



An Atlas of Accessible Chromatin in Advanced Prostate Cancer Reveals the Epigenetic Evolution during Tumor Progression

Raunak Shrestha^{1,2}, Lisa N. Chesner^{1,2}, Meng Zhang^{1,2}, Stanley Zhou^{3,4}, Adam Foye^{1,5}, Arian Lundberg^{1,2,6}, Alana S. Weinstein^{1,2}, Martin Sjöström^{1,2}, Xiaolin Zhu^{1,5}, Thaidy Moreno-Rodriguez^{1,7}, Haolong Li^{1,2}; SU2C/PCF West Coast Prostate Cancer Dream Team, Joshi J. Alumkal⁸, Rahul Aggarwal^{1,5}, Eric J. Small^{1,2,5,7}, Mathieu Lupien^{3,4,9}, David A. Quigley^{1,7,10}, and Felix Y. Feng^{1,2,5,7}

ABSTRACT

Metastatic castration-resistant prostate cancer (mCRPC) is a lethal disease that resists therapy targeting androgen signaling, the primary driver of prostate cancer. mCRPC resists androgen receptor (AR) inhibitors by amplifying AR signaling or by evolving into therapy-resistant subtypes that do not depend on AR. Elucidation of the epigenetic underpinnings of these subtypes could provide important insights into the drivers of therapy resistance. In this study, we produced chromatin accessibility maps linked to the binding of lineage-specific transcription factors (TF) by performing assay for transposase-accessible chromatin sequencing on 70 mCRPC tissue biopsies integrated with transcriptome and whole-genome sequencing. mCRPC had a distinct global chromatin accessibility profile linked to AR function. Analysis of TF occupancy

across accessible chromatin revealed 203 TFs associated with mCRPC subtypes. Notably, ZNF263 was identified as a putative prostate cancer TF with a significant impact on gene activity in the double-negative subtype (AR⁻ neuroendocrine⁻), potentially activating MYC targets. Overall, this analysis of chromatin accessibility in mCRPC provides valuable insights into epigenetic changes that occur during progression to mCRPC.

Significance: Integration of a large cohort of transcriptome, whole-genome, and ATAC sequencing characterizes the chromatin accessibility changes in advanced prostate cancer and identifies therapy-resistant prostate cancer subtype-specific transcription factors that modulate oncogenic programs.

Introduction

Prostate cancer is the second leading cause of cancer-related deaths among men (1). Although prostate cancer is initially responsive to androgen deprivation therapy (ADT), many patients develop resistance and progress to metastatic castrate-resistant

prostate cancer (mCRPC). Targeted systemic therapies with second-generation AR signaling inhibitors (ARSI), such as abiraterone or enzalutamide, prolong survival and are the standard of care for mCRPC (2–4).

Tumors can develop resistance against ADT and/or ARSI through several distinct mechanisms (5). In most mCRPCs, ARSI resistance is achieved through genetic changes that increase AR signaling (6). Up to 20% of mCRPCs lose complete AR dependence and acquire a new cellular phenotype known as treatment-emergent small-cell neuroendocrine (NE) prostate cancer (t-SCNC) or neuroendocrine prostate cancer (NEPC). This AR-NE+ subtype is associated with worse clinical outcomes (7, 8). Additional treatment-associated subtypes have been observed, including a double-negative subtype (AR-NE-) that bypasses AR dependence through FGF/MAPK signaling (9, 10) and a double-positive subtype (AR+NE+) that gains NE features while maintaining AR activity (5). A better understanding of these mCRPC subtypes is foundational for the development of new approaches to overcome resistance.

Although AR amplification is typical in AR-dependent mCRPCs (11–13) and t-SCNC often harbors TP53, RB1, and PTEN loss (14), other subtypes have no known characteristic genomic alterations. Emerging evidence suggests that epigenetic mechanisms are associated with prostate cancer progression and drug resistance (7, 15–17). Specifically, lineage plasticity plays an important role in the development of ARSI resistance (10, 18). It has been increasingly recognized that the complex interplay of epigenetic modifications including altered chromatin-binding patterns of transcription factors (TF), such as AR and FOXA1, regulates downstream gene

¹Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California. ²Department of Radiation Oncology, University of California, San Francisco, San Francisco, California. ³Princess Margaret Cancer Centre, University Health Network, Toronto, Canada. ⁴Department of Medical Biophysics, University of Toronto, Toronto, Canada. ⁵Division of Hematology and Oncology, Department of Medicine, University of California, San Francisco, San Francisco, California. ⁶The Institute of Cancer Research and The Royal Marsden Hospital, London, United Kingdom. ⁷Department of Urology, University of California, San Francisco, San Francisco, California. ⁸Division of Hematology and Oncology, University of Michigan Rogel Cancer Center, Ann Arbor, Michigan. ⁹Ontario Institute for Cancer Research, Toronto, Canada. ¹⁰Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, California.

SU2C/PCF West Coast Prostate Cancer Dream Team: A list of members appears in the Supplementary Information.

M. Lupien, D.A. Quigley, and F.Y. Feng jointly supervised this work.

Corresponding Author: Felix Y. Feng, Department of Radiation Oncology, University of California, San Francisco, Box 3110, Room 450, 1450 3rd Street, San Francisco, CA 94158. E-mail: Felix.Feng@ucsf.edu

Cancer Res 2024;84:3086–100

doi: 10.1158/0008-5472.CAN-24-0890

©2024 American Association for Cancer Research

activity, thereby driving prostate cancer progression (19). Therefore, understanding the chromatin-binding patterns of TFs that are altered in prostate cancer is crucial for the development of effective therapeutic strategies.

The assay for transposase-accessible chromatin using sequencing (ATAC-seq) has proven to be a very efficient and general epigenetic assay that yields high-quality chromatin signals from small quantities of tissue (20, 21). ATAC-seq quantifies chromatin accessibility using transposase enzymes that insert sequencing adapters at accessible chromatin sites. Prior studies (15, 21, 22) of prostate cancer have mostly used cell line and organoid models, patient-derived xenografts (PDX), or small numbers of tumor tissue biopsies. The Cancer Genome Atlas study performed ATAC-seq on multiple cancer types including 26 localized prostate cancer tumors and revealed cancer-type-specific enrichment of TF-binding elements in accessible chromatin regions (21). More recently, using ATAC-seq on CRPC organoids, PDX, and cell lines, Tang and colleagues (22) identified four mCRPC subtypes and predicted the key TF of each subtype. To the best of our knowledge, the characterization of chromatin accessibility in clinical mCRPC biopsy tissue samples using ATAC-seq has not been conducted to date.

Herein, we describe the first-in-field ATAC-seq study conducted in the largest cohort ($n = 70$) of mCRPC tissue biopsies, to date, from the Stand Up To Cancer – Prostate Cancer Foundation West Coast Prostate Cancer Dream Team (WCDT) cohort. Using a comprehensive integration of ATAC-seq and RNA-seq from matched tumor samples, we interrogated the changes in chromatin accessibility around regulatory sites to reveal transcriptional regulation associated with mCRPC subtypes. We used computational approaches to produce an exhaustive catalog of TFs that are actively occupied in mCRPC and the transcriptional programs they are predicted to regulate. Finally, we exemplified the use of these new data by characterizing ZNF263, a TF previously not associated with prostate cancer biology.

