The landscape of RNA polymerase II–associated chromatin interactions in prostate cancer

Susmita G. Ramanand,[1] Yong Chen,[2,3] Jiapei Yuan,[1] Kelly Daescu,[2] Maryou B.K. Lambros,[4] Kathleen E. Houlahan,[5,6,7,8,9,10] Suzanne Carreira,[4] Wei Yuan,[4] GuemHee Baek,[1] Adam Sharp,[4] Alec Paschalis,[4] Mohammed Kanchwala,[11] Yunpeng Gao,[1] Adam Aslam,[1] Nida Safdar,[1] Xiaowei Zhan,[12] Ganesh V. Raj,[13] Chao Xing,[8,9,14] Paul C. Boutros,[6,7,8,9,10] Johann de Bono,[4] Michael Q. Zhang,[2,15] and Ram S. Mani[1,13,16]

[1]Department of Pathology, UT Southwestern Medical Center, Dallas, Texas, USA. [2]Department of Biological Sciences, Center for Systems Biology, University of Texas at Dallas, Richardson, Texas, USA. [3]Department of Molecular and Cellular Biosciences, Rowan University, Glassboro, New Jersey, USA. [4]Prostate Cancer Targeted Therapy and Cancer Biomarkers Group, Institute of Cancer Research (ICR) and Royal Marsden NHS Foundation Trust, Sutton, United Kingdom. [5]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [6]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. [7]Vector Institute, Toronto, Ontario, Canada. [8]Department of Urology, [9]Department of Human Genetics, and [10]Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, California, USA. [11]Eugene McDermott Center for Human Growth and Development, [12]Department of Population and Data Sciences, [13]Department of Urology, and [14]Department of Bioinformatics, UT Southwestern Medical Center, Dallas, Texas, USA. [15]MOE Key Laboratory of Bioinformatics and Bioinformatics Division, Center for Synthetic and Systems Biology, TNLIST/Department of Automation, Tsinghua University, Beijing, China. [16]Harold C. Simmons Comprehensive Cancer Center, UT Southwestern Medical Center, Dallas, Texas, USA.

Transcriptional dysregulation is a hallmark of prostate cancer (PCa). We mapped the RNA polymerase II–associated (RNA Pol II–associated) chromatin interactions in normal prostate cells and PCa cells. We discovered thousands of enhancer-promoter, enhancer-enhancer, as well as promoter-promoter chromatin interactions. These transcriptional hubs operate within the framework set by structural proteins — CTCF and cohesins — and are regulated by the cooperative action of master transcription factors, such as the androgen receptor (AR) and FOXA1. By combining analyses from metastatic castration-resistant PCa (mCRPC) specimens, we show that AR locus amplification contributes to the transcriptional upregulation of the *AR* gene by increasing the total number of chromatin interaction modules comprising the *AR* gene and its distal enhancer. We deconvoluted the transcription control modules of several PCa genes, notably the biomarker *KLK3*, lineage-restricted genes (*KRT8*, *KRT18*, *HOXB13*, *FOXA1*, *ZBTB16*), the drug target *EZH2*, and the oncogene *MYC*. By integrating clinical PCa data, we defined a germline-somatic interplay between the PCa risk allele rs684232 and the somatically acquired *TMPRSS2-ERG* gene fusion in the transcriptional regulation of multiple target genes — *VPS53*, *FAM57A*, and *GEMIN4*. Our studies implicate changes in genome organization as a critical determinant of aberrant transcriptional regulation in PCa.

Introduction

Whole-genome sequencing (WGS) and exome sequencing studies have revealed that prostate cancer (PCa) is generally characterized by lower mutation rates and higher rates of recurrent genomic rearrangements. The majority of PCa-associated recurrent genomic rearrangements promote the upregulation of transcription factor genes. For example, genomic rearrangements and gene fusions resulting in the upregulation of ETS transcription factor family genes like *ERG* and *ETV1* are observed in more than 50% of PCa samples (1). The androgen receptor (AR) is a master transcription factor that is essential for normal development of the prostate gland. Intriguingly, aberrant AR signaling drives many facets of PCa etiology, including lethal metastatic castration-resistant prostate cancer (mCRPC) development (2). Copy number alterations and amplifications of the *AR* locus are observed in over 60% of mCRPC cases (3, 4). The AR and the pioneer factor FOXA1 co-occupy distant regulatory elements, such as enhancers, to regulate transcription (2, 5). The binding of ETS transcription factors to these regulatory elements adds an additional layer of complexity to the transcriptional

Authorship note: SGR, YC, and JY are co–first authors.
Conflict of interest: AS, MBKL, SC, AP, WY, and JDB are employees of the ICR, which has a commercial interest in abiraterone. AS has received travel support from Sanofi and Roche-Genentech and speakers' honoraria from Astellas Pharma. JDB has served on advisory boards and received fees from many companies including Astra Zeneca, Astellas, Bayer, Boehringer Ingelheim, CellCentric, Daiichi Sankyo, Roche-Genentech, Genmab, GSK, Janssen, Merck Serono, Merck Sharp & Dohme (MSD), Menarini Silicon Biosystems, Orion, Pfizer, QIAGEN, Sanofi-Aventis, Sierra Oncology, Taiho, and Vertex Pharmaceuticals. JDB is an employee of the ICR, which has received funding or other support for his research work from Astra Zeneca, Astellas, Bayer, CellCentric, Daiichi Sankyo, Genentech, Genmab, GSK, Janssen, Merck Serono, MSD, Menarini Silicon Biosystems, Orion, Sanofi Aventis, Sierra Oncology, Taiho, Pfizer, and Vertex Pharmaceuticals, and which has a commercial interest in abiraterone, PARP inhibition in DNA repair–defective cancers, and PI3K/AKT pathway inhibitors (no personal income). JDB was named an inventor, with no financial interest, for US patent 8,822,438 ("Methods and compositions for treating cancer"). He has been the co-investigator or principal investigator of many industry-sponsored clinical trials.
Submitted: October 16, 2019; **Accepted:** April 23, 2020; **Published:** June 22, 2020.

Reference information: *J Clin Invest*. 2020;130(8):3987–4005.
https://doi.org/10.1172/JCI134260.

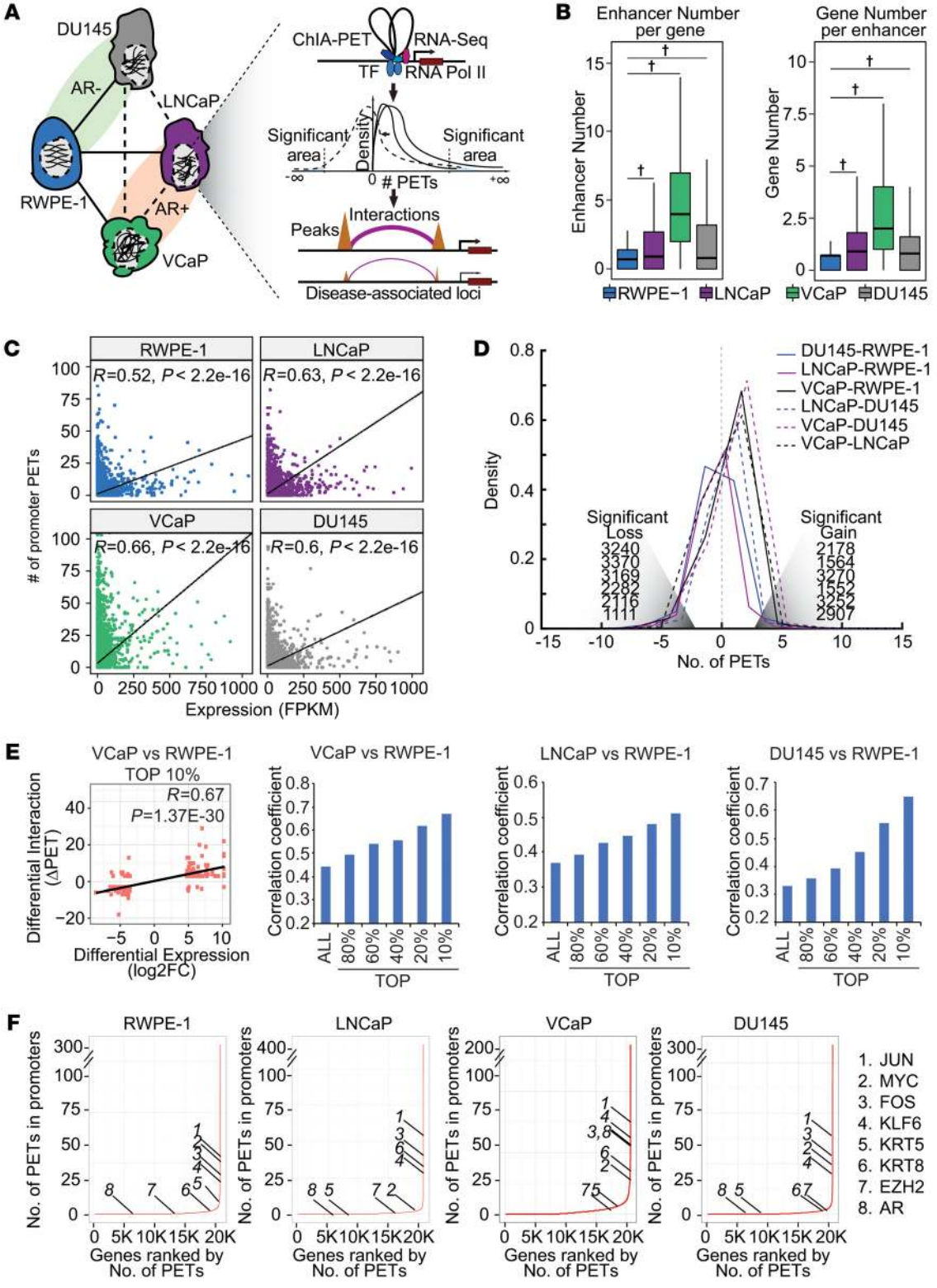

A
DU145
AR-
LNCaP
RWPE-1
AR+
VCaP
ChIA-PET
RNA-Seq
TF
RNA Pol II
Significant area
Density
Significant area
-∞
0 # PETs
+∞
Interactions
Peaks
Disease-associated loci
B
Enhancer Number per gene
Enhancer Number
Gene Number per enhancer
Gene Number
RWPE-1
LNCaP
VCaP
DU145
C
RWPE-1
R=0.52, P< 2.2e-16
LNCaP
R=0.63, P< 2.2e-16
VCaP
R=0.66, P< 2.2e-16
DU145
R=0.6, P< 2.2e-16
# of promoter PETs
Expression (FPKM)
D
DU145-RWPE-1
LNCaP-RWPE-1
VCaP-RWPE-1
LNCaP-DU145
VCaP-DU145
VCaP-LNCaP
Density
Significant Loss
3240
3370
3169
2282
2116
1111
Significant Gain
2178
1564
3270
1552
3252
2907
No. of PETs
E
VCaP vs RWPE-1
TOP 10%
R=0.67
P=1.37E-30
Differential Interaction (ΔPET)
Differential Expression (log2FC)
VCaP vs RWPE-1
LNCaP vs RWPE-1
DU145 vs RWPE-1
Correlation coefficient
ALL
80%
60%
40%
20%
10%
TOP
F
RWPE-1
LNCaP
VCaP
DU145
No. of PETs in promoters
Genes ranked by No. of PETs
1. JUN
2. MYC
3. FOS
4. KLF6
5. KRT5
6. KRT8
7. EZH2
8. AR

Figure 1. Analysis of RNA Pol II–associated chromatin interactions. (**A**) Pipeline for ChIA-PET data processing and identification of chromatin interactions. (**B**) Enhancers per gene and genes per enhancer for each cell line. For enhancers per gene, all expressed genes (fragments per kb per million mapped reads [FPKM] >1) were included; for genes per enhancer, all enhancers located in RNA Pol II peak anchor regions were included. Box plot represents the median and the 25% and 75% quantiles, with lines at 1.5 times the IQR. The significance for each pair comparison was tested using the Kolmogorov-Smirnov test ($^{\dagger}P < 2.2 \times 10^{-16}$).The analysis was normalized by adjusting for sequencing depth. (**C**) Correlation between RNA expression and chromatin interaction across the 4 cell lines. The expression level was measured by FPKM transformed by $\log_2$, and chromatin interactions were measured by the number of promoter PETs. The correlation efficient was calculated by Spearman's correlation, and *P* values are shown. (**D**) Significant gain and loss of RNA Pol II–associated chromatin interactions in the 4 cell lines. The 3 solid lines show the distributions of the 3 PCa cell lines compared with benign cells (RWPE-1), whereas the 3 broken lines show the distributions among the 3 PCa cell lines. The significant gain or loss interactions were obtained using the DNB model. The total number of interactions that were significantly altered between cell lines are listed accordingly. (**E**) Scatter plot shows the correlation between changes in gene expression and changes in promoter-associated chromatin interactions. Graphs show Spearman's correlation coefficients for the top 10%, 20%, 40%, 60%, and 80% differentially expressed genes (right panel). FC, fold change. (**F**) Total number of PETs at promoter regions in the 4 cell lines. Genes are ranked by increasing number of PETs. Eight representative genes are labeled 1 through 8 in the plots.

output, ultimately resulting in the oncogenic phenotype (6, 7). Taken together, these studies indicate that transcriptional dysregulation is a distinguishing feature of PCa development.

Two-dimensional (2D) genomics approaches like ChIP-Seq have contributed to the identification of thousands of enhancers in PCa (7, 8). However, the targets of most of these enhancers are unknown. It is unclear whether individual enhancers regulate single or multiple genes and, conversely, whether individual genes are regulated by single or multiple enhancers in PCa. 2D genomics cannot reveal whether multiple genes and enhancers are coordinately regulated in shared nuclear space. Methods like Hi-C have enabled the identification of topologically associating domains (TADs) that change in size between normal and PCa cells (9, 10). However, in general, Hi-C studies have limited utility in the precision mapping of enhancer-promoter contacts. The absence of this knowledge has stymied our understanding of the regulatory targets of somatic mutations and germline risk alleles residing in the intergenic regions.

