Combined CRISPRi and proteomics screening reveal a cohesin-CTCF-bound allele contributing to increased expression of *RUVBL1* and prostate cancer progression

Yijun Tian,[1,6] Dandan Dong,[2,6] Zixian Wang,[2] Lang Wu,[3] Jong Y Park,[4] Gong-Hong Wei,[2,5,*] Liang Wang[1,*] PRACTICAL/ELLIPSE consortium[7]

1. Department of Tumor Biology, H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Drive, Tampa, FL 33612, USA.
2. MOE Key Laboratory of Metabolism and Molecular Medicine and Department of Biochemistry and Molecular Biology of School of Basic Medical Sciences & Fudan University Shanghai Cancer Center, Shanghai Medical College of Fudan University, Shanghai, China.
3. Population Sciences in the Pacific Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, HI 96813, USA.
4. Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA.
5. Disease Networks Research Unit, Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland.
6. These authors contributed equally.
7. Check supplementary information for the full author list

***** Correspondence: gonghong_wei@fudan.edu.cn, Liang.Wang@moffitt.org

Summary

Genome-wide association studies along with expression quantitative trait loci (eQTL) mapping have identified hundreds of single nucleotide polymorphisms (SNPs) and their target genes in prostate cancer (PCa), yet functional characterization of these risk loci remains challenging. To screen for potential regulatory SNPs, we designed a CRISPRi library containing 9133 guide RNAs (gRNAs) to target 2,166 candidate SNP sites implicated in PCa and identified 117 SNPs that could regulate 90 genes for PCa cell growth advantage. Among these, rs60464856 was covered by multiple gRNAs significantly depleted in the screening (FDR<0.05). Pooled SNP association analysis in the PRACTICAL and FinnGen cohorts showed significantly higher PCa risk for the rs60464856 G allele (pvalue=1.2E-16 and 3.2E-7). Subsequent eQTL analysis revealed that the G allele is associated with increased *RUVBL1* expression in multiple datasets. Further CRISPRi and xCas9 base editing proved the rs60464856 G allele leading to an elevated *RUVBL1* expression. Furthermore, SILAC-based proteomic analysis demonstrated allelic binding of cohesin subunits at the rs60464856 region, where HiC dataset showed consistent chromatin interactions in prostate cell lines. *RUVBL1* depletion inhibited PCa cell proliferation and tumor growth in xenograft mouse model. Gene set enrichment analysis suggested an association of *RUVBL1* expression with cell-cycle-related pathways. An increased expression of *RUVBL1* and activations of cell-cycle pathways were correlated with poor PCa survival in TCGA datasets. Together, our CRISPRi screening prioritized about one hundred regulatory SNPs essential for prostate cell proliferation. In combination with proteomics and functional studies, we characterized the mechanistic role of rs60464856 and *RUVBL1* in PCa progression.

Introduction

Among all cancer types, prostate cancer (PCa) accounted for 26% of 970,250 new cancer cases and caused 11% of 319,420 cancer-related deaths in US males in 2021. [1] As a cancer type with strong genetic predispositions, PCa has been extensively investigated in genome-wide association studies (GWAS) [2] to determine susceptible variants associated with increased disease risk and aggressiveness. [2,3] It has been reported that the joint contribution of GWAS identified loci to PCa risk is nearly 20%.[4] Although GWAS have been highly productive, only a few risk loci have been functionally characterized. [5-8] Thus far, as these risk SNPs are found in non-coding parts of the genome, it is believed that many of them (or their closely linked SNPs) would alter regulatory element activities and quantitatively change gene expression rather than directly mutating protein sequences. [9-12] With tremendous large-scale GWAS findings, especially highly reproducible ones, it is believed that some non-coding variants play a subtle but profound role in PCa initiation, progression, and metastasis by modulating the expression of certain susceptible genes.

To dissect functional variants and their target genes in PCa, bioinformatic and benchtop approaches are applied to prioritize and validate the causal variants. [13-15] Curated databases such as ChIP-atlas, [16] ENCODE, [17] JASPAR, [18] and GTEx[19] provided abundant resources to estimate the genetic contribution of GWAS variants to PCa and other cancer susceptibilities. However, large-scale genetic assays are often needed to interrogate endogenous variant loci and directly characterize consequent phenotype changes to determine the biological effect. [20-22] One such assay is lentiviral-based Cas9 screening, which has emerged as a powerful tool for evaluating the biological significance of certain genes on a large scale. [23,24] Compared to canonical wildtype Cas9 based screening, dead Cas9 (dCas9) forms steric hindrance according to the gRNA sequence and induces transcription repression if fused to repressor peptides KRAB (Krüppel-associated box). The dCas9-KRAB has been proven to specifically decrease target gene expression without strand cleavage and is used for screening regulatory elements in mammalian cells. [25,26]

Most SNPs are believed to function as regulatory switches in the human genome. With the eliminated nuclease activity, [26] we hypothesized that dCas9-based CRISPRi could be used to interfere with regulatory sequences at SNP loci and faithfully mimick the transcription alteration caused by single base differences. To test this hypothesis, we first established multiple stable dCas9-KRAB prostate cell lines and designed an unbiased, highly reproducible gRNA library that targeted candidate SNPs at PCa risk loci. We then performed negative selection for

potential SNPs conferring growth advantage. We finally provided a detailed analysis of an SNP-gene pair for its functional role using prostate cell lines and mouse model and performed a successful proteomics identification of transcriptional regulators in transforming the gene regulatory effects. Our results support the CRISPRi-based approach at risk loci for regulatory SNP screening.

Method

eQTL-based SNP selection at prostate cancer risk loci. The rationale for the regulatory screening was explained in **Supplementary Figure 1A** and **1B**. In brief, to select candidate SNPs, we first retrieved cis-eQTL data from our benign prostate tissues[27] and identified all SNPs with gene-wise FDR≤1E-3. We then applied ENCODE annotations (including histone modification, common transcription factor binding, and DNase hypersensitivity in prostate cell lines) to filter for candidate functional SNPs.

gRNA selection and library pool amplification for candidate SNPs. To design gRNAs for the candidate SNPs, we retrieved DNA sequences surrounding each SNP (± 23bp). We used the CRISPOR program (http://crispor.tefor.net/) for in-silico selection. We assembled the gRNA oligo pool into the lentiGuide-Puro (Addgene #52963) backbone and transformed the pooled oligos into highly sensitive Endura ElectroCompetent Cells (Lucigen) to generate plasmid libraries.

dCas9/KRAB stable cell lines and gRNA library processing. To establish stable cell lines, we packaged lenti-dCas9-KRAB-blast (Addgene #89567) into lentiviral particles with HEK293FT cells and used ten-fold concentrated virus particles to transduce cell lines. After blasticidin selection, the stable dCas9 expressions were verified with western blots. Meanwhile, the gRNA virus library (packaged from lentiGuide-Puro) was titered in each dCas9 stable cell line. To achieve low MOI infection, we optimized the cell number and virus amount over 72-hour puromycin selection, such that the non-infected group would be eliminated by 95% to 99%, while the library group retained 30% to 40% variability. The cell viability was measured with the CellTiter-Glo® Luminescent kit (Promega G7570). After confirming low MOI transducing, we removed puromycin from the medium and continued the cell culture for 21 days. We isolated genomic DNA at baseline D1 (day 1) and endpoint D21 (day 21).

gRNA readout sequencing. We used the Illumina HiSeq platform to sequence the gRNA readout amplicons. We aimed for at least a 500-fold library size depth for each replicate to ensure quantification accuracy.

Data QC and analysis. To quantify the representation of each gRNA, we used a python script "count_spacer.py" developed by Feng Zhang lab to scan the FASTQ file for perfectly matched hits and generate raw read counts for each experiment. We then used principal components analysis (PCA) to evaluate the similarity of the plasmid library, baseline, and screening endpoint gRNA representation. To determine SNP alleles conferring growth advantage in these cell lines, we used RIGER (RNAi gene enrichment ranking) extension (https://software.broadinstitute.org/GENE-E/extensions.html) to calculate a rank list for SNPs or alleles with the most depleted gRNA representation. [28] RIGER program ranks gRNAs according to their depletion effects and then identifies the SNP targeted by the shRNAs.

Plasmid construction and siRNA design. To enable the fluorescence-based cell sorting, copGFP ORF was assembled into base editor plasmid[29] xCas9(3.7)-ABE(7.10) (Addgene #108382) with NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs). A 223-bp flanking sequencing surrounding rs60464856 was amplified from RWPE-1 cells (rs60464856 heterozygote) and further subcloned into the pGL3-basic vector between NheI and XhoI sites using NEBuilder® HiFi DNA Assembly Master Mix. Small hairpin RNA (shRNA) targeting *RUVBL1*, primers, and gRNA gblock sequences were listed in **Supplementary Table S1**.

CRISPR base editing. To change the rs60464856 A allele to G allele, we created a GFP labeled xCas9(3.7)-ABE(7.10) plasmid based on the backbone from David Liu's lab. [29] gRNA template was synthesized in a gblock fragment with a hU6 promoter and a scaffold (https://benchling.com/protocols/10T3UWFo/detailed-gblocks-based-crispr-protocol). Since rs60464856 is located in the base editing window, the adenine base editor can be directed to the SNP site and catalyze A into the G allele. We co-transfected 2.5 ug xCas9(3.7)-ABE7.10 plasmid and 1.2 ug gRNA gblock into cells 80% confluent in each well of the 6-well plate. After 48 hours, GFP-positive cells were sorted by flow cytometry and collected for allele dosage quantification. The GFP-positive cells were seeded into single clones once editing efficiencies were above 5%. After expanding for ten days, we used amplification-refractory mutation system (ARMS) PCR to genotype direct lysate from every clone. We also used Sanger sequencing to verify the germline change of each single clone.

Real-time PCR and Western blot analysis. Total RNA was extracted from cells using the Direct-zol RNA Miniprep Kit (Zymo Research). One microgram of total RNA was reverse transcribed by iScript cDNA Synthesis Kit (BioRad). Quantification reactions were performed with PowerUp SYBR Green Master Mix (Thermo Fisher Scientific) on the CFX96 Touch real-time PCR system (BioRad). The primers were listed in **Supplementary Table S1**. Total protein

was extracted and electrophoresed as described previously, [30] with minor modification using Tricine Sample Buffer and Mini-PROTEAN® Precast Gels (BioRad). SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) produced luminescent signals on the LICOR imaging system. Captured images were aligned in Photoshop and assembled in Illustrator.

Luciferase reporter assay. The cells were seeded into a 24-well plate. After 12 to 16 hours, 500 ng of pGL3 reporter plasmids were transfected to each well using Lipofectamine 3000. The media were replaced after transfection for 24 hours. After 48 hours of transfection, the cells were lysed for the luciferase assay according to Dual-Luciferase® Reporter Assay (E1960, Promega) protocol. The luminescence signals were measured with the GlowMax plate reader. After normalized to Renilla luciferase readout, relative firefly luciferase activities driven by corresponding promoters were represented by luminescence unit fold changes.

Allele-specific proteomics screening with stable isotope labeling by amino acid in cell culture (SILAC). The BPH1, DU145, and PC3 cells were grown in SILAC RPMI 1640 medium (ThermoFisher 88365) for five passages before harvesting for nuclear protein extraction (ActiveMotif 40010). After confirming that heavy amino acid labeling efficiency reached 99.9%, the nuclear extracts were applied to the desalting spin column (ThermoFisher 89882) to remove excessive ions. The DNA baits harboring rs60464856 A and G alleles were produced according to a previous publication. [31] For each binding reaction, 2ug of purified DNA baits were conjugated to 25 uL Streptavidin Dynabeads (ThermoFisher 65001). The clean conjugated beads were incubated with 12.5 uL precleared nuclear protein at 4 degrees overnight. The incubated beads were washed five times and combined for two parallel quantitative MS runs to get the allelic protein binding ratio. Qualified proteins with allelic binding were defined as **a)** concordant allele ratio changes: $\log2(A/G)\times\log2(G/A)<0$; **b)** drastic allele ratio differences: $|\log2(A/G)-\log2(G/A)|\geq2$. For proteins with allelic hits, we further narrowed down to those with known DNA binding functions according to UniProt databases (https://www.uniprot.org/).

Chromatin immunoprecipitation (ChIP) qPCR assays. Since the rs60464856 locus was demonstrated to reside in an insulation region in an analysis of previously published Hi-C data, we adapted our ChIP qPCR protocols for chromosome conformation capturing. [32] Compared with conventional ChIP assay protocol, our protocol applied dual crosslinking to maximally preserve chromatin contacts. [33] Each ChIPed DNA sample was tested in four qPCR reactions, including **a)** rs60464856 locus enrichment primer pair; **b)** rs60464856 upstream locus

enrichment primer pair; **c)** rs60464856 A allele-specific primer pair; **d)** rs60464856 G allele-specific primer pair.