Materials and Methods

Patients and samples

Human studies were approved and overseen by the University of California San Francisco Institutional Review Board. All individuals provided written informed consent to obtain fresh tumor biopsies and to perform comprehensive molecular profiling of tumor and germline samples. Fresh frozen metastatic castration-resistant tissue biopsy samples ($n = 75$) from various anatomic locations representing 69 unique patients (Supplementary Fig. S1; Supplementary Table S1) were collected through a multi-institutional image-guided prospective biopsy trial (NCT02432001), and DNA was extracted as previously described (12, 16, 17).

ATAC-seq library preparation and high-throughput sequencing

The ATAC-seq library preparation was carried out as described in the published method articles by Buenrostro and colleagues (20) and Corces and colleagues (23). Briefly, upon thawing, 30 μ L of PBS + protease inhibitor was added onto the slide containing the tissue section and subsequently scraped into a 2-mL tube containing 100 μ L of cold ATAC-Resuspension Buffer (RSB; 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin) using the tip of a scalpel blade. The sample was incubated on ice for 15 minutes, with intermittent mixing every 5 minutes. After 15 minutes, 1 mL of cold PBS + 0.1% Tween-20 was added into the tube and mixed by inversion, followed by centrifugation at 500 $\times g$ for 10 minutes at 4°C. After

centrifugation, the supernatant was aspirated, avoiding the pellet containing the cell nuclei in the process. Approximately 50 μ L of transposition mix [1 \times : 25 μ L of 2 \times TD buffer (20 mmol/L Tris-HCl pH 7.6, 10 mmol/L MgCl₂, 20% dimethyl formamide), 2.5 μ L of transposase, 16.5 μ L PBS, 0.5 μ L 1% digitonin, 0.5 μ L 10% Tween-20, 5 μ L water] was added to the nuclei for resuspension. The reaction was then incubated at 37°C for 30 minutes in a thermomixer with 1,000 RPM mixing. After the transposition reaction, the samples were purified using the Qiagen MinElute PCR Purification Kit. Upon elution of the DNA, ATAC-seq libraries were prepared. To minimize PCR biases and duplicates, library preparation was conducted on a real-time qPCR machine, in which each sample was pulled off the machine mid-exponential phase. The resulting ATAC-seq libraries were size selected with Ampure XP beads (Beckman Coulter) for 240 to 360 bp fragments. Upon successful amplification, an aliquot of the libraries was used for qPCR to calculate the fold enrichment of two accessible chromatin regions over two inaccessible chromatin regions for quality control (i.e., at least 10-fold enrichment). Samples that passed quality control were sequenced on the Illumina NovaSeq 6000 sequencing system.

POS (accessible)

GAPDH (F) 5'-GCCAATCTCAGTCCCTTCCC-3', (R) 5'-TAG-

TAGCCGGGCCCTACTTT-3'

KAT6B (F) 5'-GAAGAGGCGGACCCAGCGGT-3', (R) 5'-TTC-CTGCCGGTCATCTCGCTT-3'

NEG (closed)

SLC22A3 (F) 5'-GGAGAGGGTGGACAGATTGA-3', (R) 5'-TCA-GCCTTGCTGCTACAGTG-3'

QML_93 (F) 5'-CACTGGTTGCTTGCAGGA-3', (R) 5'-CCT-GGGTCATATTGGGACAC-3'

ATAC-seq data processing

The ATAC-seq paired-end fastq data were first trimmed to remove the Illumina Nextera adapter sequence using Cutadapt v2.6 (24) with the “-q 10 -m 20” option (Supplementary Fig. S2). After adapter trimming, FASTQC v0.11.8 (25) was used to evaluate the sequence trimming as well as overall sequence quality. Bowtie2 v2.3.5.1 (26) was then used to align the ATAC-seq reads against the human reference genome build hg38 using the “–very-sensitive” option. The uniquely mapped reads were obtained in SAM format. Samtools v1.9 (27) was used to convert SAM to BAM file as well as sort the BAM file. Picard tool (<https://broadinstitute.github.io/picard>) was then used to flag duplicate reads using the MarkDuplicates function with the “REMOVE_DUPLICATES = true” option. The resulting BAM file reads position was then corrected by a constant offset to the read start (positive stranded +4 bp, negative stranded –5 bp) using deepTools2 v3.3.2 (28) with the “alignmentSieve –ATACshift” option. This resulted in the final aligned, de-duplicated BAM file that was used in downstream analyses.

ATAC-seq peak calling was performed using MACS2 v2.2.5 (29) to obtain narrow peaks with “callpeak -f BAMPE -g hs -qvalue 0.05 -nomodel -B -keep-dup all -call-summits” option. The resulting peaks that mapped to the mitochondrial genome or genomic regions listed in the ENCODE hg38 blacklist (ENCSR636HFF) or peaks that extend beyond the ends of chromosomes were filtered out. The ATAC-seq peaks were annotated with the nearest gene and genomic region in which the peak is located using ChIPseeker (30) R-package based on hg38 GENCODE v28 annotations. Possible peak annotations are

promoter [± 3 kb from transcription start site (TSS)], exon, 5'UTR, 3'UTR, intron, and distal intergenic.

ATAC-seq quality control

Quality metrics such as fraction of reads in peak (FRIP) score and fragment length distribution were calculated as described in Corces and colleagues (21), and TSS enrichment score was calculated using ATACseqQC version 1.18.1 (31). To ensure the quality of our ATAC-seq dataset, we considered the samples that met the following criteria.

$$(FRIP\ score > 0.05) \text{ OR } (TSS\ Enrich\ score > 8) \text{ OR } (\text{ATACseq Peak counts} > 15,000)$$

Five mCRPC samples that failed to satisfy the above criteria were discarded from this study. This resulted in the final set of 70 samples representing 65 unique patients that were used throughout the study.

Consensus ATAC-seq peaks

Nonoverlapping unique ATAC-seq narrow peaks regions were obtained from the samples analyzed. Those nonoverlapping unique peak regions present in at least two samples were considered consensus peaks. Sequencing reads mapped to the consensus peak regions were counted using the “featurecount” function within Rsubread (32) R-package with the “isPairedEnd = TRUE, countMultiMappingReads = FALSE, maxFragLength = 100, autosort = TRUE” option. The read counts of the consensus peaks were normalized with the reciprocal of the size factor and variance-stabilized transform method available in the DESeq2 (33) R-package. We note that ATAC-seq peak lengths are highly variable and so are the lengths of consensus ATAC-seq peaks. Importantly, ATAC-seq read counts tend to be higher for longer peaks, which are not corrected using DESeq2. To ensure accurate comparisons of the ATAC-seq peaks, throughout the study comparison is always made between ATAC-seq peaks of the same lengths and never between two peaks of unequal lengths.

The ATAC-seq data (read count profiles) of mCRPC samples from our study were combined with those from Tang and colleagues (22) and adjusted for potential batch effects using the “ComBat” function from the “sva” R-package. Additionally, we attempted to correct potential batch effects in a larger dataset that combined ATAC-seq profiles of benign prostate, localized prostate cancer, mCRPC adenocarcinoma, and t-SCNC/NEPC from different datasets. However, this correction was not possible because some sample phenotypes (covariates) were inseparable from the dataset (batch) they came from.

Differential ATAC-seq and RNA-seq analysis

Differential ATAC-seq and RNA-seq analysis for two groups comparison was conducted using the DESeq2 (33) R-package. The normalized read counts of the consensus peaks were used in the case of the ATAC-seq data. Peaks/genes with the Benjamini–Hochberg adjusted P value ≤ 0.01 and $|\log_2 \text{folchange}| \geq 1$ were considered statistically significant. For multiple group (three or more) comparisons of the ATAC-seq dataset, we used the Kruskal–Wallis test, and the peaks with P value ≤ 0.001 were considered statistically significant.