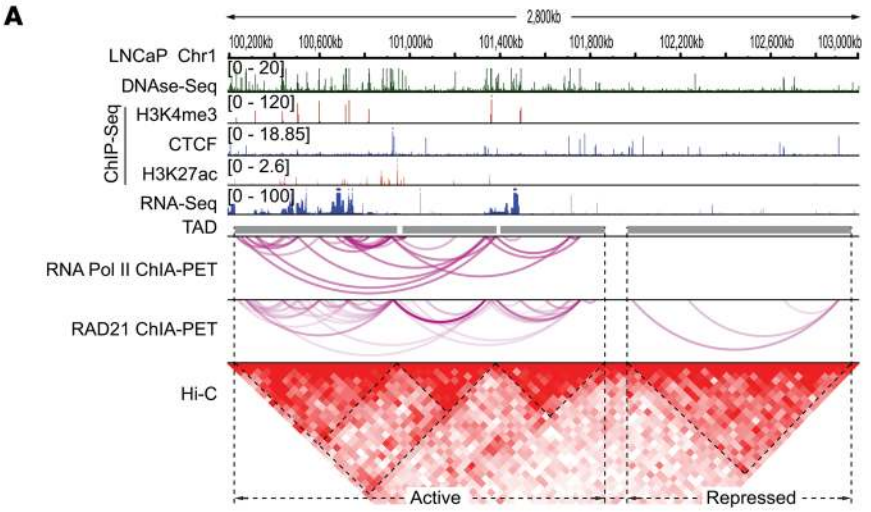
We present for the first time to our knowledge the 3D landscape of RNA polymerase II–associated (RNA Pol II–associated) chromatin interactions in normal prostate cells and PCa cells. By pairing thousands of enhancers to their target genes, we identified thousands of transcriptional network hubs operating within the framework set by structural proteins like CTCF and cohesins. Integrative analyses of these data uncovered multiple mechanisms of transcriptional regulation and its dysregulation in PCa. Finally, we demonstrate that many somatic and germline DNA alterations rewire the landscape of RNA Pol II interactions in PCa.

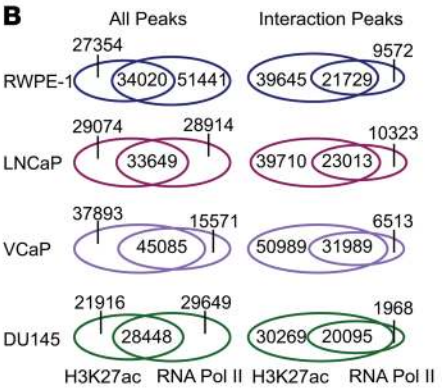
Results

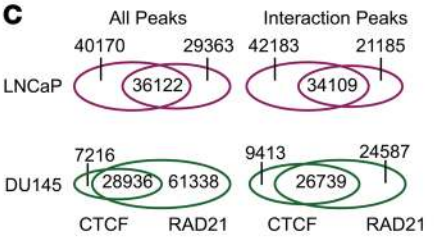
RNA Pol II–associated chromatin interactions in PCa. Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) is a genome-wide integration of ChIP and chromosome conformation capture (3C) methods to map the genomic interaction landscape of any chromatin-associated protein of interest (11, 12). We sought to identify global differences in transcriptional regulation by comparing RNA Pol II ChIA-PET, along with ChIP-Seq and RNA-Seq analyses in benign prostate cells (RWPE-1) and PCa cells (LNCaP, VCaP, and DU145) (Supplemental Tables 1 and 2; supplemental material available online with this article; https://doi.org/10.1172/JCI134260DS1). LNCaP and VCaP cells are AR positive, whereas RWPE-1 and DU145 cells are AR negative (Figure 1A).

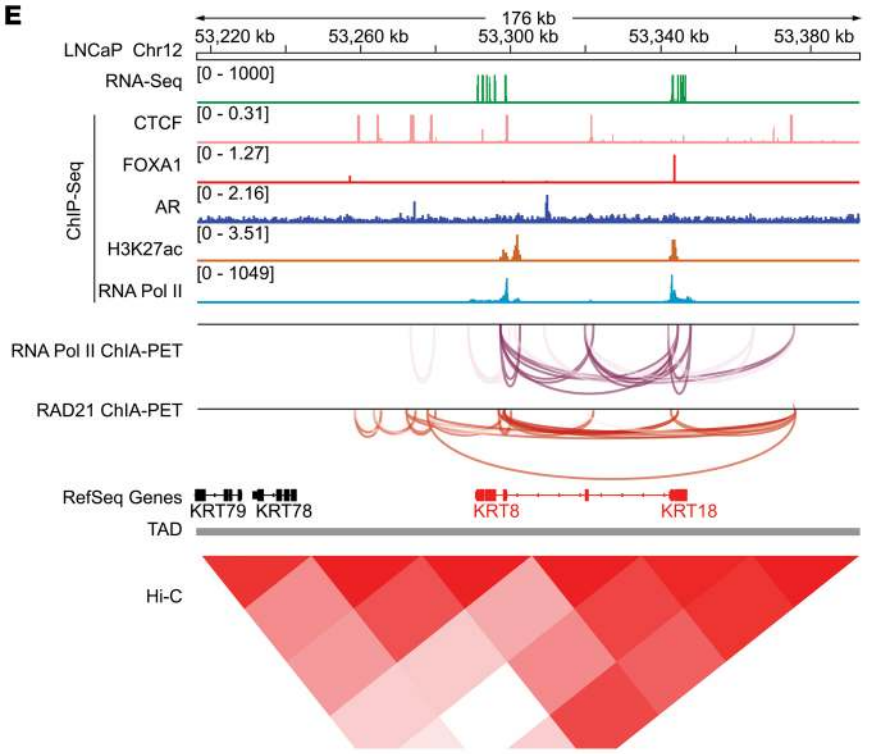
Binding peaks represent the fundamental basis of ChIA-PET analysis; peaks that interact with other peaks are called anchor peaks, whereas noninteracting ones are termed nonanchor peaks. Thus, a single RNA Pol II ChIA-PET experiment can provide both 2D data on RNA Pol II binding peaks, as well as 3D data on RNA Pol II binding anchor peaks and nonanchor peaks. Paired-end tags (PETs) are the units used for measuring the interactions between a pair of anchor peaks, with more PETs between anchors signifying stronger interactions. We identified thousands of RNA Pol II–associated chromatin interactions spanning all chromosomes in the cell lines analyzed. Saturation analysis indicated that we had sufficient sequencing depth to detect high-confidence chromatin interactions supported by 3 or more PETs. For example, by using only 50% of the sequencing reads, we were able to rediscover 64%–80% of chromatin interactions supported by 3 or more PETs and 76%–90% of chromatin interactions supported by 4 or more PETs (Supplemental Figure 1). Importantly, the saturation analysis indicated that our data were comparable to the gold-standard ENCODE ChIA-PET data sets.

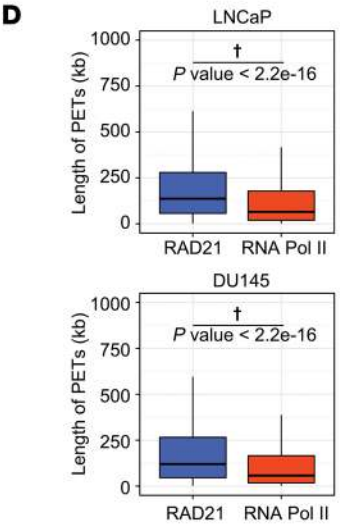
We also estimated the length distribution of RNA Pol II–associated chromatin interactions by setting multiple PET cutoffs for various interaction strengths (Supplemental Figure 1B). Overall, the bulk of chromatin interactions ranged from 3 kb to 1 Mb in all the cell lines analyzed. Interactions greater than 1 Mb were relatively rare, especially at higher PET cutoffs. We extended this association by comparing interactions between cell lines (Supplemental Figure 2A). As expected, we found a greater overlap for interactions of less than 1 Mb between any 2 or more cell lines at all measured stringencies and PET cutoffs. We observed maximal overlap in RNA Pol II–associated chromatin interactions between LNCaP and VCaP cells. This is probably because both LNCaP and VCaP cells are AR positive and are dependent on androgen signaling for survival. These results suggest that an overlap in master transcription factors is associated with a parallel overlap in RNA Pol II–associated chromatin interactions. By partitioning the genome into 3 groups — promoters, gene bodies and other regions — we observed thousands of interactions within and between these groups (Supplemental Figure 1C, top). We probed the chromatin interaction patterns of active enhancers by mapping acetylated histone H3 at lysine 27–marked (H3K27ac-marked) sites onto the RNA Pol II ChIA-PET anchor sites. In addition to interactions between enhancers and promoters, we also discovered thousands of enhancer-enhancer and promoter-promoter interactions (Supplemental Figure 1C, bottom), implying that a complex network of chromatin interactions coordinates transcriptional regulation in these cells. We leveraged our ChIA-PET data to pair enhancers and their potential target genes in the 4 cell lines (Supplemental Table 3). We observed that an increase in the RNA Pol II–associated


A
2,800kb
100,200kb
100,600kb
101,000kb
101,400kb
101,800kb
102,200kb
102,600kb
103,000kb
LNCaP Chr1
DNAse-Seq
[0 - 20]
ChIP-Seq
H3K4me3
[0 - 120]
CTCF
[0 - 18.85]
H3K27ac
[0 - 2.6]
RNA-Seq
[0 - 100]
TAD
RNA Pol II ChIA-PET
RAD21 ChIA-PET
Hi-C
Active
Repressed


B
All Peaks
Interaction Peaks
RWPE-1
27354
34020
51441
9572
39645
21729
LNCaP
29074
33649
28914
10323
39710
23013
VCaP
37893
45085
15571
6513
50989
31989
DU145
21916
28448
29649
1968
30269
20095
H3K27ac
RNA Pol II
H3K27ac
RNA Pol II


C
All Peaks
Interaction Peaks
LNCaP
40170
36122
29363
42183
34109
21185
DU145
7216
28936
61338
9413
26739
24587
CTCF
RAD21
CTCF
RAD21


E
176 kb
53,220 kb
53,260 kb
53,300 kb
53,340 kb
53,380 kb
LNCaP Chr12
RNA-Seq
[0 - 1000]
ChIP-Seq
CTCF
[0 - 0.31]
FOXA1
[0 - 1.27]
AR
[0 - 2.16]
H3K27ac
[0 - 3.51]
RNA Pol II
[0 - 1049]
RNA Pol II ChIA-PET
RAD21 ChIA-PET
RefSeq Genes
KRT79
KRT78
KRT8
KRT18
TAD
Hi-C


D
LNCaP
†
P value < 2.2e-16
Length of PETs (kb)
1000
750
500
250
0
RAD21
RNA Pol II
DU145
†
P value < 2.2e-16
Length of PETs (kb)
1000
750
500
250
0
RAD21
RNA Pol II

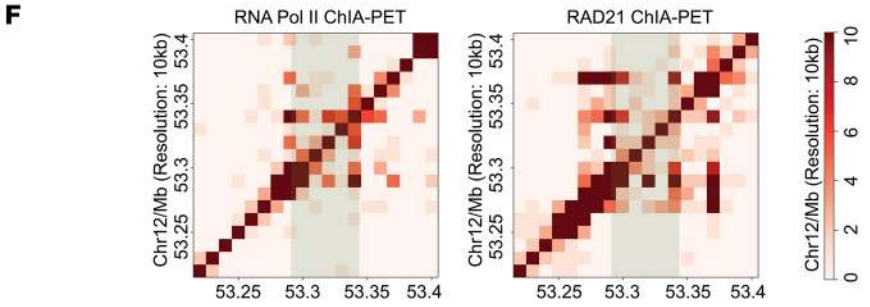

F
RNA Pol II ChIA-PET
RAD21 ChIA-PET
Chr12/Mb (Resolution: 10kb)
53.4
53.35
53.3
53.25
53.25
53.3
53.35
53.4
Chr12/Mb (Resolution: 10kb)
0
2
4
6
8
10

Figure 2. Integrative analysis of chromatin interactions. (**A**) Integrated genome view representing DNase-Seq; ChIP-Seq for CTCF, H3K4me3, and H3K27ac; RNA-Seq; ChIA-PET for RNA Pol II and RAD21; and Hi-C data from the LNCaP cell line for a representative region with both active and repressed genes. Chr1, chromosome 1. (**B**) Analysis of overlap between H3K27ac ChIP-Seq and RNA Pol II ChIA-PET data for 4 cell lines. Left: overlap in the number of total peaks; right: overlap in the number of peaks with intrachromosomal interactions. (**C**) Analysis of overlap between CTCF ChIP-Seq and RAD21 ChIA-PET data for LNCaP and DU145 cells. Left: overlap in the number of total peaks; right: overlap in the number of anchor peaks with intrachromosomal interactions. (**D**) Comparison of PET lengths between RNA Pol II ChIA-PET and RAD21 ChIA-PET data from LNCaP and DU145 cells. Box plot represents the median and the 25% and 75% quantiles, with lines at 1.5 times the IQR. $^{\dagger}P < 2.2 \times 10^{-16}$, by Kolmogorov-Smirnov test. (**E**) Integrated genome view of the *KRT8-KRT18* gene cluster and its neighborhood in LNCaP cells. The data tracks represent RNA-Seq; ChIP-Seq for CTCF, FOXA1, AR, H3K27ac, and RNA Pol II; ChIA-PET for RNA Pol II and RAD21; and Hi-C data for the LNCaP cell line. (**F**) ChIA-PET contact heatmap representing RNA Pol II– and RAD21-associated chromatin interactions for the *KRT8-KRT18* gene cluster and neighborhood regions. The *KRT8-KRT18* gene cluster is shown in light green.

chromatin interaction strength between enhancers and promoters was associated with a decrease in the 2D distances between the interacting regions (Supplemental Figure 1D). In comparison with the benign RWPE-1 cells, all the PCa cell lines displayed an overall increase in enhancer usage per gene as well as an accompanying increase in gene targets per enhancer (Figure 1B and Supplemental Figure 1, E–G). Next, we determined the relationship between enhancers per gene and gene expression and observed a very weak correlation (Supplemental Figure 1H), implying that enhancers per gene is not a useful predictor of gene expression.

We observed that RNA Pol II interaction strength, as measured by the number of PETs in gene promoters, positively correlated with mRNA abundance in all 4 cell lines (Figure 1C) and was a much better predictor of gene expression than enhancers per gene. The interaction strength between anchors followed a negative binomial (NB) distribution (Supplemental Figure 1I). By applying the difference of NB distributions (DNBs), a new statistical model for comparing 2 different negative binomial distributions, we identified thousands of significantly gained and lost interactions across all the possible 6 pair-wise comparisons among the 4 cell lines (Figure 1D and Supplemental Table 4). Next, we evaluated the relationship between changes in gene expression and changes in promoter-associated chromatin interactions between the benign RWPE-1 cells and the 3 PCa cell lines. We initiated the analyses by examining all genes that were upregulated or downregulated by 2-fold or more, followed by additional analyses by filtering the top 80%, 60%, 40%, 20%, and 10% upregulated or downregulated genes from the starting list. Remarkably, as we traversed from all genes to the top 10% of genes, we found that the correlation also increased (Figure 1E). We observed this trend for all our comparisons. This result indicates that changes in RNA Pol II–associated chromatin interactions are likely to be drivers of gene expression changes. This is particularly the case for the top dysregulated genes.