In vivo xenograft mouse model. Animal experiments were performed according to the protocol approved by the Institutional Animal Care and Use Committee of Fudan University. Nude mice (6 weeks males) were purchased from GemPharmatech Company (Jiangsu, China) and maintained in a pathogen-free environment. PC3 was grown in RPMI1640 containing 10% FBS, 100 U/mL penicillin, and 100 μg/mL streptomycin in a humidified CO_2 incubator. Before injection into the mice, the cells were harvested by trypsinization and washed two times with PBS. And then PC3 cells were resuspended in 100 μL serum-free medium, mixed with 50% Matrigel (BD Biosciences) and injected (5×10^6/site) subcutaneously into hind flank of each mouse. Tumor volume was measured using digital caliper once a week and calculated using the formula: $V = (L \times W^2)/2$ (L, length; W, width; all parameters in millimeters). After 4 weeks, the mice were sacrificed, and tumors were taken for weight measurement.

Results

Selected candidate SNPs, gRNAs and transfected libraries. We first screened ENCODE database and identified 2664 SNPs (TARGET) with strong epigenomic signals (**Supplementary Figure 1C**). We then applied a Bayesian framework using summary statistics to calculate a posterior inclusion probability (PIP) to predict SNP functionality. We included 5 SNPs with the highest PIP score for each eQTL-associated gene as another PIP SNP category (N = 194). Finally, we included 641 control SNPs (CTL) that showed strong eQTL signals but without epigenomic features. After removing duplicated SNPs, we selected 3408 SNPs for gRNA design. We scanned these SNPs with the CRISPOR program and eventually designed 9133 gRNAs, including 100 control sequences that did not target any genome loci (NCG) (**Supplementary Figure 1D**). The exact sequence design can be found in **Supplementary Table S2**.

To confirm the stable dCas9 expression, we used western blots to confirm the stable expression of dCas9 after one month of transduction (**Supplementary Figure 2A**). We ensured low MOI infection by measuring gRNA lentiviral library function titer in 72 hours (**Supplementary Figure 2B**) and visualized a 261bp gRNA amplicon for each sample on agarose gel for quantification by high-throughput sequencing (**Supplementary Figure 2C**). The sequencing summary for each sample is shown in **Supplementary Figure 2D**. We also visualized the normalized count in each replicate endpoint and observed high correlations in all three cell lines (**Supplementary Figure 2E-2G**).

CRISPRi screening identified top candidates of regulatory SNPs. We first performed PCA analysis and found tight clustering between baseline and plasmid libraries (**Supplementary Figure 3A**), suggesting faithful representations of original gRNA libraries in transfected cell lines. This analysis also found highly diverse but cell line-dependent distribution in the end libraries, indicating gRNA profile changes by the selection process. We then normalized the read count using the total perfect count in each sample and calculated fold change by dividing the read count in the end library by the read count in the baseline library. The subsequent RIGER analysis showed 779 gRNAs targeting 117 SNPs with permutation test FDR <0.1 in both replicates (**Figure 1B**). Further analysis did not find significant correlations between fold change and gRNA specificity score (**Supplementary Figure 3B-3D**), suggesting minimal off-target effects of these selected gRNAs. When comparing end and baseline libraries, we found significant gRNA depletion in BPH1, DU145, and PC3 screening experiments (**Figure 1C-1E**). When comparing gRNA fold change between different categories of candidate SNPs, we found significantly higher growth depletion only in BPH1 cells with gRNA targeting the PIP and TARGET SNPs (**Supplementary Figure 3E-3G**). We also found that the SNPs selected as screening candidates had a significantly higher proportion of residing in human gene transcription start sites (**Figure 1F**) and tended to be allelically bound to transcriptional factor binding according to ANANASTRA annotation (**Figure 1G**). [34] We plotted the gRNA representation before and after the growth selection and highlighted representative SNPs in each cell line (**Figure 1H-1J**). The raw and normalized gRNA count, and the eQTL mapped with the candidate SNP are listed in **Supplementary Table S2**. The RIGER analysis output is listed in **Supplementary Table S3**.

rs60464856 showed a regulatory role in *RUVBL1* expression underpinning prostate cancer susceptibility. Among the 117 SNPs showing significant growth inhibition over a 3-week screening, the SNP rs60464856 was consistently selected in all tested prostate cell lines. The SNP sequence was targeted by 10 gRNAs for A and 6 gRNAs for G allele. We observed significant A allele depletion (fold change ≤ 0.75) in 10 gRNAs in BPH1, 5 gRNAs in DU145 and 10 gRNAs in PC3 cells (**Figure 2A**). The rs60464856 is located in a previously identified risk locus, and the G allele was associated with a 10% increased prostate cancer risk in 107,247 cases and 127,006 controls (pvalue = 1.2E-16) (**Figure 2B**). [35] Consistently, a phenome-wide association analysis (PheWAS) in the FinnGen cohort (n = 342 499) with 2202 endpoints revealed the strongest association of rs60464856 with malignant neoplasm of prostate (11,590 cases and 110,189 controls; pvalue = 3.2E-7) (**Supplementary Figure 4A**). [36] We further showed that among the *RUVBL1* eQTL loci, rs60464856 resided in a linkage disequilibrium

block with the second-best significance (**Figure 2C**). Additionally, the rs60464856 G allele was significantly associated with elevated *RUVBL1* gene expression in GTEx, Mayo, [27,37,38] and TCGA prostate eQTL[39] datasets (**Figure 2D**). We also used three gRNAs targeting the rs60464856 locus to transiently transfect the dCas9 stable cells and observed a consistent knockdown effect on *RUVBL1* gene expressions in all three prostate cell lines (**Figure 2E**). To evaluate the nuances of endogenous allele transition contributing to downstream gene expression, we applied nickase Cas9 (xCas9) base editing technology, which featured high conversion efficiency and minimal indel rate, to faithfully mutate the rs60464856 locus in prostate cells (**Figure 2F**). Finally, we generated multiple isogenic subclones with accurate base conversion and confirmed significant increase of endogenous *RUVBL1* expression by the G allele in BPH1 and DU145 (**Figure 2G**).

rs60464856 bound to cohesion subunits allelically and mediated chromatin interactions. To evaluate potential allelic protein binding at rs60464856, we searched for ChIP-seq data collections, such as cistrome and ChIP-ATLAS, for potential transcription factor bindings. However, we did not find highly convincing evidence showing allelic TF binding on this locus. Since multiple datasets report histone modification signals on the rs60464856 locus, we then performed ChIP-qPCR to evaluate the locus-specific enrichment for active histone modification markers (H3K4me1, H3K4me3, and H3K27ac) in the base-edited subclones. We found distinct histone modification status in PC3 clones, especially for H3K4me1 on the rs60464856 locus (**Figure 3A**). Consistently, we found the enrichment of H3K4me1, H3K4me3, and H3K27ac modifications on rs60464856 in multiple cell lines (**Figure 3B**), suggesting a robust regulatory potential of this locus in the prostate. With a recent HiC dataset, [40] we also visualized the long-range interactions surrounding rs60464856 locus. By summing the normalized count from the interaction hot spot on both sides of the rs60464856 locus, we calculated the left-to-right ratio (L/R) in each cell line. This analysis showed that PC3 cells had only 60% interaction compared to DU145 cells (**Figure 3C**). In contrast, the interactions L/R ratios were roughly equal in VCaP and LNCaP cells (**Figure 3D**). To confirm the transcription activity of the rs60464856 locus, we tested its flanking sequence (chr3: 128,123,257-128,123,479, negative strand) using a reporter assay and found higher promoter activity of the G allele than the A allele in these cell lines (**Figure 3E**). We then applied SILAC-based proteomics to detect possible transcription factors or DNA-binding proteins. This assay took advantage of isotype-labeled nuclear extract in DNA pull-down reactions, thus converting the protein binding difference into the ratio of different molecular weights (**Figure 3F**). This proteomics analysis identified several cohesion subunits bonded to the rs60464856 A than the G allele with the BPH1 nuclear extract (**Figure 3G**). To

further elucidate whether the cohesion could bind to endogenous rs60464856 locus allelically, we used ChIP assays specific to the rs60464856 locus from BPH1 and PC3 base edited populations. In BPH1 cells, we found significant locus enrichment (**Figure 3H**) for SMC1A, SMC3, and CTCF, and the A allele preference for SMC3 and CTCF (**Figure 3I**). In PC3 cells, we only found minor locus enrichment for SMC3 and CTCF (**Figure 3J**), and no allele preference was observed for all tested antibodies (**Figure 3K**). We also explored the unique allelic binding role of SMC3, using existing ChIP-seq datasets[41] (GSE49402 and GSE36578) that quantified SMC1A and SMC3 binding in the human genome (**Supplementary Figure 5A**). We performed STRMEME motif scan with the private peak region of SMC1A and SMC3, and only found that only SMC3 private peaks included an outstanding significant motif in both cell lines (**Supplementary Figure 5B).** Through STRMEME -Tomtom comparison, we also found that this motif was highly similar to the CTCF binding site (MA0139.1) (**Supplementary Figure 5C**). More importantly, we found that the rs60464856 A allele is located in the CTCF zinc finger seven interaction domain and is consistently preferred by multiple versions of the CTCF motif. [42] This result implicated a potential mechanism about how the rs60464856 protective allele mediated the cohesion-CTCF complex formation and supported our observation in the HiC datasets (**Figure 3C-3D**).

***RUVBL1* knockdown inhibited prostate cell proliferation by downregulating cell cycle related pathways.** To evaluate the oncogenic role of the *RUVBL1* gene, we examined the perturbation effect of *RUVBL1* by CRISPR or RNAi screening in DepMap portal. We found that *RUVBL1* was a common essential gene in most human cell lines, with an even better dependency score than the median of all essential genes for the prostate cell lines used in our screening, including BPH1, DU145 and PC3 cells (**Figure 4A**). To further characterize the function of *RUVBL1*, we generated stable cell lines infected with small hairpin RNA (shRNA) lentiviral particles and verified that the RUVBL1 had been successfully knockdown at protein levels (**Figure 4B**). We further monitored the growth of the stable cell lines with daily Incucyte scans and found that *RUVBL1* knockdown by both shRNAs significantly suppressed proliferation in BPH1 (**Figure 4C**), DU145 (**Figure 4D**), and PC3 (**Figure 4E**). We also observed drastic reductions in colony formation for the *RUVBL1* knockdown group in BPH1, DU145, and PC3 cells (**Figure 4F**). To characterize the transcriptome alteration caused by *RUVBL1* downregulation, we quantified RNA profiling of BPH1 and PC3 cells and calculated the GSVA score for HALLMARK geneset collection. Interestingly, multiple cell-cycle-related pathways, including the MYC target, E2F target, and G2M checkpoint genes, showed significant enrichment with *RUVBL1* expression (**Figure 4G-4H**). We further visualized the gene

expression changes in these significantly enriched pathways and found consistent trends with the *RUVBL1* expression levels (**Figure 4I**). To characterize the tumorigenic effect of the *RUVBL1* in vivo, we performed xenograft mouse experiments with the PC3 stable cell lines. We found that the *RUVBL1* knockdown significantly inhibited tumor growth (**Figure 4J**) and reduced endpoint tumor weight in the mice model (**Figure 4K**).

***RUVBL1* increased aggressiveness and predicted a worse prognosis in prostate cancer.** To characterize the malignant potential associated with the *RUVBL1* gene in clinical samples, we obtained three indices of genome instabilities, including the altered fraction of the genome, mutation count, and aneuploidy score for 488 TCGA prostate cancer. We found *RUVBL1* gene expression was positively associated with these indices (**Figure 5A-5C**). Additionally, we demonstrated that prostate cancer with higher *RUVBL1* expression tended to have a more advanced tumor size stage and Gleason score (**Figure 5D**). We also calculated the GSVA score for HALLMARK geneset collection with the TCGA prostate cancer RNA profiling and found that the cell cycle related pathways showed up consistently in the best 8 significantly enriched gene sets (**Figure 5E**). We further demonstrated the enrichment score(**Figure 5F-5G**) and FDR q values for these significantly enriched pathways and found consistent positive enrichment with the *RUVBL1* levels. We performed a Kaplan-Meier analysis to see whether RUVBL1 could be a prognostic marker and found significantly worse progression-free (**Figure 5H**) survivals in those with higher *RUVBL1* expression. To determine whether the *RUVBL1* enriched gene could be prognostic, we also used k-means clustering methods with the leading-edge gene to separate the prostate cancer patients into groups with different risks (**Figure 5I**). The Kaplan-Meier analysis showed that patients with increased risk tended to have significantly worse progress-free survival (**Figure 5J-5L**). Additionally, we found *RUVBL1* gene expression was consistently upregulated in PCa primary and metastasis tumor tissue (**Supplementary Figure 4B-4D**). We also demonstrated a positive association between *RUVBL1* gene expression and elevated prediagnostic PSA level, higher Gleason score, and worse biochemical recurrence-free survival in existing cohorts (**Supplementary Figure 4E-4G**).