Pathway enrichment analysis

To test the association of signaling pathways enriched in the accessible chromatin regions (ATAC-seq peaks), we performed

GREAT (34) enrichment analysis using rGREAT R-package. In the case of enrichment analysis of a list of genes, we used a hypergeometric test-based overrepresentation analysis. We used the set of signaling pathways genesets in the Reactome, Hallmark pathway, and GO biological process present in the Molecular Signature Database (MSigDB; ref. 35) v7.5.1.

Calculation of AR and NE score

The “singscore” (36) R-package was used to calculate AR and NE scores for each mCRPC sample. The NE score was calculated using the NE genes reported by Beltran and colleagues (7), and the AR score was calculated using the gene expression profile of the “HALLMARK_ANDROGEN_RESPONSE” geneset from MSigDB (Supplementary Table S2).

Transcription factor footprinting

TF footprints were analyzed using Transcription factor Occupancy prediction By Investigation of ATAC-seq Signal (TOBIAS; ref. 37) version 0.12.11. For TF footprinting analysis, we omitted the step of shifting the position of aligned reads in the BAM file in our ATAC-seq data processing pipeline (Supplementary Fig. S2) as this step was already incorporated within the TOBIAS. We called these BAM files and resulting ATAC-seq peaks “unshifted.” The input data for TOBIAS were prepared as follows. The unshifted ATAC-seq BAM files of all samples within a subtype were merged using the “MergeSamFiles” (Picard) function. A consensus nonoverlapping set of unshifted ATAC-seq peaks present in at least two samples in the subtype was generated. A comprehensive list of 541 unique human TFs was compiled by combining TFs from the JASPAR (38) CORE database and refined AR-binding motifs (full site, half site, lenient site, and extended site) from Wilson and colleagues (39).

First, the insertion bias of the Tn5 transposase was corrected using the “ATACorrect” function taking the merged unshifted ATAC-seq BAM files and the merged unshifted ATAC-seq peak regions as inputs. The resulting bigWig files were assigned footprinting scores across all accessible chromatin regions using the function “ScoreBigwig.” Finally, the scored footprints were matched to the curated list of TF motifs described above, and then differential scores for each motif were determined for each subtype comparison using the function “BNDetect” with parameters “–motif-pvalue 0.0001; –bound-pvalue 0.001.” TOBIAS categorizes every predicted TF-binding site (for each TF motif) into bound and unbound states based on a score threshold per subtype compared. The threshold was set at the level of significance (bound P value = 0.001) of a normal distribution fit to the background distribution of scores.

By utilizing this method, we conducted an analysis of differential TF footprinting. This involved comparing the TF footprints of each mCRPC transcriptional subtype against the others. As a result, we made four comparisons for each subtype, leading to a total of 10 comparisons. From each comparison, the subtype-associated TF hits were prioritized based on their differential binding score ($s \geq |0.1|$) and associated q -value [top 80% percentile of $-\log_{10}(q \text{ value})$]. The pairwise differential TF footprint comparison results between mCRPC subtypes were further filtered using their gene expression profiles. TFs not expressed (or with negligible expression level) in relevant subtypes were omitted. These results were visualized in the form of a circularized heatmap. To generate the heatmap, we computed the TF occupancy phenotype score for each mCRPC subtype, which is determined by the product of the absolute value of the TF-binding score and the absolute value of $-\log_{10}(q \text{ value})$.

linked to that specific motif. If a TF is not enriched in a particular subtype, its score is set to 0.

For every mCRPC sample, TF footprinting analysis was also conducted in a single sample mode. For this, all TOBIAS TF footprinting functions “ATACorrect,” “ScoreBigwig,” and “BNDetect” were executed using identical parameters to that of the subtype-level analysis except that only one condition (i.e., respective mCRPC sample) was used in “BNDetect.”

Evaluation of the accuracy of TF footprint sites

To assess the accuracy of TOBIAS in predicting TF footprint sites, we compared its predictions with the TF-binding sites predicted by chromatin immunoprecipitation sequencing (ChIP-seq). We utilized publicly available ChIP-seq data for AR, FOXA1, and HOXB13 measured in mCRPC from Pomerantz and colleagues (15). The ChIP-seq TF-binding sites were considered as the reference or the ground truth. The objective was to determine if TF footprints derived from ATAC-seq peak regions could accurately capture the ChIP-seq-predicted TF-binding regions.

Given that the ChIP-seq experiment on mCRPC was limited to AR, FOXA1, and HOXB13, we restricted our TF footprint evaluation experiment to these three TFs. First, we ran TOBIAS on our ATAC-seq samples using different “-bound-pvalue” (0.5, 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001) in the “BNDetect” function. The *P* value threshold varies the bound/unbound status assignment for the predicted TF footprint sites. Each run of TOBIAS using different *P* value thresholds resulted in a different set of TF footprint sites. These TF footprint sites were compared against those observed in the ChIP-seq TF-binding sites. Overlap of at least 1 bp between the motif sites from ChIP-seq and ATAC-seq was considered a hit. For each experiment, we computed the predictions’ true positive rate and false positive rate. We then generated ROC curves and calculated the area under the ROC curve. We repeated this analysis for every subtype of mCRPC.

Reconstruction of TF-target gene regulatory network

The reconstruction of the TF-target gene regulatory network comprises two major steps: (i) prediction of TF footprints in accessible chromatin using ATAC-seq (linking TF peaks as described above) and (ii) associating ATAC-seq peaks that may potentially regulate the expression of individual genes (linking peak and gene). Finally, the TF-peak and peak-gene association results were combined to obtain the TF-gene association.

Peak-to-gene linking predictions

We used a correlation-based approach to predict potentially causal links between ATAC-seq peaks and gene expression. First, the mCRPC RNA-seq gene expression analysis was restricted to protein-coding genes and filtered out genes not expressed ($TPM = 0$) in more than 25% of all samples. The ATAC-seq peaks were restricted to those with evidence of TF footprints of at least one out of 541 TFs measured. Our analysis was further restricted to measure *cis*-interaction (peak-gene association within the same chromosome). A pairwise Spearman’s correlation was measured between every combination of *cis*-interacting peak-gene pairs across all mCRPC samples. A permutation experiment was performed to evaluate the robustness of our peak-gene correlation (see below). The majority of the pairwise peak-gene pairs were random correlations. This warranted an unbiased statistical approach to identify strong correlated or anticorrelated peak-gene pairs. Thus, to identify the most confident set of peak-gene interaction pairs, we

interrogated the enrichment of H3K27ac marks on the peak-gene pairs. We used three sets of parameters to scrutinize the peak-gene interaction pairs: Spearman’s correlation coefficient (*R*), the *P* value of peak-gene interaction pairs through the permutation experiment, and the distance of the peak to TSS. We used several combinations of values of these three parameters. The peak-gene interaction pairs generated using the parameters $R \geq |0.4|$, $P \text{ value} \leq 0.05$, and $\text{distance to TSS} \leq 500 \text{ kb}$ had the highest enrichment of H3K27ac peaks. Thus, we used these parameters to prune off weak peak-gene links.

Permutation experiment

The gene labels of the RNA-seq gene expression data were randomly permuted, and the pairwise peak-gene Spearman’s correlation was measured across all mCRPC samples. This process was repeated 1,000 different times. Using the density distribution of all permuted Spearman’s correlation coefficients, we calculated *P* values for each peak-gene interaction pair.