To extend this correlation further, we rank-ordered all the genes on the basis of the number of PETs in their promoters (Figure 1F). Genes like *FOS*, *JUN*, and *KLF6* were highly expressed in all 4 cell lines and had more than 25 independent PETS in the promoter. The luminal-like LNCaP and VCaP cells had more PETs in the luminal marker gene promoter *KRT8* than in the basal marker gene promoter *KRT5*. We observed the opposite trend in the more basal-like RWPE-1 cells. Notably, the *AR* gene promoter had approximately 50 independent PETs in VCaP cells. The *EZH2* gene promoter had a greater number of PETS in PCa cells as compared with benign RWPE-1 cells. Next, we examined PCa enriched pathways and gene ontology (GO) terms using gene lists obtained from differential transcript abundance and differential chromatin interactions. Although both these analyses indicated enrichment of common categories such as metabolic processes, cellular process, and cell cycle–related processes, among others, we also identified categories unique to differential chromatin interactions such as chromatin organization and translation (Supplemental Figure 2, B and C).

We also investigated the DNA sequence features of promoters and enhancers (Supplemental Figure 1J). In general, promoters exhibited less variability than did enhancers. ETS, CCAAT, and YY1 motifs were commonly observed in the promoters of all the cell lines. Enhancers showed considerable diversity in terms of DNA sequence features. ETS and CTCF motifs were commonly observed in the enhancers of all the cell lines. The FOXA1 forkhead motif and HOXB13 homeobox motif were unique to LNCaP and VCaP cells. The leucine zipper motif bZIP was the most significantly enriched enhancer motif in RWPE-1 and DU145 cells.

Integrative analysis of RNA Pol II–associated chromatin interactions. To better understand the regulatory concepts associated with transcriptional control, we integrated our RNA Pol II ChIA-PET data with multiple additional data sets (Supplemental Table 1 and ref. 13). We examined a representative DNA region from LNCaP cells with both transcriptionally active and inactive genes by overlaying our RNA Pol II ChIA-PET data with RAD21 ChIA-PET, Hi-C, RNA-Seq, DNase-Seq, and ChIP-Seq data for CTCF (TAD and sub-TAD boundary element mark), H3K4me3 (active promoter mark), and H3K27ac (active enhancer mark) (Supplemental Tables 1 and 2). Our RNA Pol II ChIA-PET data were consistent with the architectural features of these independent data sets (Figure 2A). Both active and repressed regions showed substantial chromatin contacts as visualized by the Hi-C track. However, the nature of the interactions within active and repressed regions revealed characteristic differences. Consistent with RNA-Seq, DNAse-Seq, H3K4me3, and H3K27ac signals, we observed that both RNA Pol II– and RAD21-associated chromatin interactions were enriched in the active regions. The prominent RNA Pol II interactions almost always involved H3K4me3 and/or H3K27ac marks. The RAD21 interactions were primarily associated with CTCF-bound regions. Next, we compared the RNA Pol II and RAD21 ChIA-PET data sets. The genome-wide total RNA Pol II peaks as well as the interaction anchor RNA Pol II peaks overlapped with H3K27ac peaks with statistical significance ($P < 0.001$, hypergeometric test), but not with CTCF or RAD21 peaks (Figure 2B and Supplemental Table 5). The genome-wide total RAD21 peaks as well as the interaction anchor RAD21 peaks overlapped with CTCF peaks with statistical significance ($P < 0.001$, hypergeometric test), but not with RNA Pol II or H3K27ac peaks (Figure 2C and Supplemental Table 5). The length of RAD21-associated chromatin interactions was

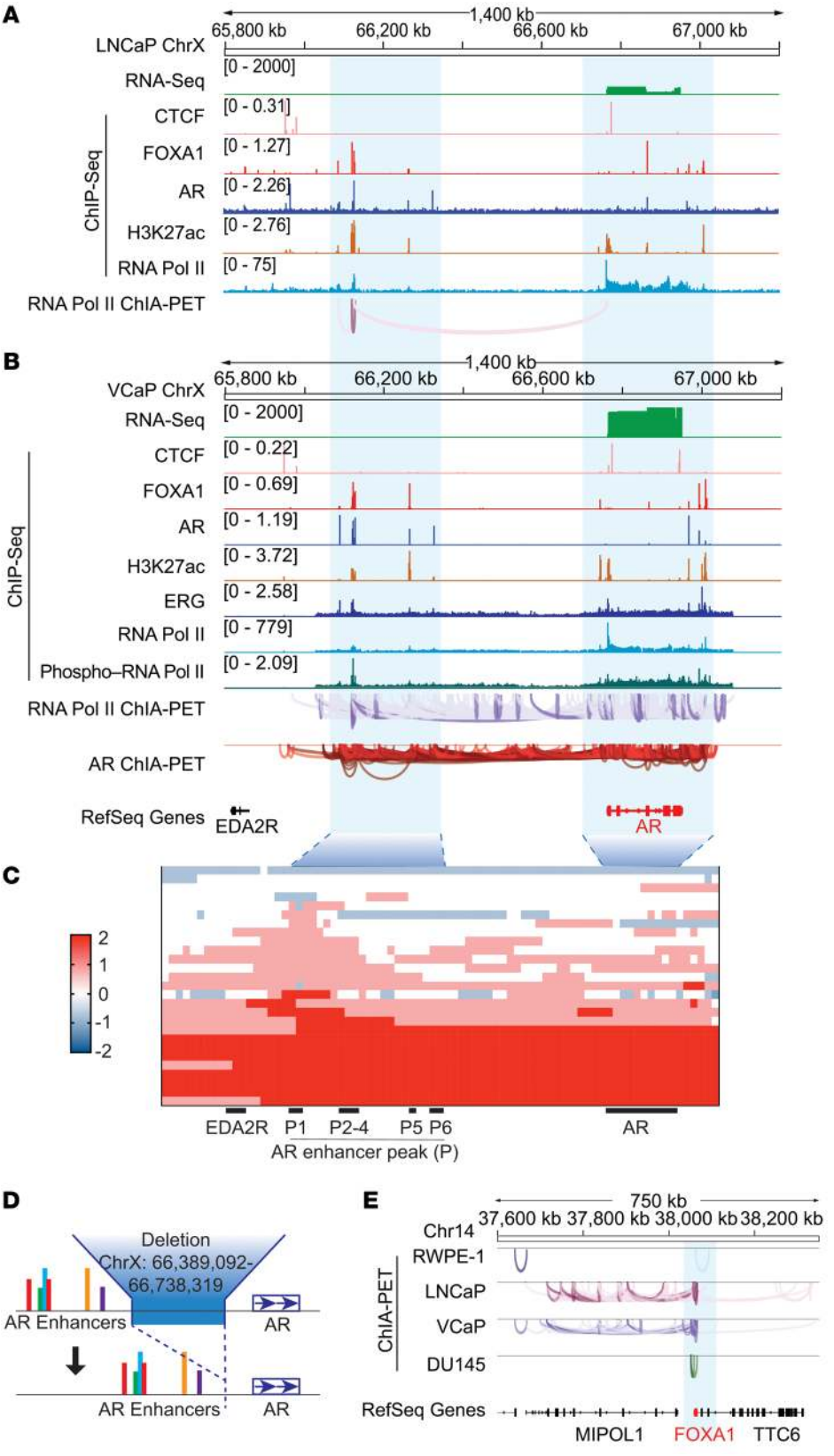


Figure 3. Transcriptional regulation of AR and FOXA1 loci. (**A** and **B**) Integrated genome view of the *AR* gene and its adjacent regions from –1400 kb to +400 kb in LNCaP and VCaP cells. RNA Pol II ChIA-PET, RNA-Seq and CTCF, FOXA1, AR, H3K27ac, and RNA Pol II ChIP-Seq data are shown. In addition, AR ChIA-PET, phospho–RNA Pol II, and ERG ChIP-Seq are shown for VCaP cells. The *AR* gene and upstream regions are highlighted in light blue. (**C**) Summary of the copy number aberrations associated with the *AR* and its enhancers. Heatmap shows shows the aCGH high-density probes for 27 patients with mCRPC. Gains are depicted in pink and losses in light blue, whereas amplifications are shown in red and deep deletions in dark blue. Each column is a probe in the aCGH platform, and each row represents a sample. Probes that cover the EDA2R, AR, and regions of the enhancer peaks are shown. (**D**) Schematic representation of a deletion between the *AR* gene and its enhancers. (**E**) Comparison of RNA Pol II–associated chromatin interactions at the FOXA1 locus and its adjacent regions in the 4 cell lines.

significantly longer than that for RNA Pol II–associated interactions (Figure 2D). On the basis of these analyses, chromatin interactions can be hierarchically classified into 2 tiers: the outer RAD21-associated interactions at CTCF-bound sites and the inner RNA Pol II–associated interactions at H3K27ac-marked sites. Given that RAD21 is a cohesin component, our results are consistent with the concept that RAD21 interactions represent boundaries of insulated neighborhoods or TADs. Functionally, the outer RAD21 interactions provide insulation for the inner RNA Pol II–associated chromatin interactions. Conceptually, although lineage-restricted enhancers trigger the transcriptional activation of their cognate target genes, this insulation draws the boundary and prevents these enhancers from activating genes outside the boundary. We present 2 examples to illustrate this point in the context of prostate epithelial lineage identity.

KRT8-KRT18 gene cluster expression is limited to prostate luminal epithelial cells and luminal PCa represented by LNCaP and VCaP cells. The *KRT8* and *KRT18* genes display divergent transcription from a bidirectional promoter/enhancer element marked by FOXA1 and H3K27ac (Figure 2E). We observed that the inner, short-range RNA Pol II–associated chromatin interactions were flanked by outer, long-range RAD21-associated chromatin interactions at sites marked by CTCF binding (Figure 2, E and F). On the basis of these data, we hypothesize that although lineage-restricted transcription factors, such as FOXA1, specify the activation of lineage-restricted enhancers, the RAD21 interactions specify the gene targets of these enhancers by creating protective moats. Although several type II keratin genes are located in this subchromosomal region, our study offers an explanation for why the expression of *KRT8* and *KRT18* genes is unique to the prostate luminal epithelial lineage, which is that RAD21 interactions insulate these 2 genes and their enhancers. The same concept can be used to explain the lineage-specific regulation of *HOXB13* and several other genes (Supplemental Figure 3).

Transcriptional regulation of AR and FOXA1 loci. Having established genomic concepts that underlie transcriptional regulation, we next examined the mechanisms associated with dysregulated transcription in PCa. The AR is the central effector of the androgen signaling pathway, which is frequently dysregulated in PCa. Recent studies have reported the identification and amplification of an *AR* enhancer in mCRPC (4, 14, 15); however, unbiased maps of RNA Pol II–associated chromatin interactions at the *AR* locus are not available to date. We report the identification of an active *AR* enhancer cluster located approximately 700 kb upstream of the *AR* promoter (Figure 3, A and B). This H3K27ac-marked enhancer cluster is occupied by the AR and FOXA1 and interacted with the *AR* promoter in LNCaP and VCaP cells. The *AR* gene and its enhancers are amplified in VCaP cells, thereby resulting in its overexpression (16). Consistent with *AR* locus amplification, VCaP cells displayed an increase in the total number of chromatin interaction modules comprising the *AR* gene and its distal enhancer compared with LNCaP cells. These results suggest that *AR* enhancers play a significant role in driving *AR* expression in both the absence and presence of *AR* locus amplification. By integrating our RNA Pol II ChIA-PET data with the AR ChIA-PET data set for VCaP cells (5), we observed that the complex chromatin interactions in the AR locus were in part mediated by AR binding (Figure 3B). VCaP cells express ERG, a transcription factor that is formed by genomic rearrangements involving *TMPRSS2* and *ERG* genes (1, 17). Interestingly, we also observed ERG binding to the *AR* enhancer cluster in VCaP cells. Thus, the enhancer cluster functions as a landing site for master regulators, which govern *AR* transcription in a combinatorial manner. Consistent with the role of bromodomain and extraterminal (BET) family proteins in controlling enhancer-driven gene expression (18), we found that treatment with a BET bromodomain inhibitor (BETi) reduced AR expression in a dose-dependent manner (Supplemental Figure 4, A and B).

Given that targeting the AR signaling axis remains the mainstay of therapy for metastatic PCa, we addressed the relevance of *AR* enhancers in a clinical setting. We conducted high-density array comparative genomic hybridization (aCGH) to evaluate the copy number status of *AR* and its enhancers in metastatic tumor biopsies from 27 patients with mCRPC (Figure 3C and Supplemental Table 6). We observed (a) 20 patients with both *AR* gene amplification/copy number gain as well as *AR* enhancer amplification/copy number gain; (b) 1 patient with *AR* gene amplification/copy number gain, but no accompanying changes in the *AR* enhancer; and (c) 4 patients with *AR* enhancer amplification/copy number gain, but no accompanying changes in the *AR* gene. Two patients in this cohort did not have amplification or copy number gain in the *AR* gene or the *AR* enhancer. In summary, *AR* enhancer alterations were more prevalent than *AR* gene alterations in our mCRPC cohort.

We validated these results by conducting droplet digital PCR (ddPCR) analysis of an independent, clinically well-annotated mCRPC cohort ($n = 46$ mCRPC patients) (ref. 19 and Supplemental Figure 4C). Again, we observed that amplification of *AR* enhancers correlated with *AR* gene amplification (Supplemental Figure 4D) and that copy number gain of *AR* enhancers correlated with *AR* gene copy number gain (Supplemental Figure 4E). To establish the relationship between *AR* gene copy number and mRNA abundance, we reanalyzed the data from 122 patients with mCRPC in the SU2C cohort (3). As hypothesized, we found that *AR* gene copy number and mRNA abundance were significantly correlated (Supplemental Figure 4F). Overall, our results point to an essential role for *AR* enhancers in promoting *AR* transcription in mCRPC.

