., Conti, D.V., Ahmed, M., Benlloch, S., Easton, D., Kraft, P., Chanock, S.J., et al. (2017). Prostate cancer meta-analysis from more than 145,000 men to identify 65 novel prostate cancer susceptibility loci. J. Clin. Oncol. *2017*, 1.
8. Al Olama, A.A., Kote-Jarai, Z., Giles, G.G., Guy, M., Morrison, J., Severi, G., Leongamornlert, D.A., Tymrakiewicz, M., Jhavar, S., Saunders, E., et al. (2009). Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat. Genet. *41*, 1058–1060.
9. Conti, D.V., Darst, B.F., Moss, L.C., Saunders, E.J., Sheng, X., Chou, A., Schumacher, F.R., Olama, A.A.A., Benlloch, S., Dadaev, T., et al. (2021). Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. Nat. Genet. *53*, 65–75.
10. Schumacher, F.R., Olama, A.A.A., Berndt, S.I., Benlloch, S., Ahmed, M., Saunders, E.J., Dadaev, T., Leongamornlert, D., Anokian, E., Cieza-Borrella, C., et al. (2019). Author Correction: Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat. Genet. *51*, 363.
11. Hua, J.T., Ahmed, M., Guo, H., Zhang, Y., Chen, S., Soares, F., Lu, J., Zhou, S., Wang, M., Li, H., et al. (2018). Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer through lncRNA PCAT19. Cell *174*, 564–575.e18.
12. Matejcic, M., Saunders, E.J., Dadaev, T., Brook, M.N., Wang, K., Sheng, X., Olama, A.A.A., Schumacher, F.R., Ingles, S.A., Govindasami, K., et al. (2018). Germline variation at 8q24 and prostate cancer risk in men of European ancestry. Nat. Commun. *9*, 4616.
13. Guo, H., Ahmed, M., Zhang, F., Yao, C.Q., Li, S., Liang, Y., Hua, J., Soares, F., Sun, Y., Langstein, J., et al. (2016). Modulation of

long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. Nat. Genet. *48*, 1142–1150.

14. Luo, Z., Rhie, S.K., Lay, F.D., and Farnham, P.J. (2017). A Prostate Cancer Risk Element Functions as a Repressive Loop that Regulates HOXA13. Cell Rep. *21*, 1411–1417.
15. Spisák, S., Lawrenson, K., Fu, Y., Csabai, I., Cottman, R.T., Seo, J.H., Haiman, C., Han, Y., Lenci, R., Li, Q., et al. (2015). CAUSEL: an epigenome- and genome-editing pipeline for establishing function of noncoding GWAS variants. Nat. Med. *21*, 1357–1363.
16. Gao, P., Xia, J.-H., Sipeky, C., Dong, X.-M., Zhang, Q., Yang, Y., Zhang, P., Cruz, S.P., Zhang, K., Zhu, J., et al. (2018). Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus. Cell *174*, 576–589.e18.
17. Gusev, A., Shi, H., Kichaev, G., Pomerantz, M., Li, F., Long, H.W., Ingles, S.A., Kittles, R.A., Strom, S.S., Rybicki, B.A., et al. (2016). Atlas of prostate cancer heritability in European and African-American men pinpoints tissue-specific regulation. Nat. Commun. *7*, 10979.
18. Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association to Function. Am. J. Hum. Genet. *102*, 717–730.
19. de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499–506.
20. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.
21. Mancuso, N., Gayther, S., Gusev, A., Zheng, W., Penney, K.L., Kote-Jarai, Z., Eeles, R., Freedman, M., Haiman, C., Pasaniuc, B.; and PRACTICAL consortium (2018). Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. Nat. Commun. *9*, 4079.
22. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.
23. Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. Nat. Genet. *51*, 592–599.
24. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.
25. Hormozdiari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. Am. J. Hum. Genet. *99*, 1245–1260.
26. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487.
27. Wen, X., Pique-Regi, R., and Luca, F. (2017). Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. PLoS Genet. *13*, e1006646.
28. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J., Im, H.K.; and GTEx Consortium (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098.
29. Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., et al. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat. Commun. *9*, 1825.
30. Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., and Knight, J.C. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science *343*, 1246949.
31. Taberlay, P.C., Achinger-Kawecka, J., Lun, A.T.L., Buske, F.A., Sabir, K., Gould, C.M., Zotenko, E., Bert, S.A., Giles, K.A., Bauer, D.C., et al. (2016). Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. Genome Res. *26*, 719–731.
32. Rhie, S.K., Perez, A.A., Lay, F.D., Schreiner, S., Shi, J., Polin, J., and Farnham, P.J. (2019). A high-resolution 3D epigenomic map reveals insights into the creation of the prostate cancer transcriptome. Nat. Commun. *10*, 4154.
33. Ramanand, S.G., Chen, Y., Yuan, J., Daescu, K., Lambros, M.B., Houlahan, K.E., Carreira, S., Yuan, W., Baek, G., Sharp, A., et al. (2020). The landscape of RNA polymerase II-associated chromatin interactions in prostate cancer. J. Clin. Invest. *130*, 3987–4005.
34. Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., et al. (2017). Enhancer connectome in primary human cells reveals target genes of disease-associated DNA elements. Nat. Genet. *49*, 1602–1612.
35. Jeng, M.Y., Mumbach, M.R., Granja, J.M., Satpathy, A.T., Chang, H.Y., and Chang, A.L.S. (2019). Enhancer Connectome Nominates Target Genes of Inherited Risk Variants from Inflammatory Skin Disorders. J. Invest. Dermatol. *139*, 605–614.
36. Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A.D., and Ren, B. (2016). Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. Cell Res. *26*, 1345–1348.
37. Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat. Genet. *47*, 598–606.
38. D'Ippolito, A.M., McDowell, I.C., Barrera, A., Hong, L.K., Leichter, S.M., Bartelt, L.C., Vockley, C.M., Majoros, W.H., Safi, A., Song, L., et al. (2018). Pre-established Chromatin Interactions Mediate the Genomic Response to Glucocorticoids. Cell Syst. *7*, 146–160.e7.
39. Greenwald, W.W., Li, H., Benaglio, P., Jakubosky, D., Matsui, H., Schmitt, A., Selvaraj, S., D'Antonio, M., D'Antonio-Chronowska, A., Smith, E.N., and Frazer, K.A. (2019). Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. Nat. Commun. *10*, 1054.
40. Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

