

RESEARCH ARTICLE

Decoding the germline genetic architecture of prostate cancer at a single cell resolution

Cheng Wang^{1,2,3,4}, Tianqi Tang^{1,2,3,4}, Yuejun Wang^{1,2,3,4}, Jingjing Li^{1,2,3,4*}

1 Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California San Francisco, San Francisco, California, United States of America, **2** Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, California, United States of America, **3** Center for Reproductive Science, School of Medicine, University of California San Francisco, San Francisco, California, United States of America, **4** Department of Neurology, School of Medicine, University of California San Francisco, San Francisco, California, United States of America

* Jingjing.Li@ucsf.edu



Abstract

Prostate cancer exhibits a strong familial association, and its heritability indicates a significant contribution from germline variants. While genome-wide association studies (GWAS) have identified common germline variants associated with prostate cancer risk, translating these statistical associations into functional mechanisms has remained a long-standing challenge. Consequently, most of our understanding of the genetic basis of prostate cancer stems from extensive studies of somatic mutations, leaving the germline genetic architecture largely unresolved. Because most germline variants lie in the noncoding genome and complex human diseases are predominantly driven by regulatory mutations, we herein asked which prostate cell types mediate the functional effects of germline variants, and thus represent the most genetically vulnerable populations. We generated paired epigenomic and transcriptomic profiles from reference human prostate tissues. Integrating these single-cell data with large-scale GWAS data identified a terminally differentiated luminal epithelial subtype that mediates the strongest germline risk in prostate cancer. We subsequently developed a deep learning model to score ~17 million GWAS variants based on their predicted impact on altering local chromatin accessibility in this vulnerable luminal epithelial subtype, and identified high-confidence candidate loci where high-risk germline variants likely alter promoter accessibility in prostate cancer. The implicated genes were involved in several pathways in tumorigenesis, displayed strong dosage sensitivity, and converged on the androgen receptor (AR)-mediated regulon, a mechanism also observed for somatic mutations. Overall, by unveiling cell types and candidate loci that mediate germline risk, our study defines the cell-type-specific germline architecture in prostate cancer and provides a comprehensive framework for understanding cancer heritability.

OPEN ACCESS

Citation: Wang C, Tang T, Wang Y, Li J (2025) Decoding the germline genetic architecture of prostate cancer at a single cell resolution. *PLoS Genet* 21(12): e1011975. <https://doi.org/10.1371/journal.pgen.1011975>

Editor: Joshua Gruber, The University of Texas Southwestern Medical Center Eugene McDermott Center for Human Growth and Development, UNITED STATES OF AMERICA

Received: July 22, 2025

Accepted: December 1, 2025

Published: December 30, 2025

Copyright: © 2025 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The data in the h5ad format is available at <https://ucsf.box.com/s/9zb2rtw4kwgcrclloavna8j7iyfqfhec>, requiring the token access with: SanFrancisco2025.

Funding: This study is supported by NIH-NIGMS award (R35 GM142983) and UCSF Prostate Cancer Program Pilot Award to J.L. We confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Prostate cancer tends to run in families, suggesting that inherited genetic factors strongly influence a man's risk of developing the disease. Although large-scale genetic studies have identified many variants associated with prostate cancer, most of these variants lie outside genes, in the noncoding regions of the genome, making it difficult to determine how they function. To address this challenge, we studied healthy human prostate tissue at single-cell resolution to identify which cell types are most affected by inherited genetic variation. By combining single-cell gene expression and chromatin accessibility with population-level genetic data, we found that a specific subtype of luminal epithelium carries the strongest inherited risk for prostate cancer. We then developed a deep learning model, DEEP+, to predict how millions of common genetic variants influence gene regulation in this cancer-relevant cell type. Our findings highlight key genetic regions and genes that affect cancer-related pathways and androgen receptor (AR) signaling, providing new insight into how inherited variants shape prostate cancer susceptibility.

Introduction

Prostate cancer is a leading malignancy among men in the United States, accounting for more than a quarter of all cancer diagnoses and ranking among the top causes of cancer-related mortality in men [1]. Its pronounced heritability (estimated at approximately 57%) [2] underscores a substantial genetic component, suggesting that germline variants contribute significantly to its molecular underpinnings. Although genome-wide association studies (GWAS) have identified a number of these germline variants [3–14], the mechanistic details remain elusive, leaving the broader germline genetic architecture of prostate cancer largely uncharacterized. By contrast, our current understanding of the disease's genetic basis has come from studies of somatic mutations [15].

One key limitation of GWAS is their tendency to pinpoint sentinel variants marking disease-associated haplotypes without offering a clear path to identifying the causal alleles. Therefore, many biologically relevant variants remain obscured within sets of statistically associated loci. Furthermore, common variants often exhibit modest effect sizes, which cannot be effectively captured by conventional GWAS relying on a uniform genome-wide significance threshold. This challenge is particularly acute in prostate cancer, where risk is not attributed to a few dominant loci but instead emerges from a highly polygenic architecture involving the aggregated effects of numerous loci across the genome. Without bridging the gap between statistical association and functional mechanism, our understanding of how germline variation drives prostate cancer risk will remain incomplete.

Our recent analyses of significant prostate cancer GWAS loci listed in the NHGRI-EBI GWAS Catalog [16] revealed their predominant localization to noncoding human

genome. This pattern aligns with growing evidence that regulatory variants play a pivotal role in complex diseases [17,18]. However, such regulatory variants further complicate the search for functional alleles: because gene regulation is both tissue- and cell-type-specific, potentially pathogenic germline variants cannot be solely inferred from genomic sequences but need to be functionally characterized in the appropriate cellular context.

The prostate gland is distinguished by its remarkable cellular heterogeneity, with glandular acini lined by luminal epithelial (LE) cells that mediate androgen signaling. These LE cells are derived from basal epithelial (BE) cells that form a protective progenitor layer beneath them [19]. Beyond the epithelial compartment, the functional integrity of the prostate and its dynamic responsiveness to hormonal cues are sustained through intricate interactions with stromal, smooth muscle, and immune cells. Despite substantial mutational heterogeneity across patients, genetic and molecular alterations in prostate cancer are convergent onto dysregulated androgen receptor (AR) signaling [20–24]. In the normal prostate, the AR activation by androgens, such as testosterone and dihydrotestosterone, enables it to function as a transcription factor regulating genes involved in cell proliferation, differentiation, and survival. In prostate cancer, alterations in AR signaling promote tumor initiation and progression [21]. While AR is highly expressed in luminal epithelial cells and the prostate stroma, their respective contributions to prostate cancer etiology remain elusive. More importantly, human luminal epithelial cells display substantial cellular heterogeneity [25], naturally raising the question of which specific cell state(s) or subtype(s) contribute the most to prostate cancer risk.

Given the pronounced heritability of prostate cancer [2], we herein asked which prostate cell types mediate the strongest germline risk in prostate cancer. Since most germline variants fall in the noncoding genome, contributing to disease through dysregulating gene expression, we set out to map, for the first time, the germline genetic architecture of prostate cancer across cellular subtypes of the prostate. By defining the cell populations in which risk variants operate, we aimed to reveal the core pathways disrupted by germline alleles and how molecular and cellular vulnerability shapes inherited cancer risk. To achieve this, we performed single-cell multiomic profiling of normal human prostate tissues, generating paired reference epigenomes and transcriptomes in single cells (Fig 1A). This approach allowed us to identify distinct cell states and subtypes, facilitating mapping disease-associated variants onto the reference single cell map. Integrating GWAS data, our study uncovered a specific luminal epithelial subtype that mediates the strongest germline risk in prostate cancer. Subsequent deep learning analysis revealed high-confident loci affecting the gene regulatory landscape within this luminal lineage, converging on a FOXA1/AR-mediated regulon. Notably, this germline-driven convergence onto FOXA1/AR pathways parallels the well-established somatic mutations that also funnel into AR-regulated networks, underscoring a shared molecular axis of prostate cancer pathogenesis. These insights underscore that the germline architecture of prostate cancer risk is both functionally coherent and cell-type-specific, carrying significant implications for clinical practice. Identifying the cell types and genes that are most genetically vulnerable is expected to refine risk stratification, guide the development of targeted interventions, and ultimately improve patient outcomes. As the genetic foundation of prostate cancer comes into sharper focus, leveraging these insights for improved diagnosis, prevention, and treatment becomes increasingly essential.

Results

Constructing the reference single-cell multiomics dataset in the human prostate

Because complex diseases are primarily driven by noncoding regulatory mutations, we began our investigation by examining germline variants from a large cohort of prostate cancer patients. Our goal was to determine whether at-risk alleles are preferentially enriched within active epigenomic regions of specific prostate cell types. For the cell types found to be (epi)genetically vulnerable, we then performed a genome-wide scan to quantify how individual germline alleles alter cell-type-specific gene regulation (Fig 1A). Central to this approach is the construction of a reference regulatory landscape for each cell type in healthy prostate tissue for downstream fine mutation mapping. Establishing such a baseline dataset allows us to evaluate how at-risk alleles shift normal gene regulatory programs toward a disease-prone state.

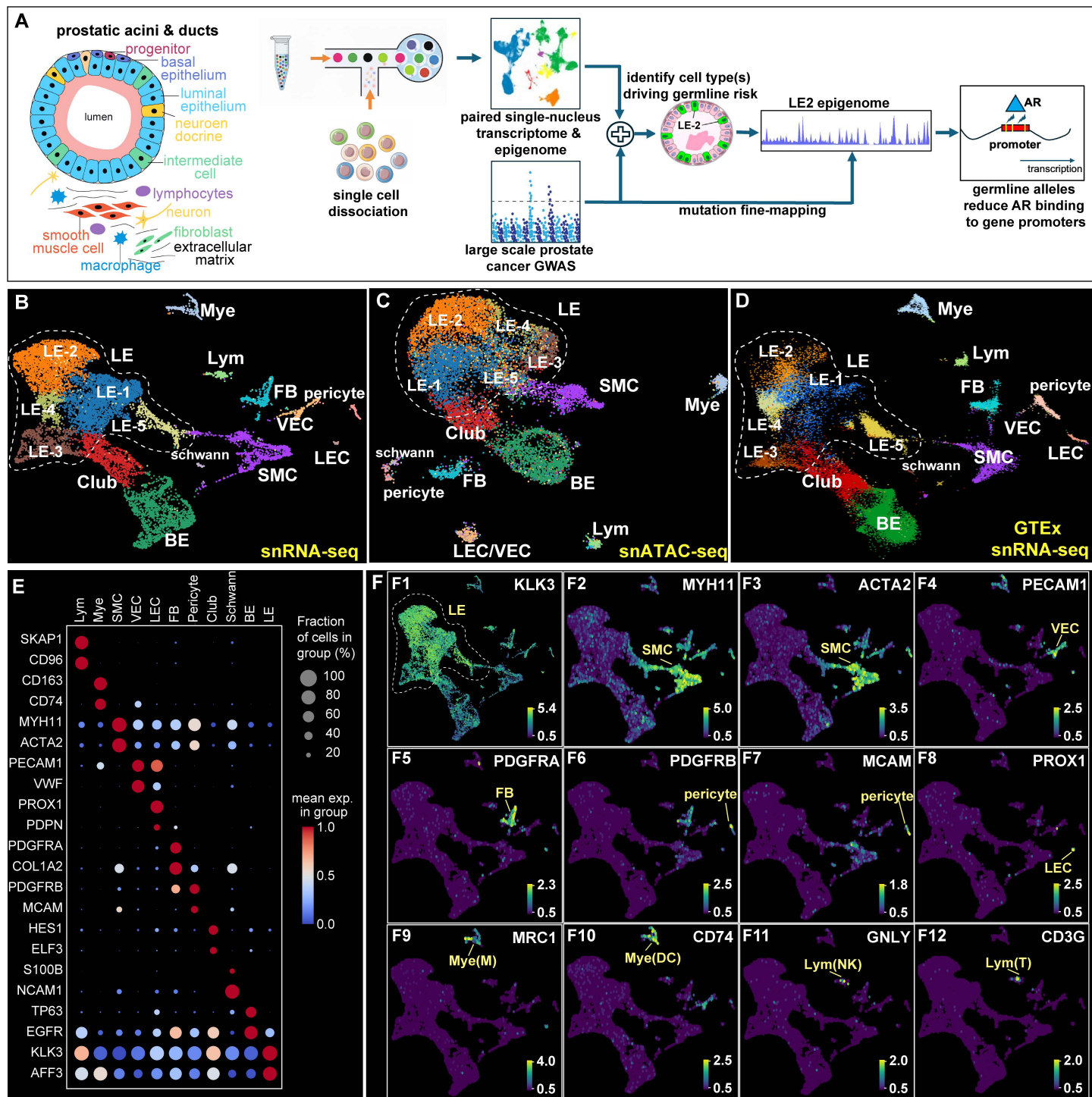


Fig 1. Integrating single-cell multiomics to uncover germline risk in prostate cancer. (A) Schematic overview of the study: Normal prostate gland tissues were analyzed using 10X Chromium snMultiome to jointly profile snRNA-seq and snATAC-seq. Single-cell data integration with large-scale prostate cancer GWAS identified a luminal epithelial subtype (LE-2) as the primary mediator of germline risk. Fine mapping of germline variants in LE-2 highlighted high-confidence risk alleles disrupting androgen receptor (AR) binding, predisposing individuals to elevated prostate cancer risk. (B–D) UMAP projections of transcriptomic (B) and epigenomic (C) profiles of normal prostate tissue, with cell types/subtypes identified via label transfer from GTEx data (D), revealing cellular diversity. (E) Dot plot displaying marker gene expression profiles across 11 cell types. Dot size and color intensity represent

the fraction of cells expressing each gene and average expression, respectively. (F1–F12) Marker gene expression visualized on the snRNA-seq UMAP. Abbreviations: LE (luminal epithelium), SMC (smooth muscle cells), VEC (vascular endothelial cells), FB (fibroblasts), Mye (myeloid cells), M (macrophages), DC (dendritic cells), Lym (lymphocytes), NK (natural killer cells), T (T lymphocytes).