We suggest that the extra copies of genomic DNA from the *AR* locus, formed by the amplification process, are further subjected to iterative rounds of mutagenesis and epigenetic alterations and ultimately filtered by treatment-induced evolutionary selection. For example, WGS analysis of a representative mCRPC specimen with amplification of *AR* and its enhancers resulted in the discovery of a subclonal deletion of approximately 350 kb that brings the entire enhancer cluster closer to the *AR* gene (Figure 3D and Supplemental Figure 4G). This deletion is supported by 139 independent unique reads (of 7775 reads) and is localized within the AR-amplified (20-copy) region. Given the negative correlation between 2D distances separating the interacting regions and RNA Pol II–associated chromatin interaction strength (Supplemental Figure 1D), as well as the positive correlation between RNA Pol II–associated chromatin interaction strength and transcript abundance (Figure 1, C and E), we suggest that such deletions are likely to further influence *AR* transcription. Intriguingly, we observed enhanced CTCF binding to the 3′ end of the *AR* gene in VCaP cells

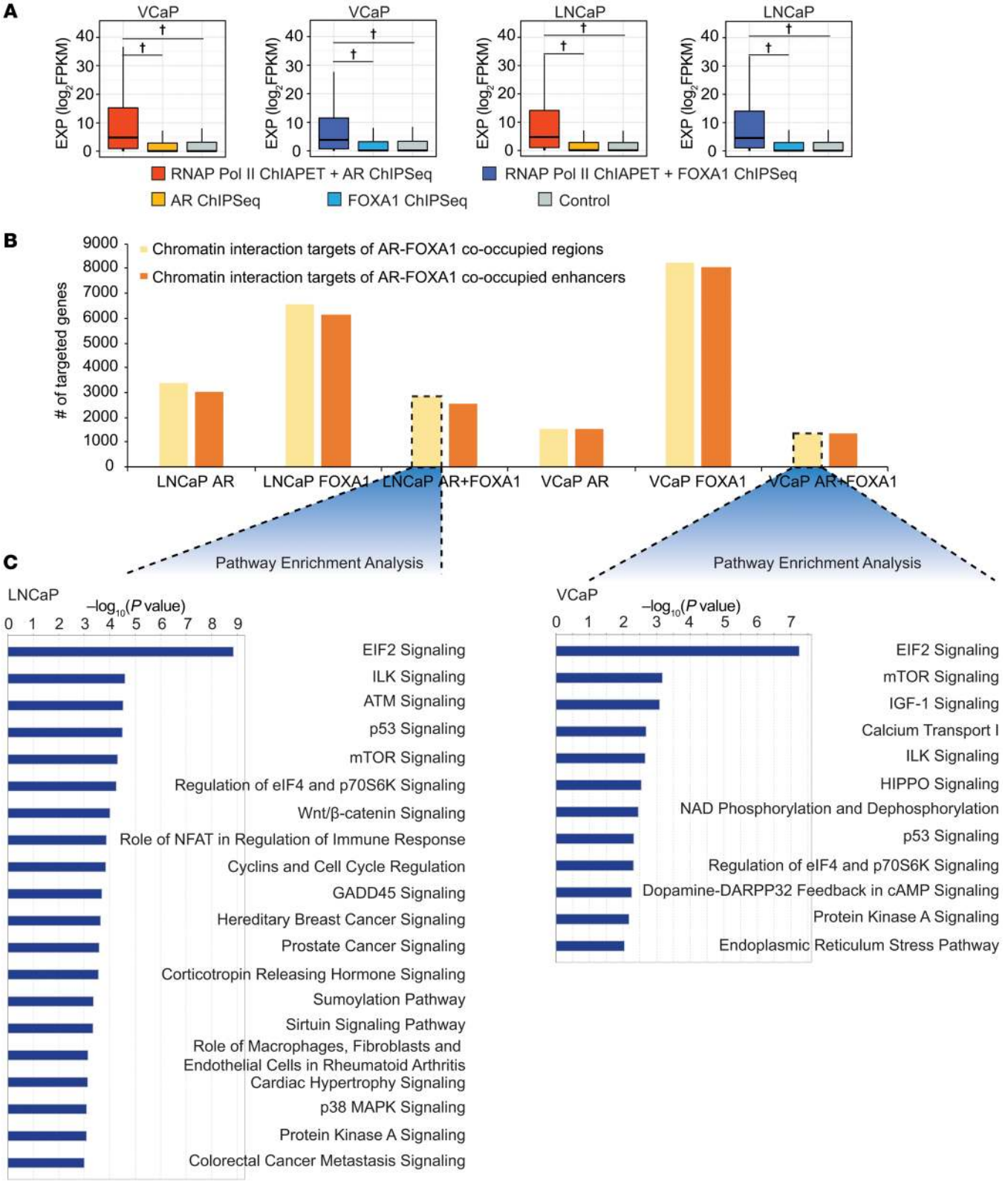


Figure 4. Chromatin interaction–associated transcriptional targets of the AR and FOXA1. (**A**) The expression of AR/FOXA1 target genes discovered by integrating RNA Pol II ChIA-PET with AR/FOXA1 ChIP-Seq was compared with the expression of genes that were nearest to AR/FOXA1 binding peaks according to ChIP-Seq data and randomly selected control genes in VCaP and LNCaP cells. The *y* axis represents expression levels, measured as FPKM transformed by $\log_2$. The box plots represent the median and 25% and 75% quantiles, with lines at 1.5 times the IQR. $^{\dagger}P < 2.2 \times 10^{-16}$, by Kolmogorov-Smirnov test). (**B**) Gene promoters that interact with AR, FOXA1, and AR-FOXA1 co-occupied regions in the RNA Pol II ChIA-PET data sets are shown in yellow. Gene promoters that interact with AR, FOXA1, and AR-FOXA1 co-occupied enhancers in the RNA Pol II ChIA-PET data sets are shown in orange. (**C**) Pathway analysis for gene promoters that interact with AR-FOXA1 co-occupied regions in LNCaP and VCaP cells in the RNA Pol II ChIA-PET data sets.

that had *AR* amplification, but not in LNCaP cells (Figure 3, A and B). VCaP cells also express AR variant proteins that lack the C-terminus ligand–binding domain (Supplemental Figure 4B). CTCF binding to a subset of *AR* alleles in the context of *AR* amplification could activate neoinsulator elements that would in turn interfere with the RNA Pol II activity to generate truncated *AR* variants. Thus, *AR* amplification can set the stage for multiple downstream regulatory processes to drive the lethal phenotype.

We also examined the transcriptional architecture of the gene encoding the pioneer factor FOXA1, which collaborates with the AR to activate enhancers (2). We noticed that multiple enhancers located within the adjacent *MIPOL1* gene interacted with the *FOXA1* gene. These enhancers were cobound by the AR and FOXA1 (Figure 3E and Supplemental Figure 5). ERG also bound to this enhancer cluster in VCaP cells. Thus, we conclude that master regulators like the AR and FOXA1 autoregulate their own expression by binding to their enhancers and that the gene fusion product ERG adds an additional layer of complexity to the regulation of and by the regulators.

Transcriptional regulation by the AR and FOXA1. The overall expression of genes nearest to the AR or FOXA1 binding peaks was not very different from that of randomly selected control genes (Figure 4A). Given our observation that multiple genes and enhancers coordinately interacted to regulate transcription, we hypothesized that chromatin-bound AR and FOXA1 are likely to regulate genes that are located beyond their nearest neighbors. For example, the *AR* enhancers are closer to the *EDA2R* gene than to the *AR* gene. By overlapping AR and FOXA1 binding peaks with RNA Pol II ChIA-PET anchor peaks, we created virtual chromatin contact maps of AR and FOXA1 occupancy and traced the target genes (Supplemental Table 7). Importantly, the expression of target genes discovered by our new approach was significantly higher than the expression of the nearest neighboring genes as well as control genes in all our comparisons (Figure 4A). These studies highlight the utility of mapping RNA Pol II–associated chromatin interactions to decipher the master transcription factor–enhancer–target gene regulatory networks.

To garner insights into the functional processes governed by the combinatorial binding of master transcription factors, we queried the RNA Pol II ChIA-PET data for chromatin interactions between AR-FOXA1 co-occupied regions and gene promoters (Figure 4B). This was followed by pathway analysis of these AR-FOXA1 chromatin interaction target genes. We were surprised to discover that the genes involved in EIF2 signaling were maximally enriched in both LNCaP and VCaP cells (Figure 4C). The EIF2 pathway regulates translation initiation. Our results are consistent with other studies reporting AR-mediated transcriptional control of translation initiation (20). We now show that translation initiation is mediated by chromatin interactions involving the AR and FOXA1. More generally, our results indicate that master transcription factors cooperatively regulate the 3D genome organization to control the transcriptome, and eventually the proteome, to establish lineage identity.

AR binding, histone acetylation, and regulation of the kallikrein gene cluster and ZBTB16. Prostate-specific antigen (PSA), the protein product of the kallikrein 3 (*KLK3*) gene, is one of the most commonly used PCa screening biomarkers (21). It is well established that the AR upregulates the expression of *KLK3*. The *KLK3* gene resides in a gene cluster that also includes multiple members of the KLK family. We leveraged our ChIA-PET data to create a portrait of transcriptional regulation at the KLK gene cluster (Figure 5A). We identified multiple H3K27ac-marked active enhancers surrounding *KLK3* and neighboring KLK genes in the AR-positive LNCaP and VCaP cells, which express KLK3. These enhancer locations were devoid of the H3K27ac mark in the AR-negative RWPE-1 and DU145 cells which do not express KLK3. The active enhancers in the KLK cluster were occupied by the AR and FOXA1 (Supplemental Figure 6, A and B). Remarkably, every H3K27ac-marked enhancer in the KLK cluster interacted with an upstream H3K27ac-marked distal enhancer located in the gene *C19ORF48*. The distal enhancer was not bound by the AR or FOXA1 but was active in all 4 of the prostate cell lines analyzed. In the AR-negative cells, this enhancer interacted with other active enhancers away from the *KLK3* gene. Analysis of this subchromosomal region using H3K27ac ChIP-Seq data sets from approximately 100 cell types indicated that the distal enhancer was constitutively active in all the cell lineages analyzed (Supplemental Figure 7). However, the activity of enhancers flanking the distal constitutive enhancer was highly cell-type specific and provided epigenetic fingerprints for every cell type. It is likely that the flanking enhancers were activated by binding of tissue-specific transcription factors to their cognate recognition sites. Thus, we conclude that complex chromatin interactions between constitutive enhancers and tissue-specific enhancers specify the expression of lineage-restricted genes like *KLK3*.

ZBTB16 is another classic AR target gene that is misregulated in mCRPC (3, 6). We observed that *ZBTB16* was expressed in AR-positive LNCaP cells but was transcriptionally silent in DU145 cells, which do not express the AR. In LNCaP cells, the binding of the AR and FOXA1 was associated with H3K27ac-marked intragenic enhancers. RNA Pol II ChIA-PET indicated interactions between these enhancers and the promoter of *ZBTB16* (Figure 5B). These RNA Pol II interactions were insulated by outer chromatin interactions involving RAD21 at sites comarked by CTCF binding. Enhancers having CTCF occupancy appeared to draw the RNA Pol II–associated interactions toward the boundaries of the insulated neighborhood. In DU145 cells, the absence of a H3K27ac mark indicated that the intragenic enhancers were inactive, presumably because these cells do not express the AR (Figure 5C). As a consequence, there were no RNA Pol II interactions and the gene was silent. However, RAD21 interactions did exist at identical sites comarked by CTCF binding. Although AR and FOXA1 binding, enhancer activation, and the associated RNA Pol II interactions upregulated lineage-restricted genes like *ZBTB16*, cohesin-associated chromatin interactions appeared to function as moats to prevent *ZBTB16* enhancers from activating other genes outside the neighborhood.

Transcriptional architecture of MYC and EZH2 loci. *MYC* is an oncogene that is misregulated in multiple cancers, including PCa. Tandem duplications involving the *MYC* neighborhood are frequently observed in mCRPC (4, 15). We queried RNA Pol II ChIA-PET data sets to examine the landscape of chromatin interactions surrounding the *MYC* gene. We observed extensive RNA Pol II–associated chromatin interactions within the 1.8-Mb