Discussion

Over the past decade, GWAS and eQTL analysis have been highly productive in finding PCa risk loci and susceptible genes. [43-45] Despite contributing to a better understanding of the biological significance of risk predisposition, these analyses did not directly demonstrate the regulatory role of individual locus and the functional consequence of each causal gene. To better delineate the functionality of these genetic findings, a large-scale functional evaluation of

target risk loci is highly warranted. Due to the non-coding nature of most risk loci, these variants are believed to play a regulatory role. Therefore, we applied the dCas9-based CRISPRi assay to target SNP sequences at risk loci, aiming to mimic the regulatory alteration caused by single base substitutions. This genome-wide screening at PCa risk loci discovered 117 SNPs showing a regulatory role in cell proliferation. Interestingly, these proliferation-related SNPs are enriched in gene transcription start sites, suggesting the majority of the phenotypical changes are related to transcription alterations caused by dCas9 interferences.

This study characterized the regulatory role of a risk SNP rs60464856. We observed consistent growth inhibition by multiple gRNAs targeting the locus. More importantly, with multiple gRNA targeting both alleles at the same genomic locations, we observed interesting phenotypical changes to seed region mismatches in real cells. As demonstrated by previous studies, [46,47] any mismatches at 7-bp seed region of the gRNA could cause a rapid rejection of these targets by the dCas9 protein. These facts may explain the A allele-specific depletion effect of gRNA on positions -5 to -7, as shown in our data (**see Figure 2A**). When the mismatch is located outside the seed region, the gRNA with SNP at -8 position showed depletion effects for both alleles. To further validate the regulatory role of rs60464856, we created multiple subclones carrying converted G alleles. As expected, the G allele-carrying subclones showed an elevated *RUVBL1* expression in BPH1 and DU145 cells. However, this elevated expression was not observed in PC3 subclones, possibly caused by the distinct status of histone modification and chromatin interaction in PC3 cells.

This study also showed allele-specific binding of SMC3 at rs60464856 locus, which is different from *SMC1A*. As the major subunits of human cohesin, both *SMC1A* and *SMC3* mediate multiple biological processes, including DNA looping, chromosome condensation, and chromosome segregation, by forming heterodimers. [48,49] Intriguingly, there have been several unique observations about the distinct phenotypical changes brought by SMC1A or SMC3 knockdown. Magdalena et al. identified that *SMC3* knockdown made *SMC1A* unstable and led to less cytoplasmic accumulation, while *SMC1A* knockdown did not influence *SMC3* stability and cytoplasmic accumulation. [50] A recent study[51] reported that *SMC1A* and *SMC3* ATPase active sites had differential effects on cohesin ATPase function, and SMC3 has a unique function in DNA tethering. Our database mining showed significant signal overlap between *SMC1A* and *SMC3* binding sites in the genome, but only private peaks from *SMC3* enriched CTCF motifs in the two cell lines. As an insulator that can block enhancers to regulate target genes, CTCF was first discovered as a transcriptional repressor and believed to execute a hub

role in controlling gene expression.[42,52] In our findings, CTCF may also play a crucial role in governing *RUVBL1* gene regulation, potentially through insulating the rs60464856 loci allelically.

RUVBL1, known as RuvB Like AAA ATPase 1 or TIP49, is a protein-coding gene possessing ATP-dependent DNA helicase activity and has been reported to regulate a wide range of cellular processes,[53] including chromatin decondensation,[54,55] misfolded protein aggregation,[56] and transcription regulation.[57] In addition to the previously reported mTORC1 pathway,[58] our enrichment analysis demonstrated that the *RUVBL1* expression was consistently correlated with cell cycle regulation and c-MYC signaling activities in both cell lines and tumor tissues. This result demonstrates a potential use of *RUVBL1* selective inhibitors in treating prostate cancer.[59,60] The result also suggests applying the rs60464856 genotype to stratify a target population for future clinical trials.

One potential issue with this study is inconsistent results in some tests among different cell lines. While these inconsistencies may attribute to genetic heterogeneity, we also want to highlight that some candidates found exclusively in DU145 cells are reported to be functional in prostate cells, for instance, the established functional variants residing in the binding sites of the transcription factors, TMPRSS2-ERG and HNF1B.[61] Additionally, to increase our gRNA library coverage, we may use novel CRISPR systems with expanded PAM site compatibility. With stringent library preparation and screening processes, the biological implications for the functional variants discovered in one cell line hits are still worth investigating, which may uncover unique SNP and gene functions specific to the cell line of interest. In summary, we applied CRISPRi screening technology to screen for survival-essential SNPs at the genome scale. We identified over a hundred functional SNPs that regulate cell proliferation. We further characterized the rs60464856 risk variant for its regulatory role in the prostate context and target gene *RUVBL1* for its functional role in prostate cancer cell proliferation and disease progression. This result will enrich our knowledge of PCa predisposition and provide insight into the cancer risk classification and potential therapeutic target for personalized treatment.

Data and code availability

The NGS data have been archived at the Gene Expression Omnibus under the reference Series GSE224654, which includes CRISPRi screening gRNA readout (GSE224653) and RNA-seq expression for *RUVBL1* knockdown experiment (GSE224646). The gRNA sequence design, raw and normalized count, and the eQTL mapped with the candidate SNP are listed in **Supplementary Table S2**. The RIGER analysis output is listed in **Supplementary Table S3**. The publicly available datasets used are listed in **Supplementary Table S4**. The SILAC proteomics sequencing result is listed in **Supplementary Table S5**. The detailed information about the rs60464856 base editing can be accessed with the box link (https://moffitt.box.com/s/353qfjsxycaztutjufjq607hok4n3kf9).

Acknowledgments

We thank Flow Cytometry, Molecular Genomic, Proteomic and Microscopy Core Facility at the Moffitt Cancer Center, an NCI-designated Comprehensive Cancer Center (P30CA076292). The authors also expect acknowledgment to the participants and investigators of the FinnGen study (www.finngen.fi). This work was supported by the National Institutes of Health (R01CA250018 and R01CA212097 to L.Wang., R01CA263494 to L. Wu) and National Natural Science Foundation of China (82073082) as well as the Jane and Aatos Erkko Foundation, the Finnish Cancer Foundation, and the Sigrid Juseliuksen Saatio to G-H.W. We thank Dr. Stephen N. Thibodeau at Mayo Clinic for providing the eQTL candidate list for design this screening. We also thank the PRACTICAL/ELLIPSE consortium for providing GWAS resources supporting this project. The full funding information for the consortium author is listed in Supplementary information.

Author contribution

Y.T. and L.W. contributed to study design and performed CRISPRi and proteomics screening, statistical analysis and functional validation. D.D., Z.W. and G.-H.W. performed functional analysis and mouse work. L.Wu. and J.P. contributed to collecting GWAS summary statistics, G-H.W. and L.W. supervised the study and contributed to study design and data interpretation. Y.T., D.D., G.-H.W., and L.W. co-wrote the manuscript. All authors read, commented and approved the final version.

Declaration of interests

The authors declare no competing interests.

Figure Legends

Figure 1. A. Study design of the current project, including eQTL target selection, CRISPRi screening, and functional validation. **B.** Overall SNP candidates in BPH1 (red), DU145 (green) and PC3 (blue) cells over 3-week screening. **C-D.** Comparison of the gRNA fold change between the selected and unselected SNPs in BPH1 (**C**), DU145 (**D**) and PC3 (**E**) cells. **F.** Enrichment of the selected SNPs in transcription start sites (TSS) compared to the unselected. **G.** Enrichment of the selected SNPs in allele-specific binding (ASB) annotation according to the ANANASTRA database compared to the unselected. **H-J.** Demonstration of gRNA representation for the selected SNPs in BPH1 (**H**), DU145 (**I**) and PC3 (**J**) cells. gRNA targeting the main validation SNP rs60464856 is highlighted in red.

Figure 2. A. Screening fold change of rs60464856 gRNA in BPH1, DU145 and PC3 cells. The Grey bar indicates rs60464856 locus in the germline, and colored strips above or below the bar indicate gRNA binding Crick or Watson strands of the germline DNA. Numbers in the center show the average fold change for two screening replicates. **B.** Locus plots for rs60464856 locus GWAS significance in the PRACTICAL cohort. **C.** Locus plots for *RUVBL1* eQTL significance in the MAYO prostate dataset. **D.** *RUVBL1* gene expressions in dCas9-KRAB stable prostate cells transfected with non-targeting control guide (NCG) and rs60464856 targeting gRNAs. The individual point indicated different gRNA clones. **E.** eQTL associations between *RUVBL1* gene expression and rs60464856 genotypes in GTEx, MAYO and TCGA prostate cohorts. **F.** Schematic figure about xCas9-based A to G base editing in prostate cell lines. **G.** *RUVBL1* gene expressions in rs60464856 base edited clones in BPH1, DU145 and PC3 cells.

Figure 3. A. H3K4me1, H3K4me3 and H3K27ac modifications on rs60464856 locus in BPH1, DU145 and PC3 clones. **B.** ChIP-seq signals of H3K4me1, H3K4me3 and H3K27ac modifications on rs60464856 locus in prostate cells. **C-D.** HiC interaction heatmap in PC3 (C top), DU145 (C bottom), VCaP (D top) and LNCaP (D bottom) cells. L and R indicate interaction ICE normalized read count on the left and right sides of rs60464856 bin. **E.** Luciferase reporter assay comparing rs60464856 allele transcription activities in BPH1, DU145 and PC3 cells.**F.** SILAC-based DNA pull-down proteomics pipeline in identifying allelic protein binding on rs60464856 locus. **G.** Allelic proteomics with BPH1 SILAC extracts. **H-K**. ChIP-qPCR measuring rs60464856 locus enrichment and allele-specific binding to SMC1A, SMC3, CTCF and CTCFL proteins in BPH1 (**H-I**) and PC3 (**J-K**) base edited populations.

Figure 4. A. Dependency score of *RUVBL1* gene for CRISPR and RNAi screening from CCLE human cell line collections. A score of 0 indicates that a gene is not essential, while a score of -1 is comparable to the median of all pan-essential genes.**B.** Measuring RUVBL1 protein expression after shRNA knockdown with lentiviral infection (low MOI). **C-E.** Cell proliferation curve comparing cell growth differences after *RUVBL1* knockdown in BPH1 (**C**), DU145 (**D**) and PC3 (**E**) cell lines. **F.** Colony formation assay comparing clonogenic potential alterations after *RUVBL1* knockdown in BPH1, DU145

and PC3 cell lines. **G.** GSVA on RNA-seq data in BPH1 and PC3 cells after *RUVBL1* knockdown on HALLMARK collection pathways. **H.** GSEA summary for top rank significant *RUVBL1*-driven gene sets in BPH1 and PC3 cells. **I.** Representative heatmap showing significantly enriched HALLMARK pathways, including MYC target, E2F target and G2M checkpoint. **J-K.** Tumor volume changes (**J**) and tumor weight (**K**) comparison in xenograft experiments with PC3 cell line.

Figure 5. A-C. Associations in TCGA prostate cancer cohort between *RUVBL1* gene expression and genome instability metrics, including the altered fraction of genome (**A**), the mutation count (**B**), and the aneuploidy score (**C**). **D.** *RUVBL1* gene expression in TCGA prostate tissue with different tumor size stage and Gleason score. **E.** GSVA on RNA-seq data in TCGA prostate cancer tissue by ranking *RUVBL1* expression on HALLMARK collection pathways. **F.** GSEA enrichment plots showing significantly enriched HALLMARK pathways in TCGA prostate cohort. G. Normalized enrichment score and FDR q-value for significantly enriched HALLMARK pathways in TCGA prostate cohort. H. Kaplan-Meier analysis on progress-free survival in TCGA prostate cancer patients stratified by *RUVBL1* gene expression. **I.** Heatmap with *RUVBL1* leading-edge genes with column side bar indicating k-means stratified groups with different risks related to significantly enriched HALLMARK pathways in TCGA prostate cohort, including MYC target V1, E2F targets and G2M checkpoint. **J-L.** Kaplan-Meier analysis on progress-free survival in TCGA prostate cancer patients stratified by GSEA edge gene enriched with *RUVBL1* in each pathway, including MYC target V1 (**J**), E2F targets (**K**), and G2M checkpoint (**L**).