<https://doi.org/10.1371/journal.pgen.1011975.g001>

We performed single-cell profiling in reference prostate samples from healthy individuals obtained from the ENCODE consortium and characterized cell-type-specific gene regulation and expression. Using the 10X Genomics single-nucleus multiomic platform (Chromium Single Cell Multiome ATAC + Gene Expression), we profiled the paired ATAC-seq (the open chromatin epigenome) and RNA-seq (the transcriptome) data for total 19,182 nuclei from human prostate tissues. After stringent quality control (S1 Table), we retained 12,020 high-quality single cells with matched RNA-seq and ATAC-seq profiles. We performed clustering analyses in both the epigenomic (Fig 1B) and transcriptomic (Fig 1C) spaces. To optimize the number of clusters, we systematically varied clustering resolution and evaluated two metrics—silhouette score (cluster cohesion) and within-cluster inertia (cluster compactness) (S1A–S1B Fig) (Methods and Materials). The optimal parameters yielded 15 distinct clusters representing the major prostate cell types. To confirm biological relevance, we applied label transfer to align our single-cell transcriptomes with those from healthy prostate samples in the GTEx cohort [26]. The resulting one-to-one correspondence between clusters (Fig 1D) confirmed that our dataset recapitulates the canonical prostate cellular landscape at the population level. Because GTEx provides only transcriptomic without matched epigenomic data, we used our dataset as the reference for subsequent fine-scale variant mapping.

Using curated cell-type markers, we unambiguously annotated each cluster to a specific prostate cell type (Fig 1E). The concordance between epigenomic and transcriptomic cluster identities suggested that prostate cell types are primarily defined at the epigenomic level and manifested at the transcriptomic level. We systematically examined cell-type-specific marker expression: as expected, luminal epithelial (LE) cells were readily identified by the prostate-specific antigen *KLK3* (PSA) (Fig 1F1). Smooth muscle cells displayed canonical markers *MYH11* and *ACTA2*, whereas vascular endothelial cells were distinguished by *PECAM1*(CD31), and lymphatic endothelial cells were marked by *PROX1*. Fibroblasts were characterized by *PDGFRA* expression, and perivascular cells were marked by *PDGFRB* and *MCAM*. Among immune-associated cell types, myeloid populations included macrophages (*MRC1*⁺) and dendritic cells (*CD74*⁺), and lymphoid populations encompassed NK cells (*GNLY*⁺) and T cells (*CD3G*⁺) (Fig 1F). Overall, these observations underscore the high quality of our dataset and confirm that our single-cell profiling captures the expected repertoire of cell types present in the normal human prostate gland.

Dissecting luminal epithelium into five distinct cell subtypes/states

We observed notable heterogeneity within the epithelial compartment, extending from basal epithelial (BE) cells and club cells to five distinct luminal epithelial (LE) states or subtypes (LE-1–LE-5; Fig 1B, 1C). The presence of these subtypes in the independent GTEx dataset (Fig 1D) confirmed that they are representative components of normal prostate tissue. To further elucidate their molecular identities, we performed differential gene expression analyses. Aligning these results with established epithelial markers, we annotated each subtype and demonstrated their distinct transcriptomic profiles (Fig 2A). As expected, BE cells were uniquely marked by the known marker *TP63*, whereas club cells were distinguished by *RARRES1* or *KRT7* (Fig 2A, 2B2, 2B3). Club cells also displayed heterogeneity based on a few recently suggested markers like *CP*, *WFDC2* and *PIGR* [27,28] (S2A–S2C Fig). We noted specific expression of *ELF3* in club cells (S2D Fig).

Among the LE subtypes, LE-1 cells displayed progenitor-like characteristics by specifically expressing *TENM1* (teneurin transmembrane protein 1), which has been implicated in prostate cancer differentiation [29]. LE-2 cells specifically expressed *TMEFF2* (Fig 2B6), which is an androgen-dependent suppressor of prostate cancer cell growth. Interestingly, androgen receptor (*AR*) expression was strongest and most specific in the progenitor-like LE-1 and the more differentiated LE-2 subtypes, while its expression was moderate in other LE subtypes (Fig 2B7). This observation suggests significant

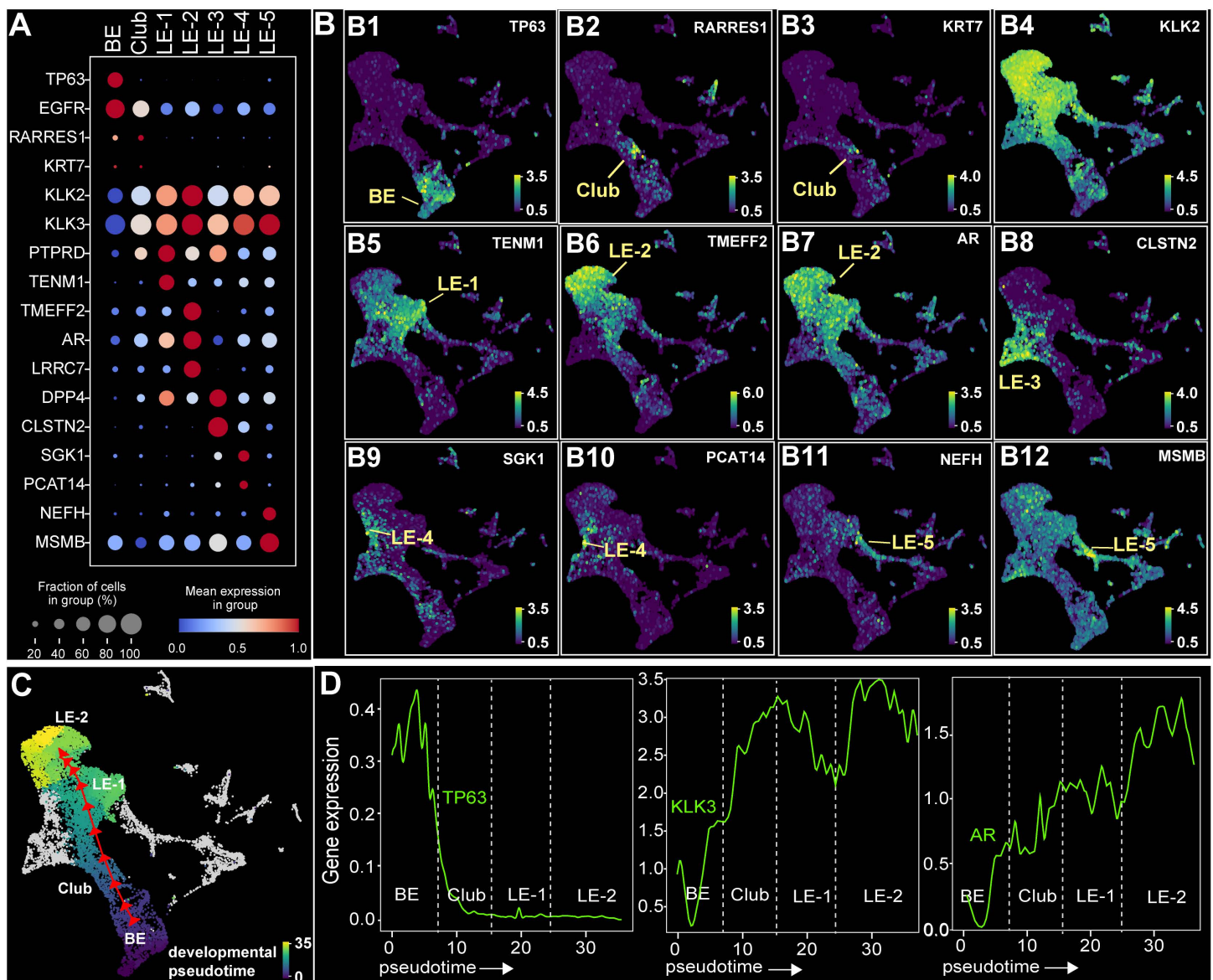


Fig 2. Molecular characteristics of luminal epithelial cell subtypes. (A) Dot plot highlighting marker genes that define BE (basal epithelium), club cells, and five LE subtypes (LE-1–LE-5). (B1–B12) Specific expression of LE subtype markers on the single cell transcriptome UMAP. (C) Differentiation trajectory reconstructed from BE cells to LE-2, with pseudotime indicating cell developmental progression. Arrow marks the primary differentiation direction. (D) Expression dynamics of *TP63*, *KLK3*, and *AR* during transitions from BE to LE-2, showing *TP63* down-regulation and *KLK3*/*AR* up-regulation along the developmental trajectory.

<https://doi.org/10.1371/journal.pgen.1011975.g002>

roles for LE-1 and LE-2 in mediating *AR*-dependent functions. In sharp contrast to LE-1 and LE-2, LE-3 cells were specifically marked by *DPP4* (Dipeptidyl peptidase 4), an *AR*-stimulated tumor suppressor in prostate cancer. In parallel, LE-4 cells were uniquely characterized by *SGK1* (glucocorticoid-induced protein kinase 1) (Fig 2B9), which is upregulated in primary prostate tumors [30], as well as the well-established RNA gene *PCAT14* (prostate cancer associated transcript 14, Fig 2B), a biomarker for prostate cancer diagnosis [31] and response to androgen-deprivation therapy [32]. LE-5 cells, while expressing common luminal markers such as *KLK3*, exhibited a more divergent transcriptome compared to LE-1–4

(Fig 2B). Notably, LE-5 specifically expressed *NEFH* and *MSMB*. *NEFH* functions as a tumor suppressor [33], while *MSMB* encodes a prostate-secreted protein and biomarker for prostate cancer [34]. Both *NEFH* and *MSMB* are downregulated in prostate cancer [34,35]. These observations collectively suggest that LE-1–5 subtypes are pervasively involved in prostate cancer, each contributing distinct features to disease pathogenesis.

As LEs are developmentally derived from BEs, we leveraged the single-cell transcriptomic data to infer cell developmental trajectories. We reconstructed the differentiation continuum from BE, club cells, the progenitor-like LE-1 to the mature LE-2 subtype (Fig 2C). Though the precise developmental origin of the club cell remains unresolved, our analysis suggested a transitioning nature of club cells and LE-1 subtype, and confirmed that LE-2 represents a terminally differentiated LE subtype. Along the LE-2 maturation trajectory, we observed a progressive decline in the BE marker *TP63*, accompanied by a steady increase in the PSA gene *KLK3* and *AR* along cell development progression (Fig 2D). Notably, LE-3/4 and LE-5 were in divergent developmental trajectories from LE-2, suggesting distinct branches of the cell fates (S2E–S2F Fig).

Constructing single cell germline risk map for prostate cancer

After generating paired single-cell chromatin epigenomes and transcriptomes from our reference prostate tissues, we next aimed to identify the cell types most strongly associated with germline genetic risk for prostate cancer. To this end, we leveraged a large-scale prostate cancer genome-wide association study (GWAS) dataset comprising 79,194 cases and 61,112 controls of European ancestry [10]. This dataset provided summary statistics (including risk Z-scores) for approximately 17 million germline single nucleotide polymorphisms (SNPs). To integrate these GWAS results with our single-cell profiles, we applied the SCAVENGE algorithm [36]. Briefly, SCAVENGE aggregates the GWAS risk Z-scores of SNPs located within each cell's active regulatory elements (i.e., open chromatin regions), conceptually generating a chromatin-informed polygenic risk score for each cell. Because each cell exhibits a unique epigenomic configuration, the resulting aggregated risk associations differs from cell to cell. However, cells belonging to the same cell type share highly similar epigenomic landscapes, and therefore tend to exhibit similar genetic risk. Thus, by comparing SCAVENGE scores across single cells and across cell types, we can pinpoint cell type(s) most vulnerable to germline genetic risk for prostate cancer.