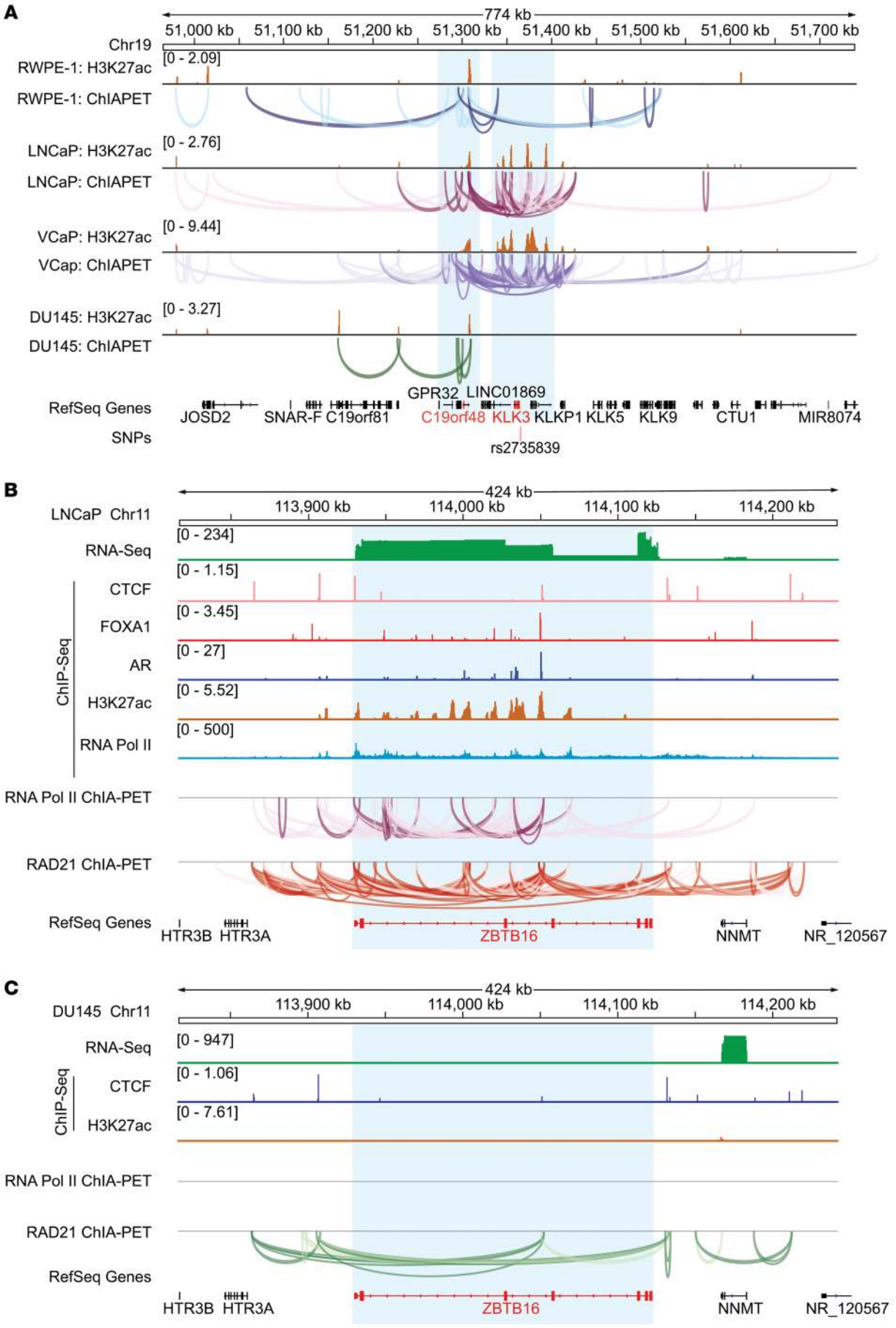

A
774 kb
51,000 kb
51,100 kb
51,200 kb
51,300 kb
51,400 kb
51,500 kb
51,600 kb
51,700 kb
Chr19
RWPE-1: H3K27ac
[0 - 2.09]
RWPE-1: ChIAPET
LNCaP: H3K27ac
[0 - 2.76]
LNCaP: ChIAPET
VCaP: H3K27ac
[0 - 9.44]
VCap: ChIAPET
DU145: H3K27ac
[0 - 3.27]
DU145: ChIAPET
GPR32 LINC01869
RefSeq Genes
JOSD2
SNAR-F C19orf81
C19orf48 KLK3 KLKP1 KLK5 KLK9
CTU1
MIR8074
SNPs
rs2735839
B
424 kb
113,900 kb
114,000 kb
114,100 kb
114,200 kb
LNCaP Chr11
RNA-Seq
[0 - 234]
ChIP-Seq
CTCF
[0 - 1.15]
FOXA1
[0 - 3.45]
AR
[0 - 27]
H3K27ac
[0 - 5.52]
RNA Pol II
[0 - 500]
RNA Pol II ChIA-PET
RAD21 ChIA-PET
RefSeq Genes
HTR3B HTR3A
ZBTB16
NNMT
NR_120567
C
424 kb
113,900 kb
114,000 kb
114,100 kb
114,200 kb
DU145 Chr11
RNA-Seq
[0 - 947]
ChIP-Seq
CTCF
[0 - 1.06]
H3K27ac
[0 - 7.61]
RNA Pol II ChIA-PET
RAD21 ChIA-PET
RefSeq Genes
HTR3B HTR3A
ZBTB16
NNMT
NR_120567

Figure 5. Transcriptional regulation the *KLK* gene cluster and *ZBTB16*. (**A**) Comparison of RNA Pol II–associated chromatin interactions and H3K27ac ChIP-Seq signals in the *KLK3* gene and its neighborhood regions from –350 Kb to +350 Kb in 4 cell lines. (**B** and **C**) Transcriptional regulation of the *ZBTB16* gene in (**B**) LNCaP and (**C**) DU145 cells.

subchromosomal region harboring the *MYC* oncogene (Figure 6A). The anchor peaks were characterized by the H3K27ac mark, which is indicative of transcriptional regulation by enhancers, and were highly cell-type specific. In RWPE-1, LNCaP, and VCaP cells, the interactions spanned both the 5′ upstream and 3′ downstream regions of *MYC*, whereas in DU145 cells, the interactions were comparatively restricted to the 3′ downstream region of *MYC*. Interestingly, in LNCaP and VCaP cells, the RNA Pol II–associated chromatin interactions surrounding *MYC* exhibited substantial overlap and also displayed AR and FOXA1 occupancy (Supplemental Figure 8, A and B). Thus, it is conceivable that binding of the AR and FOXA1 facilitates the site-specific recruitment of coactivators, resulting in the formation of H3K27ac mark and enhancer activation. These activated enhancers in turn interact with the *MYC* promoter and transmit the signal to modulate RNA Pol II activity. Consistent with our study, recent Hi-ChIP studies of H3K27ac in leukemia and lymphoma cells and of cohesin in HCT-116 cells reported that the enhancer landscape of the *MYC* locus is cell-type specific (22, 23). Our analysis of the active enhancer mark H3K27ac in approximately 100 cell types revealed distinct patterns of enhancer activity on both sides of the *MYC* gene (Supplemental Figure 8E). Given the unavailability of RNA Pol II ChIA-PET data for most of these cell lines, we assigned the maximum enhancer peak score in the gene neighborhood as a surrogate for enhancer activity. We observed a positive correlation between enhancer activity and MYC transcript abundance (Supplemental Figure 8F). As *MYC* is expressed in most human cancers, we propose a working model that depicts cell-type–specific transcription factors specifying enhancer activation at their cognate binding sites in the neighboring subchromosomal region and resulting in transcriptional regulation of *MYC* via chromatin interactions.

Another interesting feature of the *MYC* locus is that the region is devoid of protein coding genes but has an abundance of noncoding RNAs. We noticed that several *MYC* interacting anchor regions represented genes that encode noncoding RNAs. For example, the long, intergenic, noncoding RNA PCAT-1, which is overexpressed in PCa (24), was associated with AR and FOXA1 binding and was enriched with the H2K27ac mark (Supplemental Figure 8, A and B). PCAT-1 and its neighboring enhancers directly interacted with the *MYC* promoter in VCaP cells. PCAT-1 and its neighboring enhancers showed indirect interaction with the *MYC* promoter in LNCaP cells; these interactions converged at an intervening DNA region that was flanked by germline PCa risk–predisposing SNPs. The chromosome 8q24 harbors risk alleles for multiple cancers, including PCa (25). Several PCa risk SNPs from GWAS ($n = 16$) lay in the region upstream of *MYC*, 7 of which were located within enhancer elements in 1 or more of the cell types analyzed (Figure 6B, left). Four of these 7 SNP loci were also associated with chromatin interactions (Figure 6B, right). We conclude that many PCa risk loci in the *MYC* subchromosomal region were located on or near enhancers and were associated with RNA Pol II–associated chromatin interactions.

The histone methyl transferase *EZH2* is overexpressed in several solid tumors, including PCa, however, the mechanisms underlying its transcriptional regulation are not completely understood (26). By integrative analysis of RNA Pol II ChIA-PET data, we identified chromatin interaction involving a H3K27ac-marked enhancer and the EZH2 promoter in RWPE-1, LNCaP, VCaP, and DU145 cells (Figure 6, C and D, and Supplemental Figure 8, C and D). The strength of this interaction was higher in VCaP and DU145 cells. Conversely, the *EZH2* promoter interacted with a single enhancer in benign RWPE-1 cells, but with multiple additional distant enhancers in the cancerous LNCaP, VCaP, and DU145 cells. These enhancers in turn interacted with additional modules of enhancers, promoters, and gene bodies, resulting in complex transcriptional hubs. Thus, the strength of individual enhancer-promoter interactions and the modular assembly of interacting enhancers govern *EZH2* regulation in PCa cells. We analyzed the enhancer mark H3K27ac in the *EZH2* neighborhood in a panel of cell lines (Supplemental Figure 8G). Again, by assigning the maximum enhancer peak score in the gene neighborhood as a surrogate for enhancer activity, we observed a positive correlation between enhancer activity and *EZH2* transcript abundance (Supplemental Figure 8H).

Germline PCa risk variants and transcriptional regulation. In most GWAS, disease-associated SNPs map to noncoding regions, creating bottlenecks in interpreting their functional role. We leveraged our RNA Pol II ChIA-PET data sets to interpret the function of PCa risk SNPs identified from GWAS. We analyzed a curated list of 122 PCa risk SNPs from multiple GWAS (27) to assess how risk alleles impact transcription control. We hypothesized that many risk alleles would be located in enhancers and/or promoters and hence influence the transcriptional regulation of target genes by altering chromatin interactions. To understand the functional relevance of PCa risk loci in the context of genome architecture, we applied hypothesis testing using both simulation experiments and enrichment analysis.

We computationally partitioned the genome into several thousand bins and conducted simulations to test the observed versus expected values for (a) RNA Pol II peaks that overlap with PCa risk SNPs and (b) chromatin interaction–associated RNA Pol II peaks that overlap with PCa risk SNPs. The observed overlap between RNA Pol II peaks and PCa risk SNPs in RWPE-1, LNCaP, and VCaP cells was significantly greater than the expected values derived from 10,000 simulations (Figure 7A). We also observed the same trend for the subset of RNA Pol II peaks involved in chromatin interactions in RWPE-1, LNCaP, and VCaP cells (Figure 7B). Interestingly, neither RNA Pol II peaks nor chromatin interaction–associated RNA Pol II peaks in AR-nonexpressing DU145 cells showed a statistically significant enrichment for PCa risk SNPs. The results obtained with computational simulations were largely recapitulated by conducting enrichment analysis using Fisher's exact test (Figure 7C). These analyses indicated that RNA Pol II–bound regions and chromatin interactions were enriched for PCa risk SNPs in both normal prostate epithelial cells as well as AR-positive PCa cells. The absence of this enrichment in the AR-negative

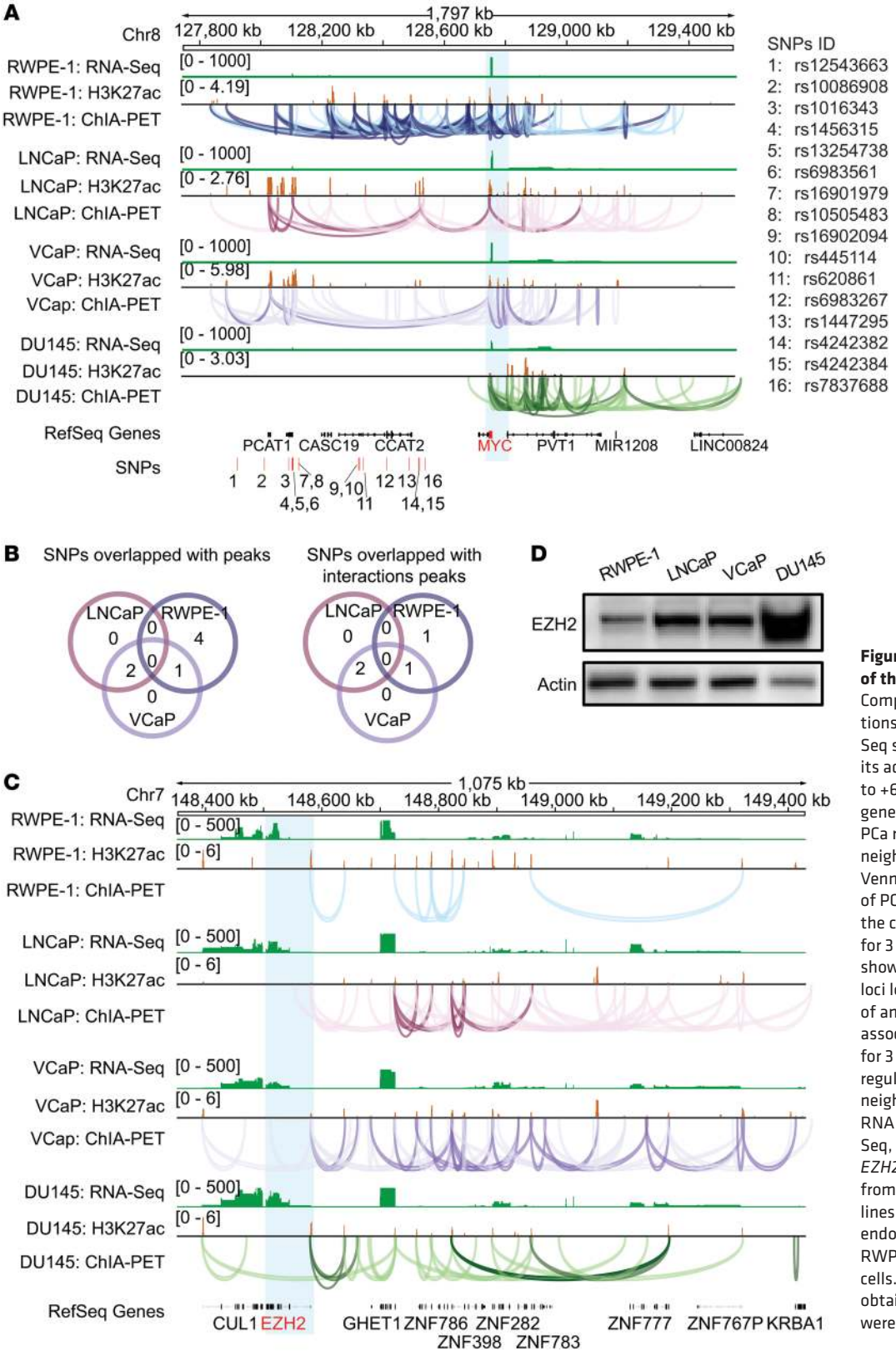


Figure 6. Transcriptional regulation of the *MYC* and *EZH2* genes. (**A**) Comparison of chromatin interactions, H3K27ac ChIP-Seq, and RNA-Seq signals of the *MYC* gene and its adjacent regions from –1000 kb to +600 kb in 4 cell lines. The *MYC* gene is highlighted in light blue. PCa risk SNP loci located in the *MYC* neighborhood are shown. (**B**) Left: Venn diagram describes the number of PCa risk SNP loci located within the coordinates of RNA Pol II peaks for 3 cell lines. Right: Venn diagram shows the number of PCa risk SNP loci located within the coordinates of anchor regions of RNA Pol II–associated chromatin interactions for 3 cell lines. (**C**) Transcriptional regulation of the *EZH2* gene and its neighboring genes. Comparison of RNA Pol II ChIA-PET, H3K27ac ChIP-Seq, and RNA-Seq signals at the *EZH2* locus and its adjacent regions from –150 kb to +800 kb in 4 cell lines. (**D**) Immunoblot representing endogenous EZH2 expression in RWPE-1, LNCaP, VCaP, and DU145 cells. The EZH2 and actin blots were obtained from separate gels that were run contemporaneously.

DU145 cells provided both a valuable internal control and deeper biological insight. These results also imply that PCa risk alleles are not enriched in transcription-associated *cis* elements for all cell types; rather, this enrichment is restricted to certain lineages.