Reference

1. Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2021). Cancer Statistics, 2021. CA Cancer J Clin *71*, 7-33. 10.3322/caac.21654.
2. Schaid, D.J., Sinnwell, J.P., Batzler, A., and McDonnell, S.K. (2022). Polygenic risk for prostate cancer: Decreasing relative risk with age but little impact on absolute risk. Am J Hum Genet *109*, 900-908. 10.1016/j.ajhg.2022.03.008.
3. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res *47*, D1005-D1012. 10.1093/nar/gky1120.
4. Stadler, Z.K., Thom, P., Robson, M.E., Weitzel, J.N., Kauff, N.D., Hurley, K.E., Devlin, V., Gold, B., Klein, R.J., and Offit, K. (2010). Genome-wide association studies of cancer. J Clin Oncol *28*, 4255-4267. 10.1200/JCO.2009.25.7816.
5. Farashi, S., Kryza, T., Clements, J., and Batra, J. (2019). Post-GWAS in prostate cancer: from genetic association to biological contribution. Nat Rev Cancer *19*, 46-59. 10.1038/s41568-018-0087-3.
6. Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association to Function. Am J Hum Genet *102*, 717-730. 10.1016/j.ajhg.2018.04.002.
7. Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am J Hum Genet *93*, 779-797. 10.1016/j.ajhg.2013.10.012.
8. Tian, P., Zhong, M., and Wei, G.H. (2021). Mechanistic insights into genetic susceptibility to prostate cancer. Cancer Lett *522*, 155-163. 10.1016/j.canlet.2021.09.025.
9. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43-49. 10.1038/nature09906.
10. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190-1195. 10.1126/science.1222794.
11. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., and Raychaudhuri, S. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet *45*, 124-130. 10.1038/ng.2504.
12. Shetty, A., Seo, J.H., Bell, C.A., O'Connor, E.P., Pomerantz, M.M., Freedman, M.L., and Gusev, A. (2021). Allele-specific epigenetic activity in prostate cancer and normal prostate tissue implicates prostate cancer risk mechanisms. Am J Hum Genet *108*, 2071-2085. 10.1016/j.ajhg.2021.09.008.
13. Zhu, Q., Ge, D., Heinzen, E.L., Dickson, S.P., Urban, T.J., Zhu, M., Maia, J.M., He, M., Zhao, Q., Shianna, K.V., and Goldstein, D.B. (2012). Prioritizing genetic variants for causality on the basis of preferential linkage disequilibrium. Am J Hum Genet *91*, 422-434. 10.1016/j.ajhg.2012.07.010.
14. Fritsche, L.G., Patil, S., Beesley, L.J., VandeHaar, P., Salvatore, M., Ma, Y., Peng, R.B., Taliun, D., Zhou, X., and Mukherjee, B. (2020). Cancer PRSweb: An Online Repository with Polygenic Risk Scores for Major Cancer Traits and Their Evaluation in Two Independent Biobanks. Am J Hum Genet *107*, 815-836. 10.1016/j.ajhg.2020.08.025.
15. Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am J Hum Genet *104*, 65-75. 10.1016/j.ajhg.2018.11.008.
16. Oki, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., Kawaji, H., Nakaki, R., Sese, J., and Meno, C. (2018). ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. EMBO Rep *19*. 10.15252/embr.201846255.
17. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res *46*, D794-D801. 10.1093/nar/gkx1081.
18. Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranasic, D., et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res *48*, D87-D92. 10.1093/nar/gkz1001.
19. Consortium, G.T. (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet *45*, 580-585. 10.1038/ng.2653.

20. Gurumurthy, C.B., Grati, M., Ohtsuka, M., Schilit, S.L., Quadros, R.M., and Liu, X.Z. (2016). CRISPR: a versatile tool for both forward and reverse genetics research. Hum Genet *135*, 971-976. 10.1007/s00439-016-1704-4.
21. Yang, L., Chan, A.K.N., Miyashita, K., Delaney, C.D., Wang, X., Li, H., Pokharel, S.P., Li, S., Li, M., Xu, X., et al. (2021). High-resolution characterization of gene function using single-cell CRISPR tiling screen. Nat Commun *12*, 4063. 10.1038/s41467-021-24324-0.
22. Yan, F., Li, J., Milosevic, J., Petroni, R., Liu, S., Shi, Z., Yuan, S., Reynaga, J.M., Qi, Y., Rico, J., et al. (2022). KAT6A and ENL Form an Epigenetic Transcriptional Control Module to Drive Critical Leukemogenic Gene-Expression Programs. Cancer Discov *12*, 792-811. 10.1158/2159-8290.CD-20-1459.
23. Gasperini, M., Findlay, G.M., McKenna, A., Milbank, J.H., Lee, C., Zhang, M.D., Cusanovich, D.A., and Shendure, J. (2017). CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. Am J Hum Genet *101*, 192-205. 10.1016/j.ajhg.2017.06.010.
24. Wojtal, D., Kemaladewi, D.U., Malam, Z., Abdullah, S., Wong, T.W., Hyatt, E., Baghestani, Z., Pereira, S., Stavropoulos, J., Mouly, V., et al. (2016). Spell Checking Nature: Versatility of CRISPR/Cas9 for Developing Treatments for Inherited Disorders. Am J Hum Genet *98*, 90-101. 10.1016/j.ajhg.2015.11.012.
25. Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647-661. 10.1016/j.cell.2014.09.029.
26. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442-451. 10.1016/j.cell.2013.06.044.
27. Thibodeau, S.N., French, A.J., McDonnell, S.K., Cheville, J., Middha, S., Tillmans, L., Riska, S., Baheti, S., Larson, M.C., Fogarty, Z., et al. (2015). Identification of candidate genes for prostate cancer-risk SNPs utilizing a normal prostate tissue eQTL data set. Nat Commun *6*, 8653. 10.1038/ncomms9653.
28. Luo, B., Cheung, H.W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X., Hinkle, G., Boehm, J.S., Beroukhim, R., Weir, B.A., et al. (2008). Highly parallel identification of essential genes in cancer cells. Proc Natl Acad Sci U S A *105*, 20380-20385. 10.1073/pnas.0810485105.
29. Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., and Liu, D.R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature *556*, 57-63. 10.1038/nature26155.
30. Tian, Y., Liu, Q., Yu, S., Chu, Q., Chen, Y., Wu, K., and Wang, L. (2020). NRF2-Driven KEAP1 Transcription in Human Lung Cancer. Mol Cancer Res *18*, 1465-1476. 10.1158/1541-7786.MCR-20-0108.
31. Butter, F., Davison, L., Viturawong, T., Scheibe, M., Vermeulen, M., Todd, J.A., and Mann, M. (2012). Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS Genet *8*, e1002982. 10.1371/journal.pgen.1002982.
32. Lafontaine, D.L., Yang, L., Dekker, J., and Gibcus, J.H. (2021). Hi-C 3.0: Improved Protocol for Genome-Wide Chromosome Conformation Capture. Curr Protoc *1*, e198. 10.1002/cpz1.198.
33. Akgol Oksuz, B., Yang, L., Abraham, S., Venev, S.V., Krietenstein, N., Parsi, K.M., Ozadam, H., Oomen, M.E., Nand, A., Mao, H., et al. (2021). Systematic evaluation of chromosome conformation capture assays. Nat Methods *18*, 1046-1055. 10.1038/s41592-021-01248-7.
34. Boytsov, A., Abramov, S., Aiusheeva, A.Z., Kasianova, A.M., Baulin, E., Kuznetsov, I.A., Aulchenko, Y.S., Kolmykov, S., Yevshin, I., Kolpakov, F., et al. (2022). ANANASTRA: annotation and enrichment analysis of allele-specific transcription factor binding at SNPs. Nucleic Acids Res *50*, W51-56. 10.1093/nar/gkac262.
35. Schumacher, F.R., Al Olama, A.A., Berndt, S.I., Benlloch, S., Ahmed, M., Saunders, E.J., Dadaev, T., Leongamornlert, D., Anokian, E., Cieza-Borrella, C., et al. (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet *50*, 928-936. 10.1038/s41588-018-0142-8.
36. Kurki, M.I., Karjalainen, J., Palta, P., Sipila, T.P., Kristiansson, K., Donner, K.M., Reeve, M.P., Laivuori, H., Aavikko, M., Kaunisto, M.A., et al. (2023). FinnGen provides genetic insights from a well-phenotyped isolated population. Nature *613*, 508-518. 10.1038/s41586-022-05473-8.

37. Tian, Y., Soupir, A., Liu, Q., Wu, L., Huang, C.C., Park, J.Y., and Wang, L. (2022). Novel role of prostate cancer risk variant rs7247241 on PPP1R14A isoform transition through allelic TF binding and CpG methylation. Hum Mol Genet *31*, 1610-1621. 10.1093/hmg/ddab347.
38. Larson, N.B., McDonnell, S., French, A.J., Fogarty, Z., Cheville, J., Middha, S., Riska, S., Baheti, S., Nair, A.A., Wang, L., et al. (2015). Comprehensively evaluating cis-regulatory variation in the human prostate transcriptome by using gene-level allele-specific expression. Am J Hum Genet *96*, 869-882. 10.1016/j.ajhg.2015.04.015.
39. Gong, J., Mei, S., Liu, C., Xiang, Y., Ye, Y., Zhang, Z., Feng, J., Liu, R., Diao, L., Guo, A.Y., et al. (2018). PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. Nucleic Acids Res *46*, D971-D976. 10.1093/nar/gkx861.
40. San Martin, R., Das, P., Dos Reis Marques, R., Xu, Y., Roberts, J.M., Sanders, J.T., Golloshi, R., and McCord, R.P. (2022). Chromosome compartmentalization alterations in prostate cancer cell lines model disease progression. J Cell Biol *221*. 10.1083/jcb.202104108.
41. Yan, J., Enge, M., Whitington, T., Dave, K., Liu, J., Sur, I., Schmierer, B., Jolma, A., Kivioja, T., Taipale, M., and Taipale, J. (2013). Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. Cell *154*, 801-813. 10.1016/j.cell.2013.07.034.
42. Yin, M., Wang, J., Wang, M., Li, X., Zhang, M., Wu, Q., and Wang, Y. (2017). Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. Cell Res *27*, 1365-1377. 10.1038/cr.2017.131.
43. Eeles, R.A., Durocher, F., Edwards, S., Teare, D., Badzioch, M., Hamoudi, R., Gill, S., Biggs, P., Dearnaley, D., Ardern-Jones, A., et al. (1998). Linkage analysis of chromosome 1q markers in 136 prostate cancer families. The Cancer Research Campaign/British Prostate Group U.K. Familial Prostate Cancer Study Collaborators. Am J Hum Genet *62*, 653-658. 10.1086/301745.
44. Xu, J. (2000). Combined analysis of hereditary prostate cancer linkage to 1q24-25: results from 772 hereditary prostate cancer families from the International Consortium for Prostate Cancer Genetics. Am J Hum Genet *66*, 945-957. 10.1086/302807.
45. Schaid, D.J., McDonnell, S.K., Blute, M.L., and Thibodeau, S.N. (1998). Evidence for autosomal dominant inheritance of prostate cancer. Am J Hum Genet *62*, 1425-1438. 10.1086/301862.
46. Boyle, E.A., Andreasson, J.O.L., Chircus, L.M., Sternberg, S.H., Wu, M.J., Guegler, C.K., Doudna, J.A., and Greenleaf, W.J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. Proc Natl Acad Sci U S A *114*, 5461-5466. 10.1073/pnas.1700557114.
47. Zheng, T., Hou, Y., Zhang, P., Zhang, Z., Xu, Y., Zhang, L., Niu, L., Yang, Y., Liang, D., Yi, F., et al. (2017). Profiling single-guide RNA specificity reveals a mismatch sensitive core sequence. Sci Rep *7*, 40638. 10.1038/srep40638.
48. Sun, M., Nishino, T., and Marko, J.F. (2013). The SMC1-SMC3 cohesin heterodimer structures DNA through supercoiling-dependent loop formation. Nucleic Acids Res *41*, 6149-6160. 10.1093/nar/gkt303.
49. Deardorff, M.A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J., Gil-Rodriguez, C., Arnedo, M., Loeys, B., Kline, A.D., et al. (2007). Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. Am J Hum Genet *80*, 485-494. 10.1086/511888.
50. Laugsch, M., Seebach, J., Schnittler, H., and Jessberger, R. (2013). Imbalance of SMC1 and SMC3 cohesins causes specific and distinct effects. PLoS One *8*, e65149. 10.1371/journal.pone.0065149.
51. Camdere, G., Guacci, V., Stricklin, J., and Koshland, D. (2015). The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. Elife *4*. 10.7554/eLife.11315.
52. Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V.V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol *16*, 2802-2813. 10.1128/MCB.16.6.2802.
53. Dauden, M.I., Lopez-Perrote, A., and Llorca, O. (2021). RUVBL1-RUVBL2 AAA-ATPase: a versatile scaffold for multiple complexes and functions. Curr Opin Struct Biol *67*, 78-85. 10.1016/j.sbi.2020.08.010.
54. Gentili, C., Castor, D., Kaden, S., Lauterbach, D., Gysi, M., Steigemann, P., Gerlich, D.W., Jiricny, J., and Ferrari, S. (2015). Chromosome Missegregation Associated with RUVBL1 Deficiency. PLoS One *10*, e0133576. 10.1371/journal.pone.0133576.

55. Magalska, A., Schellhaus, A.K., Moreno-Andres, D., Zanini, F., Schooley, A., Sachdev, R., Schwarz, H., Madlung, J., and Antonin, W. (2014). RuvB-like ATPases function in chromatin decondensation at the end of mitosis. Dev Cell *31*, 305-318. 10.1016/j.devcel.2014.09.001.
56. Zaarur, N., Xu, X., Lestienne, P., Meriin, A.B., McComb, M., Costello, C.E., Newnam, G.P., Ganti, R., Romanova, N.V., Shanmugasundaram, M., et al. (2015). RuvbL1 and RuvbL2 enhance aggresome formation and disaggregate amyloid fibrils. EMBO J *34*, 2363-2382. 10.15252/embj.201591245.
57. Wang, H., Li, B., Zuo, L., Wang, B., Yan, Y., Tian, K., Zhou, R., Wang, C., Chen, X., Jiang, Y., et al. (2022). The transcriptional coactivator RUVBL2 regulates Pol II clustering with diverse transcription factors. Nat Commun *13*, 5703. 10.1038/s41467-022-33433-3.
58. Shin, S.H., Lee, J.S., Zhang, J.M., Choi, S., Boskovic, Z.V., Zhao, R., Song, M., Wang, R., Tian, J., Lee, M.H., et al. (2020). Synthetic lethality by targeting the RUVBL1/2-TTT complex in mTORC1-hyperactive cancer cells. Sci Adv *6*, eaay9131. 10.1126/sciadv.aay9131.
59. Assimon, V.A., Tang, Y., Vargas, J.D., Lee, G.J., Wu, Z.Y., Lou, K., Yao, B., Menon, M.K., Pios, A., Perez, K.C., et al. (2019). CB-6644 Is a Selective Inhibitor of the RUVBL1/2 Complex with Anticancer Activity. ACS Chem Biol *14*, 236-244. 10.1021/acschembio.8b00904.
60. Zhang, G., Wang, F., Li, S., Cheng, K.W., Zhu, Y., Huo, R., Abdukirim, E., Kang, G., and Chou, T.F. (2022). Discovery of small-molecule inhibitors of RUVBL1/2 ATPase. Bioorg Med Chem *62*, 116726. 10.1016/j.bmc.2022.116726.
61. Giannareas, N., Zhang, Q., Yang, X., Na, R., Tian, Y., Yang, Y., Ruan, X., Huang, D., Yang, X., Wang, C., et al. (2022). Extensive germline-somatic interplay contributes to prostate cancer progression through HNF1B co-option of TMPRSS2-ERG. Nat Commun *13*, 7320. 10.1038/s41467-022-34994-z.

Figure 1

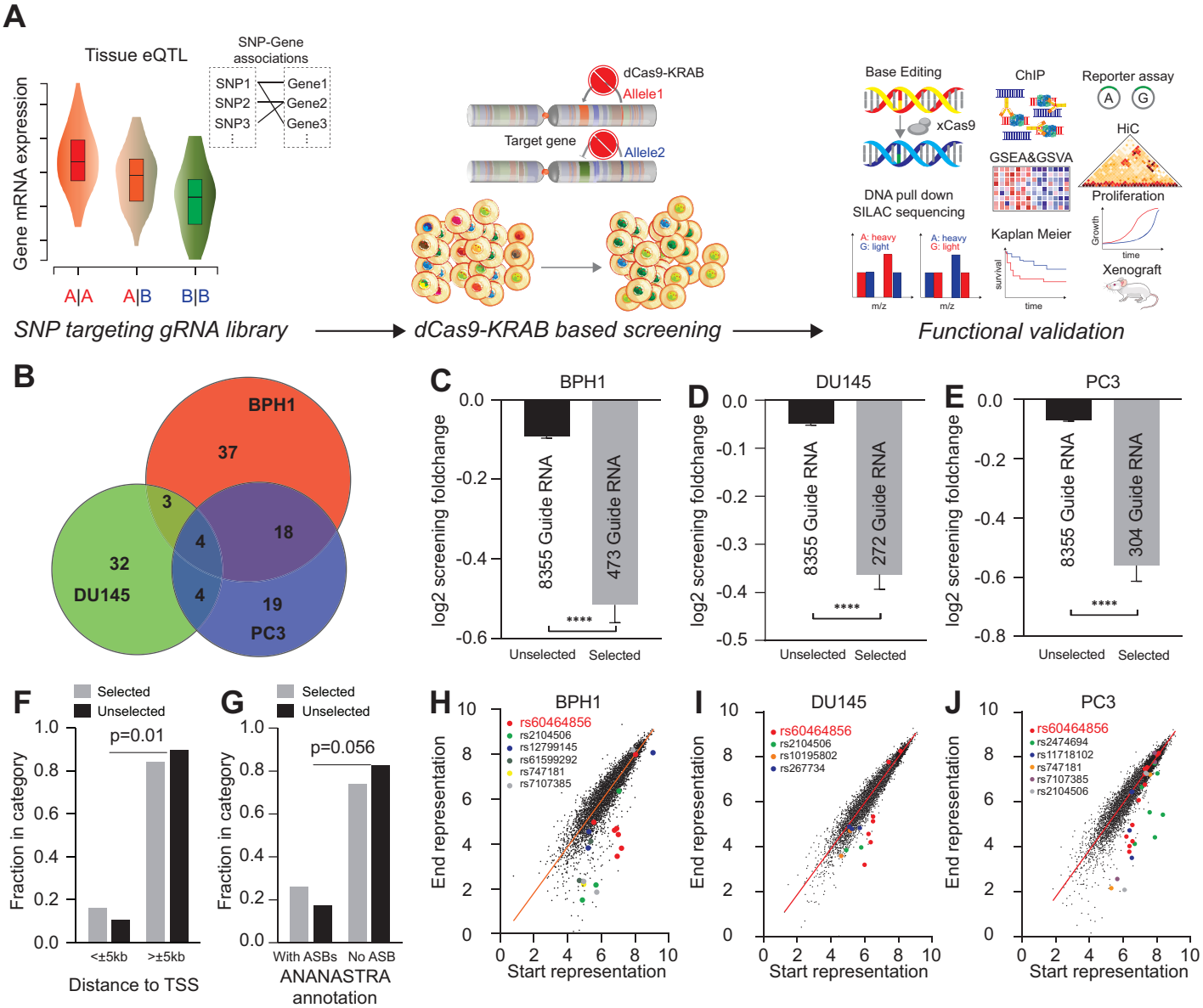

A
Tissue eQTL
SNP-Gene associations
SNP1
SNP2
SNP3
Gene1
Gene2
Gene3
Gene mRNA expression
A|A
A|B
B|B
dCas9-KRAB
Allele1
Target gene
Allele2
Base Editing
xCas9
ChIP
Reporter assay
HiC
GSEA&GSVA
Proliferation
DNA pull down
SILAC sequencing
Kaplan Meier
Xenograft
SNP targeting gRNA library
dCas9-KRAB based screening
Functional validation
B
BPH1
DU145
PC3
37
3
18
4
32
4
19
C
BPH1
log2 screening foldchange
8355 Guide RNA
473 Guide RNA
****
Unselected
Selected
D
DU145
8355 Guide RNA
272 Guide RNA
E
PC3
8355 Guide RNA
304 Guide RNA
F
Selected
Unselected
p=0.01
Fraction in category
<±5kb
>±5kb
Distance to TSS
G
p=0.056
With ASBs
No ASB
ANANASTRA annotation
H
BPH1
rs60464856
rs2104506
rs12799145
rs61599292
rs747181
rs7107385
End representation
Start representation
I
DU145
rs60464856
rs2104506
rs10195802
rs267734
J
PC3
rs60464856
rs2474694
rs11718102
rs747181
rs7107385
rs2104506

Figure 2

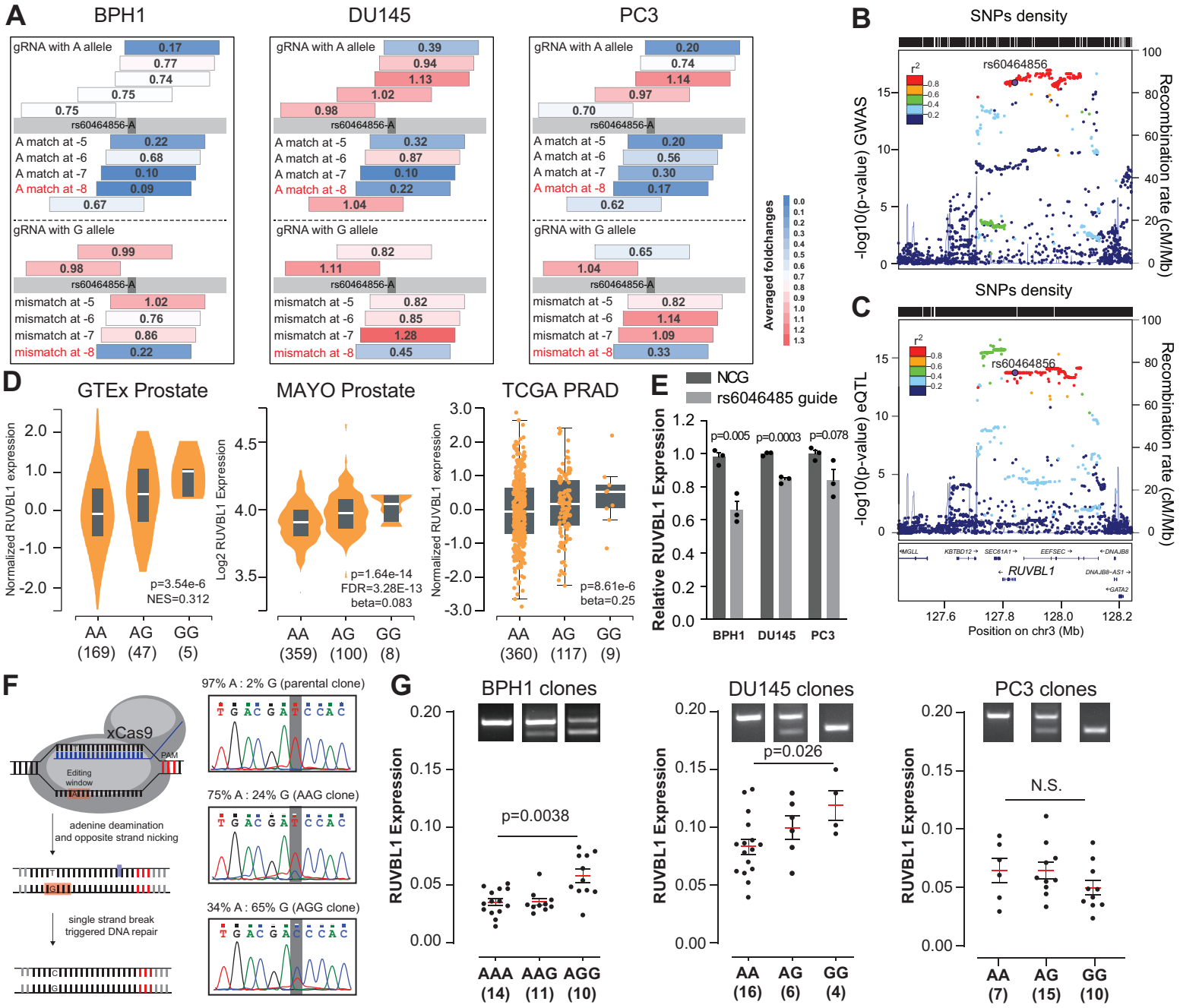

A
BPH1
DU145
PC3
gRNA with A allele
rs60464856-A
A match at -5
A match at -6
A match at -7
A match at -8
gRNA with G allele
mismatch at -5
mismatch at -6
mismatch at -7
mismatch at -8
Averaged foldchanges
B
SNPs density
rs60464856
-log10(p-value) GWAS
Recombination rate (cM/Mb)
C
SNPs density
-log10(p-value) eQTL
RUVBL1
Position on chr3 (Mb)
D
GTEx Prostate
MAYO Prostate
TCGA PRAD
Normalized RUVBL1 expression
Log2 RUVBL1 Expression
p=3.54e-6
NES=0.312
p=1.64e-14
FDR=3.28E-13
beta=0.083
p=8.61e-6
beta=0.25
AA (169)
AG (47)
GG (5)
AA (359)
AG (100)
GG (8)
AA (360)
AG (117)
GG (9)
E
NCG
rs6046485 guide
Relative RUVBL1 Expression
p=0.005
p=0.0003
p=0.078
BPH1
DU145
PC3
F
xCas9
PAM
Editing window
adenine deamination and opposite strand nicking
single strand break triggered DNA repair
97% A : 2% G (parental clone)
75% A : 24% G (AAG clone)
34% A : 65% G (AGG clone)
G
BPH1 clones
DU145 clones
PC3 clones
RUVBL1 Expression
p=0.0038
p=0.026
N.S.
AAA (14)
AAG (11)
AGG (10)
AA (16)
AG (6)
GG (4)
AA (7)
AG (15)
GG (10)