Reconstruction of the mCRPC subtype-specific regulatory network

For each mCRPC subtype, we extracted TF associated with the subtype and genes linked to the subtype-associated ATAC-seq peaks. This resulted in the TF-target gene links associated with the mCRPC subtype.

Inferring the effects of ZNF263 binding on transcription of downstream target genes

To investigate how ZNF263 binding influences the expression of the nearest gene in mCRPC, we evaluated the changes in gene expression levels that occurred when ZNF263 was present or absent in the gene’s promoter region. To conduct this analysis, we performed TF footprinting on individual mCRPC samples. For each gene, we grouped the 70 mCRPC samples based on whether ZNF263 was present or absent in the promoter region of the gene. Genes with at least two samples in each group and at least one sample with expression level $\log_2 \text{ TPM} > 5$ were included for further analysis. Then we performed the Wilcoxon rank-sum test and measured the fold change in the gene expression levels between the two groups. Genes with the Benjamini-Hochberg adjusted *P* value ≤ 0.05 and $|\log_2 \text{ folchange}| \geq 1$ were considered as statistically significant. The above analysis was also repeated using samples within each subtype (AR+NE-, AR_{low}NE-, and AR-NE-). We note that due to limited samples in AR-NE+, AR-NE-, and AR+NE- subtypes, none of the genes were statistically significant.

Tumor purity estimation

Tumor purity of the mCRPC samples used in this study was calculated using the PURPLE tool (<https://github.com/hartwigmedical/hmftools>) based on whole-genome sequencing profiles of the corresponding tumor samples. In brief, PURPLE combines B-allele frequency, read depth ratios, somatic variants, and structural variants to estimate the purity of a tumor sample.

Data availability

The ATAC sequencing data generated in this study are available from the European Genome-Phenome Archive (EGA) under the accession number EGAS00001006698. RNA sequencing from the same tumors is available under the accession numbers EGAD00001008487, EGAD00001008991, and EGAD00001009065.

The RNA-seq data from the WCDT mCRPC cohort were aligned with STAR and quantified at the gene level for Gencode v28

transcripts as previously described (16). The raw RNA-seq fastq files from the PAIR (40) cohort (GSE115414) and CPCG (41) cohort (EGAD00001004424) were processed and analyzed as described above. The mRNA expression data of non-diseased tissues were obtained from the GTEx portal (<https://www.gtexportal.org/>).

The publicly available ATAC-seq data used in this study were obtained from Pomerantz and colleagues [Sequence Read Archive (SRA) BioProject; accession number PRJNA540151; ref. 15], Cejas and colleagues (PRJNA691927; ref. 42), and Tang and colleagues (PRJNA818767; ref. 22). The aligned BAM files from The Cancer Genome Atlas localized prostate cancer ATAC-seq data of Corces and colleagues (21) were obtained from the NIH Genomic Data Commons portal (<https://portal.gdc.cancer.gov/>). These ATAC-seq data were processed and analyzed using the same ATAC-seq data processing pipeline described above.

mCRPC PDX ChIP-seq data for AR, FOXA1, HOXB13, and H3K27ac were obtained from Pomerantz and colleagues (15). Briefly, raw ChIP-seq data were downloaded from SRA (PRJNA540151). Reads with base quality scores more than 30 across all bases were aligned using bwa-mem v0.7.17 to build hg38. The aligned reads were deduplicated, and peaks were called using MACS2 v.2.2.5, with an FDR threshold of 0.01. Peaks in the ENCODE hg38 blacklist (ENCSR636HFF) were excluded, and only peaks that were enriched at least 10-fold more than background were kept for further analysis.

The ZNF263 ChIP-seq peak genomic regions were obtained from Imbeault and colleagues (GSE78099; ref. 43), Frierze and colleagues (GSE19235; ref. 44), and Pope and colleagues (GSE31477; ref. 45), and MYC ChIP-seq peak genomic regions were obtained from Barfield and colleagues (GSE73994; ref. 46), See and colleagues (GSE164777; ref. 47), and Guo and colleagues (GSE157105; ref. 48). The ChIP-seq peak regions were uplifted to hg38 before comparing against the TF footprint regions from mCRPC samples. Further, ChIP-seq profiles of 157 TFs were downloaded from the ChIP-Atlas database (49).

Code availability

Code used in this manuscript is available at https://github.com/DavidQuigley/WCDT_ATAC_mCRPC.

Results

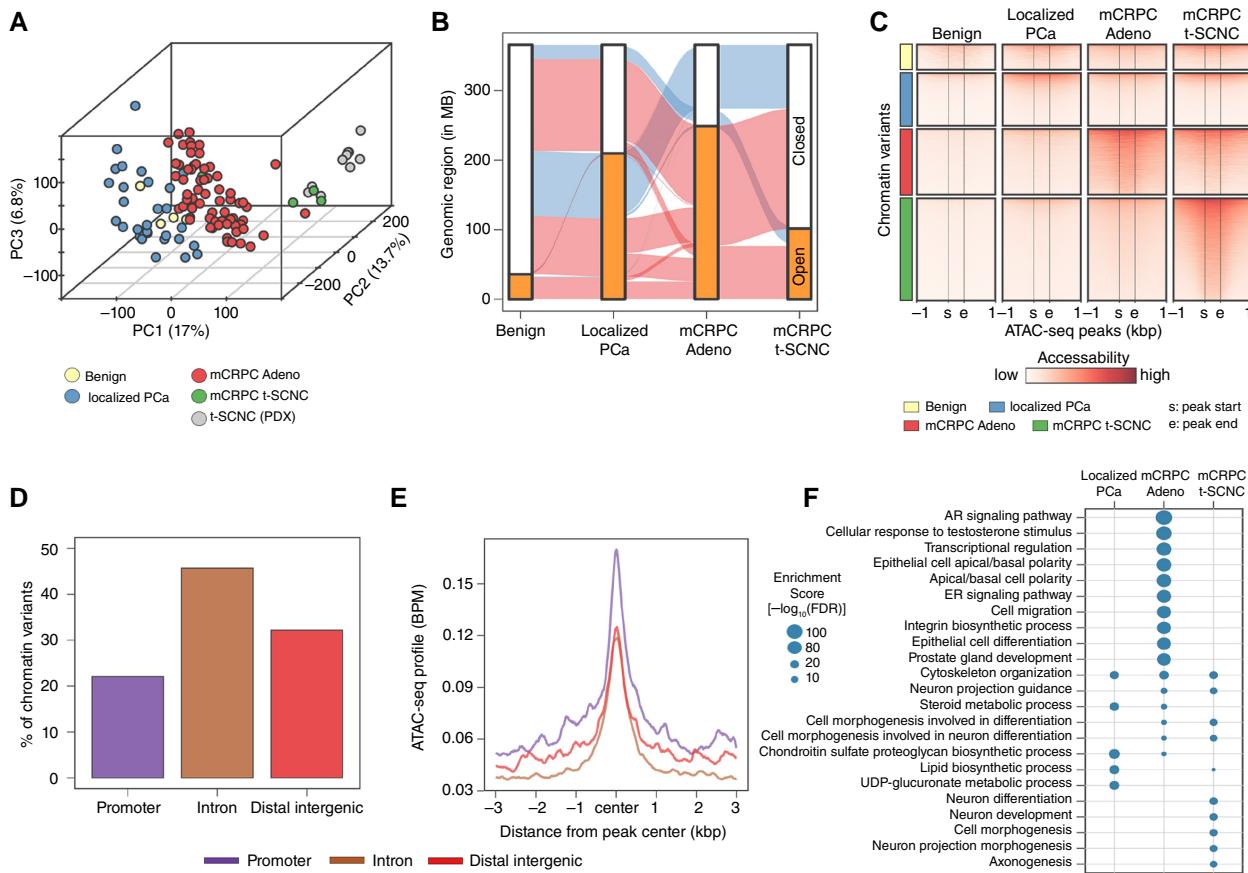
A prospective multi-institution Institutional Review Board-approved study (NCT02432001) obtained fresh frozen core biopsies of metastases from patients with mCRPC as described previously (12, 16, 17). To create a rigorous atlas of accessible regulatory DNA elements active in metastatic prostate cancer, we performed ATAC-seq on 70 mCRPC tissue biopsies obtained from various anatomic locations representing 65 unique patients (Supplementary Fig. S1; Supplementary Table S1). The ATAC-seq data was processed using an in-house pipeline (Materials and Methods; Supplementary Fig. S2). Sequencing was performed to a remarkably high depth of 204 to 411 million reads (mean 308 million; Supplementary Fig. S3A). This resulted in an average of 81,215 (range of 22,497–157,071) ATAC-seq peaks per sample, marking accessible chromatin regions. We inspected our ATAC-seq dataset using quality control metrics such as the FRIP and TSS enrichment score. Sample sequencing depth was not significantly correlated with the number of peaks detected or any quality metrics. The number of peaks in the sample was significantly correlated with the sample FRIP score (Spearman's correlation coefficient, $R = 0.79$, P value