Transcriptional regulation by the PCa risk SNP rs684232. The PCa risk SNP rs684232 has been reported to function as an expression quantitative trait locus (eQTL) (28–31). We delved deeper into the mechanism by which this SNP affects PCa risk by using

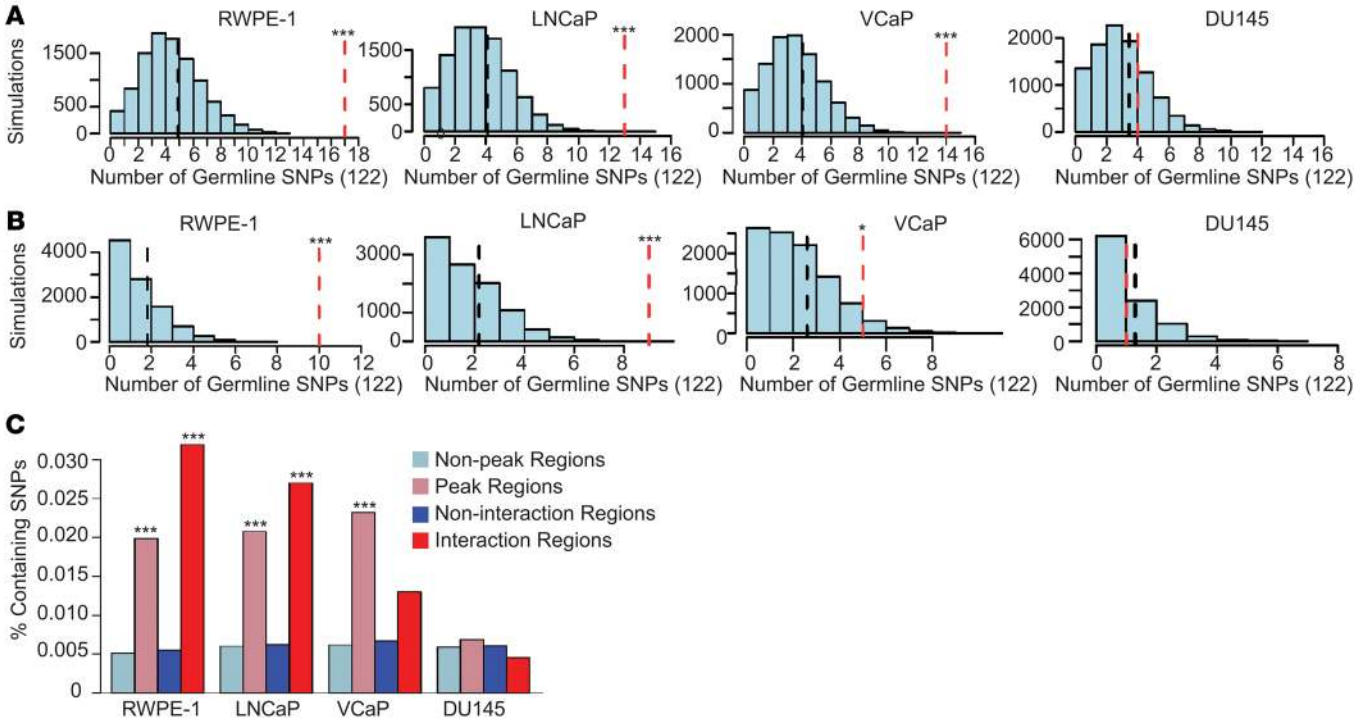


Figure 7. Evaluation of RNA Pol II–associated peaks and interaction with 122 prostate-specific germline SNP locations. (**A**) Peak analysis. The red dashed lines indicate the observed number of peaks containing SNPs for each cell line. The histograms illustrate the results from 10,000 simulations that assessed the expected number of peaks containing SNPs. The mean of the simulations is shown with a dashed black line. ***$P < 0.001$, for significant differences between the expected and observed values. RWPE-1, LNCaP, VCaP, and DU145 have 17, 13, 14, and 4 peaks that overlap SNPs, respectively. The black dashed lines indicate the expected number of overlapping peaks (the mean of all the simulations). The expected values for RWPE-1, LNCaP, VCaP, and DU145 cells are 4.88, 4.09, 4.03, and 3.44, respectively. (**B**) Interaction analysis. The same procedure was repeated except using only the peaks involved in interactions. The red dashed lines indicate the observed number of SNPs, and the black dashed lines show the expected values. *$P < 0.05$ and ***$P < 0.001$, for significant differences between the expected and observed values. RWPE-1, LNCaP, VCaP, and DU145 cells had an observed value of 10, 9, 5, and 1, respectively, as indicated by the red dashed lines. The expected values for RWPE-1, LNCaP, VCaP, and DU145 cells are indicated by the black dashed lines and were 1.81, 2.17, 2.60, and 1.30, respectively. (**C**) Enrichment analysis. Fisher's enrichment analysis was performed to compare the number of SNP-positive peaks with the rest of the genome as well as to compare the number of SNP-positive and interaction-positive peaks with the rest of the genome. ***$P < 0.001$, for significant enrichment. Fisher's exact test was used to determine the P values for **C**. Empirical method was used to determine the P values for **A** and **B**.

our RNA Pol II ChIA-PET data. The risk SNP rs684232 is located approximately 1 kb upstream of the VPS53 transcriptional start site (TSS) in a proximal enhancer element. RNA Pol II ChIA-PET analysis revealed chromatin interactions between the VPS53 promoter and the adjacent genes *FAM57A* and *GEMIN4*, indicating that these genes are coregulated transcriptionally (Figure 8A). To assess the clinical significance of the chromatin interactions, we examined patient tumor data from 3 clinical cohorts from The Cancer Genome Atlas (TCGA), the Canadian Prostate Cancer Genome (CPC-GENE), and the Porto cohort (8, 32, 33). By stratifying the tumors into AA, AB, and BB genotypes, with A and B representing the WT allele and the risk allele, respectively, we observed that the presence of the risk allele was associated with decreased expression of *VPS53*, *FAM57A*, and *GEMIN4* genes in all the 3 clinical cohorts with statistical significance (Figure 8B). These data indicate that the risk SNP rs684232 is likely to downregulate these 3 neighboring genes. The locus harboring the risk SNP rs684232 has the H3K27ac enhancer mark and is co-occupied by the AR, FOXA1, and ERG in VCaP cells (Figure 8C). We examined the epigenetic correlates of the germline risk SNP rs684232 in the Porto cohort by stratifying the patient tumors into AA, AB, and BB genotypes (8). We found that the presence of the risk allele was associated with decreased H3K27ac levels in the locus (Figure 8D). Furthermore, the presence of the risk allele was also associated with decreased H3K27ac levels in the *FAM57A* and *GEMIN4* genes (Figure 8E). These results suggest that the risk allele manifests transcriptional downregulation by reducing enhancer activity. Next, we studied the effect of the WT and risk alleles in enhancing gene transcription by conducting dual-luciferase reporter assays in LNCaP and VCaP cells (Figure 8, F and G). The WT allele markedly increased the reporter luciferase activity, and the magnitude of the effect was significantly reduced in the presence of the risk allele in both cell lines. Thus, by integrating cell-based assays and population genomics, we have unraveled mechanistic features of transcriptional regulation by the PCa risk SNP rs684232.

The regulation and clinical significance of the transcriptional targets of the PCa risk SNP rs684232. We next hypothesized that master transcription factors cooperatively act on enhancer elements to upregulate expression of the risk SNP rs684232 target genes *VPS53*, *FAM57A*, and *GEMIN4*. siRNA-based knockdown of the AR resulted in the downregulation of *VPS53*, *FAM57A*, and *GEMIN4* transcription in LNCaP cells (Figure 9A). To explore the relation-

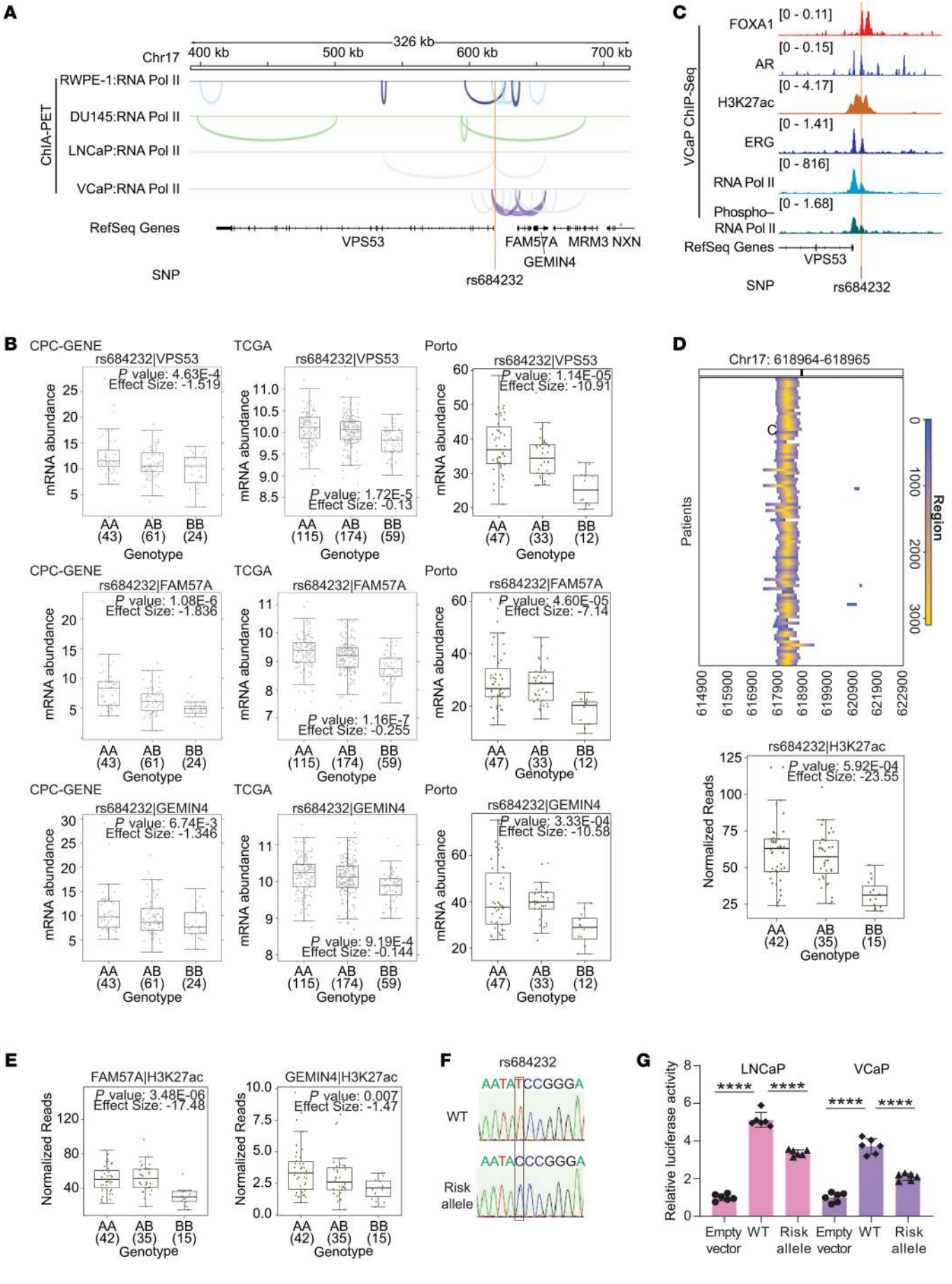

A
Chr17
400 kb
500 kb
326 kb
600 kb
700 kb
ChIA-PET
RWPE-1:RNA Pol II
DU145:RNA Pol II
LNCaP:RNA Pol II
VCaP:RNA Pol II
RefSeq Genes
VPS53
FAM57A
MRM3
NXN
GEMIN4
SNP
rs684232
C
VCaP ChIP-Seq
FOXA1
[0 - 0.11]
AR
[0 - 0.15]
H3K27ac
[0 - 4.17]
ERG
[0 - 1.41]
RNA Pol II
[0 - 816]
Phospho–RNA Pol II
[0 - 1.68]
RefSeq Genes
VPS53
SNP
rs684232
B
CPC-GENE
TCGA
Porto
rs684232|VPS53
P value: 4.63E-4
Effect Size: -1.519
P value: 1.72E-5
Effect Size: -0.13
P value: 1.14E-05
Effect Size: -10.91
rs684232|FAM57A
P value: 1.08E-6
Effect Size: -1.836
P value: 1.16E-7
Effect Size: -0.255
P value: 4.60E-05
Effect Size: -7.14
rs684232|GEMIN4
P value: 6.74E-3
Effect Size: -1.346
P value: 9.19E-4
Effect Size: -0.144
P value: 3.33E-04
Effect Size: -10.58
mRNA abundance
AA (43) AB (61) BB (24)
AA (115) AB (174) BB (59)
AA (47) AB (33) BB (12)
Genotype
D
Chr17: 618964-618965
Patients
Region
614900
615900
616900
617900
618900
619900
620900
621900
622900
rs684232|H3K27ac
P value: 5.92E-04
Effect Size: -23.55
Normalized Reads
AA (42) AB (35) BB (15)
Genotype
E
FAM57A|H3K27ac
P value: 3.48E-06
Effect Size: -17.48
GEMIN4|H3K27ac
P value: 0.007
Effect Size: -1.47
Normalized Reads
AA (42) AB (35) BB (15)
Genotype
F
rs684232
WT
AATATCCGGGA
Risk allele
AATACCCGGGA
G
LNCaP
VCaP
****
Relative luciferase activity
Empty vector
WT
Risk allele

Figure 8. Transcriptional regulation by the PCa risk SNP rs684232. (**A**) Integrated genome view of RNA Pol II–associated chromatin interactions in the genomic region harboring the PCa risk SNP rs684232. (**B**) rs684232 was significantly associated with mRNA abundance of *FAM57A*, *VPS53*, and *GEMIN4* in the CPC-GENE, TCGA, and Porto cohorts. Box plots represent the median and the 0.25 and 0.75 quantiles, with whiskers at 1.5 times the IQR. mRNA abundance was measured in FPKM. The numbers below the genotypes indicate the number of samples in each group. *P* values and effect size are from a linear model. (**C**) Epigenetic features of the PCa risk SNP rs684232 locus in VCaP cells are described using ChIP-Seq analysis. (**D**) rs684232 falls in an active enhancer region, and the alternative allele was found to be significantly associated with decreased H3K27ac binding. Heatmap shows H3K27ac ChIP-Seq signal within chr17:614900-622900 (*x* axis) for 92 patients (*y* axis). Color indicates ChIP-Seq signal intensity, and the black bar in the covariate along the top indicates the location of rs684232. Box plot shows H3K27ac signal intensity stratified by genotype in the Porto cohort (Mann-Whitney *U* test for the recessive model). The *y* axis represents the number of H3K27ac ChIP-Seq read counts mapped to the SNP rs684232 region, which were normalized by the trimmed mean of M values (TMM). (**E**) Box plots show H3K27ac signal intensity in the promoter regions of *FAM57A* and *GEMIN4* stratified by genotype in the Porto cohort (Mann-Whitney *U* test for the recessive model). *P* values and effect size are from a linear model. (**F**) Sequence analysis to confirm the cloning of the WT and risk (rs684232) alleles in the pGL2 luciferase reporter plasmid. (**G**) Luciferase reporter assays in LNCaP and VCaP cells. Cells were cotransfected with pSV-*Renilla* and the luciferase reporter encoding the WT or risk (rs684232) allele and processed 48 hours after transfection. Firefly Luc/*Renilla* Luc activity was determined (mean ± SD, $n = 6$; ****$P < 0.0001$, by 2-tailed Student's *t* test).

ship between the presence of the risk SNP rs684232 and AR occupancy in the locus, we examined the Porto cohort by stratifying the patient tumors into AA, AB, and BB genotypes. The presence of the risk allele was associated with decreased AR occupancy in the locus (Figure 9B). These results indicated that, mechanistically, the risk SNP rs684232 diminished AR binding and enhancer activation, resulting in downregulation of the 3 target genes.