Figure 3

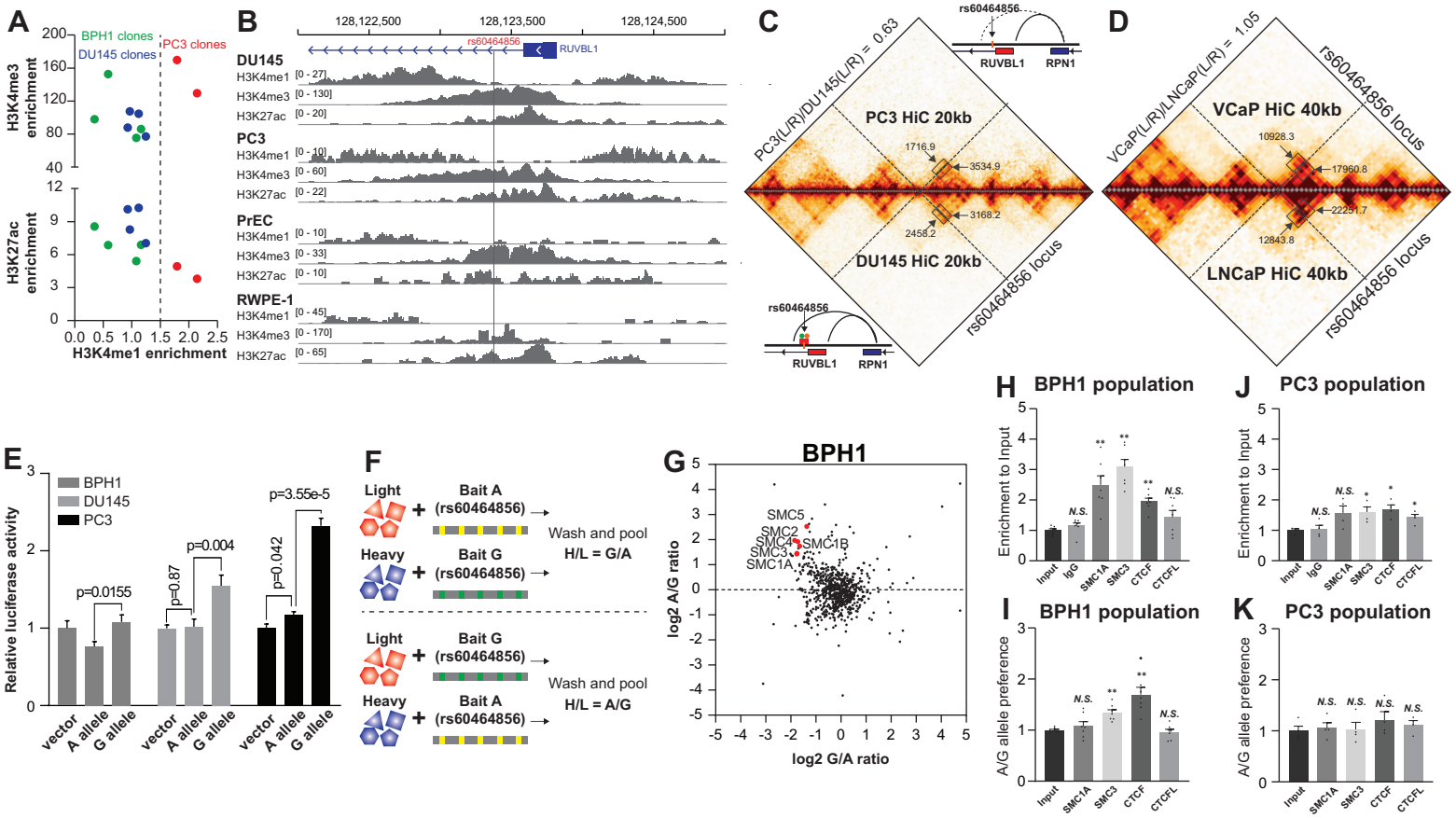

A
BPH1 clones
DU145 clones
PC3 clones
H3K4me3 enrichment
H3K27ac enrichment
H3K4me1 enrichment
B
128,122,500
128,123,500
128,124,500
rs60464856
RUVBL1
DU145
PC3
PrEC
RWPE-1
H3K4me1
H3K4me3
H3K27ac
C
PC3(L/R)/DU145(L/R) = 0.63
PC3 HiC 20kb
DU145 HiC 20kb
RPN1
D
VCaP(L/R)/LNCaP(L/R) = 1.05
VCaP HiC 40kb
LNCaP HiC 40kb
rs60464856 locus
E
BPH1
DU145
PC3
Relative luciferase activity
p=0.0155
p=0.87
p=0.004
p=0.042
p=3.55e-5
vector
A allele
G allele
F
Light
Heavy
Bait A (rs60464856)
Bait G (rs60464856)
Wash and pool
H/L = G/A
H/L = A/G
G
BPH1
SMC5
SMC2
SMC4
SMC3
SMC1B
SMC1A
log2 A/G ratio
log2 G/A ratio
H
BPH1 population
Enrichment to Input
Input
IgG
SMC1A
SMC3
CTCF
CTCFL
N.S.
I
BPH1 population
A/G allele preference
J
PC3 population
K
PC3 population

Figure 4

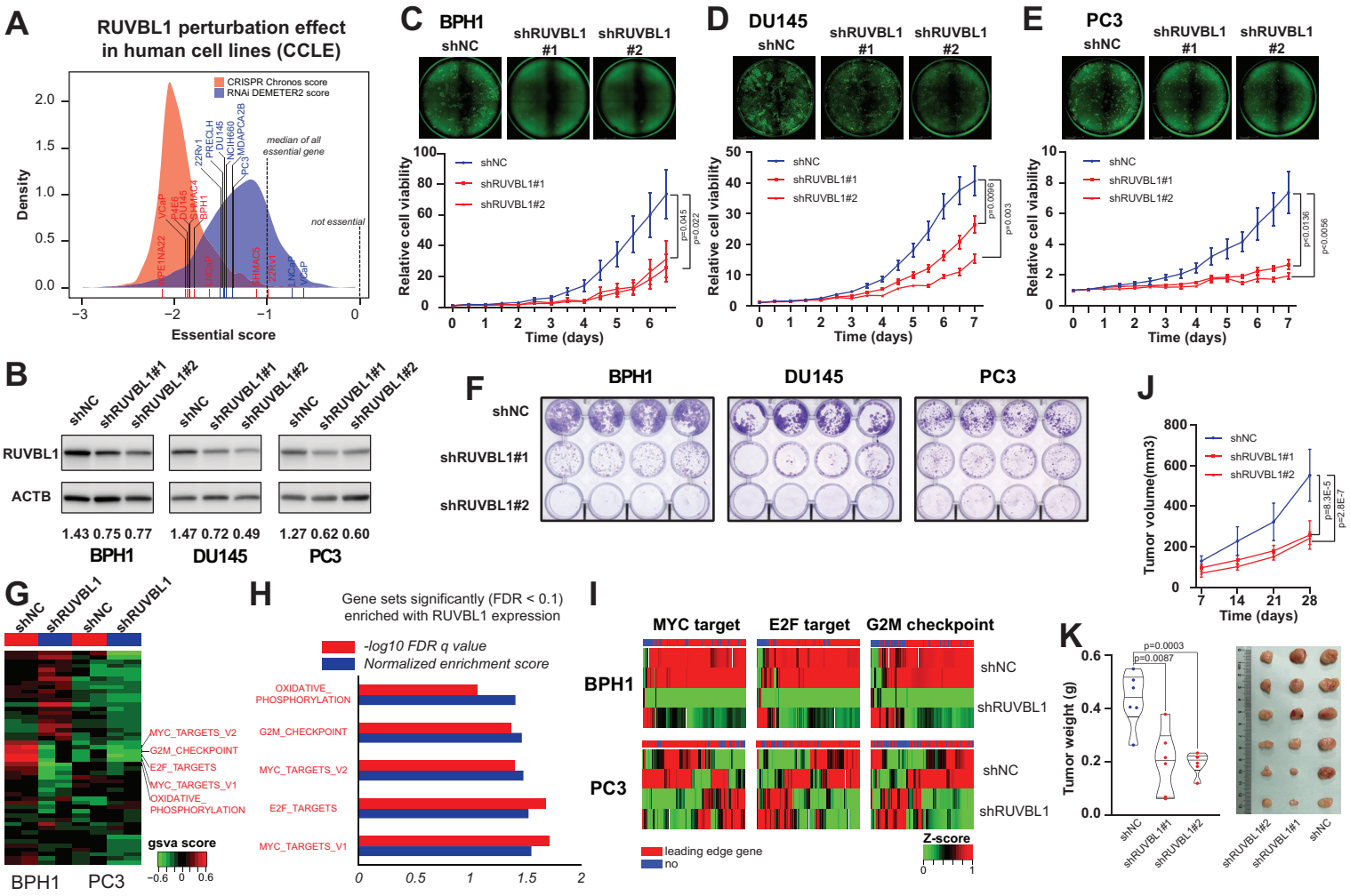
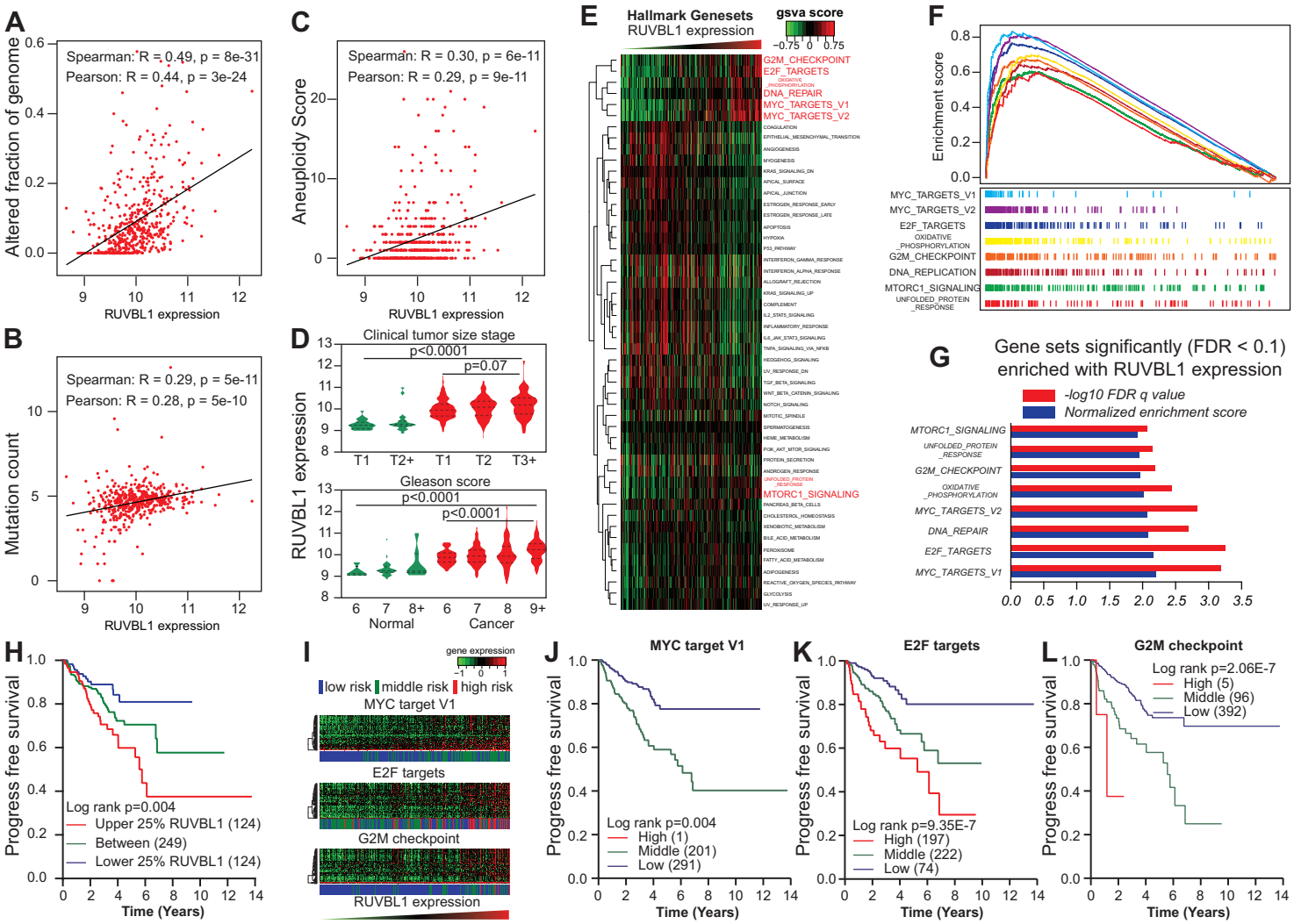
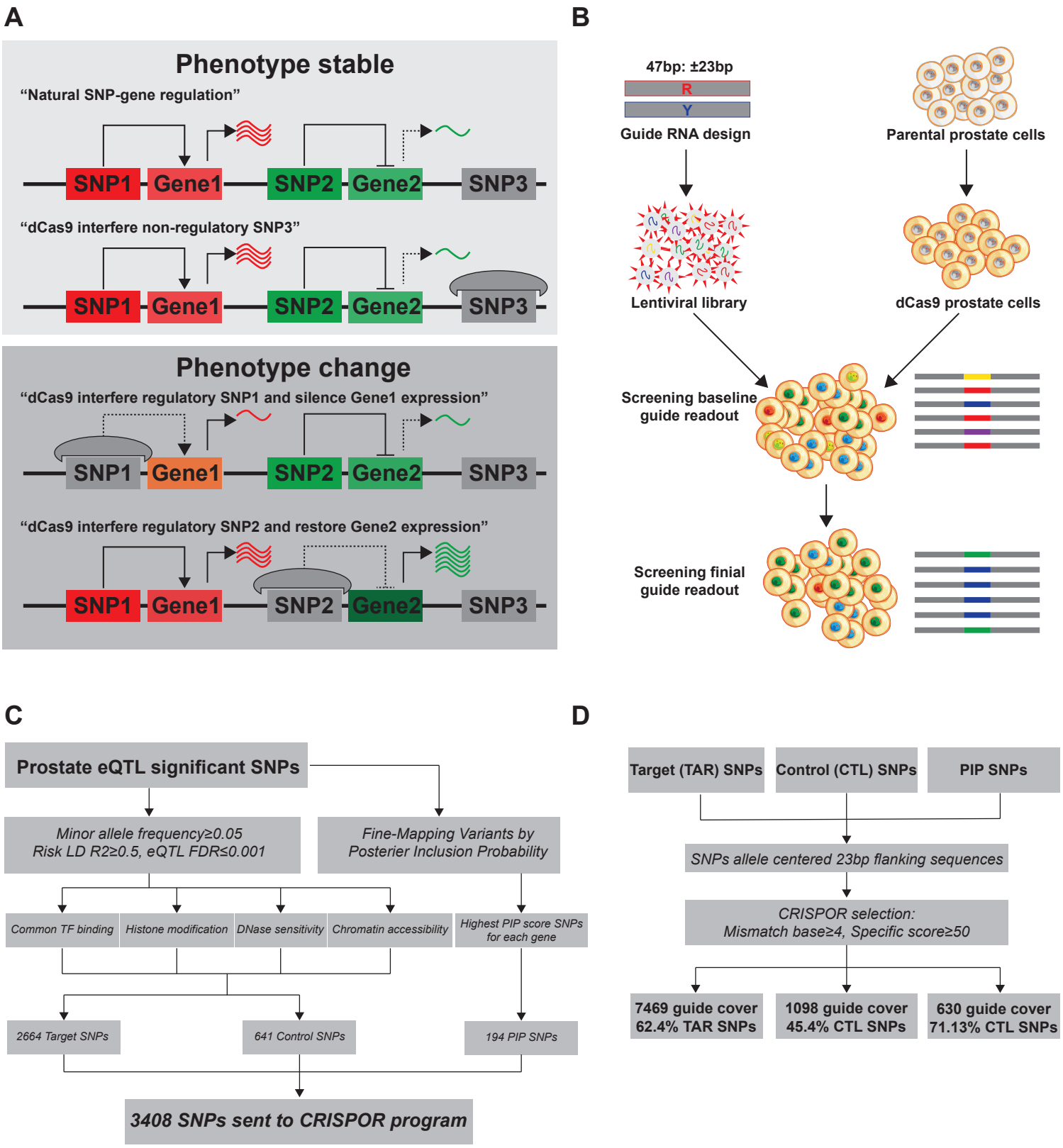

A
RUVBL1 perturbation effect in human cell lines (CCLE)
CRISPR Chronos score
RNAi DEMETER2 score
Density
Essential score
median of all essential gene
not essential
B
shNC
shRUVBL1#1
shRUVBL1#2
RUVBL1
ACTB
1.43 0.75 0.77
1.47 0.72 0.49
1.27 0.62 0.60
BPH1
DU145
PC3
C
BPH1
D
DU145
E
PC3
shNC
shRUVBL1 #1
shRUVBL1 #2
Relative cell viability
Time (days)
p=0.045
p=0.022
p=0.0096
p=0.003
p<0.0136
p<0.0056
F
BPH1
DU145
PC3
shNC
shRUVBL1#1
shRUVBL1#2
J
Tumor volume(mm3)
Time (days)
p=8.3E-5
p=2.8E-7
G
shNC
shRUVBL1
MYC_TARGETS_V2
G2M_CHECKPOINT
E2F_TARGETS
MYC_TARGETS_V1
OXIDATIVE_ PHOSPHORYLATION
gsva score
BPH1
PC3
H
Gene sets significantly (FDR < 0.1) enriched with RUVBL1 expression
-log10 FDR q value
Normalized enrichment score
OXIDATIVE_ PHOSPHORYLATION
G2M_CHECKPOINT
MYC_TARGETS_V2
E2F_TARGETS
MYC_TARGETS_V1
I
MYC target
E2F target
G2M checkpoint
BPH1
PC3
shNC
shRUVBL1
leading edge gene
no
Z-score
K
p=0.0003
p=0.0087
Tumor weight (g)
shNC
shRUVBL1#1
shRUVBL1#2

Figure 5

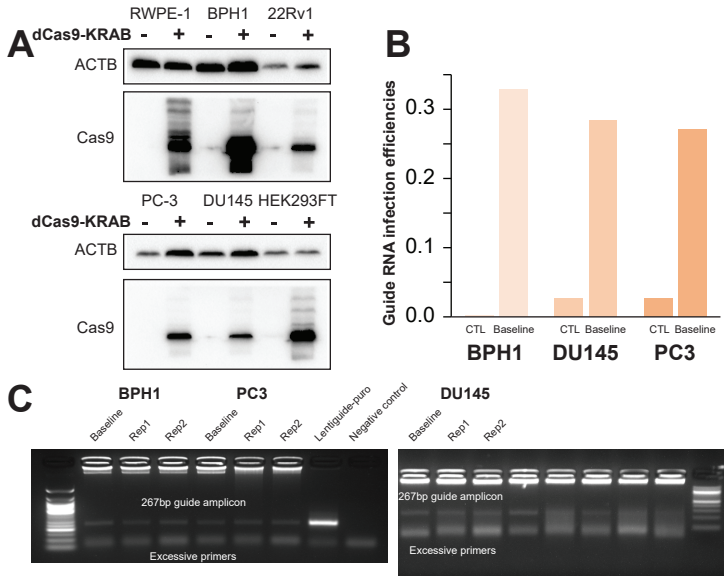

A
Spearman: R = 0.49, p = 8e-31
Pearson: R = 0.44, p = 3e-24
Altered fraction of genome
RUVBL1 expression
B
Spearman: R = 0.29, p = 5e-11
Pearson: R = 0.28, p = 5e-10
Mutation count
RUVBL1 expression
C
Spearman: R = 0.30, p = 6e-11
Pearson: R = 0.29, p = 9e-11
Aneuploidy Score
RUVBL1 expression
D
Clinical tumor size stage
p<0.0001
p=0.07
T1 T2+ T1 T2 T3+
Gleason score
p<0.0001
p<0.0001
6 7 8+ 6 7 8 9+
Normal
Cancer
RUVBL1 expression
E
Hallmark Genesets
RUVBL1 expression
gsva score
G2M_CHECKPOINT
E2F_TARGETS
DNA_REPAIR
MYC_TARGETS_V1
MYC_TARGETS_V2
MTORC1_SIGNALING
F
Enrichment score
MYC_TARGETS_V1
MYC_TARGETS_V2
E2F_TARGETS
OXIDATIVE _PHOSPHORYLATION
G2M_CHECKPOINT
DNA_REPLICATION
MTORC1_SIGNALING
UNFOLDED_PROTEIN _RESPONSE
G
Gene sets significantly (FDR < 0.1) enriched with RUVBL1 expression
-log10 FDR q value
Normalized enrichment score
MTORC1_SIGNALING
UNFOLDED_PROTEIN _RESPONSE
G2M_CHECKPOINT
OXIDATIVE _PHOSPHORYLATION
MYC_TARGETS_V2
DNA_REPAIR
E2F_TARGETS
MYC_TARGETS_V1
H
Progress free survival
Log rank p=0.004
Upper 25% RUVBL1 (124)
Between (249)
Lower 25% RUVBL1 (124)
Time (Years)
I
gene expression
low risk
middle risk
high risk
MYC target V1
E2F targets
G2M checkpoint
RUVBL1 expression
J
MYC target V1
Progress free survival
Log rank p=0.004
High (1)
Middle (201)
Low (291)
Time (Years)
K
E2F targets
Progress free survival
Log rank p=9.35E-7
High (197)
Middle (222)
Low (74)
Time (Years)
L
G2M checkpoint
Progress free survival
Log rank p=2.06E-7
High (5)
Middle (96)
Low (392)
Time (Years)

Supplementary Figure 1



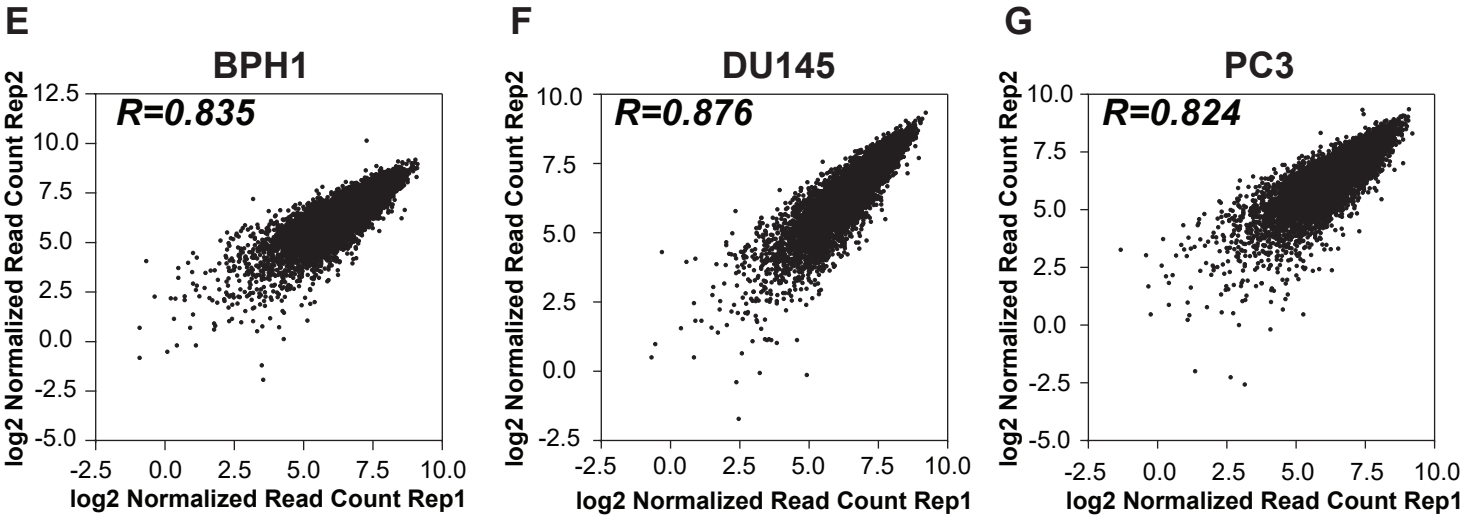
Supplementary Figure S1: Overall CRISPRi screening projection and design. A. Rationale for phenotype oriented CRISPRi-SNPs-seq screening. **B.** Procedure of CRISPRi screening in dCas9 stable prostate cell lines. **C.** Candidate eQTL risk SNPs selection before gRNA design. **D.** gRNA design with CRISPOR program

Supplementary Figure 2



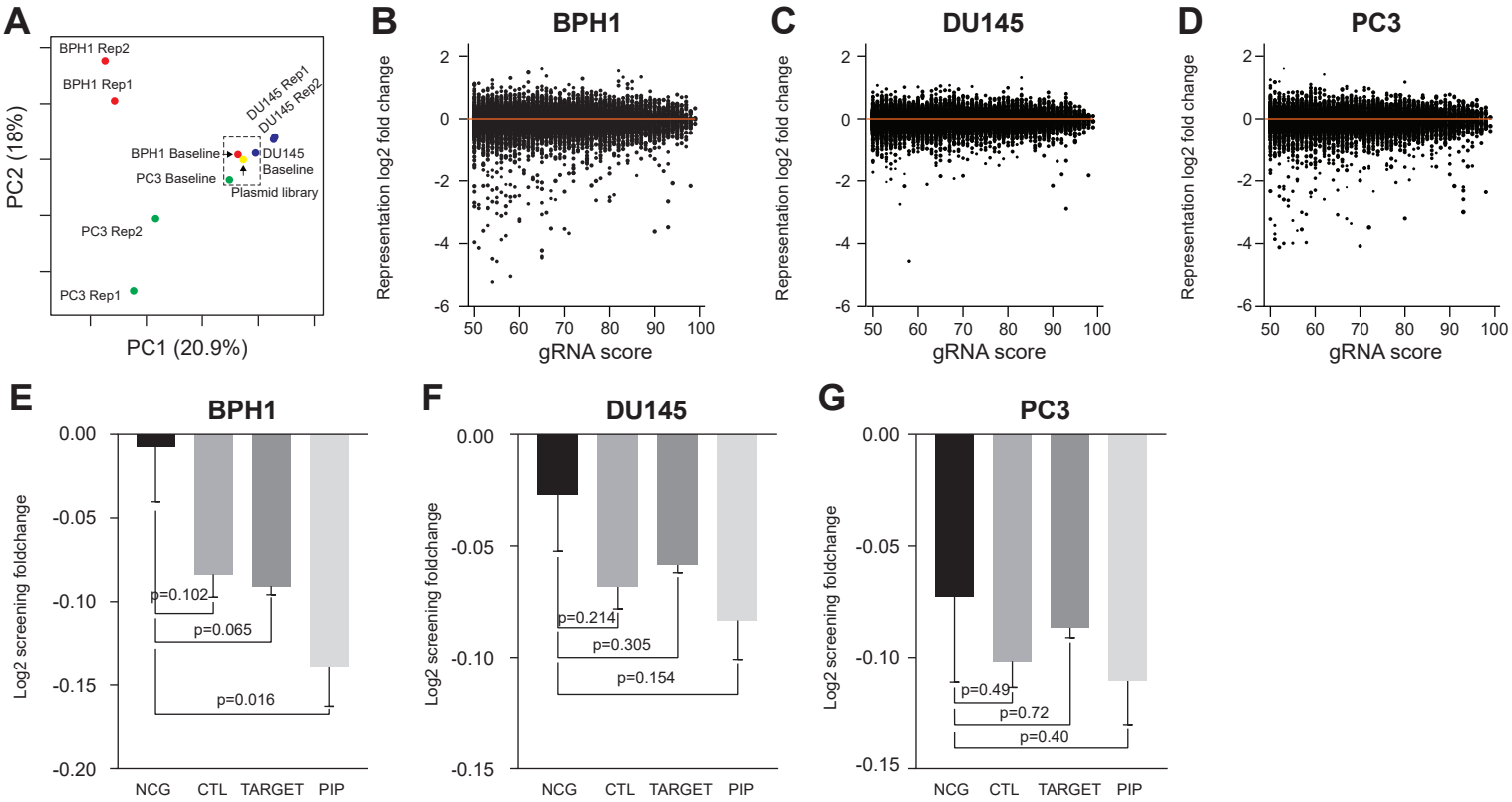
D

	Total read count	*Perfect guide count*	*Skew ratio*
Lentiguide puro	*24.9M*	*17.8M*	*5.32*
BPH1-Baseline	*24.5M*	*17.3M*	*5.40*
BPH1-Rep1	*28.6M*	*20.7M*	*6.85*
BPH1-Rep2	*32.2M*	*22.9M*	*7.17*
DU145-Baseline	*37.9M*	*27.7M*	*5.93*
DU145-Rep1	*32.5M*	*22.6M*	*6.71*
DU145-Rep2	*62.7M*	*43.9M*	*6.82*
PC3-Baseline	*42.6M*	*29.5M*	*5.47*
PC3-Rep1	*31.7M*	*22.6M*	*7.27*
PC3-Rep2	*33.6M*	*26.8M*	*6.98*



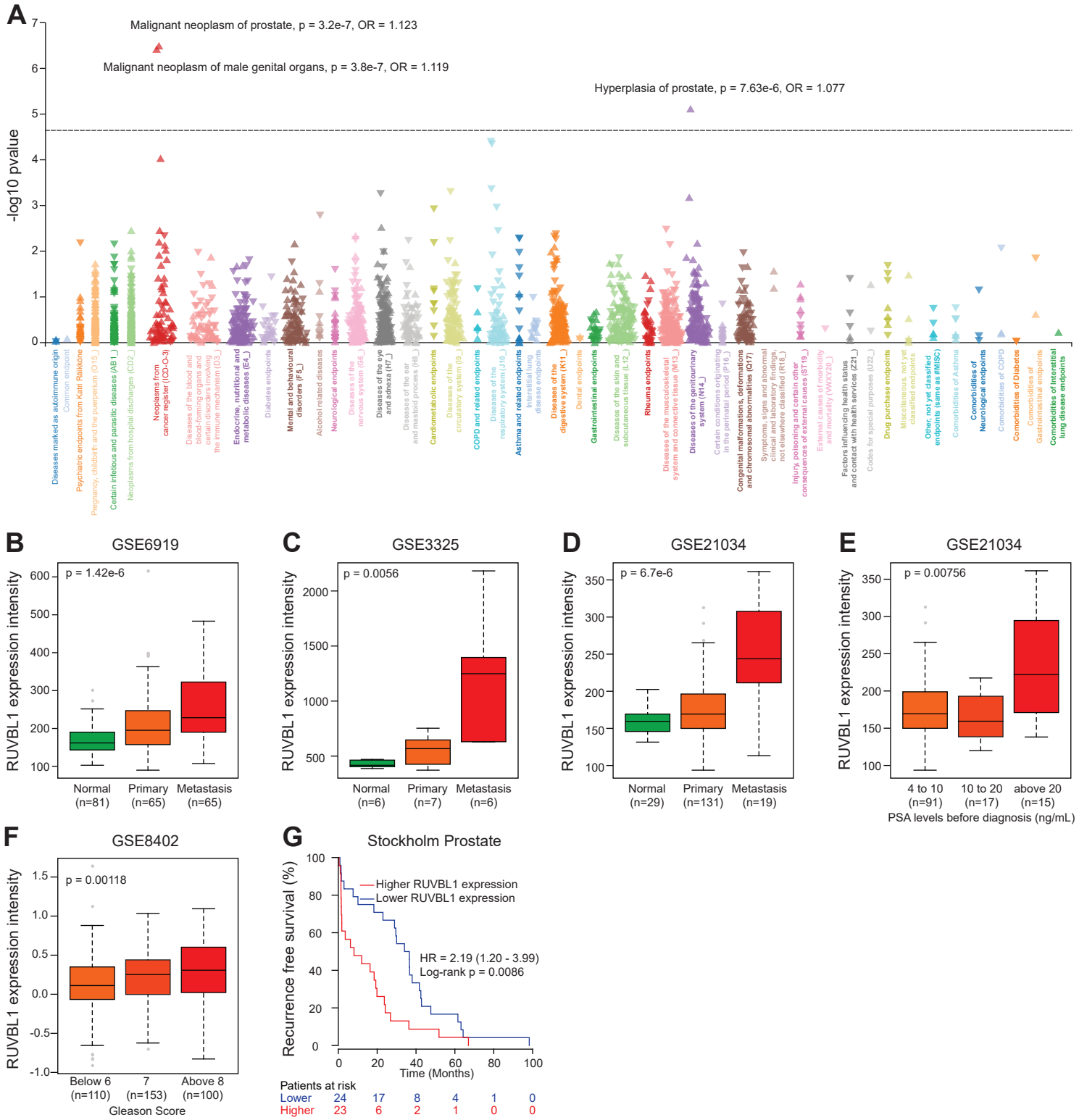
Supplementary Figure S2: CRISPRi screening quality control. A. Westernblot of Cas9 protein in dCas9 stable prostate cell lines. **B.** Functional titration of lentiGuide-Puro virus to ensure low MOI integration. **C.** Gel image of gRNA readout amplicon from each screening after PCR. Template gDNA amount for each sample is equivalent to at least 5 million cells. **D.** gRNA read count summary after count_spacers.py quantification **E-G.** Reproducibility of technical replicates in BPH1 (**E**), DU145 (**F**), and PC3 (**G**) cells.

Supplementary Figure 3



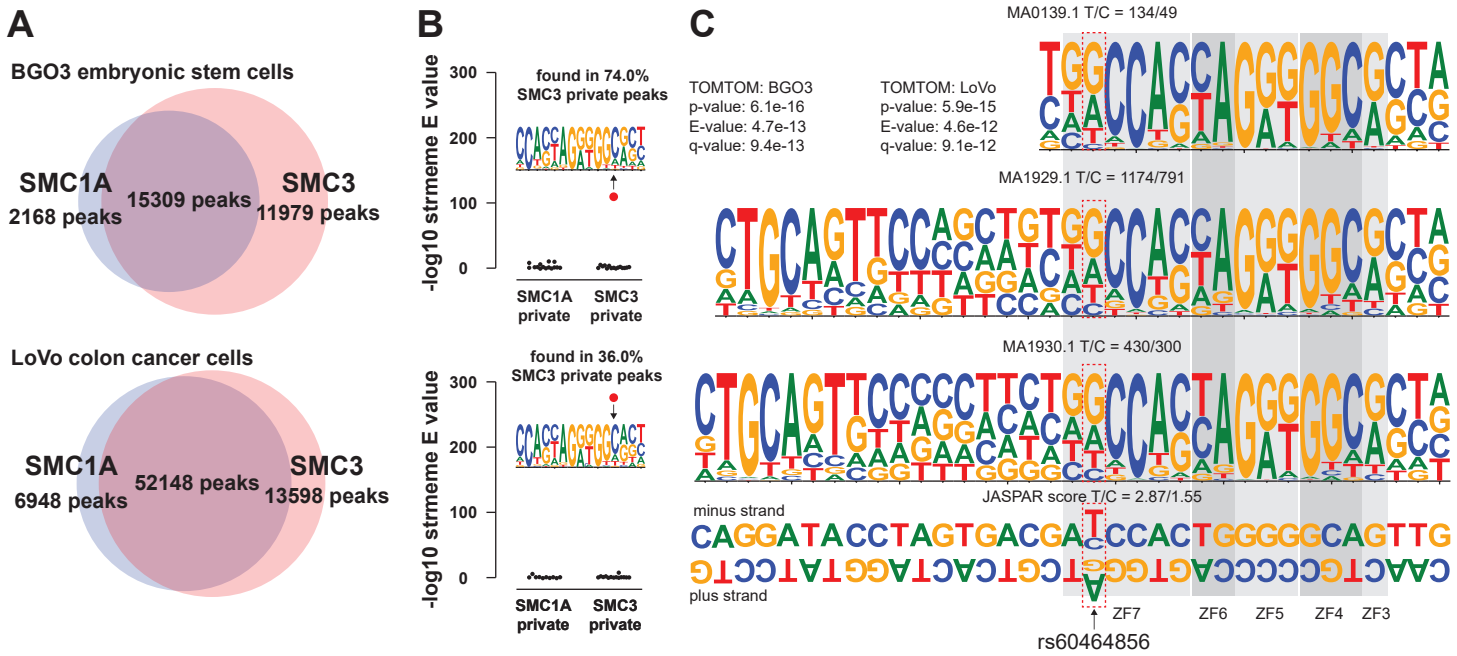
Supplementary Figure S3: Screening candidate characterization. A. PCA analysis of raw gRNA count in lentiGuide-Puro plasmid and each screening sample. **B-D.** Correlation between average gRNA foldchange and specificity score in BPH1 (**B**), DU145 (**C**), and PC3 (**D**) cells. **E-G.** gRNA foldchange comparison between predefined categories in BPH1 (**E**), DU145 (**F**), and PC3 (**G**) cells.

Supplementary Figure 4



Supplementary Figure S4: Clinical associations with RUVBL1 in other prostate cohorts. A. PheWAS results between rs60464856 and 2202 endpoints in the FinnGen study (n = 342,499). **B-C.** Associations between RUVBL1 gene expression and prostate cancer tissue type in Yu (**B**), Varambally (**C**), and Taylor's (**D**) prostate cancer cohorts. **E.** Associations between RUVBL1 gene expression and the pre-diagnosis prostate-specific antigen (PSA) levels in Taylor's prostate cancer cohorts. **F.** Associations between RUVBL1 gene expression and the Gleason score in Setlur's prostate cancer cohorts. **G.** Kaplan-Meier analysis on biochemical recurrence-free survival in Stockholm prostate cohort stratified by RUVBL1 expression. The Mann-Whitney U test was used to calculate the p-value for B to F.

Supplementary Figure 5



Supplementary Figure S4: Motif analysis for SMC1A and SMC3 in human cell lines. A. Proportional venn diagram of SMC1A and SMC3 ChIP-seq peak overlapping in BGO3 and LoVo cell lines. **B.** STREME motif scan in BGO3 and LoVo cell lines. **C.** TOMTOM comparison of SMC3 private motif to CTCF(MA0139.1) and base composition on SNP location in different CTCF motifs. The shaded blocks highlighted DNA binding sites interacting with CTCF zinc finger (ZF) domains.