We applied SCAVENGE to integrate our single-cell multiomics data with large-scale prostate cancer GWAS results, generating a single-cell germline risk map (Fig 3A, 3B). Notably, cells exhibiting significant prostate cancer risk associations were concentrated within the LE-2 subtype, a striking contrast to their relative depletion in non-LE cell types (Fig 3A, 3B). This observation was statistically validated at the cell-type level, where LE-2 was the only subtype displaying a significant risk association with prostate cancer (Fig 3C). In other words, the active gene regulatory elements defining the LE-2 subtype are strongly enriched for high-risk germline variants. Moreover, we examined the developmental trajectory from BE to LE-2 (Fig 2C). Grouping cells by pseudotime intervals revealed a progressive increase in germline risk, culminating at the terminally differentiated LE-2 state (Fig 3D). Notably, the risk increase is accompanied by the increasing *AR* expression during LE-2 maturation, suggesting a regulatory convergence between germline risk and *AR*-driven transcriptional programs (Fig 2D). Taken together, our analysis, for the first time, demonstrated that germline genetic risk for prostate cancer is mediated by the terminally differentiated LE-2 cells in the human prostate.

Developing a deep learning model for fine mutation mapping in LE-2 cells

We next sought to identify potentially functional germline alleles in LE-2, with the goal of revealing key molecular pathways rendered vulnerable to germline risk, thereby predisposing individuals to prostate cancer. Given the strong enrichment of high-risk variants in the LE-2-specific epigenome (Fig 3B), we scanned all ~17 million SNPs from the same prostate cancer GWAS cohort to evaluate the regulatory effects of each allele on perturbing local chromatin accessibility in LE-2 cells.

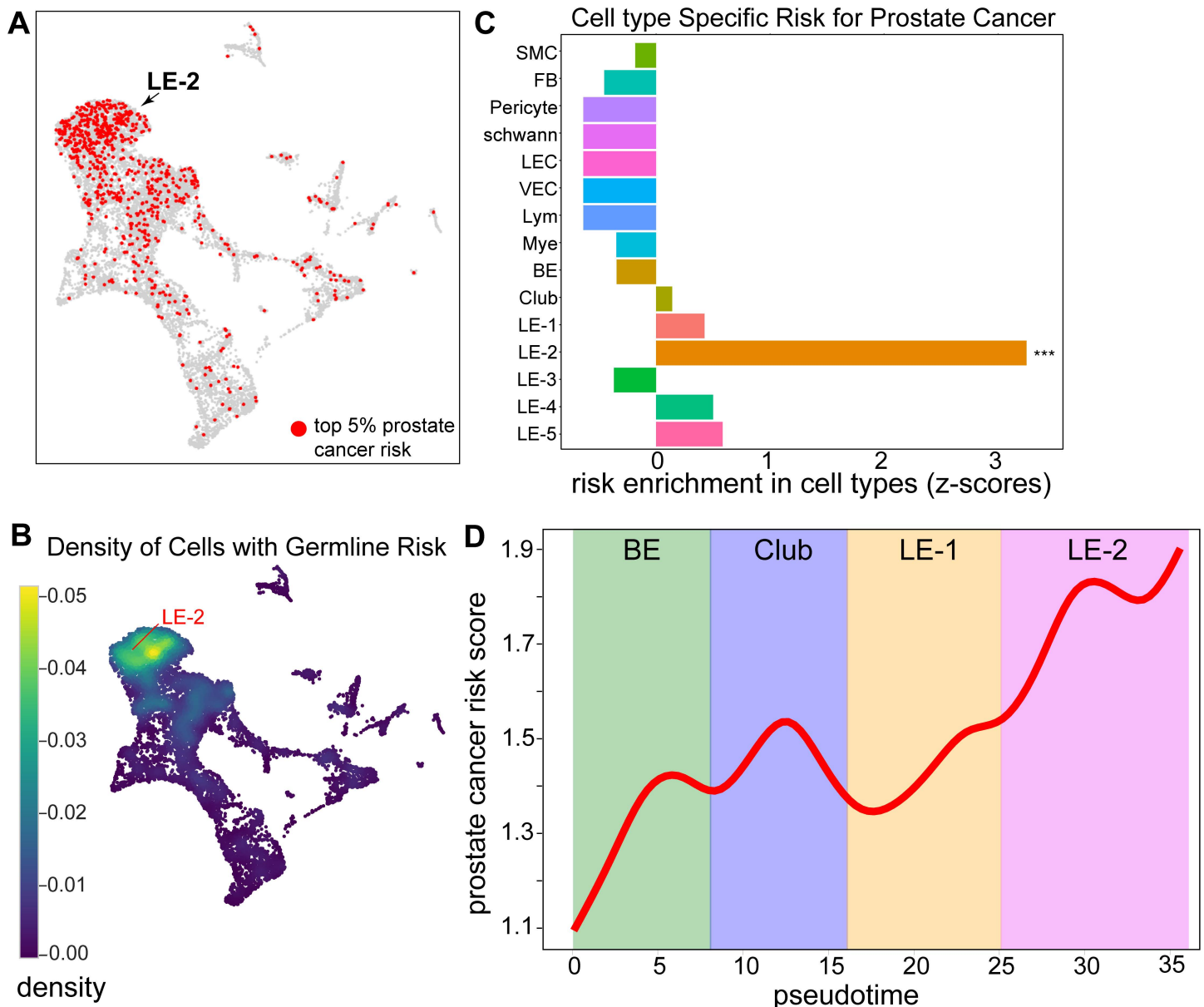


Fig 3. Single-cell germline risk map in prostate cancer. (A, B) Integrative analysis of single-cell multiomics and GWAS data identified vulnerable cells (highlighted in red) to germline risk in prostate cancer. These vulnerable cells (top 5% in red, A) were predominantly concentrated in the LE-2 subtype. Nebulosa plot showing the density of vulnerable cells on the UMAP (B). (C) Statistical permutation analysis confirmed that the LE-2 cell subtype was the only cell population significantly associated with germline risk in prostate cancer. *** denotes $p < 0.001$ (permutation test). Risk enrichment was assessed by Z-scores indicating the enrichment of genetically vulnerable cells in each cell type. (D) Grouping cells into pseudotime bins along the developmental trajectory from basal epithelium (BE) to the terminally differentiated LE-2 subtype revealed a progressive elevation of germline risk along the cell developmental trajectory.

<https://doi.org/10.1371/journal.pgen.1011975.g003>

To achieve this, we employed our recently developed deep learning framework, DEEP (Deep Estimation from Epigenome Prediction) [37,38], which was originally designed to identify regulatory somatic alleles in prostate cancer. As outlined in Fig 4A, the model operates in two sequential stages: [1] It is first trained on human genomic sequences encompassing ATAC-seq peaks from tissues or cell types of interest and is subsequently validated in blind tests to ensure

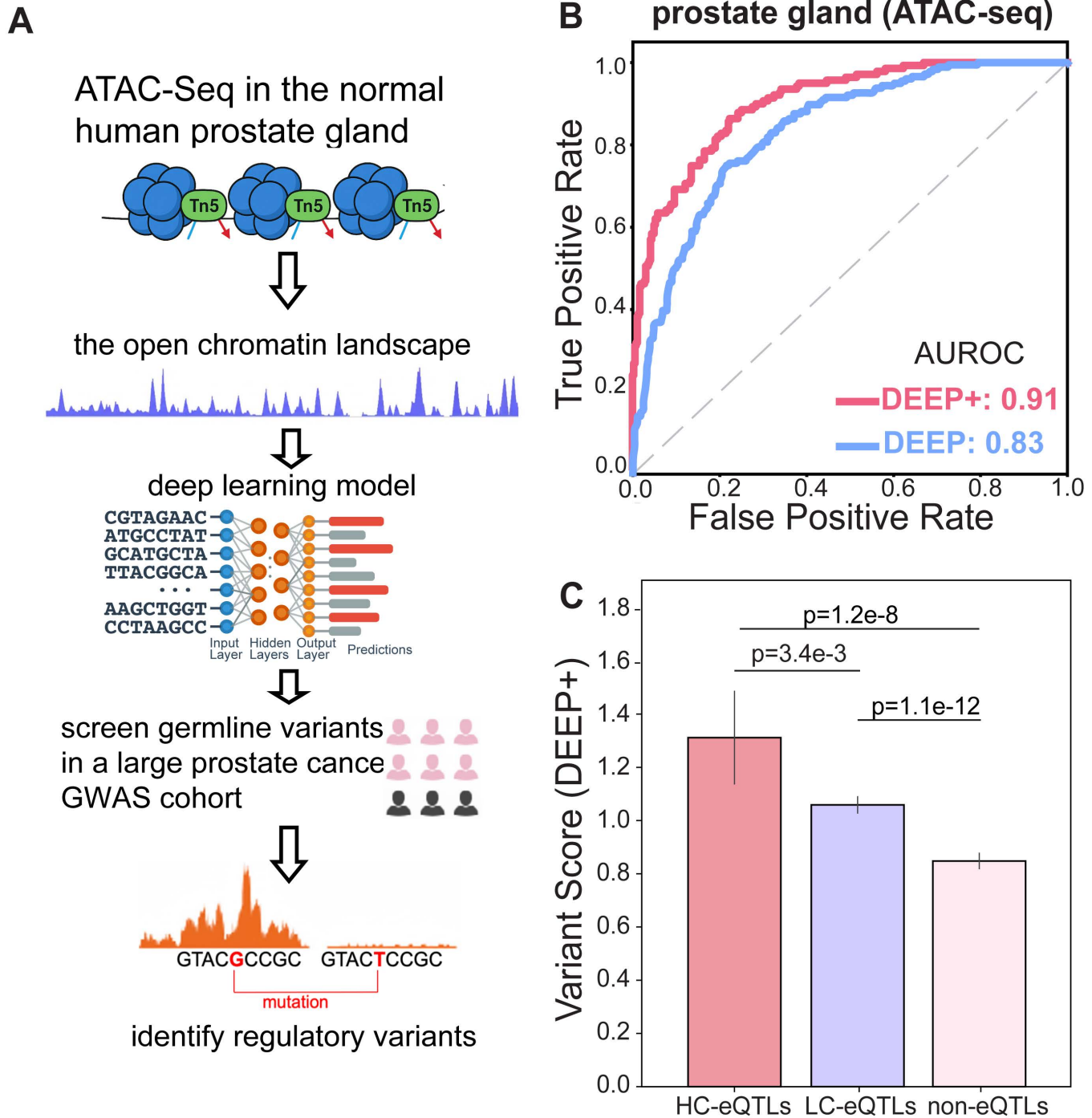


Fig 4. A deep learning framework for fine mutation mapping in LE-2 cells. (A) A deep learning framework was trained to predict chromatin accessibility from genomic sequences. The trained model was then applied to screen the genome, enabling the identification of allelic effects on local chromatin accessibility. **(B)** The enhanced DEEP+ model outperformed the previously developed DEEP model, demonstrating improved accuracy in predicting chromatin accessibility. **(C)** Validation using GTEx prostate eQTL data showed that high-confidence eQTLs (HC-eQTLs) were consistently assigned the highest scores by DEEP+, followed by low-confidence eQTLs (LC-eQTLs) and non-eQTL loci. These results validated the model by demonstrating that loci identified by DEEP+ as having regulatory effects are more likely to impact gene expression.

<https://doi.org/10.1371/journal.pgen.1011975.g004>

robust prediction of chromatin accessibility for any given genomic sequence. [2] Once trained, the model is used to score genomic sequences containing candidate variants by estimating the chromatin accessibility for both the reference and alternative alleles. The difference between these predictions reflects the allelic impact on local chromatin structure, where positive and negative scores, respectively, indicate increased and decreased local chromatin structure for a given allele. In this study, we refined our original DEEP model [35,36] by introducing the DEEP+ framework, which replaces the convolutional neural network (CNN) backbone with a residual neural network architecture (S3 Fig). Since DEEP was initially trained on ATAC-seq data from bulk prostate tissues (from ENCODE) [39], we adopted the same training strategy and dataset for DEEP+. The updated DEEP+ model outperformed the original DEEP model in predicting prostate chromatin accessibility on previously unseen genomic sequences (Fig 4B). To ensure that the chromatin-altering alleles identified by DEEP+ also influence gene expression, we applied the trained DEEP+ model to score 356,598 common variants from the GTEx project [40]. We observed that the prostate eQTLs identified by GTEx were consistently received significantly higher DEEP+ scores (Fig 4C), confirming that regulatory alleles identified by DEEP+ have a significant impact on gene expression in prostate tissue.