We also addressed the role of ERG in the framework of transcription control by the risk SNP rs684232. As ERG expression is a gene fusion–mediated acquired somatic event, although the risk allele is inherited from the germline, the simultaneous presence of these 2 properties may have opposing effects in terms of expression of the 3 target genes. We further stratified the CPC-GENE cohort into ERG-positive and -negative groups. Remarkably, the association between the presence of the risk allele and the downregulation of the target genes was more profound in ERG-positive tumors than in the ERG-negative tumors (Figure 9C). Thus, we conclude that the effect of the risk allele rs684232 on its target genes is further modulated by acquired somatic events like ERG gene fusions.

We observed a decrease in the transcript abundance of *VPS53*, *FAM57A*, and *GEMIN4* genes with an increase in the PCa International Society of Urological Pathology (ISUP) grade group, and the effect was most prominent for *GEMIN4* (Figure 9D). We also stratified the PCa patient tumors on the basis of *VPS53*, *FAM57A*, and *GEMIN4* transcript abundance. We found that lower *GEMIN4* expression in tumors was associated with shorter biochemical recurrence–free (BCR-free) survival (Figure 9E). Although we also observed a weak trend with *VPS53* and *FAM57A*, it did not reach statistical significance. Consistent with these findings, siRNA-based knockdown of *VPS53*, *FAM57A*, and *GEMIN4* in LNCaP cells resulted in a modest increase in cell viability, both individually and in combination (Supplemental Figure 9).

Discussion

In this study, we describe the landscapes of transcription in PCa and identify thousands of RNA Pol II–associated long-range chromatin interactions — with implications ranging from basic biology to cancer etiology and risk. We have shown that RNA Pol II–associated chromatin interactions strongly overlap with the H3K27ac marks, which are indicative of active enhancers. The enrichment of transcription factor binding motifs in enhancers provides a flexible mechanism for cells to change their state by altering their pool of transcription factors and/or their DNA binding sites — a feature that is constantly exploited in cancers.

The physical nature and temporal dynamics of chromatin interactions are still elusive. These chromatin interactions can also be promoted by bromodomain and extraterminal (BET) proteins. Because of the presence of tandem acetyl lysine binding bromodomains, BET proteins can function as adaptors to connect enhancers with promoters, and both enhancers and promoters are rich in acetylated proteins. In addition to cooperative interactions via specific structured bromodomains, it has not escaped our notice that the low-complexity regions/disordered domains of BET proteins and perhaps other transcriptional regulators can also enhance chromatin interactions via liquid-liquid phase separation, and this could be further regulated by post-translational modifications (34–36). At present, it is not clear if the DNA regions in between enhancers and promoters loop out or if these regions are associated with weak, nonspecific chromatin interactions. Clearly, there is much complexity associated with such chromatin interactions. This complexity is necessary, as a single genotype gives rise to multiple phenotypes and cell types at the organismal level. However, underlying this complexity is a simple design principle for transcriptional regulation conserved from bacteria to humans: genome-wide binding of *trans*-acting factors and cognate *cis*-regulatory elements to specify which genes are turned on and which are turned off by RNA polymerase at any given time. This design and the genome organization are progressively altered during cancer development.

By integrative analysis, we have outlined the architectural features of transcription control. The RAD21 interactions associated with CTCF binding provide a structural framework for RNA Pol II interactions associated with the H3K27ac mark. RAD21 interactions are in general longer than RNA Pol II interactions. RAD21 interactions may form closed genomic neighborhoods (also called insulated neighborhoods, sub-TADs, or chromatin contact domains) to contain RNA Pol II interactions. A significant proportion of enhancers contain CTCF elements, but promoters are not enriched for these elements. Therefore, it could be speculated that enhancers with CTCF elements connect with both promoters and CTCF-containing boundary elements; the promoters will be indirectly drawn toward boundary elements because of their interaction with enhancers. Thus, multiple layers of chromatin interactions specify the transcriptional output of individual cellular lineages: enhancer-promoter interactions direct the specificity of transcriptional regulation, whereas enhancer-boundary inter-

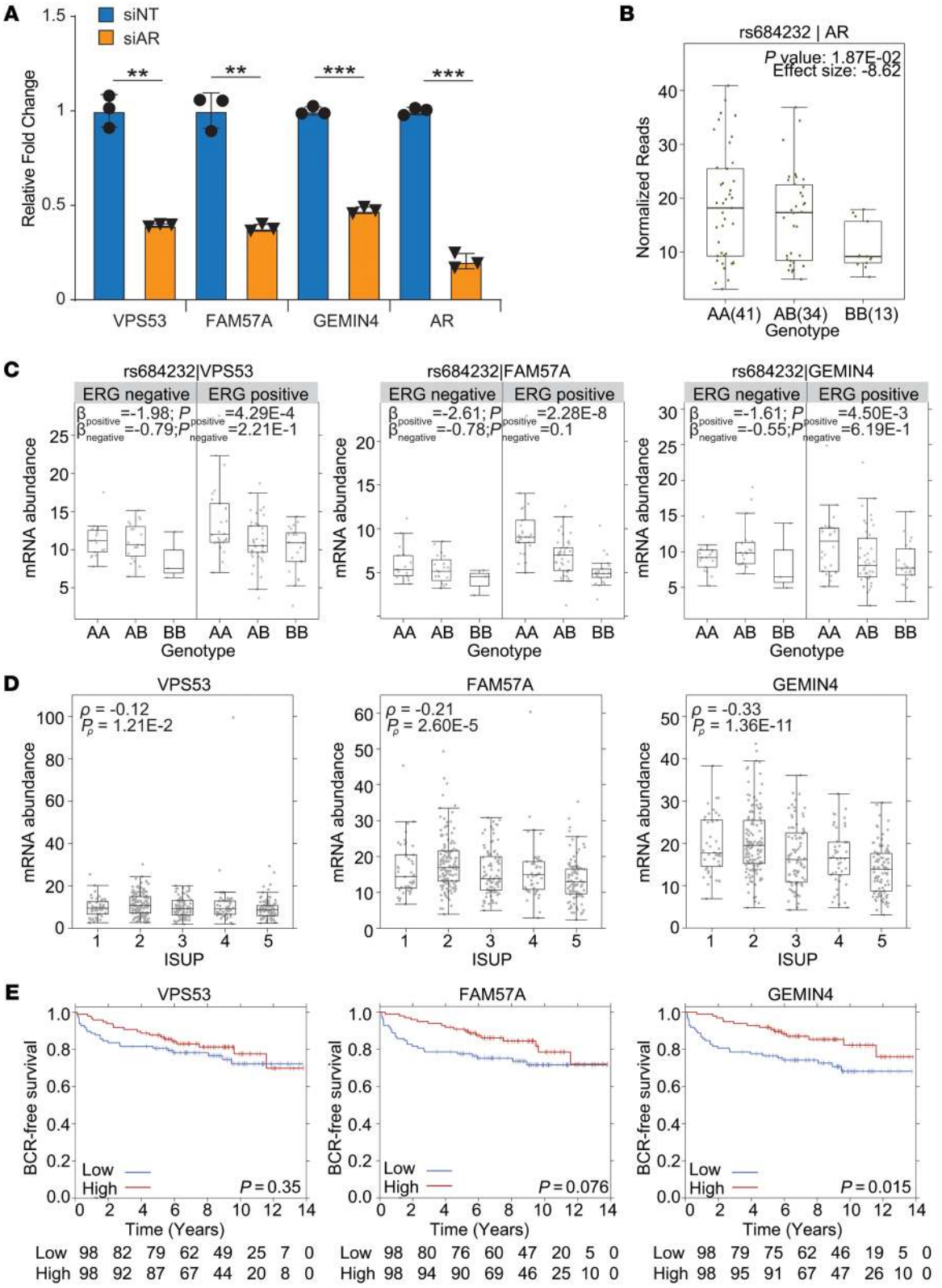

A
siNT
siAR
Relative Fold Change
VPS53
FAM57A
GEMIN4
AR
B
rs684232 | AR
P value: 1.87E-02
Effect size: -8.62
Normalized Reads
AA(41) AB(34) BB(13)
Genotype
C
rs684232|VPS53
rs684232|FAM57A
rs684232|GEMIN4
ERG negative
ERG positive
mRNA abundance
Genotype
D
VPS53
FAM57A
GEMIN4
mRNA abundance
ISUP
E
BCR-free survival
Low
High
P = 0.35
P = 0.076
P = 0.015
Time (Years)
Low 98 82 79 62 49 25 7 0
High 98 92 87 67 44 20 8 0
Low 98 80 76 60 47 20 5 0
High 98 94 90 69 46 25 10 0
Low 98 79 75 62 46 19 5 0
High 98 95 91 67 47 26 10 0

Figure 9. Transcriptional regulation and clinical correlates of the chromatin interaction targets of the PCa risk SNP rs684232. (**A**) qRT-PCR validation of AR knockdown and the expression of *VPS53*, *GEMIN4*, and *FAM57A* genes upon treatment of LNCaP cells with AR siRNA. $^{**}P < 0.01$, and $^{***}P < 0.001$, by 2-tailed Student's *t* test. Error bars indicate the SD of 3 technical replicates. (**B**) Box plot shows AR ChIP-Seq signal intensity stratified by genotype in the Porto cohort (Mann-Whitney *U* test for the recessive model). The *y* axis shows the number of AR ChIP-Seq read counts mapped to the SNP rs684232 region, which were normalized by the TMM method. Box plot represents the median and the 0.25 and 0.75 quantiles, with whiskers at 1.5 times the IQR. (**C**) The regulatory impact of rs684232 was enhanced in the presence of the *TMPRSS2-ERG* fusion. Box plots show the mRNA abundance (FPKM) of each gene stratified by genotype and further split by ERG status in the CPC-GENE cohort. $\beta_{positive}$ and $\beta_{negative}$, and the associated *P* values quantify the eQTL within ERG-positive and -negative patients, respectively (linear model). (**D**) Box plots show the mRNA abundance of *FAM57A*, *GEMIN4*, and *VPS53* genes in PCa specimens from TCGA cohort. Tumors were classified into various ISUP grade groups. Relationship between mRNA abundance and ISUP Grade Group was quantified using Spearman's correlation and represented as Spearman's rho and corresponding *P* values. The *P* values for Spearman's correlation were computed using the algorithm AS 89 (47). (**E**) BCR-free survival curves for PCa patient groups defined by transcript abundance for *FAM57A, GEMIN4*, and *VPS53* genes in the CPC-GENE cohort. *P* values in **E** were determined by log-rank test.

actions and boundary-boundary interactions serve as protective moats to reduce off-target chromatin interactions and control transcriptional noise. Genomic rearrangements, gene fusions, and CTCF mutations enable cancer cells to breach the boundaries of lineage-dependent transcriptional regulation (1, 17, 37).

Misregulation of chromatin interactions is a distinguishing feature of most cancers. For example, *TMPRSS2-ERG* gene fusions are observed in 50% of human PCas (1). As a result of the gene fusion, *ERG* is transcriptionally activated by the promoter and enhancer elements of the AR target gene *TMPRSS2*. Most PCa treatment strategies target the AR signaling axis. Hence, it is critical to understand the transcriptional regulation of and by the AR. In this study, we have identified and described long-range *AR* enhancer clusters. We found that *AR* locus amplification was associated with increased interaction between *AR* enhancers and the promoter. Analysis of mCRPC specimens suggested that *AR* locus amplification frequently involved AR enhancer amplification. Thus, enhancers regulated AR function in cell line models and clinical PCa specimens.

We delineate general principles of oncogene activation. The genomic location of the *MYC* locus facilitates its activation by multiple enhancers in various cancers because of the availability of multiple transcription factor binding sites. Remarkably, we found that several PCa GWAS alleles lay on *MYC* enhancers, suggestive of altered chromatin interactions and transcriptional misregulation. The cooperative action of multiple enhancers underlies *EZH2* upregulation in PCa. More generally, we suggest that RNA Pol II network hubs provide multiple nodes for receiving, processing, and transmitting regulatory signals, thereby contributing to coordinated transcription as well as regulation of dynamic processes like transcriptional bursting (38, 39). Multinodal transcriptional hubs are likely to be associated with subclonal transcriptional variability in tumors, resulting in the selection of drug-tolerant clones and emergence of resistance to targeted therapies (40, 41).

We have shown that germline PCa risk alleles were enriched in the enhancers found in nonmalignant, basal-like RWPE-1 prostate epithelial cells, as well as in luminal AR-positive cancer cells, but not in AR-negative cancer cells that had diverged transcriptionally. Unlike mutations in exons, which typically affect individual genes, PCa risk alleles like rs684232 can simultaneously alter the expression of multiple genes. Importantly, our discovery of a genetic interaction between the germline risk allele rs684232 and the somatically acquired TMPRSS2-ERG indicates a potential role of epistasis and modifier genes in modulating PCa risk. These results have implications for the development of intervention strategies to prevent or delay the onset of PCa.

In summary, we have described the genome architectural features of transcription control in multiple PCa cellular models, integrated the results with existing data sets, expanded our discoveries to clinical specimens, and, finally, made relevant connections to PCa germline susceptibility alleles. We anticipate that this work will create fertile avenues for future research in transcriptional regulation and cancer development.