41. Hawley, J.R., Zhou, S., Arlidge, C., Grillo, G., Kron, K., Hugh-White, R., van der Kwast, T., Fraser, M., Boutros, P.C., Bristow, R.G., et al. (2021). Cis-regulatory Element Hijacking by Structural Variants Overshadows Higher-Order Topological Changes in Prostate Cancer. bioRxiv. https://doi.org/10.1101/2021.01.05.425333.
42. Takeda, D.Y., Spisák, S., Seo, J.-H., Bell, C., O'Connor, E., Korthauer, K., Ribli, D., Csabai, I., Solymosi, N., Szállási, Z., et al. (2018). A Somatically Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver in Advanced Prostate Cancer. Cell *174*, 422–432.e13.
43. Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.
44. Qin, Q., Mei, S., Wu, Q., Sun, H., Li, L., Taing, L., Chen, S., Li, F., Liu, T., Zang, C., et al. (2016). ChiLin: a comprehensive ChIP-seq and DNase-seq quality control and analysis pipeline. BMC Bioinformatics *17*, 404.
45. Cornwell, M., Vangala, M., Taing, L., Herbert, Z., Köster, J., Li, B., Sun, H., Li, T., Zhang, J., Qiu, X., et al. (2018). VIPER: Visualization Pipeline for RNA-seq, a Snakemake workflow for efficient and complete RNA-seq analysis. BMC Bioinformatics *19*, 135.
46. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
47. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.
48. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
49. Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods *13*, 919–922.
50. Krueger, F. (2015). Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
51. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259.
52. Bhattacharyya, S., Chandra, V., Vijayanand, P., and Ay, F. (2019). Identification of significant chromatin contacts from HiChIP data by FitHiChIP. Nat. Commun. *10*, 4221.
53. Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., et al. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. Nat. Genet. *51*, 1664–1669.
54. Schumacher, F.R., Al Olama, A.A., Berndt, S.I., Benlloch, S., Ahmed, M., Saunders, E.J., Dadaev, T., Leongamornlert, D., Anokian, E., Cieza-Borrella, C., et al. (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat. Genet. *50*, 928–936.
55. Kichaev, G., Yang, W.-Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P., and Pasaniuc, B. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. *10*, e1004722.
56. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. J. R. Stat. Soc. Series B Stat. Methodol. *82*, 1273–1300.
57. Sey, N.Y.A., Hu, B., Mah, W., Fauni, H., McAfee, J.C., Rajarajan, P., Brennand, K.J., Akbarian, S., and Won, H. (2020). A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. Nat. Neurosci. *23*, 583–593.
58. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.
59. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.
60. Thibodeau, S.N., French, A.J., McDonnell, S.K., Cheville, J., Middha, S., Tillmans, L., Riska, S., Baheti, S., Larson, M.C., Fogarty, Z., et al. (2015). Identification of candidate genes for prostate cancer-risk SNPs utilizing a normal prostate tissue eQTL data set. Nat. Commun. *6*, 8653.
61. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M.; and Cancer Genome Atlas Research Network (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. *45*, 1113–1120.
62. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358.
63. Armenia, J., Wankowicz, S.A.M., Liu, D., Gao, J., Kundra, R., Reznik, E., Chatila, W.K., Chakravarty, D., Han, G.C., Coleman, I., et al. (2018). The long tail of oncogenic drivers in prostate cancer. Nat. Genet. *50*, 645–651.
64. Schlomm, T. (2016). Re: The Molecular Taxonomy of Primary Prostate Cancer. Eur. Urol. *69*, 1157.
65. Robinson, D., Van Allen, E.M., Wu, Y.-M., Schultz, N., Lonigro, R.J., Mosquera, J.-M., Montgomery, B., Taplin, M.-E., Pritchard, C.C., Attard, G., et al. (2015). Integrative Clinical Genomics of Advanced Prostate Cancer. Cell *162*, 454.
66. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.
67. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.-R., Lareau, C., Shoresh, N., et al.; Brainstorm Consortium (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet. *50*, 621–629.
68. Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res. *45* (W1), W98–W102.
69. Giambartolomei, C., Zhenli Liu, J., Zhang, W., Hauberg, M., Shi, H., Boocock, J., Pickrell, J., Jaffe, A.E., Pasaniuc, B., Roussos, P.; and CommonMind Consortium (2018). A Bayesian framework for multiple trait colocalization from summary association statistics. Bioinformatics *34*, 2538–2545.
70. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

71. Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. *29*, e45.

72. Wasserman, N.F., Aneas, I., and Nobrega, M.A. (2010). An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome Res. *20*, 1191–1197.

73. Sur, I.K., Hallikas, O., Vähärautio, A., Yan, J., Turunen, M., Enge, M., Taipale, M., Karhu, A., Aaltonen, L.A., and Taipale, J. (2012). Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. Science *338*, 1360–1363.

74. Ahmadiyeh, N., Pomerantz, M.M., Grisanzio, C., Herman, P., Jia, L., Almendro, V., He, H.H., Brown, M., Liu, X.S., Davis, M., et al. (2010). 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proc. Natl. Acad. Sci. USA *107*, 9742–9746.

75. Taylor, B.S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B.S., Arora, V.K., Kaushik, P., Cerami, E., Reva, B., et al. (2010). Integrative genomic profiling of human prostate cancer. Cancer Cell *18*, 11–22.

76. Lonergan, P.E., and Tindall, D.J. (2011). Androgen receptor signaling in prostate cancer development and progression. J. Carcinog. *10*, 20.

77. Brooke, G.N., and Bevan, C.L. (2009). The role of androgen receptor mutations in prostate cancer progression. Curr. Genomics *10*, 18–25.

78. Bu, H., Narisu, N., Schlick, B., Rainer, J., Manke, T., Schäfer, G., Pasqualini, L., Chines, P., Schweiger, M.R., Fuchsberger, C., and Klocker, H. (2016). Putative Prostate Cancer Risk SNP in an Androgen Receptor-Binding Site of the Melanophilin Gene Illustrates Enrichment of Risk SNPs in Androgen Receptor Target Sites. Hum. Mutat. *37*, 52–64.

79. Asatiani, E., Huang, W.-X., Wang, A., Rodriguez Ortner, E., Cavalli, L.R., Haddad, B.R., and Gelmann, E.P. (2005). Deletion, methylation, and expression of the NKX3.1 suppressor gene in primary human prostate cancer. Cancer Res. *65*, 1164–1173.

80. Bowen, C., and Gelmann, E.P. (2010). NKX3.1 activates cellular response to DNA damage. Cancer Res. *70*, 3089–3097.

81. Bhatia-Gaur, R., Donjacour, A.A., Sciavolino, P.J., Kim, M., Desai, N., Young, P., Norton, C.R., Gridley, T., Cardiff, R.D., Cunha, G.R., et al. (1999). Roles for Nkx3.1 in prostate development and cancer. Genes Dev. *13*, 966–977.

82. Comoglio, F., Park, H.J., Schoenfelder, S., Barozzi, I., Bode, D., Fraser, P., and Green, A.R. (2018). Thrombopoietin signaling to chromatin elicits rapid and pervasive epigenome remodeling within poised chromatin architectures. Genome Res., gr.227272.117.

83. Yao, D.W., O'Connor, L.J., Price, A.L., and Gusev, A. (2020). Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat. Genet. *52*, 626–633.

84. Mora, A., Sandve, G.K., Gabrielsen, O.S., and Eskeland, R. (2015). In the loop: promoter–enhancer interactions and bioinformatics. Brief. Bioinform. *17*, 980–995.

85. Li, D. (2019). Epigenome Browser update 2019. Nucleic Acids Res. https://doi.org/10.1093/nar/gkz348.