$= 5.8 \times 10^{-16}$; Supplementary Fig. S3B–D). Tumors with higher estimated purity had higher FRIP scores ($R = 0.4$, P value $= 4.4 \times 10^{-5}$; Supplementary Fig. S3E), suggesting that the tumor content in the tissue sample influenced the ATAC-seq sample quality. ATAC-seq peaks have previously been reported to occur most frequently in intronic and distal intergenic regions followed by gene promoter regions, to be enriched at the TSS, and to demonstrate read fragment size periodicity correlated to the integer multiples of the nucleosome (20, 21). Our ATAC-seq data were consistent with those reports (Supplementary Fig. S3A, S3F, and S3G). These findings confirmed the high quality of our ATAC-seq data, consistent with the reports from other groups (20, 21).

Chromatin accessibility changes during prostate cancer progression affect stage-specific regulatory elements

To define how chromatin accessibility is altered during prostate cancer progression, we compared the ATAC-seq profile of our mCRPC samples with publicly available ATAC-seq datasets of benign prostate ($n = 4$; ref. 15), localized prostate cancer ($n = 26$; ref. 21), and NE PDX models ($n = 6$ with replicates; ref. 42). The ATAC-seq quality metrics, including the FRIP scores, of our mCRPC samples were comparable to the publicly available datasets (Supplementary Fig. S4). We first created a consolidated dataset of ATAC-seq peaks by merging the complete dataset. This produced a nonoverlapping set of 348,799 consensus genomic regions that were accessible in at least two samples. To test whether this analysis would show systematic differences in ATAC-seq profiles corresponding to tumor stage, we performed principal component analysis (PCA) using the normalized read counts of consensus accessible regions. Our analysis revealed that mCRPC had distinct chromatin accessibility profiles compared with localized prostate cancer and benign prostate tissue (Fig. 1A; Supplementary Fig. S5). Importantly, among the mCRPC cohort, t-SCNC/NEPC samples were found to have distinct chromatin accessibility profiles. Corroborating with the prior observations (15–17), benign prostate and localized prostate cancer had similar chromatin accessibility profiles. A comparison of the accessible chromatin regions identified that mCRPC had the highest number of genomic regions with accessible chromatin conformation (Fig. 1B). This corresponds to our knowledge of the genome-wide loss of methylation in mCRPC (16, 17) and the observation that increased AR expression results in genome-wide chromatin relaxation (50). Accessible ATAC-seq peaks were extensively shared across localized prostate cancer, mCRPC, and t-SCNC/NEPC samples (Fig. 1B) and were enriched in genes involved in DNA damage repair, apoptosis, and immune system signaling processes (Supplementary Fig. S6).

Chromatin accessibility is known to significantly affect the transcription of nearby genes (21). We noted that there was a robust correlation between the chromatin accessibility of a gene's promoter and the expression of its corresponding gene (mean Spearman's correlation coefficient, $R = 0.4$; Supplementary Fig. S7A and S7B). We found that the differentially expressed genes that were expressed at higher levels were more likely to have accessible chromatin nearby, suggesting that changes in chromatin accessibility can contribute to differences in gene expression (Supplementary Fig. S7C). Motivated by this finding, to identify chromatin variants, the regions of the genome that differ in chromatin accessibility (i.e., differentially accessible ATAC-seq peaks), between various stages of prostate cancer progression, we performed a differential accessibility analysis comparing each stage. We further annotated these peaks with the nearest gene and genomic region in which the

**Figure 1.**

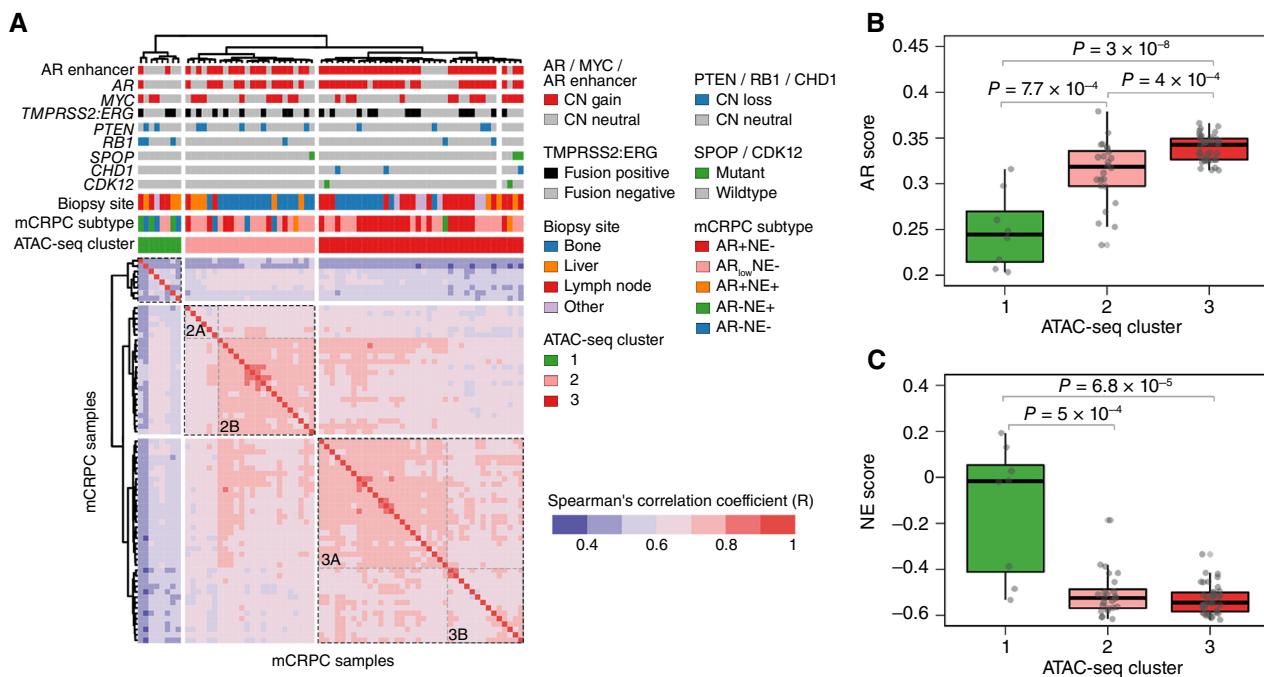
Chromatin accessibility changes during prostate cancer progression affect stage-specific regulatory elements. **A**, PCA of the ATAC-seq profiles comparing different stages of prostate cancer including benign prostate, localized prostate cancer, mCRPC Adeno, and mCRPC t-SCNC. The normalized read counts of these consensus accessible regions were used for the PCA analysis. Each dot in the plot represents an individual sample. **B**, Alluvial plot demonstrating changes in accessible chromatin regions in various stages of prostate cancer (PCa). Each bar corresponds to a distinct prostate cancer stage, with the orange and white sections indicating accessible and inaccessible chromatin regions, respectively. The shaded areas connecting the bars represent changes in the accessibility of these chromatin regions. The pink and blue shaded regions represent accessible and inaccessible chromatin regions in mCRPC Adeno, respectively. **C**, Heatmap representation of the chromatin between different stages. The rows are segregated by the chromatin variants in each stage. **D**, The percentage of chromatin variants in mCRPC Adeno. The ATAC-seq peaks are grouped by the genomic regions (promoter, intron, or distal intergenic) to which they are mapped. **E**, ATAC-seq profile plot illustrating potential regulatory regions in chromatin variants in mCRPC Adeno. The profile plot represents the overlapping region between the chromatin variants and publicly available H3K27ac ChIP-seq data from mCRPC PDX. **F**, Enrichment of chromatin regions exclusively accessible in localized prostate cancer, mCRPC Adeno, or mCRPC t-SCNC against GO biological processes.

peak is located (Materials and Methods). We identified 76,311 (21.9%) peaks that exhibit a significant change in accessibility in at least one stage (Fig. 1C). The majority of these peaks were exclusively detected in mCRPC including the t-SCNC samples (Fig. 1B and C). These peaks included promoters of prostate cancer-relevant genes such as *AR*, but most chromatin variants were observed in introns and distal intergenic regions rather than in promoters (Fig. 1D). ATAC-seq peaks in distal intergenic regions of genes such as *AR*, *CHGA*, *DNMT3A*, and *PIK3R1* were exclusively detected in mCRPC samples. To assess whether chromatin variants were more likely to harbor regulatory DNA, we intersected these regions with measures of H3K27ac, a histone mark associated with active enhancers, in mCRPC (15). We observed increased chromatin accessibility of intergenic regions, introns, and promoters within regions of the H3K27ac signal (Fig. 1E). We predicted the gene pathways activated by these chromatin variants using the GREAT

enrichment tool (34) and found that the chromatin regions exclusively accessible in mCRPC Adeno were enriched in the AR signaling pathway, cell migration, and prostate development processes (Fig. 1F). Accessible chromatin regions in t-SCNC samples were enriched in neuronal development and differentiation processes. Similarly, chromatin regions accessible in localized prostate cancer were enriched in lipid biosynthetic and metabolic processes and changes in cytoskeleton organization. These results imply that the chromatin variants associated with tumor progression potentially deregulate the oncogenic signaling required for malignant transformation.

Chromatin accessibility in mCRPC is associated with subtypes linked to androgen signaling

To investigate the global patterns of chromatin accessibility in mCRPC, we performed unsupervised hierarchical clustering on

**Figure 2.**

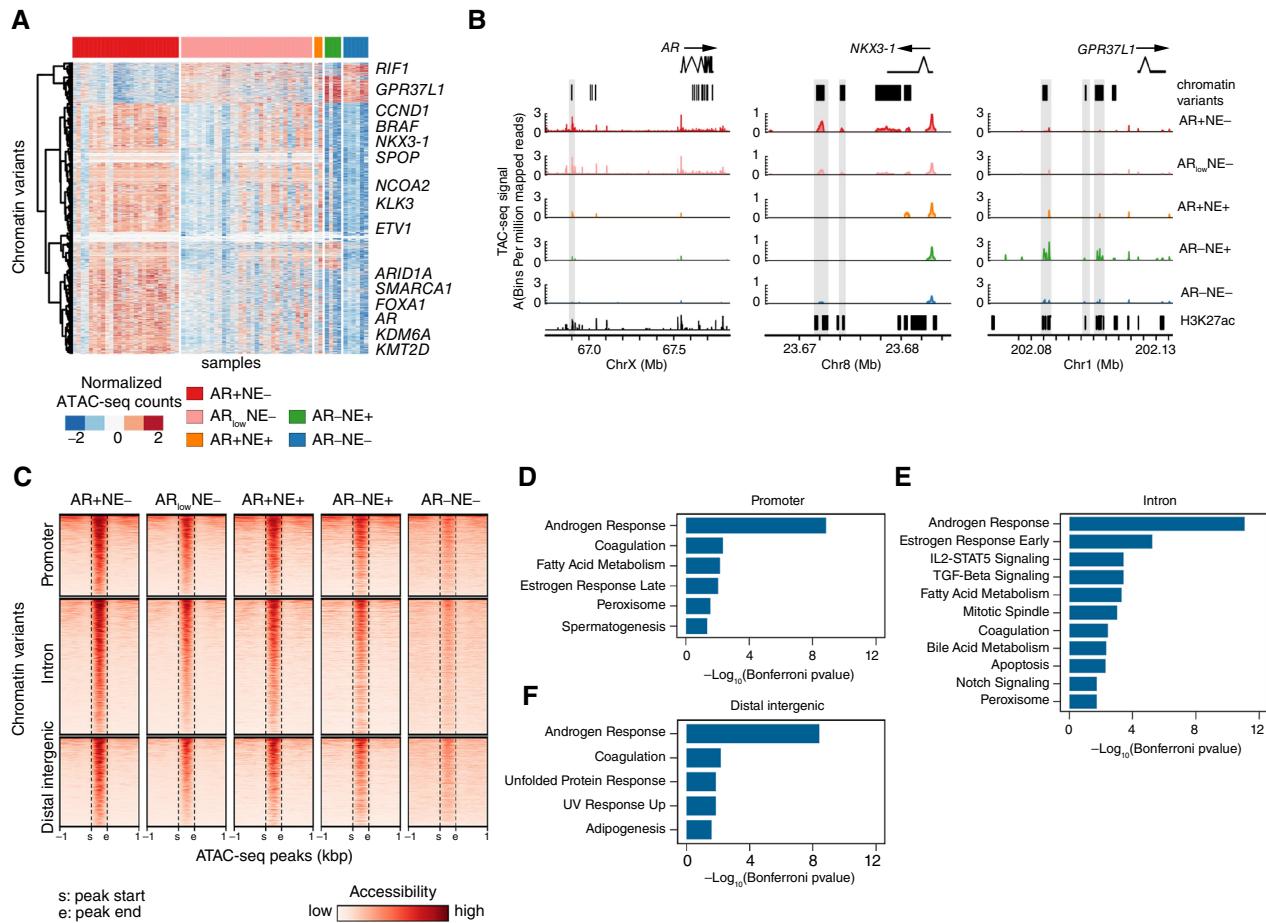
Chromatin accessibility in mCRPC is associated with subtypes linked to androgen signaling. **A**, Unsupervised hierarchical clustering of pairwise sample Spearman correlation based on the normalized read counts of consensus ATAC-seq peaks of mCRPC. **B** and **C**, Distribution of AR pathway score (**B**) and NE score (**C**) calculated based on RNA-seq gene expression profiles of mCRPCs classified into individual clusters in **A**. Statistically significant Wilcoxon rank-sum test *P* values between the clusters are indicated in the plot.

ATAC-seq profiles from 70 mCRPC samples. This was achieved by applying pairwise Spearman's correlation to the normalized read counts of consensus accessible peaks. Our analysis revealed three distinct clusters of mCRPC samples (Fig. 2A; Supplementary Fig. S8). These clusters were not associated with metastatic tissue sites or alterations of driver genes such as *PTEN* or *RB1*. However, cluster assignments were correlated with tumor AR pathway and NE scores assessed by gene expression signature analysis (Materials and Methods; Fig. 2B and C; Supplementary Table S2). Tumors in cluster 3 had significantly higher AR scores and lower NE scores (Wilcoxon rank-sum test *P* value: 3×10^{-8} and 4×10^{-4}), whereas tumors in cluster 1 had low AR scores but high NE scores (*P* value: 6.8×10^{-5} and 5×10^{-4}). Furthermore, leveraging AR and NEPC gene signatures from Labrecque and colleagues (5), we recently stratified our mCRPC samples into five subtypes (AR+NE, AR_{low}-NE-, AR+NE+, AR-NE+, and AR-NE-) based on their RNA-seq gene expression profiles (51). Out of 70 mCRPCs analyzed in this study, 26 tumors were classified as AR+NE-, 32 tumors as AR_{low}-NE-, two tumors as AR+NE+, four tumors as AR-NE+, and six tumors as AR-NE- (Supplementary Table S1). Cluster 1 was associated with the AR-NE+ subtype (Fisher's exact test; *P* value = 0.003), whereas cluster 3 was associated with AR+NE- subtype (*P* value = 0.005). Cluster 2 had mixed sample phenotypes, with almost 58% of the samples in the group associated with AR_{low}-NE- subtypes and 25% of the samples associated with the AR+NE- subtype. We further observed that clusters 2 and 3 each contained two subclusters designated as 2A and 2B and 3A and 3B, respectively. These subclusters were primarily distinguished by variations at the tissue level. Notably, both subclusters 2B (*P* value = 0.002) and 3A

(*P* value = 0.002) were enriched with bone metastatic tissue biopsies. A recent study of chromatin availability in cell lines and organoid models by Tang and colleagues (22) reported the existence of four ATAC-seq subtypes of CRPC: CRPC-AR, CRPC-NE, CRPC-WNT, and stem cell-like (CRPC-SCL) subtypes. To compare these subtypes with our cohort of mCRPC samples, we analyzed ATAC-seq data from both studies together. Unsupervised hierarchical clustering of ATAC-seq profiles showed that the majority of samples in CRPC-AR and AR+NE- subtypes were a part of the same cluster (Supplementary Fig. S9). Similarly, samples in CRPC-NE and CRPC-WNT subtypes clustered together with the AR-NE+ subtype. Samples in the CRPC-SCL subtype were clustered with AR_{low}-NE- and AR-NE- subtypes. These results indicate that the chromatin accessibility in our mCRPC data was linked most strongly with mCRPC transcriptional subtypes.

mCRPC transcriptional subtypes are associated with chromatin variants of prostate cancer signaling pathways

After establishing that androgen signaling was significantly associated with chromatin accessibility status genome-wide, we performed a supervised analysis to identify chromatin loci in which accessibility status was correlated with the five mCRPC transcriptional subtypes from Labrecque and colleagues (5). This analysis was motivated by our observation that chromatin accessibility proximal to the genes that defined mCRPC transcriptional subtype signatures was correlated with the gene expression of the corresponding gene (Supplementary Fig. S10). Moreover, there was no significant difference between the

**Figure 3.**

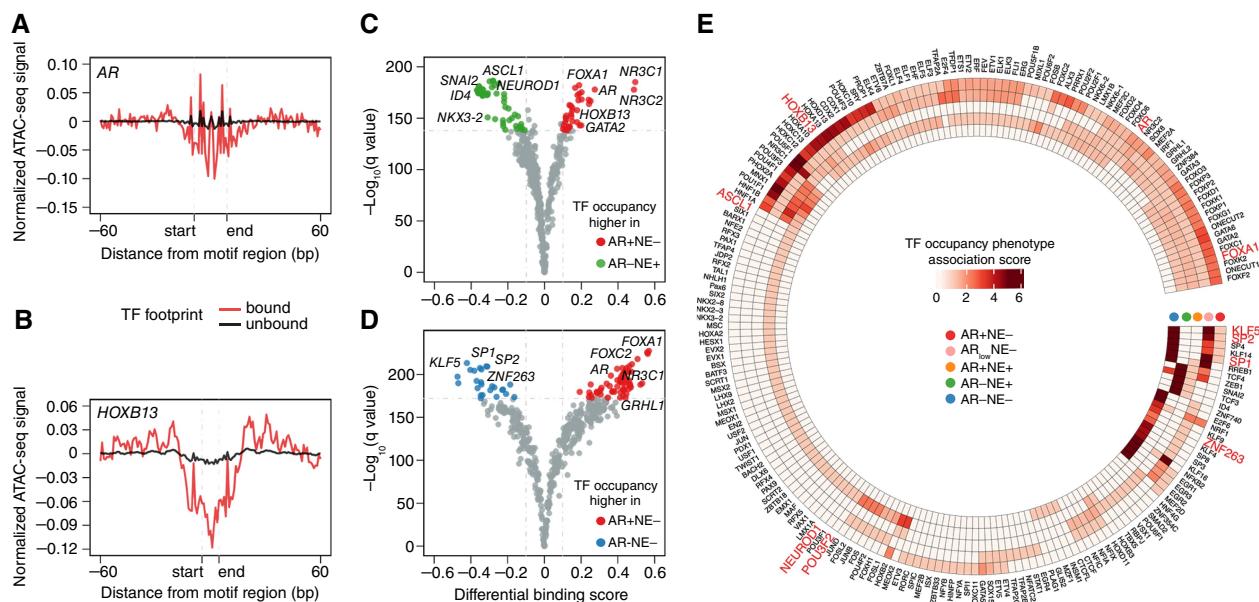
mCRPC transcriptional subtypes are associated with chromatin variants of prostate cancer signaling pathways. **A**, Heatmap of chromatin variants between mCRPC transcriptional subtypes. **B**, ATAC-seq peaks around *AR*, *NFKX3-1*, and *GPR37L1* gene regions. The highlighted vertical strip illustrates the presence of ATAC-seq peaks at the enhancer region. **C**, Heatmap representation of the chromatin variants between mCRPC transcriptional subtypes. The rows are segregated by the differential regions mapped to gene promoter, intron, or distal intergenic regions. **D–F**, Hallmark pathway enrichment of chromatin variants between the mCRPC transcriptional subtypes mapped to promoter (**D**), intron (**E**), and distal intergenic (**F**) regions identified using GREAT enrichment analysis (see Materials and Methods).

numbers of peaks and FRIP score among the mCRPC transcriptional subtypes (Supplementary Fig. S11).

Therefore, we tested for differential chromatin accessibility across the five mCRPC subtypes. This analysis identified 6,704 ATAC-seq peaks with significant differences in accessibility across all samples in each of the subtypes (Supplementary Table S3). AR+NE- tumors most frequently harbored increased accessibility, followed by AR_{low}NE- tumors (Fig. 3A), in prostate cancer-associated genes such as *AR*, *KLK3*, *FOXA1*, *NFKX3-1*, *SPOP*, *ZBTB16*, and *NCOA2*. Similarly, regions around several epigenetic drivers of prostate cancer such as *ARID1A*, *SMARCA1*, *KMT2D*, and *KDM6A* were more accessible in AR+NE- and AR_{low}NE- tumors compared with the remaining subtypes. Interestingly, most chromatin variants between the mCRPC subtypes were annotated distal to the TSSs (i.e., distal intergenic peaks); about 75% of chromatin variants were located more than 3 kb from TSS (Supplementary Fig. S12A). These regions may represent enhancer regions active in specific subtypes of the disease. One such example was a distal accessible region upstream of the AR that we and others have previously identified as

a driver of ADT and ARSI resistance in AR-positive disease (12, 13). We additionally identified chromatin variants in distal regions near *NFKX3-1* in AR+NE- and AR_{low}NE- subtypes and *GPR37L1* in the AR+NE- subtype as compared with the remaining subtypes (Fig. 3B). *GPR37L1* encodes for a G protein-coupled receptor protein almost exclusively expressed in the nervous system, and studies in murine models have suggested it to have a neuro-protective function (52). *AR* and *NFKX3-1* mRNA were expressed in AR+NE- and AR_{low}NE- subtypes and *GPR37L1* mRNA was highly expressed in the AR-NE+ subtype (Supplementary Fig. S12B–D). Moreover, these chromatin variants in *AR*, *NFKX3-1*, and *GPR37L1* were correlated with their corresponding gene expression (Supplementary Fig. S12E–G).

To gain insight into the functional role of accessible chromatin, we extracted the chromatin variants across all mCRPC transcriptional subtypes that mapped to the promoter, intron, and distal intergenic regions (Fig. 3C). Chromatin variants in the promoter, intron, and distal intergenic regions were enriched for the Hallmark androgen response pathway (Fig. 3D–F). Chromatin regions

**Figure 4.**

mCRPC transcriptional subtypes are defined by DNA accessibility-guided patterns of TF regulation. **A** and **B**, ATAC-seq TF (AR and HOXB13) footprint signal difference between TF-bound and TF-unbound sites. **C** and **D**, Volcano plot of differential TF footprint occupancy analysis comparing the AR+NE- and AR-NE+ subtypes (**C**) and AR+NE- and AR-NE- subtypes (**D**). Each dot in the plot represents a TF motif. The colored dots indicate a significantly differentially bound TF motif (data available as Supplementary Table S4). **E**, Heatmap of genome-wide active TF occupancy, determined by TF footprints, associated with different mCRPC transcriptional subtypes. Each rim of the circular heatmap represents an individual mCRPC transcriptional subtype, and the sector represents TF. The darker color shade indicates the strong association of the TF with the respective mCRPC subtype. See Materials and Methods for details on TF occupancy phenotype score calculation.

mapped to intron and intergenic were also enriched in oncogenic and proliferative signaling pathways. This underscores the functional importance of intronic and distal intergenic chromatin variants in potentially modulating the epigenetic landscape of mCRPC transcriptional subtypes.

mCRPC transcriptional subtypes are defined by DNA accessibility-guided patterns of transcription factor regulation

Following our observation that the regulatory elements are enriched in the accessible chromatin region in mCRPC, we built a comprehensive catalog of TF occupancy across mCRPC transcriptional subtypes. TF binding to chromatin prevents Tn5 cleavage within the binding site and generates depletion in ATAC-seq coverage known as “TF footprints” (53). We interrogated TF footprint signals in the accessible chromatin of mCRPCs using the TOBIAS (37) software tool. First, to evaluate the reliability of the predicted TF footprints, we compared the AR, FOXA1, and HOXB13 footprints predicted in our mCRPC samples against the respective ChIP-seq binding sites observed in mCRPC PDX samples obtained from Pomerantz and colleagues (Materials and Methods; ref. 15). Our analysis found that TOBIAS-predicted TF footprint sites were correlated to experimentally observed ChIP-seq binding sites (Supplementary Fig. S13A–F). For each TF, TOBIAS classifies every predicted TF-binding site as a bound or unbound state based on a user-defined footprint score threshold (Materials and Methods). The bound TF footprint sites were observed to have a depleted ATAC-seq accessibility signal as compared with the unbound sites (Fig. 4A and B). Moreover, a significant proportion (82%–96%) of

the predicted TF-bound footprint sites were found to coincide with the regions identified by ChIP-seq peaks as compared with the unbound sites (Supplementary Fig. S13G–I). This suggests that our predictions of TF footprints are strong and reliable.

Next, we performed an unbiased genome-wide active TF occupancy analysis using 541 human TFs (Supplementary Fig. S14) from the JASPAR (54) database and examined their association with different mCRPC transcriptional subtypes. We performed an unbiased differential TF footprint occupancy analysis comparing each mCRPC transcriptional subtype to the other (Fig. 4C and D; Supplementary Fig. S15). Subtype-associated TF hits were prioritized based on their differential binding score and associated probability score (Materials and Methods; Supplementary Table S4). For example, AR, FOXA1, HOXB13, and NR3C1 were identified to have differential TF footprint occupancy in the AR+NE- subtype compared with the AR-NE+ or AR-NE- subtype (Fig. 4C and D; Supplementary Table S4). Similarly, TFs such as ASCL1, NEUROD1, SNAI2, ID4, and NKX3-2 were identified to have differential TF footprint occupancy in the AR-NE+ compared with the AR+NE- subtype. AR-NE- tumors were enriched for high TF footprint occupancy of SP1, SP2, ZNF263, and KLF5 (Fig. 4D).

We aggregated the results of significantly differential occupied TFs from all 10 pairwise comparisons between the mCRPC subtypes and identified 203 TFs enriched in distinct mCRPC subtypes (Fig. 4E; Supplementary Table S5). To further validate our predictions through computational analysis, we compared our predicted TF footprints to publicly available data on binding sites for 120 TFs obtained using ChIP-seq (Supplementary Fig. S16). Our analysis revealed that for 94 of these factors, at least 25% of our predicted

footprint locations overlapped with the ChIP-seq binding sites. For TFs known to be associated with prostate cancer, such as FOXA1, AR, ERG, HOXB13, NR3C1, ASCL1, and GATA2, the overlap between our predicted footprints and the ChIP-seq data was even greater, with more than 80% concordance.

We observed that TFs such as AR, FOXA1, HOXB13, GATA2, SP1, SP2, and KLF5 were enriched in multiple subtypes. Eighty-four TFs were associated with both AR+NE- and AR_{low}NE- subtypes, indicating similarity in gene transcription regulation between these two subtypes. About 80% (20 of 25) of AR-NE- TFs were also associated with the AR_{low}NE- subtype corroborating previous (5) observations of transcriptomic similarity between AR_{low}NE- and AR-NE-. Interestingly, each subtype of mCRPC shows a distinct affinity toward certain subtype-specific TFs. For example, a large set of 49 TFs, comprising ASCL1, NEUROD1, NKX3-2, and POU3F2, was found to be exclusively associated with the AR-NE+ subtype. ASCL1 is a proneural TF that acts as a driver of the neuronal transcriptional program to support treatment resistance in the AR-NE+ subtype (42, 55). Similarly, POU3F2 is a neural TF that is directly suppressed by AR and mediates NE differentiation and treatment resistance in the AR-NE+ subtype (56). NEUROD1 is a neuronal TF associated with neuronal development in both NEPC and small-cell lung cancer (42). Our results provide a unique opportunity to comprehensively interrogate several TFs associated with mCRPCs.

Knowing where TFs bind on the genome is important because it can provide valuable insights into the gene expression regulatory mechanisms. