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

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

1. Department of Tumor Biology, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612, United States
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. Division of Cancer Epidemiology, Population Sciences in the Pacific Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Hawaii, HI 96822, USA.
4. Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA.
5. Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland.

Abstract

Introduction: Genome-wide association studies along with expression quantitative trait loci (eQTL) have identified hundreds of single nucleotide polymorphisms (SNPs) and their target genes in prostate cancer (PrCa). Although these genetic associations to PrCa have been widely reported, functional characterization of these risk loci remains challenging.

Methods: To screen for regulatory SNPs, we designed a library containing 9133 guide RNAs (gRNAs) to target 2,166 candidate SNP sites implicated in PrCa. We performed negative screening in dCas9-KRAB stable prostate cell lines and applied the RIGOR program to identify the essential SNPs for cell proliferation. We further characterized the regulatory role of a selected single nucleotide polymorphim (SNP, rs60464856) using luciferase reporter assay, ChIP-qPCR, and xCas9 base editing in prostate cells. Finally, we investigated the biological impact of the SNP-regulated gene *RUVBL1* on cell proliferation and tumor growth via gene knockdown using in vitro and in vivo assays.

Results: From interference of 2,166 candidate SNPs via CRISPR interference screening, the RIGOR program identified 117 SNPs that could regulate genes to promote growth advantage in prostate cancer cell lines. Compared to unselected SNPs, the 117 candidates tended to reside near 5 kb flanking the transcription start sites ($p = 0.01$). To characterize the regulatory role of these SNPs, we selected one SNP (rs60464856) for detailed analysis. This SNP was covered by multiple gRNAs significantly depleted in the screening (FDR<0.05). Pooled SNP association analysis in the PRACTICAL cohort showed significantly higher PrCa risk for the G allele (pvalue=1.2E-16). eQTL analysis showed that the G allele is associated with an increased expression of *RUVBL1* in multiple datasets. To further validate the CRISPR interference effect, we transfected a gRNA targeting the rs60464856 site in the dCas9 stable cell lines and observed significant inhibition of the *RUVBL1* expression. We also applied the xCas9 adenine base editor to convert the rs60464856 A into G allele and observed an increased *RUVBL1* expression in subclones carrying the rs60464856 G allele in prostate cell lines. To test if any protein showed allele-specific binding at rs60464856, we used SILAC-based DNA pull-down proteomics and observed that cohesin subunits (including SMC3) preferred the A allele. ChIP qPCR assays showed significant enrichment of CTCF and SMC3 signals at the rs60464856 site with preferential binding to the A allele. To evaluate the potential role of this locus in maintaining long-range chromatin structure, we analyzed a HiC dataset and found that the rs60464856 locus enriched consistent chromatin interactions in prostate cell lines. To determine the potential role of the rs60464856 target gene, we knocked down the *RUVBL1* via shRNA and observed

significant proliferation inhibition in prostate cell lines. We also tested PC3 cells with *RUVBL1* knockdown in nude mice xenografts and observed reduced tumorigenesis. Gene set enrichment analysis showed that *RUVBL1* expression was associated with the enrichment of cell cycle related pathways in both cell line and TCGA prostate cancer cohorts. Lastly, we showed that an increased *RUVBL1* expression and its relevant pathway activation were associated with poor survival.

Conclusion: We applied the CRISPR interference screening at selected prostate cancer risk loci and identified over a hundred regulatory SNPs essential for prostate cell proliferation. Further analysis confirmed the important role of rs60464856 and its target gene *RUVBL1* in prostate cell growth and tumorigenesis.

Introduction

Among all cancer types, in 2021, prostate cancer (PrCa) accounted for 26% of 970,250 new cancer cases and caused 11% of 319,420 cancer-related deaths in US males (1). As a cancer type with a strong genetic predisposition, PrCa has been extensively investigated in genome-wide association studies (GWAS). The GWAS have been widely used in determining susceptible variants associated with increased disease risk and aggressiveness (2). It has been reported that the joint contribution of GWAS identified loci to PrCa risk approaches 20% (3). Although GWAS have been highly productive, only a few of risk loci have been functionally characterized (4, 5). Thus far, as these risk SNPs are found in noncoding 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 (6-8). With tremendous large-scale GWAS findings, especially highly reproducible ones, it is believed that some of these noncoding variants play a subtle but profound role in prostate cancer initiation, progression, and metastasis by modulating the expression of certain susceptible genes.

To dissect functional variants and their target genes in PrCa, bioinformatic and benchtop approaches are applied to prioritize and validate the causal variants. Curated databases such as ChIP-atlas (9), ENCODE (10), SRA, JASPAR (11), and GTEx provided abundant resources to estimate the genetic contribution of GWAS variants to PrCa susceptibility. However, to determine the biological effect, large-scale genetic assays are often needed to interrogate endogenous variant loci and directly characterize consequent phenotype changes (12). 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. Compared with 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 (13) and used for screening regulatory elements in mammalian cells (14).

It is believed that most SNPs function as a regulatory switch in the human genome. With the eliminated nuclease activity (13), 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 dCas9-KRAB stable prostate cell lines and designed an unbiased, highly reproducible gRNA library that

targeted candidate SNPs at PrCa risk loci. We then performed negative selection for potential SNPs conferring growth advantage. We finally provided detail analysis of a SNP-gene pair for their functional role in prostate cell lines and animal model. 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 (15) 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 further 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 backbone and transformed the pooled oligos into highly sensitive Endura ElectroCompetent Cells to generate plasmid libraries.

dCas9/KRAB stable cell lines and gRNA library processing. To establish stable cell lines, we packaged lenti-dCas9-KRAB-blast into lentiviral particles with HEK293FT cells and used ten-fold concentrated virus particles to transduce cell lines. After 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. To ensure quantification accuracy, we aimed for at least a 500-fold library size depth for each replicate.

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 generated 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 with the most depleted gRNA representation16). RIGER program ranks gRNAs according to their differential effects between two classes of samples, then identifies the SNP targeted by the shRNAs at the top of the list.

Plasmid construction and siRNA design. To enable the fluorescence-based cell sorting, copGFP ORF was assembled into base editor plasmid (17) 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 (17). 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 edit 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 single clone. We also used sanger sequencing to further verify every 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 (18), with minor modification using Tricine Sample Buffer and Mini-PROTEAN® Precast Gels (BioRad). SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) was used to produce 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 sequencing 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 (19). 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 previous Hi-C data, we adapted our ChIP qPCR protocols for chromatin conformation capturing (20). Compared with conventional ChIP assay protocol, our protocol applied dual crosslinking to maximally preserve chromatin contacts (21). 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.

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) for the prediction of SNP functionality. We included 5 SNPs with the highest PIP score for each eQTL-associated gene as another category named PIP SNPs (N=194). Finally, we included additional 641 control SNPs (CTL) that showed strong eQTL signals but without epigenomic signals. After removing duplicated SNPs, we selected 3408 SNPs for gRNA design. We scanned these SNPs with the CRISPOR program and 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 correlation 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 representation 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 read count in the end library to 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 didn't find significant correlations between fold change and gRNA specificity score (**Supplementary Figure 3B-3D**), suggesting minimun off-target effect of these selected gRNAs. When compared end libraries to baseline libraries, we found significant gRNA depletion in BPH1, DU145, and PC3 screening experiments (**Figure 1C-1E**). When compared 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**). We plotted the gRNA representation before and after the growth selection and highlighted representative SNPs in each cell line (**Figure 1H-1J**).

rs60464856 showd a regulatory role in *RUVBL1* expression and contributed to prostate cancer risk. 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**). We further showed that among the *RUVBL1* eQTL loci, rs60464856 resided in a linkage disequilibrium block with second best significance (**Figure 2C**). Additionally, the rs60464856 G allele was significantly associated with elevated RUVBL1 gene expression in GTEx, Mayo (22, 23), and TCGA prostate eQTL (24) datasets (**Figure 2D**). We also used three gRNAs targeting 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 downstream gene expression, we applied nickase Cas9 (xCas9) base editing technology, which featured on 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 ChIP-seq data collections, such as cistrome and ChIP-ATLAS, for potential transcription factor bindings. However, we didn't find highly convincing evidence showing allelic factor TF binding on this locus. Since there are multiple datasets reporting histone modification signals on rs60464856 locus, we then performed ChIP-qPCR to evaluate the locus-specific enrichment for 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 rs60464856 locus (**Figure 3A**). Consistently, we found H3K4me1, H3K4me3, and H3K27ac motification on rs60464856 in multiple cell lines (**Figure 3B**), suggesting a conservative regulatory potential of this locus in prostate. With a recent HiC datasets (25), we also visualized the long range interactions surrounding rs60464856 locus. By suming 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% interactions compared to DU145 cells (**Figure 3C**). In contrast, the interactions LR 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 3F**). 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 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 reactions to pull down 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 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 two existing ChIP-seq datasets [(26) and GSE36578] that quantified SMC1A and SMC3 binding in the human genome (**Supplementary Figure 4A**., 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 4B).** Through strmeme-tomtom comparison, we also found this motif was highly similar to the CTCF binding site (**Supplementary Figure 4C**). This result implicated a unique DNA binding role of SMC3, which is worth further investigation.

***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 effection of *RUVBL1* by CRISPR or RNAi screening in DepMap portal. We found that RUVBL1 was a common essential gene in majority of human cell lines, with 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 (**Figure 4B**) levels. 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 the 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 expressions (**Figure 4G-4H**). We further visualized the gene expression changes in these significantly enriched pathways and found consistent trends with the RUVBL1 levels (**Figure 4I**). To characterize the tumorigenic effect of the RUVBL1 gene, we performed xenograft experiments with the PC3 stable cell lines. We found that the RUVBL1 knockdown significantly inhibited tumor growth (**Figure 4K**) and endpoint tumor weight in the mice model (**Figure 4L**).

***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 expressions 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 8significant 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. To see whether RUVBL1 could be a prognostic marker, we performed a Kaplan-Meier analysis and found significantly worse progression-free (**Figure 5H**) survivals in those with higher RUVBL1 expressions. 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 risks tended to have significantly worse progress-free survival (**Figure 5J-5L**).

Discussion

Over the past decade, GWAS and eQTL analysis have been highly productive in finding PrCa risk loci and causal genes. Despite contributing to a better understanding of the biological significance of risk predisposition, these analyses didn't 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, large scale functional evaluation of target risk loci is highly warranted. Due to the non-coding nature in most risk loci, these variants are believed to play a regulatory role. Therefore, we applied the dCas9-based CRISPR interference assay to target SNP sequences at risk loci, aiming to mimic the regulatory alteration caused by single base substitutions. This genomewide screening at PrCa 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 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 (27, 28), any mismatches at 7-bp seed region of the gRNA could cause a rapid rejection of these targets by dCas9 protein. This may explain the A allele specific depletion effect of gRNA on position -5 to -7, as is 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 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 this 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 (29). 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 didn't influence *SMC3* stability and cytoplasmic accumulation (30). More recently (31), *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* detected sequence motifs in the two cell lines.

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 (32), including chromatin decondensation (33, 34), misfolded protein aggregation (35), and transcription regulation (36). In addition to the previously reported mTORC1 (37) pathway, our enrichment analysis demonstrated that the *RUVBL1* expression

was consistently correlated to 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 (38, 39) in treating prostate cancer. This result also suggests a potential application of rs60464856 genotype to select a target population for future clinical trials.

One potential issue of this study is inconsistent results in some tests among different cell lines. While these inconsistence 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 (40). Additionally, to increase our gRNA library coverage, we may use a 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 in-depth investigation, which may uncover unique SNP and gene function specific to the cell line of interest. In summary, we applied CRISPRi screening technology to screen for survival-essential SNPs at genome scale. We identified over a hundred of functional SNPs that regulate cell proliferation. We further functionally characterized a rs60464856-RUVBL1 pair for regulatory and functional role in the prostate context. This result will enrich our knowledge in PrCa predisposition and provide insight into the cancer risk classification and potential therapeutic target for personalized treatment.

Acknowledgment

We thank Flow Cytometry, Molecular Genomic, Proteomic and Microscopy Core Facility at the Moffitt Cancer Center, an NCI designated Comprehensive Cancer Center (P30CA076292). This work was supported by National Institutes of Health (R01CA250018 and R01CA212097 to L.Wang., R01CA263494 to L. Wu).

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 starting sites (TSS) compared to the unselected. **G.** Enrichment of the selected SNPs in allele specific binding (ASB) annotation according to 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 targtting 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 binds Crick or Watson strands of the germline DNA, with numbers in the center showing 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 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 expressions 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 modification on rs60464856 locus in BPH1, DU145 and PC3 clones. **B.** ChIP-seq signals of H3K4me1, H3K4me3 and H3K27ac modification 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 CRSIPR and RNAi screening from CCLE human cell line collections. A score of 0 indicate 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 in related tosignificantly 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.
2. 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.
3. 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.
4. 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.
5. Gallagher, M.D. and Chen-Plotkin, A.S. (2018) The Post-GWAS Era: From Association to Function. *Am J Hum Genet*, **102**, 717-730.
6. 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.
7. 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.
8. 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.
9. 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. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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. *et al.* (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, **556**, 57-63.
18. 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.
19. 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.
20. 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.

21. 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.
22. 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.
23. 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.
24. 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.
25. 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**.
26. Yan, J., Enge, M., Whitington, T., Dave, K., Liu, J., Sur, I., Schmierer, B., Jolma, A., Kivioja, T., Taipale, M. *et al.* (2013) Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell*, **154**, 801-813.
27. 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.
28. 28 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.
29. 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.
30. 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.
31. 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**.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.
40. 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.

Figure 1

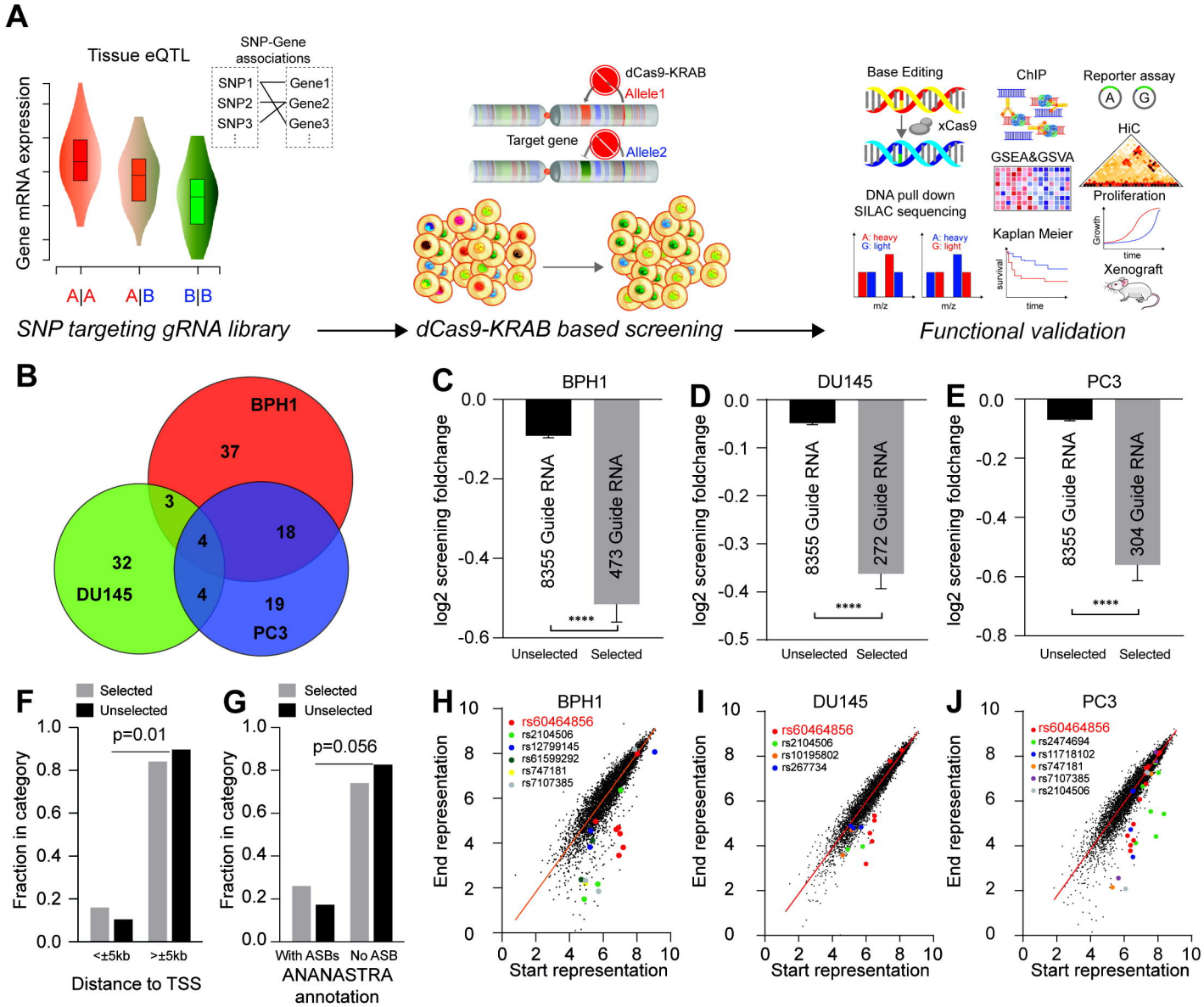

A
Tissue eQTL
SNP-Gene associations
SNP1
SNP2
SNP3
Gene1
Gene2
Gene3
Gene mRNA expression
A|A
A|B
B|B
dCas9-KRAB
Allele1
Target gene
Allele2
Base Editing
xCas9
ChIP
Reporter assay
HiC
GSEA&GSVA
DNA pull down
SILAC sequencing
Proliferation
Kaplan Meier
Xenograft
SNP targeting gRNA library
dCas9-KRAB based screening
Functional validation
B
BPH1
DU145
PC3
37
3
18
32
4
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

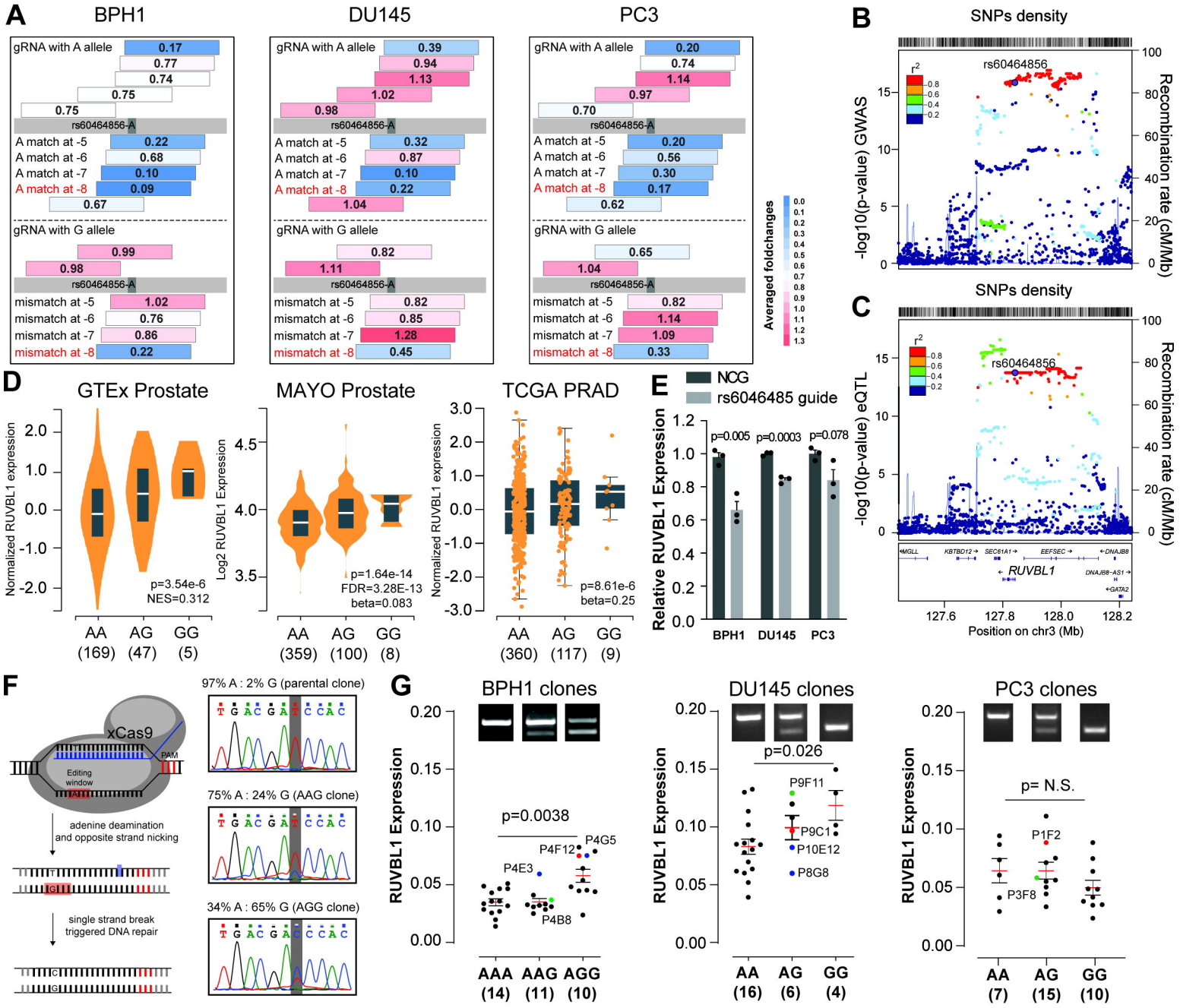

A
BPH1
DU145
PC3
gRNA with A allele
0.17
0.77
0.74
0.75
0.75
rs60464856-A
A match at -5
0.22
A match at -6
0.68
A match at -7
0.10
A match at -8
0.09
0.67
gRNA with G allele
0.99
0.98
rs60464856-A
mismatch at -5
1.02
mismatch at -6
0.76
mismatch at -7
0.86
mismatch at -8
0.22
gRNA with A allele
0.39
0.94
1.13
1.02
0.98
rs60464856-A
A match at -5
0.32
A match at -6
0.87
A match at -7
0.10
A match at -8
0.22
1.04
gRNA with G allele
0.82
1.11
rs60464856-A
mismatch at -5
0.82
mismatch at -6
0.85
mismatch at -7
1.28
mismatch at -8
0.45
gRNA with A allele
0.20
0.74
1.14
0.97
0.70
rs60464856-A
A match at -5
0.20
A match at -6
0.56
A match at -7
0.30
A match at -8
0.17
0.62
gRNA with G allele
0.65
1.04
rs60464856-A
mismatch at -5
0.82
mismatch at -6
1.14
mismatch at -7
1.09
mismatch at -8
0.33
Averaged foldchanges
B
SNPs density
rs60464856
-log10(p-value) GWAS
Recombination rate (cM/Mb)
C
SNPs density
rs60464856
-log10(p-value) eQTL
Recombination rate (cM/Mb)
RUVBL1
Position on chr3 (Mb)
D
GTEx Prostate
Normalized RUVBL1 expression
p=3.54e-6
NES=0.312
AA (169)
AG (47)
GG (5)
MAYO Prostate
Log2 RUVBL1 Expression
p=1.64e-14
FDR=3.28E-13
beta=0.083
AA (359)
AG (100)
GG (8)
TCGA PRAD
Normalized RUVBL1 expression
p=8.61e-6
beta=0.25
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
RUVBL1 Expression
p=0.0038
P4E3
P4F12
P4G5
P4B8
AAA (14)
AAG (11)
AGG (10)
DU145 clones
RUVBL1 Expression
p=0.026
P9F11
P9C1
P10E12
P8G8
AA (16)
AG (6)
GG (4)
PC3 clones
RUVBL1 Expression
p= N.S.
P1F2
P3F8
AA (7)
AG (15)
GG (10)

Figure 3

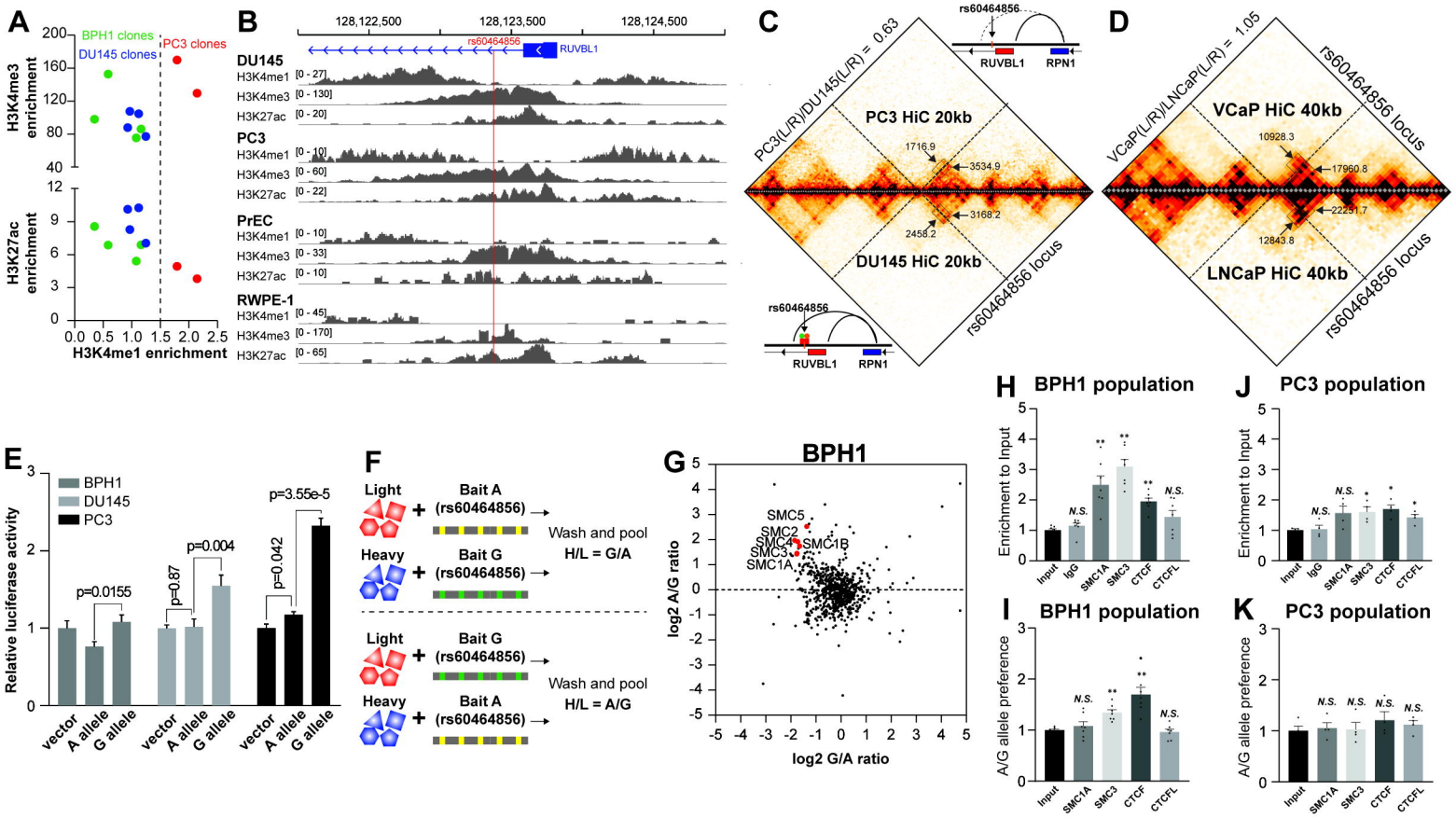

A
BPH1 clones
DU145 clones
PC3 clones
H3K4me3 enrichment
H3K27ac enrichment
H3K4me1 enrichment
B
128,122,500
128,123,500
128,124,500
rs60464856
RUVBL1
DU145
PC3
PrEC
RWPE-1
H3K4me1
H3K4me3
H3K27ac
C
PC3(L/R)/DU145(L/R) = 0.63
PC3 HiC 20kb
DU145 HiC 20kb
rs60464856 locus
RPN1
D
VCaP(L/R)/LNCaP(L/R) = 1.05
VCaP HiC 40kb
LNCaP HiC 40kb
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
SMC1B
SMC3
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

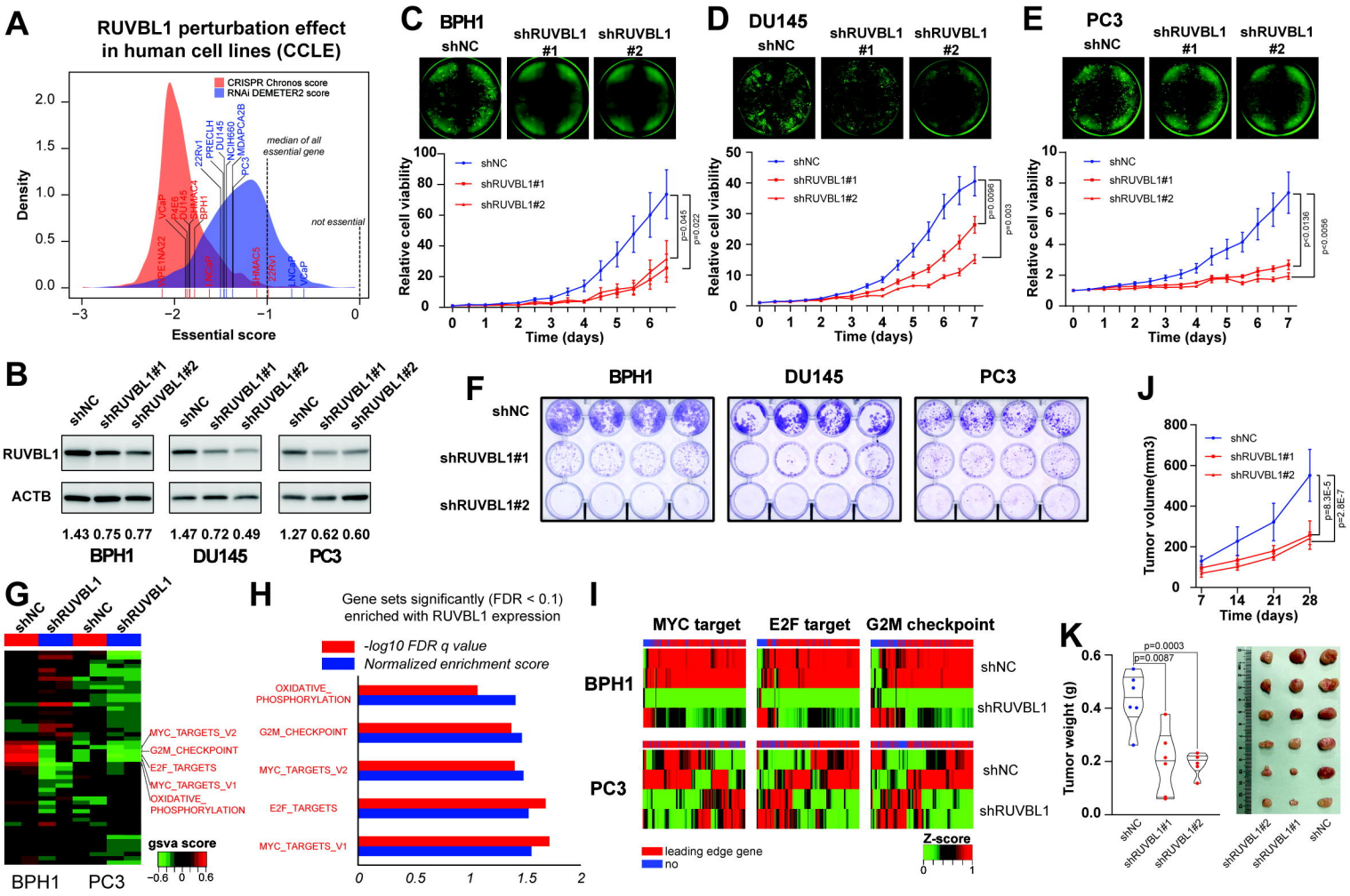
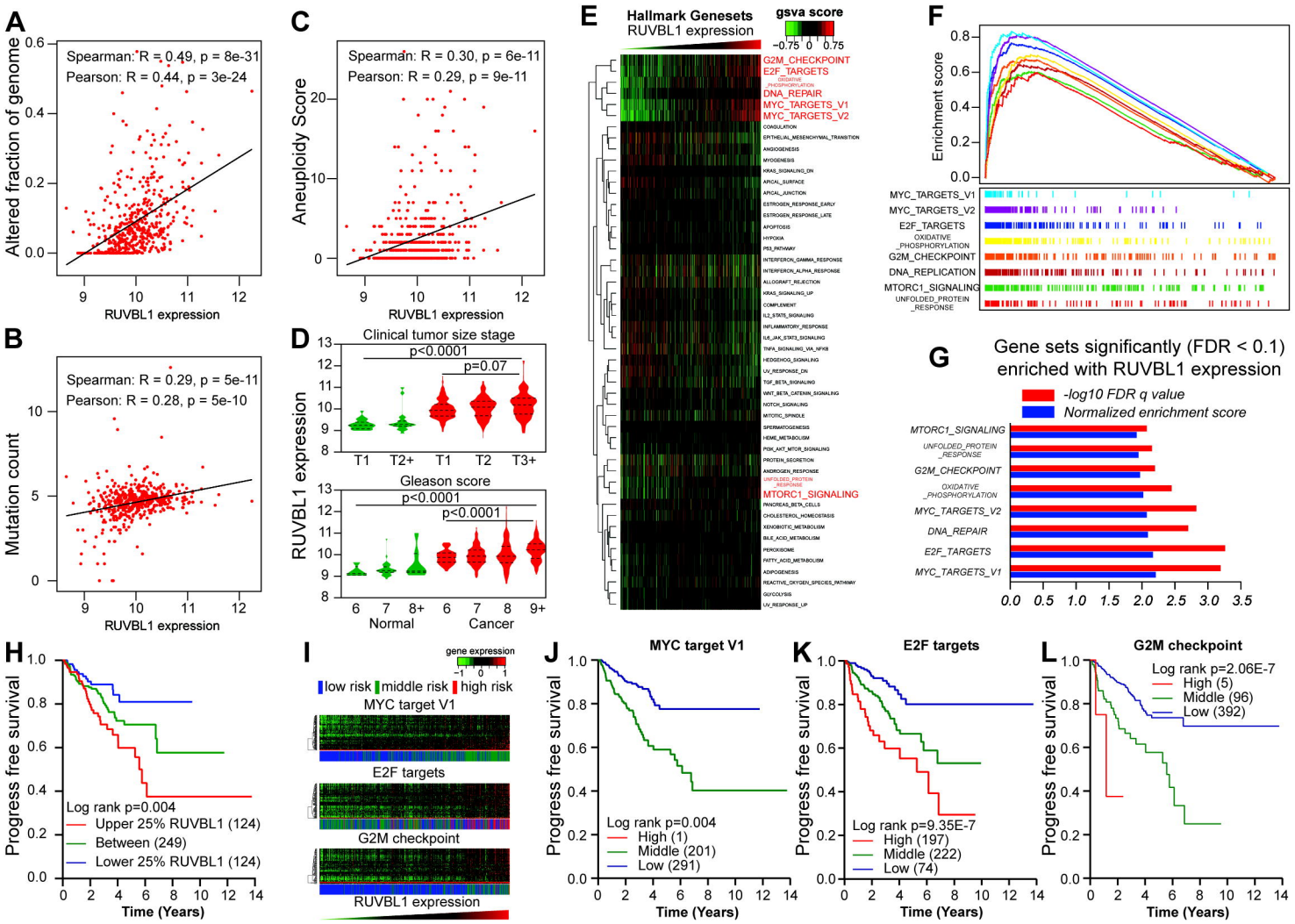
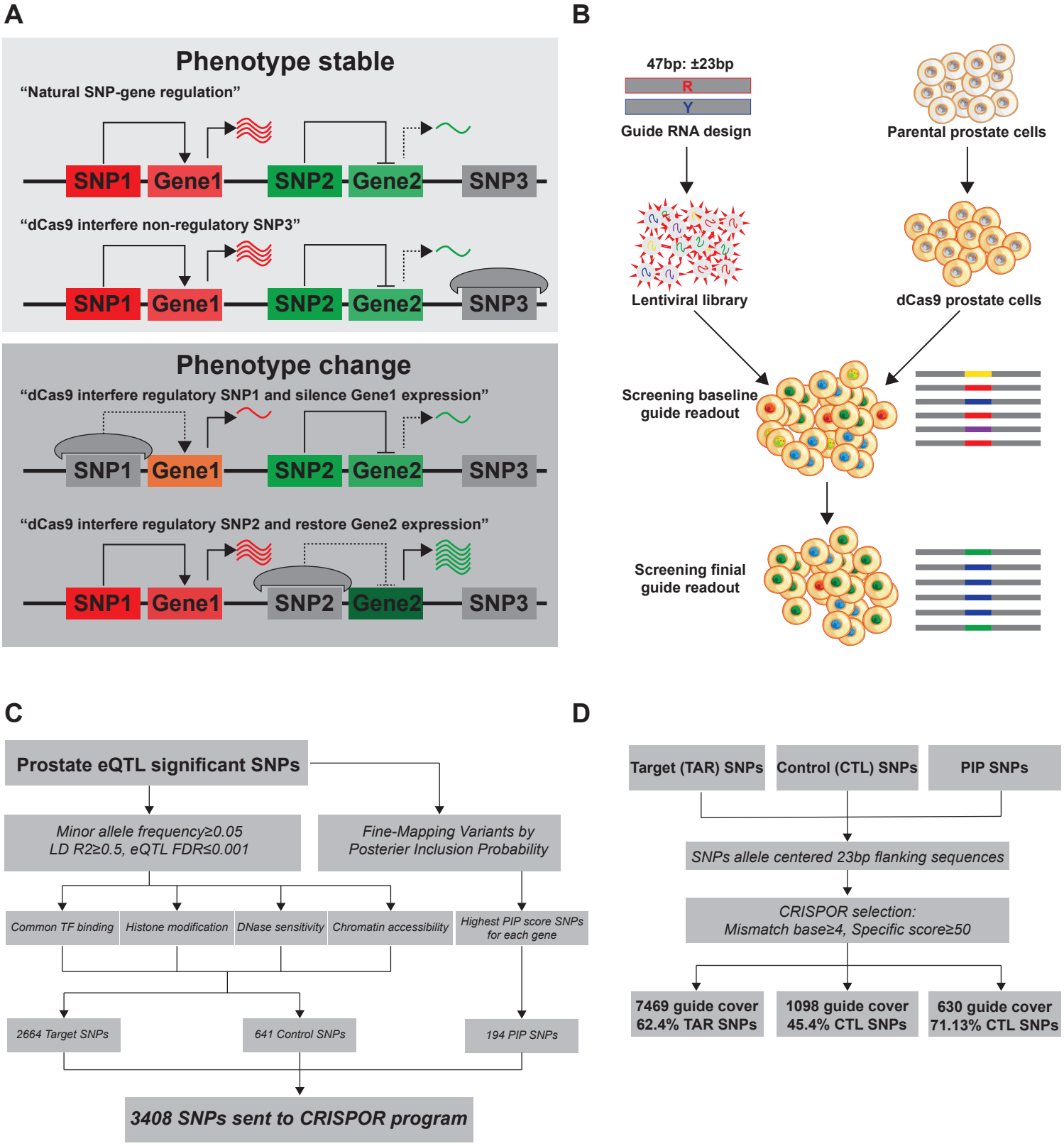

A
RUVBL1 perturbation effect in human cell lines (CCLE)
CRISPR Chronos score
RNAi DEMETER2 score
median of all essential gene
not essential
Density
Essential score
B
shNC
shRUVBL1#1
shRUVBL1#2
RUVBL1
ACTB
1.43 0.75 0.77
1.47 0.72 0.49
1.27 0.62 0.60
BPH1
DU145
PC3
C
BPH1
D
DU145
E
PC3
Relative cell viability
Time (days)
p=0.045
p=0.022
p=0.0096
p=0.003
p=0.0138
p<0.0056
F
J
Tumor volume(mm3)
p=8.3E-5
p=2.8E-7
G
gsva score
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
leading edge gene
no
Z-score
K
Tumor weight (g)
p=0.0003
p=0.0087

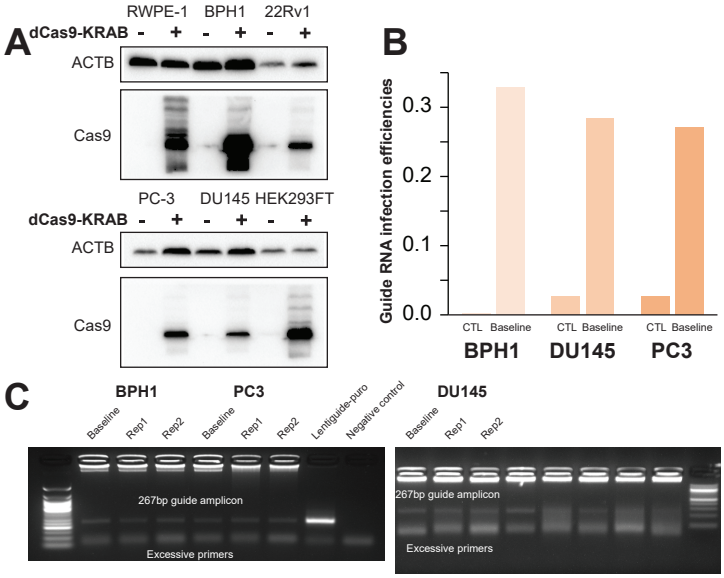
Figure 5


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

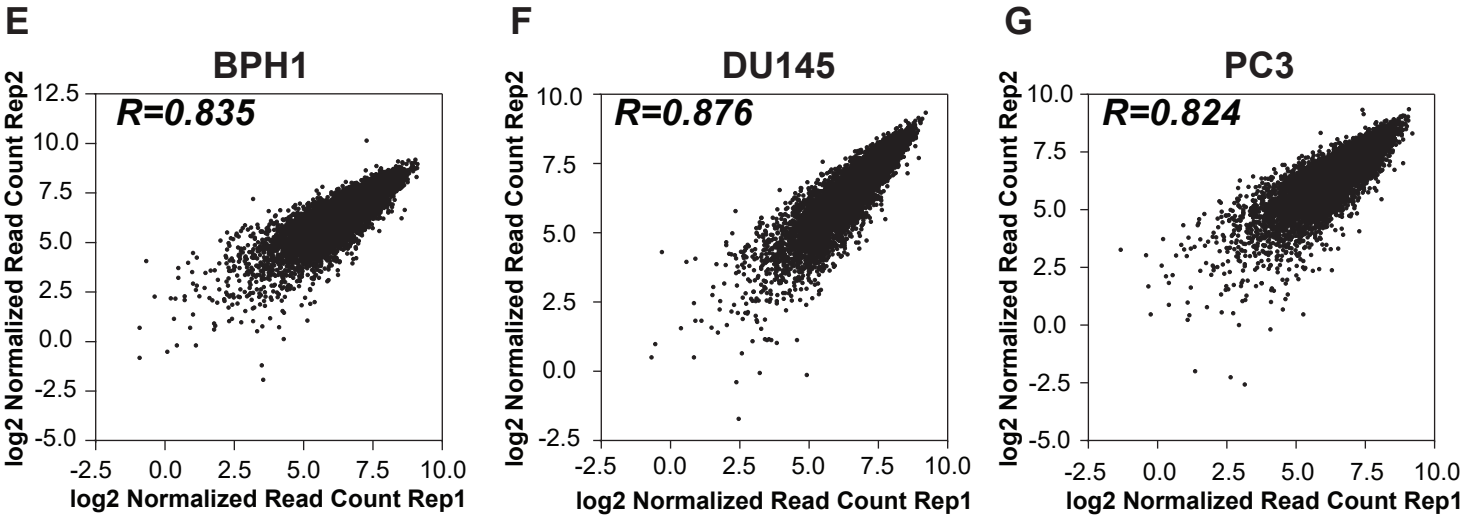


Supplementary Figure 2

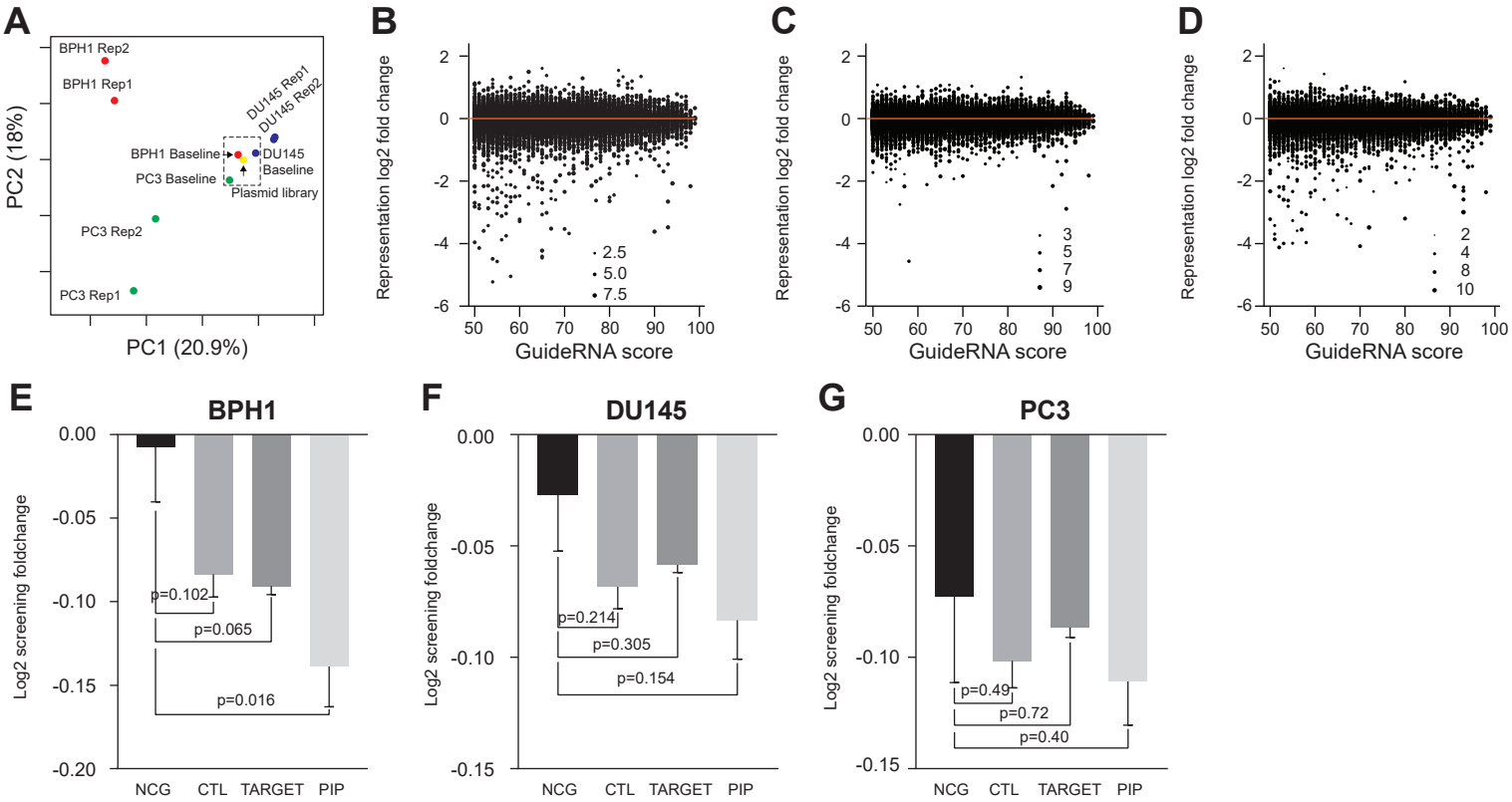
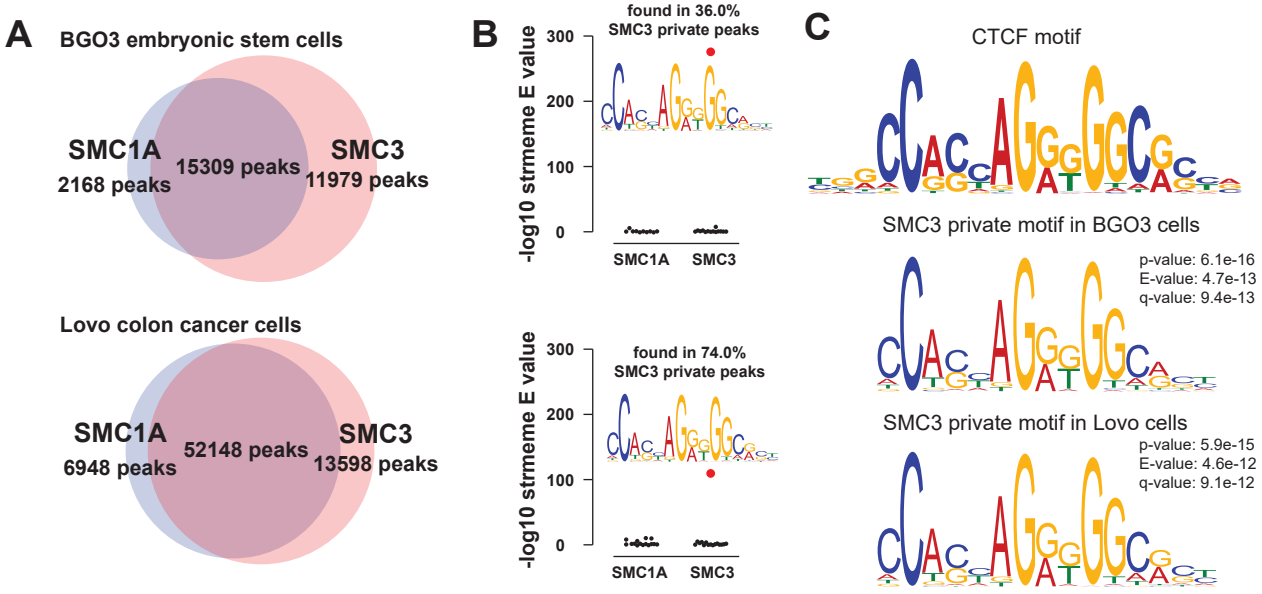


D

	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 3



Supplementary Figure 4