Methods

Cell lines. The cell lines used in this study (RWPE-1, LNCaP, VCaP, and DU145,) were procured from the American Type Culture Collection (ATCC) and grown according to their recommendations. All cell lines were verified by genotyping.

Chromatin interaction analysis by ChIA-PET. ChIA-PET analysis was performed using previously published methods with a few modifications (42). Briefly, approximately 200 million cells per cell line were cross-linked using 1.5 mM EGS (35 minutes) and 1% formaldehyde (8 minutes) and then quenched using 0.125M glycine (5 minutes) at room temperature. The cells were lysed and the chromatin was sheared using a HighCell ChIP Kit Protein G (Diagenode) and used for ChIP. The anti–Pol II monoclonal antibody 8WG16 (BioLegend) was used. A quality check for the ChIP product was carried out by quantitative PCR (qPCR) using the primers GAPDH promoter, forward: 5′-TACTAGCGGTTTTACGGGCG-3′, GAPDH promoter, reverse: 5′-TCGAACAGGAGGAGCAGAGAGCGA-3′ as a positive control and GAPDH exon 8, forward: 5′-CCATCACTGCCACCCAGAAG-3′, GAPDH exon 8, reverse: 5′-AGCTTCCCGTTCAGCTCAGG-3′ as the negative control. The ChIP DNA was quantified by fluorometry (Qubit, Invitrogen, Thermo Fisher Scientific) and then end-blunted, A-tailed, and proximity-ligated overnight with a biotinylated bridge linker. The DNA was reverse crosslinked, fragmented using transposase (Nextera DNA Library Preparation Kit, Illumina), and purified (Genomic DNA Clean & Concentrator, Zymo Research). Libraries were constructed using a minimal number of cycles for PCR amplification, and the PCR products were purified, size selected (Agencourt Ampure XP, Beckman Coulter), and quality checked using Tape Station (High Sensitivity D1000 Screentape, Agilent Technologies). Paired-end sequencing (2 × 150 bp) in NextSeq 500 (Illumina) was carried out to yield approximately 70–80 million reads per sample.

ChIA-PET data processing and analysis. Data processing of ChIA-PET reads was performed using ChIA-PET2 software (43) to obtain the binding peaks and the interactions among peaks. Pair-end read (PET) sequences were scanned to filter the bridge linker sequence 5′-CGCGATATCTTATCTGACT-3′. The filtered reads were mapped to the human reference genome (hg19), and only uniquely aligned

(MAPQ ≥30) PETs were retained. By evaluating the genomic locations of the 2 ends of a PET, each PET was categorized as either intrachromosomal PET (2 ends in the same chromosome) or interchromosomal PET (2 ends in different chromosomes). All intrachromosomal PETs were further separated into self-ligation PETs (2 ends in the same peak region) or regular intrachromosomal PETs (between 2 different peak regions). These data were used to define the interaction strength between 2 peak regions. The RNA Pol II binding peaks were then compared between different cell lines to identify commonalities, and the R package Vennerable (https://sourceforge.net/projects/vennerable) was used to generate the Venn diagram.

DNB model. For a pair of interacting peaks, the number of PETs between them is defined as their interaction strength. We found that the interaction strengths of a cell line followed a NB distribution, which is commonly observed in analysis of RNA-Seq and ChIP-Seq data (44, 45). Since there is a lack of computational methods for directly comparing 2 samples of ChIA-PET data that are fitted as NB distributions, we introduced a statistical model, named the DNB model, to compare 2 general NB distributions whose parameters were real positive numbers. Suppose that X and Y are the random variables of 2 NB distributions, where $X \sim NB(r_1, p_1)$ and $Y \sim NB(r_2, p_2)$. We calculated the probability function of their difference $Z = X - Y$. Using the convolution formula, we have:

$$P(Z=k)=\sum_{y=0}^{\infty}P(Y=y,X=k+y)=\sum_{y=0}^{\infty}P(Y=y)P(X=k+y)$$

(Equation 1)

We then calculated the probability mass function (PMF) of Z as:

$$P(Z=k)=\frac{p_1^k(1-p_1)^{r_1}(1-p_2)^{r_2}}{(r_1-1)!(r_2-1)!}\sum_{y=0}^{\infty}(p_1p_2)^y\frac{(y+r_2-1)!(k+y+r_1-1)!}{y!(k+y)!}$$

(Equation 2)

For a peak pair i, we could calculate their difference of interaction strength X_i and Y_i for 2 samples. We then calculated the probability of $P(Z > X_i - Y_i)$ to determine whether the difference was significant. Clearly, the smaller the P value, the more significant was the difference of the interaction strengths between 2 samples. With a significance level of 0.05 as a threshold, we obtained a large number of peak pairs whose interactions were significantly changed by comparing LNCaP, VCap, and DU145 cells with RWPE-1 cells. For the peak pair with significant different interactions, their nearest genes were extracted for functional analysis of GO terms and pathways using the PANTHER database (46).

Data and software availability. The ChIA-PET, ChIP-Seq, and RNA-Seq data have been deposited in the NCBI's Gene Expression Omnibus (GEO) database under the accession numbers GSE121020, GSE121021, and GSE121022, respectively. The aCGH data have been deposited in the EMBL-EBI's ArrayExpress archive under accession number E-MTAB-7326, and the WGS data were deposited and the European Nucleotide Archive (ENA) under accession number ERS2773662.

Statistics. All quantitative reverse transcription PCR (qRT-PCR) experiments were performed in triplicate, and quantitative data are presented as the mean ± SD. All reporter luciferase assays were performed in 6 replicates, and quantitative data are presented as the mean ± SD. Two-tailed Student's t tests were used to determine significance for qRT-PCR and reporter luciferase assays, and P values of less than 0.05 were considered statistically significant.

Study approval. All procedures involving human subjects were approved by the ethics review committee of the Royal Marsden NHS Foundation Trust Hospital (reference no. 04/Q0801/60).

Author contributions

RSM, MQZ, JDB, PCB, SGR, YC, and JY conceived and designed the study. SGR, YC, JY, KD, MBKL, KEH, SC, WY, PCB, JDB, MQZ, and RSM developed methodology. SGR, YC, JY, KD, MBKL, KEH, SC, WY, GB, AS, AP, AA, and NS acquired data. SGR, YC, JY, KD, MBKL, KEH, SC, WY, GB, AS, AP, MK, YG, CX, PCB, JDB, MQZ, and RSM analyzed and interpreted data. RSM and SGR wrote, reviewed, and revised the manuscript with input from all authors. GVR and XZ provided administrative and/or technical support. RSM supervised the study. JDB is a National Institute for Health Research (NIHR) Senior Investigator. The views expressed in this article are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR, or the UK Department of Health. The order of the co–first authors was assigned on the basis of their relative contributions to the study.

Acknowledgments

We thank James Malter, Diego Castrillon, and James Brugarolas for insightful comments and discussions, Weibo Luo for assistance with luciferase assays, and Xiangyi Li for technical assistance. This work was supported by a NIH Pathway to Independence (PI) Award (R00CA160640), a National Cancer Institute (NCI), NIH grant (R01CA245294), a Cancer Prevention and Research Institute of Texas (CPRIT) Individual Investigator Research Award (RP190454), a US Department of Defense Prostate Cancer Research Program (PCRP) Impact Award (W81XWH-17-1-0675), a Kidney Cancer Specialized Programs of Research Excellence (SPORE) Career Enhancement Program (CEP) award (P50CA196516), and UT Southwestern startup funds (to RSM). MQZ receives support from the University of Texas at Dallas (UTD) startup fund, the Cecil H. and Ida Green Endowment, the CPRIT (RP180826), and the National Natural Science Foundation of China (NSFC) (grants 31671384 and 91329000). Work in the JDB laboratory was supported by funding from the Movember Foundation, Prostate Cancer UK, the US Department of Defense, the Prostate Cancer Foundation, Stand Up To Cancer, Cancer Research UK, and the UK Department of Health through an Experimental Cancer Medicine Centre grant. YC was supported by the Seed Funding Program of Rowan University. AS is supported by the Medical Research Council, the Academy of Medical Sciences, Prostate Cancer UK, and the Prostate Cancer Foundation. PCB was supported by the NCI via award number P30CA016042, an operating grant from the National Cancer Institute Early Detection Research Network (1U01CA214194-01), Prostate Cancer Canada, a Movember Prostate Cancer Canada Rising Star Award, and by Canadian Institutes of Health Research (CIHR) and Terry Fox Research Institute (TFRI) New Investigator Awards. GVR acknowledges support from the US Department of Defense, the Wilson Foundation, and the Mimi and John Cole Foundation. The results described here are in part based on data

generated by TCGA (http://cancergenome.nih.gov), managed by the NCI and the National Human Genome Research Institute (NHGRI).

Address correspondence to: Ram S. Mani, Department of Pathology, UT Southwestern Medical Center, 5323 Harry Hines Blvd NB6.444, Dallas, Texas 75390, USA. Phone: 214.645.7007; Email: ram.mani@utsouthwestern.edu. Or to: Michael Q. Zhang, Department of Biological Sciences, Center for Systems Biology, University of Texas at Dallas, 800 West Campbell Road, Richardson, Texas 75080, USA. Phone: 972.883.2523. Email: michael.zhang@utdallas.edu.

1. Tomlins SA, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*. 2005;310(5748):644–648.
2. Dai C, Heemers H, Sharifi N. Androgen sgnaling in prostate cancer. *Cold Spring Harb Perspect Med*. 2017;7(9):a030452.
3. Robinson D, et al. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015;161(5):1215–1228.
4. Quigley DA, et al. Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell*. 2018;174(3):758–769.
5. Zhang Z, et al. An AR-ERG transcriptional signature defined by long-range chromatin interactomes in prostate cancer cells. *Genome Res*. 2019;29(2):223–235.
6. Yu J, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell*. 2010;17(5):443–454.
7. Kron KJ, et al. TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. *Nat Genet*. 2017;49(9):1336–1345.
8. Stelloo S, et al. Integrative epigenetic taxonomy of primary prostate cancer. *Nat Commun*. 2018;9(1):4900.
9. Rhie SK, et al. A high-resolution 3D epigenomic map reveals insights into the creation of the prostate cancer transcriptome. *Nat Commun*. 2019;10(1):4154.
10. Taberlay PC, et al. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res*. 2016;26(6):719–731.
11. Fullwood MJ, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*. 2009;462(7269):58–64.
12. Tang Z, et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*. 2015;163(7):1611–1627.
13. Bert SA, et al. Regional activation of the cancer genome by long-range epigenetic remodeling. *Cancer Cell*. 2013;23(1):9–22.
14. Takeda DY, et al. A somatically acquired enhancer of the androgen receptor is a noncoding driver in advanced prostate vancer. *Cell*. 2018;174(2):422–432.e13.
15. Viswanathan SR, et al. Structural alterations driving castration-resistant prostate cancer revealed by linked-read genome sequencing. *Cell*. 2018;174(2):433–447.e19.
16. Liu W, et al. Homozygous deletions and recurrent amplifications implicate new genes involved in prostate cancer. *Neoplasia*. 2008;10(8):897–907.
17. Mani RS, et al. Induced chromosomal proximity and gene fusions in prostate cancer. *Science*. 2009;326(5957):1230.
18. Xu Y, Vakoc CR. Targeting cancer cells with BET bromodomain inhibitors. *Cold Spring Harb Perspect Med*. 2017;7(7):a026674.
19. Mateo J, et al. DNA-repair defects and olaparib in metastatic prostate cancer. *N Engl J Med*. 2015;373(18):1697–1708.
20. Liu Y, et al. The androgen receptor regulates a druggable translational regulon in advanced prostate cancer. *Sci Transl Med*. 2019;11(503):eaaw4993.
21. Shen MM, Abate-Shen C. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev*. 2010;24(18):1967–2000.
22. Mumbach MR, et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat Genet*. 2017;49(11):1602–1612.
23. Schuijers J, et al. Transcriptional dysregulation of MYC reveals common enhancer-docking mechanism. *Cell Rep*. 2018;23(2):349–360.
24. Prensner JR, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol*. 2011;29(8):742–749.
25. Eeles R, et al. The genetic epidemiology of prostate cancer and its clinical implications. *Nat Rev Urol*. 2014;11(1):18–31.
26. Varambally S, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*. 2002;419(6907):624–629.
27. Guo H, et al. Modulation of long noncoding RNAs by risk SNPs underlying genetic predispositions to prostate cancer. *Nat Genet*. 2016;48(10):1142–1150.
28. Eeles RA, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet*. 2013;45(4):385–391, 391e1.
29. Schumacher FR, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet*. 2018;50(7):928–936.
30. Whitington T, et al. Gene regulatory mechanisms underpinning prostate cancer susceptibility. *Nat Genet*. 2016;48(4):387–397.
31. Larson NB, et al. Comprehensively evaluating *cis*-regulatory variation in the human prostate transcriptome by using gene-level allele-specific expression. *Am J Hum Genet*. 2015;96(6):869–882.
32. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell*. 2015;163(4):1011–1025.
33. Chen S, et al. Widespread and functional RNA circularization in localized prostate cancer. *Cell*. 2019;176(4):831–843.e22.
34. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A phase separation model for transcriptional control. *Cell*. 2017;169(1):13–23.
35. Sabari BR, et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. 2018;361(6400):eaar3958.
36. Gibson BA, et al. Organization of chromatin by intrinsic and regulated phase separation. *Cell*. 2019;179(2):470–484.e21.
37. Hnisz D, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. 2016;351(6280):1454–1458.
38. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. Mammalian genes are transcribed with widely different bursting kinetics. *Science*. 2011;332(6028):472–474.
39. Fukaya T, Lim B, Levine M. Enhancer control of transcriptional bursting. *Cell*. 2016;166(2):358–368.
40. Shaffer SM, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*. 2017;546(7658):431–435.
41. Sharma SV, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*. 2010;141(1):69–80.
42. Li X, et al. Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. *Nat Protoc*. 2017;12(5):899–915.
43. Li G, Chen Y, Snyder MP, Zhang MQ. ChIA-PET2: a versatile and flexible pipeline for ChIA-PET data analysis. *Nucleic Acids Res*. 2017;45(1):e4.
44. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106.
45. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
46. Mi H, et al. PANTHER version 11: expanded annotation data from Gene Ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. 2017;45(D1):D183–D189.
47. Best DJ, Roberts DE.Algorithm AS 89: The upper tail probabilities of Spearman's Rho. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*. 1975;4(3):377–379.