Functional characterization of LE-2-specific germline regulatory alleles in prostate cancer

Establishing the model's performance, we next leveraged DEEP+ to screen each germline allele for its regulatory impact in specific cell types. We trained DEEP+ models using cell type-specific open chromatin regions reconstructed from snATAC-seq data via pseudobulk analysis. Independent DEEP+ models were trained for each identified cell type, with particular emphasis on LE-2, which exhibited the strongest genetic predisposition to prostate cancer. Cross-validation confirmed high predictive accuracy for the LE-2 model on open chromatin regions from given genomic sequences (AUROC = 0.86 ± 0.02). We then applied DEEP+ to score each of the ~17 million SNPs in the prostate cancer GWAS dataset described above [10]. Each SNP was assigned a DEEP+ score quantifying its allelic effect on local chromatin accessibility, with positive and negative values indicating increased or decreased chromatin accessibility, respectively. To assess the magnitude of these regulatory perturbation, we examined the absolute DEEP+ scores (regardless of increasing or decreasing chromatin accessibility) for each SNP. We found that most germline variants were functionally neutral, exhibiting minimal impact on prostate chromatin structure (peaked at 0, Fig 5A). However, the long-tailed distribution revealed a distinct subset of germline variants that exerted strong regulatory influences in LE-2 (Fig 5A). Because the presence of regulatory effects alone does not necessarily indicate a relevance to prostate cancer susceptibility, we integrated the absolute DEEP+ scores with the prostate cancer GWAS risk Z-scores. This approach ensured identifying alleles that both significantly alter LE-2 chromatin accessibility and are directly linked to prostate cancer susceptibility. **Notably, GWAS variants were not pruned for linkage disequilibrium (LD) because each variant was independently assessed for its regulatory impact on gene expression.** Consequently, each GWAS variant was jointly represented by its risk Z-score from GWAS and its DEEP+ regulatory effect score, enabling mechanistic interpretation of how germline variants may influence gene regulation in LE-2 to modulate prostate cancer risk.

The joint distribution of these two scores for ~17 million GWAS variants is shown in Fig 5B. Our primary interest focused on the upper-right corner in Fig 5B, where variants not only exert strong regulatory effects in LE-2 (high absolute DEEP+ scores) but also display the highly significant prostate cancer GWAS risk (large |Z|-scores). These variants represented high-confidence functional germline variants with both mechanistic and statistical evidence for involvement in prostate cancer. To systematically prioritize these candidates, we employed a 2D Gaussian framework to model the genome-wide joint distribution of DEEP+ and GWAS risk Z-scores, which were statistically independent (the correlation coefficient $R \approx 0$). We then selected loci with 2-D joint probability exceeding 0.90 (shown in black in Fig 5B). This practice prioritized 1,732 loci, representing approximately 0.01% of the ~17 million variants screened. These loci constitute a high-confidence set of germline variants exhibiting both prostate cancer risk association and putative functional consequences in the LE-2 regulatory landscape (Fig 5B).

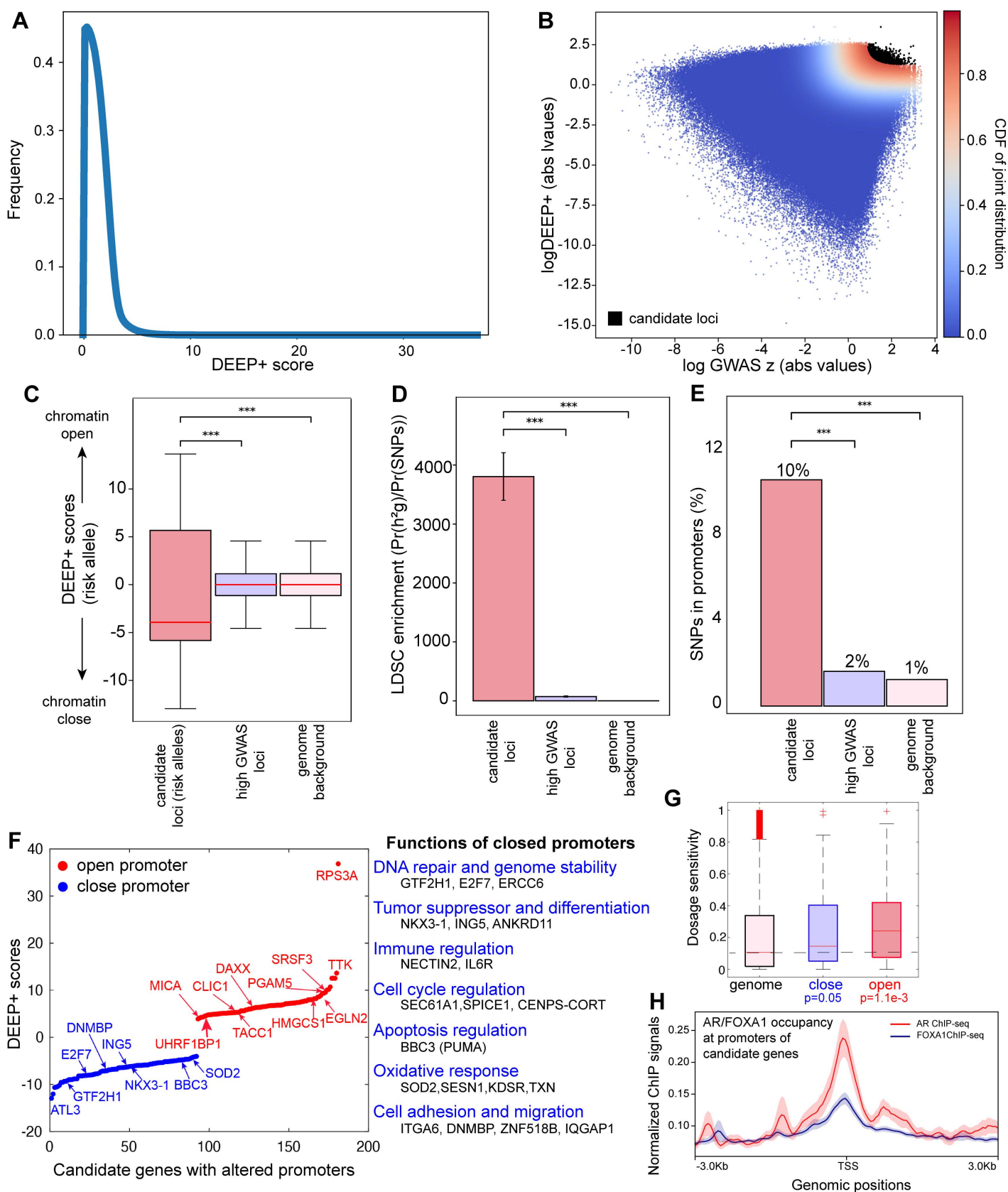


Fig 5. Functional characterization of high-confidence germline alleles in prostate cancer. (A) Screening ~17 million SNPs in the prostate cancer GWAS dataset using DEEP+ identified an excess of alleles with strong regulatory effects on chromatin structure in LE-2 cells, forming a distinct long tail in the score distribution. (B) Plotting the DEEP+ scores (absolute values, y-axis) against GWAS risk Z-scores (x-axis, logarithmic scale) for each SNP revealed high-confidence loci with extreme risk associations and regulatory effects. These loci, representing the top 0.01% of all SNPs scanned, are highlighted as high-confidence candidates (data points in black). (C) The risk alleles (log Z-scores greater than 0) of the identified high-confidence candidate loci were more likely to decrease their local chromatin accessibility in LE-2, reflected by their reduced DEEP+ scores. (D) Linkage disequilibrium score regression (LDSC) analysis confirmed that the high-confidence loci were significantly enriched for prostate cancer heritability compared to loci with high GWAS risk Z-scores alone (without considering DEEP+ regulatory effects) or background genomic loci. Error bars represent standard deviation. (E) The high-confidence loci showed significant enrichment in gene promoter regions, revealing mutational convergence on gene promoters. P values were derived by Chi-square test. (F) Candidate genes with altered promoter accessibility due to the identified germline risk alleles (those with log Z-scores greater 0) are shown. Genes with decreased promoter accessibility (negative DEEP+ scores) are highlighted in blue, while genes with increased promoter accessibility (positive DEEP+ scores) are highlighted in red. Representative genes and functional categories (for those exhibiting reduced promoter accessibility) are also indicated. (G) Genes with perturbed promoters by germline risk alleles (open or close chromatin accessibility) display significantly increased gene dosage sensitivity. P values were derived from Wilcoxon rank-sum test. (H) The identified candidate genes showed co-occupancy of FOXA1 and AR at their promoters. ChIP-seq data for FOXA1 and AR binding were normalized based on benign prostate tissue samples. This co-occupancy underscores the convergence of germline risk onto FOXA1/AR-mediated regulon. All P values were corrected by the Benjamini-Hochberg method.

<https://doi.org/10.1371/journal.pgen.1011975.g005>

Our variant prioritization above considered the magnitude of chromatin structure alterations (the absolute values of DEEP+ scores). For each variant, we then identified its risk-associated allele (log Z score > 0), and observed that these risk alleles predominantly reduced chromatin accessibility (i.e., receiving negative DEEP+ scores, $p = 3e-3$, Wilcoxon rank-sum test, Fig 5C). Targeting these 1,732 top candidate loci (S3 Table), we next performed linkage disequilibrium score regression (LDSC) [41] and observed a striking 3,806-fold enrichment for prostate cancer risk relative to the genomic background ($P = 2.52e-29$). This enrichment far exceeded the more modest 73-fold risk enrichment observed when selecting loci solely by high GWAS risk scores (with the same risk Z-score threshold applied to the top candidate loci) without considering their regulatory allelic effects (Fig 5D). This contrast was expected due to that many high risk GWAS variants were merely driven by linkage disequilibrium proxies rather than direct biological activity, underscoring the importance of incorporating functional insights when interpreting GWAS signals. To gain deeper functional insights, we annotated genomic locations of the identified top candidate loci and found that 181 of the 1,732 loci (~10.4%) resided in gene promoters. This finding demonstrated a significant enrichment for promoter-associated variants, a trend absent when loci were selected solely based on GWAS risk Z-scores (Fig 5E). Interestingly, promoter enrichment further increased under more stringent thresholds (S4A Fig), highlighting the disproportionate representation of promoter-associated variants among the prioritized risk loci. Therefore, by integrating regulatory effects with GWAS risk, we were able to identify functional variants from the statistically associated loci. The pronounced promoter enrichment further indicates that disruption of LE-2 promoter activity is likely a recurrent mechanism underlying germline risk for prostate cancer.

We closely examined the 181 genes with affected promoters (Fig 5E). The magnitude of the alterations on their promoter accessibility is shown in Fig 5F, where at-risk alleles opening promoters (in red) or closing promoters (in blue) are indicated (S4 Table). One prominent example is *rs11781886* at the *NKX3-1* promoter locus, where the risk allele T was predicted to disrupt the chromatin accessibility in LE-2 cells (S4B Fig). Strikingly, genes with risk alleles that increase promoter accessibility were overwhelmingly enriched for oncogenic or prostate-cancer-upregulated genes. For example, *RPS3A* stands out as an extreme outlier, driven by a risk-associated variant that significantly enhanced its promoter accessibility. Overexpression of *RPS3A* induces tumor formation by suppressing apoptosis [42], and its expression is markedly elevated in prostate adenocarcinoma [43]. Similarly, *TTK*, a mitotic checkpoint kinase and recognized oncogene, is positively correlated with higher Gleason scores in prostate cancer [44,45]. Another mitosis-related gene, *TACC1*, which stabilizes microtubules during cell division and facilitates unchecked proliferation, is strongly up-regulated in hormone-refractory prostate cancer with distant metastasis [46]. In the context of hypoxia-driven pathways, *EGLN2* (egl-9 family hypoxia-inducible factor 2), a key regulator of cellular adaptation to low oxygen, shows elevated expression in

prostate cancer, and its ubiquitination by the tumor suppressor *SPOP* constrains cancer cell growth [47]. Hypoxia-induced reprogramming of RNA splicing machinery up-regulates the oncogenic splicing factor *SRSF3*, which promotes prostate cancer progression [48,49]. Similarly, the mitochondrial protein *PGAM5*, which regulates oxidative stress responses and the ROS-induced cell death pathway (known as oxelptosis), exhibits increased expression in high-grade prostate tumors and down-regulated following castration therapy [50–52]. Likewise, the metabolic gene *HMGCS1* is oncogenic in prostate cancer [53], and its overexpression significantly stimulates tumor cell growth [54]. The apoptosis regulator *DAXX* consistently displays up-regulation across cancers, including prostate cancer [55]. Additional cancer-associated genes affected by promoter-associated risk alleles (Fig 5F) include *CLIC1*, which facilitates prostate cancer proliferation and migration [56]; *MICA*, which enhances immune evasion mechanisms [57]; and *UHRF1 BP1*, which supports prostate tumor progression through epigenetic regulation [58]. Taken together, genes up-regulated in prostate cancer tend to harbor germline risk alleles that increase their promoter accessibility, thereby predisposing normal luminal epithelial cells to a cancer-prone state (Fig 5F, genes in red).

Likewise, genes with germline alleles reducing promoter accessibility (Fig 5F, genes in blue) tended to be down-regulated in prostate cancer or function as tumor suppressors. Representative genes in related functional categories are summarized in Fig 5F. We particularly note the well-known prostate tumor suppressor *NKX3–1* [59], which harbors a germline variant that markedly decreases its promoter accessibility, potentially attenuates its transcriptional activity. Several additional tumor suppressors were similarly affected by promoter-closing variants, including *BBC3* (*PUMA*), a pro-apoptotic gene that suppresses prostate cancer cell growth [60], and known tumor suppressors, *ATL3* [61] and *ING5* [62]. Promoter-closing variants further targeted genes essential for genome stability. *GTF2H1* and *ERCC6* regulate nucleotide excision repair [63,64], and their loss is expected to result in the compromised ability of cells in DNA repair, a hallmark of cancer progression. Additionally, *DNMBP*, which supports epithelial structure by regulating apical junctions [65,66], is consistently down-regulated at the protein level across human tumor types (the Human Protein Atlas) [67]. Reduced promoter accessibility in LE-2 may destabilize epithelial integrity and promote metastasis. We particularly note *E2F7*, whose promoter is occupied by p53 [68]. In response to DNA damage, p53-dependent up-regulation of *E2F7* serves as a critical brake on proliferation by repressing cell cycle progression genes [68]. Therefore, germline variants that close the *E2F7* promoter likely obstruct p53 binding, compromising this critical checkpoint and resulting in uncontrolled proliferation. Taken together, the identified promoter-opening alleles tended to affect oncogenes, whereas promoter-closing alleles tended to target tumor suppressors. This directional disruption predisposes LE-2 cells to elevate cancer risk by compromising critical cellular processes such as DNA repair, apoptosis, and cell cycle regulation.

The identified candidate genes are sensitive to dosage alterations

We hypothesized that if the identified germline variants contribute to tumorigenesis by disrupting gene regulation, then the affected genes would likely exhibit dosage sensitivity. To evaluate this, we utilized gene haploinsufficiency scores [69] as a proxy for dosage sensitivity. Haploinsufficient genes are known to be intolerant to dosage imbalance and can be toxic when overexpressed [70] due to the need for tightly controlled expression levels [71]. Our analysis revealed that the identified genes showed significantly higher dosage sensitivity compared to the genomic background (Fig 5G). This enrichment was especially pronounced among genes whose promoter accessibility was increased by germline risk alleles ($P = 1.1 \times 10^{-3}$, Wilcoxon rank-sum test). These findings indicate that our identified allelic perturbations preferentially affect dosage-sensitive gene promoters, supporting their functional consequences to prostate cancer susceptibility.

The at-risk germline alleles are convergent onto the AR-mediated regulon

Since the identified genes and their associated allelic effects were specific to the LE-2 subtype, which is characterized by the highest expression of the androgen receptor (AR, Fig 2B and 2D), we investigated whether these genes, despite their involvement in diverse functional groups (Fig 5F), converge on AR-associated pathways, a hallmark of prostate cancer

etiology [21,72]. In the normal prostate, AR recognizes its target genes through cooperation with the pioneering chromatin remodeler FOXA1, which opens chromatin to facilitate AR binding at prostate-specific loci. Disruption of FOXA1 guidance leads to promiscuous AR binding to androgen response elements, even in the absence of androgen, leading to aberrant transcriptional activation of oncogenic programs [73] and activating genes associated with castration-resistant prostate cancers (CRPC) [24]. Therefore, the delicate balance of transcriptional co-regulation by FOXA1 and AR defines the AR-mediated regulatory program in prostate cells, and perturbation of this equilibrium is a key driver of tumorigenesis [22].

Interestingly, the high-confidence candidate genes identified in this study, which are affected by germline variants in prostate cancer, displayed strong enrichment for FOXA1 targets based on ChIP-seq data from non-cancerous LHSAR cells [23], an immortalized prostate epithelial line expressing exogenous AR [74] (FDR=8.23e-7, Methods and Materials). To validate these findings, we first confirmed that *FOXA1* is co-expressed with *AR* in LE-2 cells (S5 Fig). We then re-analyzed ChIP-seq data for FOXA1 and AR in benign human prostate samples [23] and observed significant co-binding peaks at the promoters of the identified genes (Fig 5H). This co-occupancy suggests that these promoters are likely targets of FOXA1-assisted AR recruitment, establishing prostate-specific AR-mediated regulatory networks. This co-binding was also evident for genes with promoters either opened or closed by germline variants (blue or red genes, Fig 5F) (S6 Fig). These observations imply that increased promoter accessibility likely hyperactivates AR function, while decreased accessibility likely impairs FOXA1/AR binding to prostate-specific genes. Together, these germline risk variants would dysregulate the FOXA1/AR equilibrium via remodeling promoter accessibility, potentially converting LE-2 cells into a cancer-prone state.

Discussion

This study presents the first germline risk architecture of prostate cancer at single-cell resolution, unveiling the cellular basis of heritability that has remained elusive in literature. Whereas most existing studies have focused on somatic mutations, germline variants have largely been explored through GWAS, which, despite its utility, has struggled to translate statistical associations into functional mechanisms. By integrating GWAS with single-cell data, our study systematically partitioned germline risk across distinct cell types, enabling the identification of the cell populations most genetically susceptible to prostate cancer. Furthermore, our deep learning framework provided a novel approach to quantify the regulatory effects of individual alleles genome-wide, offering a scalable strategy to functionally interpret GWAS risk loci. Importantly, the germline risk loci and their target genes uncovered in our study converge on AR pathways, consistent with well-established somatic drivers of prostate cancer. This highlights a shared mechanism between germline and somatic alterations, emphasizing the pivotal role of AR signaling in prostate tumorigenesis. Overall, our findings provide critical cellular and molecular insights into the heritability of prostate cancer. This integration of genomics, single-cell analysis, and deep learning establishes a generalizable blueprint for dissecting germline predisposition across diverse cancer types.

Our single-cell analysis identified distinct luminal epithelial subtypes, and many well-known prostate cancer-associated genes are uniquely expressed in specific luminal subtypes (Fig2). This observation highlights that different luminal subtypes may collectively contribute to prostate cancer susceptibility and progression. Our findings were validated independently using GTEx dataset, which recapitulated the characteristics of luminal subtypes in the prostate. Notably, we examined published single-cell data from primary prostate tumors [28] but many of these luminal subtypes were not detectable. For instance, LE-2 cells showing high *LRRC7* expression were absent in primary prostate tumors (S7 Fig). This likely reflects the massive dysregulation in malignant luminal epithelial cells that disrupts normal cell-type identities. Therefore, sequencing normal prostate samples will not only provide a reference baseline for fine mapping mutations but also offer unique insights into the early events of tumor initiation. Although this study focused primarily on the LE-2 subtype, this does not preclude the contribution from other luminal subtypes or their progenitors (e.g., the club-like cells) through distinct mechanisms. Thus, the uncovered luminal subtypes warrant further investigation to elucidate their respective roles in tumor initiation, progression, and treatment responses.

Our study is particularly focused on the terminally differentiated luminal epithelial cells (LE-2), which exhibit the strongest enrichment of germline risk in prostate cancer. Thus, we propose that terminally differentiated LE-2 cells represent

the cell of origin for germline predisposition in prostate cancer. In early studies, basal epithelial cells were considered the likely cell of origin in prostate cancer [75–77], and luminal progenitor cells have also been implicated [78,79]. Our findings suggest a developmental transition from luminal progenitors, including club cells, to terminally differentiated luminal epithelial subtypes prior to malignant transformation. This model is consistent with *in vivo* lineage-tracing studies that identified luminal epithelial cells as the likely cells of origin for prostate cancer [77], but further investigation is needed to delineate the full differentiation trajectory, particularly the transient club-cell state. Prostate cancer predominantly display a luminal phenotype and expresses luminal markers [79,80]. Our single-cell germline mapping now provides genetic evidence that the terminally differentiated luminal epithelial cells (the LE-2 subtype) is the cell type of origin in prostate cancer given the substantially amplified germline risk in this subtype (Fig 3). At the molecular level, the chromatin landscape of stem-like and progenitor cells is more poised or repressed [81,82], limiting the ability of germline variants to affect regulatory activity. Germline effects are expected to manifest when these regulatory elements are progressively activated in a more differentiated state [83–85]. This notion is well supported by our analysis, where germline risk progressively increased along the cell developmental trajectory, culminating in the terminally differentiated LE-2 cells (Fig 3C). Furthermore, mature luminal cells are exposed to various physiological stress, such as oxidative stress and inflammation, due to their metabolic activity and role in prostatic fluid secretion [86,87]. These stressors may amplify the regulatory effects of germline variants, thereby increasing susceptibility to malignant transformation. While our data showed that LE-2 cells as the primary mediators of germline risk in prostate cancer, we emphasize that this does not exclude the contributions from other cell types driven by somatic mutations, environmental stimulation or other pathogenic mechanisms.

Our study next leveraged our deep learning model to screen each germline allele across the genome for its potential regulatory effects in LE-2. Unlike most of the existing generalized models that learn DNA sequence patterns from heterogeneous datasets (such as local genomic features like ChIP-seq and chromatin accessibility) from different tissues, our framework is task-oriented and specifically tailored to prostate tissues and LE-2 cells by training on our own single-cell multiomic profiles. Thus, our design would help identify pathogenic alleles with cell-type-specific regulatory consequences instead of merely capturing conserved functional loci. By assessing allelic influences on gene regulation, we established mechanistic links between GWAS associations and underlying functional consequences (Fig 5B). This approach thus offers a powerful framework to derive functional insights from statistical associations and underscores the substantial role that the noncoding genome may play in prostate cancer etiology. Notably, these findings appear to contrast with previous interpretations. While it is commonly accepted that noncoding regulatory variants drive complex diseases [88], the recent Pan-Cancer Genome Analysis proposed that noncoding regulatory mutations may not significantly influence cancer development [89]. However, the conclusion relied solely on comparing raw mutation counts between tumors and controls, without considering the functional impact of individual variants on gene regulation. By incorporating a deep-learning-based framework to move beyond statistical associations and directly quantify the regulatory consequences of every single allele, our study paves the way for a more refined and mechanistic understanding of the genetic architecture of cancer.

For the identified germline alleles in prostate cancer, we identified a prevalent mechanism of perturbing gene promoters. Examination of individual affected genes revealed a strong pattern: alleles that increase promoter accessibility tend to affect oncogenes, whereas alleles that decrease promoter accessibility predominantly target tumor suppressors. Additional regulatory elements such as enhancers and silencers influenced by genetic perturbations likely play significant roles in prostate cancer susceptibility. Definitive functional interpretation of these elements as oncogenic or tumor-suppressive will require future experimental validation, such as CRISPR-based genome editing in prostate-relevant cell types. Thus, our observations suggest that germline risk variants adopt a tumor-like epigenomic configuration on LE-2 cells, preconditioning them toward a pre-malignant state. Notably, many of the re-programmed promoters belong to genes involved in DNA repair (Fig 5F), implying that germline perturbations could diminish cells genome stability and sensitize cells to the accumulation of additional somatic alterations, thereby accelerating malignant progression. We particularly highlight that the genes affected by risk germline variants are convergent onto the FOXA1/AR-mediated regulon, but the direct allelic

effects on FOXA1 binding affinity will require experimental validation through CRISPR base editing or massive parallel reporter assays. Our finding is in consistency with prior work showing that non-coding mutations accumulating in prostate cancer are likely perturb the binding affinity of FOXA1 as well as alter the AR cisomes [90]. Moreover, extensive studies [22–24,73,91–98], including our previous work [37], demonstrated that somatic mutations in prostate cancer frequently converge on the FOXA1/AR pathways. Thus, regardless of mutation origin, perturbation of the FOXA1/AR-mediated regulatory program emerges as a unifying mechanism that drives prostate tumor formation and progression.

Germline variants have traditionally been considered as having weak effect sizes. However, the effect sizes are derived from allele frequency differences between case and control populations, and may therefore underestimate the biological impact of individual alleles on gene regulation (e.g., the outlier affecting *RPS3A* in Fig 5F). Even if single variants have weak regulatory effects, each individual typically carries ~4 million germline variants, and the cumulative impacts would substantially predispose individuals to cancer. Thus, identifying germline mutations and genetic predisposition to prostate cancer has significant implications for early detection and intervention. Moreover, because the germline variants we identified converge on the AR-mediated regulon, our findings raise the possibility that epigenetic therapies, including chromatin-modifying agents or AR-targeting drugs, could be tailored to reverse or mitigate the tumor-like epigenomic configuration in genetically high-risk individuals prior to malignant transformation.

In summary, our study, for the first time, revealed a key cell type mediating germline risk in prostate cancer and uncovered high-confidence loci in promoters of genes involved in AR-mediated regulons. This study offers a unifying framework for understanding prostate cancer heritability and establishes a direct connection between inherited genetic risk and epigenetic priming of tumorigenesis, paving the way for innovative screening, diagnosis, and therapeutic intervention aimed at intercepting cancer at its earliest stages.

Methods and materials

Ethics statement

Normal human prostate gland tissues were collected from two deceased donors through the ENCODE Consortium under the existing protocols met the relevant IRB standards. All materials from decedents were fully de-identified. The study was deemed not to involve human subjects and did not require additional IRB approval.

Nuclei Isolation from normal prostate tissues

Normal prostate gland tissues were collected from two deceased donors with written informed consent. Samples were snap-frozen according to GTEx protocols [26] prior to single-nuclei isolation. Nuclei extraction was performed with slight modifications to previously described methods [99]. Briefly, flash-frozen tissues were mechanically ground on dry ice, and 30–50 mg of the ground tissue was transferred to a pre-chilled 7 mL PYREX Dounce homogenizer (Corning Life Science) containing 2 mL of cold homogenization buffer (250 mM sucrose, 0.3% NP40, 5 mM MgCl₂, 25 mM KCl, 10 mM Tris pH 7.8, 1 × protease inhibitors (Roche, cOmplete), and 0.6 U/μL Ribolock RNase inhibitor (Thermo Fisher)). After homogenization, debris and connective tissues were removed using a 40 μm cell strainer (pluriSelect). Intact nuclei were purified using a three-layered OptiPrep Iodixanol (Sigma-Aldrich) gradient (25%, 30%, and 40% from top to bottom). Following centrifugation at 3000 r.c.f. for 30 minutes in a swinging-bucket centrifuge (Eppendorf 5810R), nuclei bands visible at the 30%–40% interface were collected. Nuclei purification and integrity were verified under a microscope, and approximately 10,000 nuclei from each sample were used as input for Chromium Next GEM Single Cell Multiome ATAC + Gene Expression (10x Genomics Inc.).

Single-nucleus multiome library construction and sequencing

Isolated nuclei were processed for single-nucleus multiome library construction using the Chromium Next GEM Single Cell Multiome ATAC & Gene Expression kit (10x Genomics Inc.), following the manufacturer's protocols. Libraries for snATAC-seq and snRNA-seq were sequenced on the Illumina NovaSeq 6000 platform with parameters recommended by

the manufacturer. Raw sequencing signals (BCL format) were demultiplexed into FASTQ files using the CellRanger-ARC suite (v2.0.0, 10x Genomics Inc.). Each snATAC-seq or snRNA-seq library yielded an average of 250 million paired-end reads. CellRanger-ARC was also used for barcode calling, read alignment, and quality assessment using the human reference genome (GRCh38 v3.0.0), following the protocols described by 10X Genomics (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/advanced/references>).

Quality control and doublet detection

Data quality for each sample is summarized in [S1 Table](#). The CellRanger-ARC suite (v2.0.0) was used for initial barcode calling and quality assessment, retaining only high-quality sequenced nuclei. Nuclei passing the following criteria were included in downstream analyses: (a) Gene expression counts (gex_n_counts) between 1,000 and 50,000. (b) Detection of >400 genes (n_genes). (c) Less than 20% mitochondrial gene expression (percent_mito). (d) Total ATAC fragment counts (atac_n_counts) between 1,000 and 100,000. (e) Transcription Start Site (TSS) enrichment score >1. (f) Nucleosome signal strength <2. Among the retained nuclei, we identified potential doublets by Scrublet, and we excluded nuclei that were designated as potential doublets by Scrublet from our downstream analyses [\[100\]](#).

Data Integration and batch alignment

For snATAC-seq, open chromatin peaks were identified in individual samples using MACS2 v2.2.7 [\[101\]](#). Peaks from all samples were unified into genomic intervals, excluding regions blacklisted by ENCODE. The ATAC-seq fragments in peaks were counted as an input matrix for downstream analysis. Peak-by-nuclei count matrices were integrated using reciprocal LSI projection functions in the Signac package [\[102\]](#). For snRNA-seq, gene-by-nuclei count matrices were integrated using reciprocal PCA projections with Seurat. Data normalization, scaling, and variable feature detection were performed following best practices [\[102,103\]](#).

Data integration and batch alignment

For snATAC-seq analysis, open chromatin region peaks were called on individual sample using MACS2 v2.2.7 [\[101\]](#). Peaks from all samples were unified into genomic intervals, and the intervals falling in the ENCODE blacklisted regions were excluded from downstream analyses. The ATAC-seq fragments in peaks were counted as the input matrix for downstream analysis. The peak-by-nuclei count matrices from each sample were integrated by reciprocal LSI projection functions using Signac package. For snRNA-seq data, the gene-by-nuclei count matrices for all nuclei passing quality control were integrated by reciprocal PCA projections between different samples using Seurat. Normalization, data scaling and variable features detection were subsequently performed by Signac/Seurat following the best practice described in Stuart et al. [\[102,103\]](#).

Pseudotime trajectory inference

Pseudotime analysis was conducted with Monocle 3 [\[104\]](#). Pseudotime values for individual cells were projected onto a two-dimensional UMAP embedding and smoothed using neighboring cell information. Differentiation trajectories were visualized with directional arrows that represent the maximum differentiation potential to illustrate progression from initial to adjacent cell states.

Identification of cell type-specific genetic risk

SCAVENGE [\[36\]](#) was implemented to integrate snATAC-seq data with GWAS data [\[10\]](#) to quantify cell type-specific genetic risk for prostate cancer. GWAS summary statistics were obtained from Schumacher et al. [\[10\]](#) via the GWAS Catalog (accession: GCST006085). All GWAS variants were subjected to genome-wide conditional and joint association

analyses using GCTA-COJO [105]. Sentinel loci were identified as those with the highest posterior probability (PP) within each linkage disequilibrium (LD) block. For each cell, the PP values of all sentinel SNPs overlapping open chromatin regions were aggregated to compute a trait-relevant score (TRS) for prostate cancer. Permutation tests were performed to assess the robustness of genetic vulnerability at the single-cell level, providing confidence in the TRS calculations. Cells with TRS values in the top 5% were classified as risk associated. To further refine the analysis, additional cell type-specific permutation tests using label shuffling were conducted to identify significant enrichment of risk-associated cells within specific cell types. The normalized proportion of risk-associated cells in each cell type was then calculated and visualized, offering insights into cell type-specific genetic predisposition to prostate cancer. All calculations were implemented by SCAVENGE [34].

The DEEP+ model

The DEEP+ model was extended from our earlier DEEP model [106]. In this study, we trained the model using bulk prostate ATAC-seq data as we performed in our previous work [106] or using the ATAC-seq data in this study only from the LE-2 cells. Pseudobulk ATAC-seq peaks for each cell type were called using MACS2 v2.2.7 [101]. A residual-network-based DEEP+ neural network [106,107] was trained to capture sequence contexts defining cell type-specific open chromatin. Cross-validation (20-fold) was performed to optimize the model, and a held-out chromosome was used for independent validation. The model was applied to quantify the allelic effects of GWAS variants on chromatin accessibility in different cell types. Model performance was assessed using ROC curves on the held-out test chromosome.

eQTL analysis

Prostate gland eQTLs were identified using CAVIAR fine-mapping data from GTEx Analysis V8 (dbGaP Accession phs000424.v8.p2) [38]. Variants were categorized based on their posterior probability (PP) as follows: those with $PP > 0.9$ were designated as high-confidence eQTLs (HC-eQTLs), those with $PP < 0.1$ were classified as non-eQTLs, and variants with PP values falling between these thresholds were considered low-confidence eQTLs (LC-eQTLs).

Identification of DEEP+ candidate loci and genes

Variants were prioritized using a two-dimensional Gaussian model that integrated genome-wide DEEP+ scores and GWAS Z scores. To ensure consistency and account for score magnitude, the absolute values of the scores were logarithmically transformed prior to analysis. Variants with a joint probability exceeding 0.90 were designated as candidate loci. Associated genes were subsequently annotated using Homer [108], enabling the identification of regulatory elements and promoter regions linked to these high-confidence variants.

Partitioned heritability analysis

Partitioned heritability analysis was implemented by LD score regression as previously described [109,110]. Briefly, heritability was partitioned across loci grouped into three categories: (A) prioritized candidate loci ($n = 1,732$) receiving high DEEP+ and GWAS Z-scores (minimum $Z = 2.47$); (B) genetic loci with GWAS Z-scores exceeding the threshold from group A ($Z = 2.47$) but showing insignificant regulatory effects; and (C) all remaining loci in GWAS datasets. Heritability enrichment was estimated by using the LDSC software package [41,110].

Gene ontology and dosage sensitivity analysis

Enriched functional categories for genes linked to risk alleles were identified using Enrichr [111]. FDR values were computed using the Benjamini-Hochberg correction. Dosage sensitivity was approximated using the haploinsufficiency scores [69], and the Wilcoxon rank-sum test was used to determine statistical significance.

ChIP-seq analysis

ChIP-seq data AR and FOXA1 in benign prostate tissue samples were downloaded from GEO (accession # GSE56288). For a given list of genes, occupancy of FOXA1 or AR at the candidate promoters was quantified by averaging the ChIP-seq signals intensity from multiple bigwig files before aligning the 6Kb flanking regions of the transcriptional starting sites (TSS). Enrichment of the ChIP-seq signals were compared to the genome background and visualized using deepTools and [112].

Supporting information

S1 Fig. Optimization of clustering parameters for Louvain-based cell type identification. Line plots depict metrics used to evaluate clustering performance across a range of Louvain resolution values: (A) mean silhouette scores across clusters, (B) mean within-cluster inertia, and (C) total number of clusters identified. The clustering resolution selected for downstream analyses is highlighted in red.

(PDF)

S2 Fig. (A-D) Expression of club cell markers *CP*, *WFDC2*, *PIGR*, and *ELF3* on the UMAP. (E-F) Reconstructed pseudo-time trajectory of luminal epithelium differentiation from BE to terminal LE states LE-4 (E) and LE-5 (F). Pseudotime of individual cells of each lineage is color-coded.

(PDF)

S3 Fig. Schematic of DEEP+ model for allelic specific chromatin activity analysis. Input DNA sequences are first one-hot encoded, where each nucleotide (A, T, C, G) is represented as a binary vector. The encoded sequences are passed through multiple residual blocks, each consisting of a series of convolutional layers followed by batch normalization and activation functions. Skip connections (labeled as “Identity”) add the input of each residual block to its output to facilitate gradient flow and improve training stability. The final representation is passed through a fully connected layer to generate the output. This architecture enables effective learning of complex regulatory syntax from DNA sequences.

(PDF)

S4 Fig. (A) Bar plots showing the percentage of promoter SNPs with different thresholds on joint probability ($p = 0.05$ – 0.95) from the 2D Gaussian distribution. Risk alleles with extreme regulatory functions are enriched in promoter regions. (B) Genome track showing one candidate promoter SNP *rs11781886* at the *NKX3-1* promoter locus with top DEEP+ score. Prostate cancer risk allele T was predicted to disrupt chromatin accessibility.

(PDF)

S5 Fig. (A) The concordant dynamics of *FOXA1* and *AR* expression along the transition from BE (basal epithelium) to LE-2 (luminal epithelium 2). (B-C) Expression of *AR* (B) and *FOXA1* (C) on UMAP embeddings of snRNA-seq indicates the co-expression of two genes across all luminal epithelial cells.

(PDF)

S6 Fig. (A-B) Co-occupancy of AR (A) and FOXA1 (B) at the candidate promoters in benign prostate tissues across six donors identified by ChIP-seq. Candidate promoters that are closed or opened by the identified risk alleles were highlighted in dark or light blue.

(PDF)

S7 Fig. (A-C) UMAP projection showing *LRRC7* expression in single-cell/nucleus data from (A) normal prostate gland in this study, (B) normal prostate gland in GTEx data, and (C) primary prostate tumors in Song et al.

(PDF)

S1 Table. Quality measurement of snMultiome data in normal prostate gland tissues.
(XLSX)

S2 Table. List of identified marker genes for cell type/subtypes.
(XLSX)

S3 Table. List of identified candidate risk loci showing strongest regulatory effects on local chromatin structure.
(XLSX)

S4 Table. Summary of functional categories of genes associated with candidate risk loci located in promoter regions.
(XLSX)

Author contributions

Conceptualization: Cheng Wang, Jingjing Li.

Data curation: Cheng Wang.

Formal analysis: Cheng Wang, Tianqi Tang, Yuejun Wang.

Funding acquisition: Jingjing Li.

Investigation: Cheng Wang, Tianqi Tang.

Methodology: Cheng Wang, Tianqi Tang.

Project administration: Jingjing Li.

Resources: Jingjing Li.

Software: Cheng Wang, Tianqi Tang.

Supervision: Cheng Wang, Jingjing Li.

Visualization: Cheng Wang, Tianqi Tang, Yuejun Wang.

Writing – original draft: Cheng Wang, Jingjing Li.

Writing – review & editing: Cheng Wang, Tianqi Tang, Jingjing Li.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin.* 2024;74(1):12–49.
2. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial risk and heritability of cancer among twins in Nordic countries. *JAMA.* 2016;315(1):68–76. <https://doi.org/10.1001/jama.2015.17703> PMID: [26746459](#)
3. Cook MB, Wang Z, Yeboah ED, Tettey Y, Biritwum RB, Adjei AA, et al. A genome-wide association study of prostate cancer in West African men. *Hum Genet.* 2014;133(5):509–21. <https://doi.org/10.1007/s00439-013-1387-z> PMID: [24185611](#)
4. Du Z, Hopp H, Ingles SA, Huff C, Sheng X, Weaver B, et al. A genome-wide association study of prostate cancer in Latinos. *Int J Cancer.* 2020;146(7):1819–26. <https://doi.org/10.1002/ijc.32525> PMID: [31226226](#)
5. Xu J, Mo Z, Ye D, Wang M, Liu F, Jin G, et al. Genome-wide association study in Chinese men identifies two new prostate cancer risk loci at 9q31.2 and 19q13.4. *Nat Genet.* 2012;44(11):1231–5. <https://doi.org/10.1038/ng.2424> PMID: [23023329](#)
6. Eeles RA, Kote-Jarai Z, Al Olama AA, Giles GG, Guy M, Severi G, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet.* 2009;41(10):1116–21. <https://doi.org/10.1038/ng.450> PMID: [19767753](#)
7. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet.* 2007;39(5):631–7. <https://doi.org/10.1038/ng1999> PMID: [17401366](#)
8. Haiman CA, Chen GK, Blot WJ, Strom SS, Berndt SI, Kittles RA, et al. Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nat Genet.* 2011;43(6):570–3. <https://doi.org/10.1038/ng.839> PMID: [21602798](#)
9. Hoffmann TJ, Van Den Eeden SK, Sakoda LC, Jorgenson E, Habel LA, Graff RE, et al. A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences. *Cancer Discov.* 2015;5(8):878–91. <https://doi.org/10.1158/2159-8290.CD-15-0315> PMID: [26034056](#)

10. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet.* 2018;50(7):928–36. <https://doi.org/10.1038/s41588-018-0142-8> PMID: [29892016](#)
11. Takata R, Akamatsu S, Kubo M, Takahashi A, Hosono N, Kawaguchi T, et al. Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet.* 2010;42(9):751–4. <https://doi.org/10.1038/ng.635> PMID: [20676098](#)
12. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet.* 2008;40(3):310–5. <https://doi.org/10.1038/ng.91> PMID: [18264096](#)
13. Wang A, Shen J, Rodriguez AA, Saunders EJ, Chen F, Janivara R, et al. Characterizing prostate cancer risk through multi-ancestry genome-wide discovery of 187 novel risk variants. *Nat Genet.* 2023;55(12):2065–74. <https://doi.org/10.1038/s41588-023-01534-4> PMID: [37945903](#)
14. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet.* 2007;39(5):645–9. <https://doi.org/10.1038/ng2022> PMID: [17401363](#)
15. Cotter K, Rubin MA. The evolving landscape of prostate cancer somatic mutations. *Prostate.* 2022;82 Suppl 1(Suppl 1):S13–24. <https://doi.org/10.1002/pros.24353> PMID: [35657155](#)
16. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* 2023;51(D1):D977–85. <https://doi.org/10.1093/nar/gkac1010> PMID: [36350656](#)
17. Corradin O, Scacheri PC. Enhancer variants: evaluating functions in common disease. *Genome Med.* 2014;6(10):85.
18. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet.* 2015;16(4):197–212. <https://doi.org/10.1038/nrg3891> PMID: [25707927](#)
19. Pitzen SP, Dehm SM. Basal epithelial cells in prostate development, tumorigenesis, and cancer progression. *Cell Cycle.* 2023;22(11):1303–18.
20. Schweizer MT, Yu EY. AR-signaling in human malignancies: prostate cancer and beyond. *Cancers (Basel).* 2017;9(1).
21. Heinlein CA, Chang C. Androgen receptor in prostate cancer. *Endocr Rev.* 2004;25(2):276–308. <https://doi.org/10.1210/er.2002-0032> PMID: [15082523](#)
22. Jin H-J, Zhao JC, Wu L, Kim J, Yu J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nat Commun.* 2014;5:3972. <https://doi.org/10.1038/ncomms4972> PMID: [24875621](#)
23. Pomerantz MM, Li F, Takeda DY, Lenci R, Chonkar A, Chabot M, et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat Genet.* 2015;47(11):1346–51. <https://doi.org/10.1038/ng.3419> PMID: [26457646](#)
24. Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J, et al. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell.* 2009;138(2):245–56. <https://doi.org/10.1016/j.cell.2009.04.056> PMID: [19632176](#)
25. Crowley L, Shen MM. Heterogeneity and complexity of the prostate epithelium: New findings from single-cell RNA sequencing studies. *Cancer Lett.* 2022;525:108–14. <https://doi.org/10.1016/j.canlet.2021.10.035> PMID: [34728312](#)
26. Eraslan G, Drokhlyansky E, Anand S, Fiskin E, Subramanian A, Slyper M, et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science.* 2022;376(6594):eabl4290. <https://doi.org/10.1126/science.abl4290> PMID: [35549429](#)
27. Kiviahio A, Eerola SK, Kallio HML, Andersen MK, Hoikka M, Tiihonen AM, et al. Single cell and spatial transcriptomics highlight the interaction of club-like cells with immunosuppressive myeloid cells in prostate cancer. *Nat Commun.* 2024;15(1):9949. <https://doi.org/10.1038/s41467-024-54364-1> PMID: [39550375](#)
28. Song H, Weinstein HNW, Allegakoen P, Wadsworth MH 2nd, Xie J, Yang H, et al. Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nat Commun.* 2022;13(1):141. <https://doi.org/10.1038/s41467-021-27322-4> PMID: [35013146](#)
29. Peppino G, Ruii R, Arigoni M, Riccardo F, Iacoviello A, Barutello G. Teneurins: role in cancer and potential role as diagnostic biomarkers and targets for therapy. *Int J Mol Sci.* 2021;22(5).
30. Szmulewitz RZ, Chung E, Al-Ahmadie H, Daniel S, Kocherginsky M, Razmaria A, et al. Serum/glucocorticoid-regulated kinase 1 expression in primary human prostate cancers. *Prostate.* 2012;72(2):157–64. <https://doi.org/10.1002/pros.21416> PMID: [21563193](#)
31. Yan Y, Liu J, Xu Z, Ye M, Li J. lncRNA PCAT14 is a diagnostic marker for prostate cancer and is associated with immune cell infiltration. *Dis Markers.* 2021;2021:9494619. <https://doi.org/10.1155/2021/9494619> PMID: [35003397](#)
32. White NM, Zhao SG, Zhang J, Rozycki EB, Dang HX, McFadden SD, et al. Multi-institutional analysis shows that low PCAT-14 expression associates with poor outcomes in prostate cancer. *Eur Urol.* 2017;71(2):257–66. <https://doi.org/10.1016/j.eururo.2016.07.012> PMID: [27460352](#)
33. Revill K, Wang T, Lachenmayer A, Kojima K, Harrington A, Li J, et al. Genome-wide methylation analysis and epigenetic unmasking identify tumor suppressor genes in hepatocellular carcinoma. *Gastroenterology.* 2013;145(6):1424–35.e1-25. <https://doi.org/10.1053/j.gastro.2013.08.055> PMID: [24012984](#)
34. Bergström SH, Järemo H, Nilsson M, Adamo HH, Bergh A. Prostate tumors downregulate microseminoprotein-beta (MSMB) in the surrounding benign prostate epithelium and this response is associated with tumor aggressiveness. *Prostate.* 2018;78(4):257–65. <https://doi.org/10.1002/pros.23466> PMID: [29250809](#)
35. Wang T-Y, Liu Q, Ren Y, Alam SK, Wang L, Zhu Z, et al. A pan-cancer transcriptome analysis of exon splicing identifies novel cancer driver genes and neopeptides. *Mol Cell.* 2021;81(10):2246–2260.e12. <https://doi.org/10.1016/j.molcel.2021.03.028> PMID: [33861991](#)

36. Yu F, Cato LD, Weng C, Liggett LA, Jeon S, Xu K, et al. Variant to function mapping at single-cell resolution through network propagation. *Nat Biotechnol.* 2022;40(11):1644–53. <https://doi.org/10.1038/s41587-022-01341-y> PMID: 35668323
37. Wang C, Li J. A deep learning framework identifies pathogenic noncoding somatic mutations from personal prostate cancer genomes. *Cancer Res.* 2020;80(21):4644–54. <https://doi.org/10.1158/0008-5472.CAN-20-1791> PMID: 32907840
38. Wang C, Wang YJ, Ying L, Wong RJ, Quaintance CC, Hong X, et al. Integrative analysis of noncoding mutations identifies the druggable genome in preterm birth. *Sci Adv.* 2024;10(3):eadk1057. <https://doi.org/10.1126/sciadv.adk1057> PMID: 38241369
39. ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature.* 2020;583(7818):699–710. <https://doi.org/10.1038/s41586-020-2493-4> PMID: 32728249
40. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying causal variants at loci with multiple signals of association. *Genetics.* 2014;198(2):497–508. <https://doi.org/10.1534/genetics.114.167908> PMID: 25104515
41. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet.* 2018;50(4):621–9. <https://doi.org/10.1038/s41588-018-0081-4> PMID: 29632380
42. Naora H, Takai I, Adachi M, Naora H. Altered cellular responses by varying expression of a ribosomal protein gene: sequential coordination of enhancement and suppression of ribosomal protein S3a gene expression induces apoptosis. *J Cell Biol.* 1998;141(3):741–53. <https://doi.org/10.1083/jcb.141.3.741> PMID: 9566973
43. Pogue-Geile K, Geiser JR, Shu M, Miller C, Wool IG, Meisler AI, et al. Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein. *Mol Cell Biol.* 1991;11(8):3842–9. <https://doi.org/10.1128/mcb.11.8.3842-3849.1991> PMID: 1712897
44. Zhang L, Jiang B, Zhu N, Tao M, Jun Y, Chen X, et al. Mitotic checkpoint kinase Mps1/TTK predicts prognosis of colon cancer patients and regulates tumor proliferation and differentiation via PKC α /ERK1/2 and PI3K/Akt pathway. *Med Oncol.* 2019;37(1):5. <https://doi.org/10.1007/s12032-019-1320-y> PMID: 31720873
45. Chen S, Wang J, Wang L, Peng H, Xiao L, Li C, et al. Silencing TTK expression inhibits the proliferation and progression of prostate cancer. *Exp Cell Res.* 2019;385(1):111669. <https://doi.org/10.1016/j.yexcr.2019.111669> PMID: 31605696
46. Devilard E, Bladou F, Ramuz O, Karsenty G, Dalès J-P, Gravis G, et al. FGFR1 and WT1 are markers of human prostate cancer progression. *BMC Cancer.* 2006;6:272. <https://doi.org/10.1186/1471-2407-6-272> PMID: 17137506
47. Zhang L, Peng S, Dai X, Gan W, Nie X, Wei W, et al. Tumor suppressor SPOP ubiquitinates and degrades EglN2 to compromise growth of prostate cancer cells. *Cancer Lett.* 2017;390:11–20. <https://doi.org/10.1016/j.canlet.2017.01.003> PMID: 28089830
48. Bowler E, Porazinski S, Uzor S, Thibault P, Durand M, Lapointe E, et al. Hypoxia leads to significant changes in alternative splicing and elevated expression of CLK splice factor kinases in PC3 prostate cancer cells. *BMC Cancer.* 2018;18(1):355. <https://doi.org/10.1186/s12885-018-4227-7> PMID: 29606096
49. Xiong J, Chen Y, Wang W, Sun J. Biological function and molecular mechanism of SRSF3 in cancer and beyond. *Oncol Lett.* 2022;23(1):21. <https://doi.org/10.3892/ol.2021.13139> PMID: 34858525
50. Pallichankandy S, Thayyullathil F, Cheratta AR, Subburayan K, Alakkal A, Sultana M, et al. Targeting oxeiptosis-mediated tumor suppression: a novel approach to treat colorectal cancers by sanguinarine. *Cell Death Discov.* 2023;9(1):94. <https://doi.org/10.1038/s41420-023-01376-3> PMID: 36914635
51. Xu H, Liu W, He C, Mirza M, Li B. Aberrant expression of multiple glycolytic enzyme genes is significantly associated with disease progression and survival outcomes in prostate cancers. *Am J Clin Exp Urol.* 2023;11(6):530–41.
52. Liu S, Kumari S, Hu Q, Senapati D, Venkadakrishnan VB, Wang D, et al. A comprehensive analysis of coregulator recruitment, androgen receptor function and gene expression in prostate cancer. *Elife.* 2017;6:e28482. <https://doi.org/10.7554/eLife.28482> PMID: 28826481
53. Zhou C, Wang Z, Cao Y, Zhao L. Pan-cancer analysis reveals the oncogenic role of 3-hydroxy-3-methylglutaryl-CoA synthase 1. *Cancer Rep (Hoboken).* 2022;5(9):e1562. <https://doi.org/10.1002/cnr2.1562> PMID: 34549901
54. Ashida S, Kawada C, Inoue K. Stromal regulation of prostate cancer cell growth by mevalonate pathway enzymes HMGCS1 and HMGR. *Oncol Lett.* 2017;14(6):6533–42. <https://doi.org/10.3892/ol.2017.7025> PMID: 29163687
55. Mahmud I, Liao D. DAXX in cancer: phenomena, processes, mechanisms and regulation. *Nucleic Acids Res.* 2019;47(15):7734–52. <https://doi.org/10.1093/nar/gkz634> PMID: 31350900
56. Tian Y, Guan Y, Jia Y, Meng Q, Yang J. Chloride intracellular channel 1 regulates prostate cancer cell proliferation and migration through the MAPK/ERK pathway. *Cancer Biother Radiopharm.* 2014;29(8):339–44. <https://doi.org/10.1089/cbr.2014.1666> PMID: 25279971
57. Zhang J, Luo Q, Li X, Guo J, Zhu Q, Lu X, et al. Novel role of immune-related non-coding RNAs as potential biomarkers regulating tumour immunoresponse via MICR/NG2D pathway. *Biomark Res.* 2023;11(1):86. <https://doi.org/10.1186/s40364-023-00530-4> PMID: 37784183
58. Babbio F, Pistore C, Curti L, Castiglioni I, Kunderfranco P, Brino L, et al. The SRA protein UHRF1 promotes epigenetic crosstalks and is involved in prostate cancer progression. *Oncogene.* 2012;31(46):4878–87. <https://doi.org/10.1038/ncr.2011.641> PMID: 22330138
59. Gurel B, Ali TZ, Montgomery EA, Begum S, Hicks J, Goggins M, et al. NKX3.1 as a marker of prostatic origin in metastatic tumors. *Am J Surg Pathol.* 2010;34(8):1097–105. <https://doi.org/10.1097/PAS.0b013e3181e6cbf3> PMID: 20588175
60. Shan Z, Liu Q, Li Y, Wu J, Sun D, Gao Z. PUMA decreases the growth of prostate cancer PC-3 cells independent of p53. *Oncol Lett.* 2017;13(3):1885–90. <https://doi.org/10.3892/ol.2017.5657> PMID: 28454339

61. Wang Z, Xu D, Ding H-F, Kim J, Zhang J, Hai T, et al. Loss of ATF3 promotes Akt activation and prostate cancer development in a Pten knockout mouse model. *Oncogene*. 2015;34(38):4975–84. <https://doi.org/10.1038/onc.2014.426> PMID: [25531328](#)
62. Zheng H-C, Xue H, Jiang H-M. The roles of ING5 in cancer: a tumor suppressor. *Front Cell Dev Biol*. 2022;10:1012179. <https://doi.org/10.3389/fcell.2022.1012179> PMID: [36425530](#)
63. Ribeiro-Silva C, Sabatella M, Helfricht A, Martijn JA, Theil AF, Vermeulen W, et al. Ubiquitin and TFIIH-stimulated DDB2 dissociation drives DNA damage handover in nucleotide excision repair. *Nat Commun*. 2020;11(1):4868. <https://doi.org/10.1038/s41467-020-18705-0> PMID: [32985517](#)
64. Cheng L, Spitz MR, Hong WK, Wei Q. Reduced expression levels of nucleotide excision repair genes in lung cancer: a case-control analysis. *Carcinogenesis*. 2000;21(8):1527–30. <https://doi.org/10.1093/carcin/21.8.1527> PMID: [10910954](#)
65. Otani T, Ichii T, Aono S, Takeichi M. Cdc42 GEF Tuba regulates the junctional configuration of simple epithelial cells. *J Cell Biol*. 2006;175(1):135–46. <https://doi.org/10.1083/jcb.200605012> PMID: [17015620](#)
66. Rajabian T, Gavicherla B, Heisig M, Müller-Altroch S, Goebel W, Gray-Owen SD, et al. The bacterial virulence factor InlC perturbs apical cell junctions and promotes cell-to-cell spread of *Listeria*. *Nat Cell Biol*. 2009;11(10):1212–8. <https://doi.org/10.1038/ncb1964> PMID: [19767742](#)
67. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhorji G, et al. A pathology atlas of the human cancer transcriptome. *Science*. 2017;357(6352):eaan2507. <https://doi.org/10.1126/science.aan2507> PMID: [28818916](#)
68. Carvajal LA, Hamard P-J, Tonnessen C, Manfredi JJ. E2F7, a novel target, is up-regulated by p53 and mediates DNA damage-dependent transcriptional repression. *Genes Dev*. 2012;26(14):1533–45. <https://doi.org/10.1101/gad.184911.111> PMID: [22802528](#)
69. Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet*. 2010;6(10):e1001154. <https://doi.org/10.1371/journal.pgen.1001154> PMID: [20976243](#)
70. Morrill SA, Amon A. Why haploinsufficiency persists. *Proc Natl Acad Sci U S A*. 2019;116(24):11866–71. <https://doi.org/10.1073/pnas.1900437116> PMID: [31142641](#)
71. Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci U S A*. 2012;109(37):14746–53.
72. Tan MH, Li J, Xu HE, Melcher K, Yong EL. Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol Sin*. 2015;36(1):3–23.
73. Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*. 2011;474(7351):390–4. <https://doi.org/10.1038/nature10006> PMID: [21572438](#)
74. Berger R, Febbo PG, Majumder PK, Zhao JJ, Mukherjee S, Signoretti S, et al. Androgen-induced differentiation and tumorigenicity of human prostate epithelial cells. *Cancer Res*. 2004;64(24):8867–75. <https://doi.org/10.1158/0008-5472.CAN-04-2938> PMID: [15604246](#)
75. Lawson DA, Zong Y, Memarzadeh S, Xin L, Huang J, Witte ON. Basal epithelial stem cells are efficient targets for prostate cancer initiation. *Proc Natl Acad Sci U S A*. 2010;107(6):2610–5. <https://doi.org/10.1073/pnas.0913873107> PMID: [20133806](#)
76. Xin L. Cells of origin for cancer: an updated view from prostate cancer. *Oncogene*. 2013;32(32):3655–63. <https://doi.org/10.1038/onc.2012.541> PMID: [23178496](#)
77. Lee SH, Shen MM. Cell types of origin for prostate cancer. *Curr Opin Cell Biol*. 2015;37:35–41. <https://doi.org/10.1016/j.ceb.2015.10.002> PMID: [26506127](#)
78. Korsten H, Ziel-van der Made A, Ma X, van der Kwast T, Trapman J. Accumulating progenitor cells in the luminal epithelial cell layer are candidate tumor initiating cells in a Pten knockout mouse prostate cancer model. *PLoS One*. 2009;4(5):e5662. <https://doi.org/10.1371/journal.pone.0005662> PMID: [19461893](#)
79. Wang X, Kruihof-de Julio M, Economides KD, Walker D, Yu H, Halili MV, et al. A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature*. 2009;461(7263):495–500. <https://doi.org/10.1038/nature08361> PMID: [19741607](#)
80. Okada H, Tsubura A, Okamura A, Senzaki H, Naka Y, Komatz Y, et al. Keratin profiles in normal/hyperplastic prostates and prostate carcinoma. *Virchows Arch A Pathol Anat Histopathol*. 1992;421(2):157–61. <https://doi.org/10.1007/BF01607049> PMID: [1381129](#)
81. Lesch BJ, Page DC. Poised chromatin in the mammalian germ line. *Development*. 2014;141(19):3619–26. <https://doi.org/10.1242/dev.113027> PMID: [25249456](#)
82. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518(7539):331–6. <https://doi.org/10.1038/nature14222> PMID: [25693564](#)
83. Hansen T, Fong S, Capra JA, Hodges E. Human gene regulatory evolution is driven by the divergence of regulatory element function in both cis and trans. *bioRxiv*. 2023.
84. Chatrath A, Ratan A, Dutta A. Germline variants that affect tumor progression. *Trends Genet*. 2021;37(5):433–43. <https://doi.org/10.1016/j.tig.2020.10.005> PMID: [33203571](#)
85. Yang J, Adli M. Mapping and making sense of noncoding mutations in the genome. *Cancer Res*. 2019;79(17):4309–14. <https://doi.org/10.1158/0008-5472.CAN-19-0905> PMID: [31387922](#)
86. Khandrika L, Kumar B, Koul S, Maroni P, Koul HK. Oxidative stress in prostate cancer. *Cancer Lett*. 2009;282(2):125–36. <https://doi.org/10.1016/j.canlet.2008.12.011> PMID: [19185987](#)

87. Mani RS, Amin MA, Li X, Kalyana-Sundaram S, Veeneman BA, Wang L, et al. Inflammation-induced oxidative stress mediates gene fusion formation in prostate cancer. *Cell Rep*. 2016;17(10):2620–31. <https://doi.org/10.1016/j.celrep.2016.11.019> PMID: [27926866](#)
88. Chen Y, Paramo MI, Zhang Y, Yao L, Shah SR, Jin Y, et al. Finding needles in the haystack: strategies for uncovering noncoding regulatory variants. *Annu Rev Genet*. 2023;57:201–22. <https://doi.org/10.1146/annurev-genet-030723-120717> PMID: [37562413](#)
89. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature*. 2020;578(7793):82–93. <https://doi.org/10.1038/s41586-020-1969-6> PMID: [32025007](#)
90. Teng M, Zhou S, Cai C, Lupien M, He HH. Pioneer of prostate cancer: past, present and the future of FOXA1. *Protein Cell*. 2021;12(1):29–38. <https://doi.org/10.1007/s13238-020-00786-8> PMID: [32946061](#)
91. Adams EJ, Karthaus WR, Hoover E, Liu D, Gruet A, Zhang Z, et al. FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes. *Nature*. 2019;571(7765):408–12. <https://doi.org/10.1038/s41586-019-1318-9> PMID: [31243370](#)
92. Annala M, Taavitsainen S, Vandekerckhove G, Bacon JVW, Beja K, Chi KN, et al. Frequent mutation of the FOXA1 untranslated region in prostate cancer. *Commun Biol*. 2018;1:122. <https://doi.org/10.1038/s42003-018-0128-1> PMID: [30272002](#)
93. Del Giudice M, Foster JG, Peirone S, Rissone A, Caizzi L, Gaudino F, et al. FOXA1 regulates alternative splicing in prostate cancer. *Cell Rep*. 2022;40(13):111404. <https://doi.org/10.1016/j.celrep.2022.111404> PMID: [36170835](#)
94. McMullin RP, Dobi A, Mutton LN, Orosz A, Maheshwari S, Shashikant CS, et al. A FOXA1-binding enhancer regulates Hoxb13 expression in the prostate gland. *Proc Natl Acad Sci U S A*. 2010;107(1):98–103. <https://doi.org/10.1073/pnas.0902001107> PMID: [20018680](#)
95. Sahu B, Laakso M, Ovaska K, Mirtti T, Lundin J, Rannikko A, et al. Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *EMBO J*. 2011;30(19):3962–76. <https://doi.org/10.1038/emboj.2011.328> PMID: [21915096](#)
96. Sahu B, Laakso M, Pihlajamaa P, Ovaska K, Sinielnikov I, Hautaniemi S, et al. FoxA1 specifies unique androgen and glucocorticoid receptor binding events in prostate cancer cells. *Cancer Res*. 2013;73(5):1570–80. <https://doi.org/10.1158/0008-5472.CAN-12-2350> PMID: [23269278](#)
97. Shah N, Brown M. The sly oncogene: FOXA1 mutations in prostate cancer. *Cancer Cell*. 2019;36(2):119–21. <https://doi.org/10.1016/j.ccell.2019.07.005> PMID: [31408618](#)
98. Zhou S, Hawley JR, Soares F, Grillo G, Teng M, Madani Tonekaboni SA, et al. Noncoding mutations target cis-regulatory elements of the FOXA1 plexus in prostate cancer. *Nat Commun*. 2020;11(1):441. <https://doi.org/10.1038/s41467-020-14318-9> PMID: [31974375](#)
99. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*. 2017;14(10):959–62. <https://doi.org/10.1038/nmeth.4396> PMID: [28846090](#)
100. Wolock SL, Lopez R, Klein AM. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst*. 2019;8(4):281–291.e9. <https://doi.org/10.1016/j.cels.2018.11.005> PMID: [30954476](#)
101. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137. <https://doi.org/10.1186/gb-2008-9-9-r137> PMID: [18798982](#)
102. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. *Nat Methods*. 2021;18(11):1333–41. <https://doi.org/10.1038/s41592-021-01282-5> PMID: [34725479](#)
103. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048> PMID: [34062119](#)
104. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. 2019;566(7745):496–502. <https://doi.org/10.1038/s41586-019-0969-x> PMID: [30787437](#)
105. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of Anthropometric Traits (GIANT) Consortium, DIABetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44(4):369–75, S1–3. <https://doi.org/10.1038/ng.2213> PMID: [22426310](#)
106. Wang C, Li J. A deep learning framework identifies pathogenic noncoding somatic mutations from personal prostate cancer genomes. *Cancer Res*. 2020;80(21):4644–54. <https://doi.org/10.1158/0008-5472.CAN-20-1791> PMID: [32907840](#)
107. He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. *IEEE*; 2016:770–8.
108. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89. <https://doi.org/10.1016/j.molcel.2010.05.004> PMID: [20513432](#)
109. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291–5. <https://doi.org/10.1038/ng.3211> PMID: [25642630](#)
110. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228–35. <https://doi.org/10.1038/ng.3404> PMID: [26414678](#)
111. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90–7. <https://doi.org/10.1093/nar/gkw377> PMID: [27141961](#)
112. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42(Web Server issue):W187–91. <https://doi.org/10.1093/nar/gku365> PMID: [24799436](#)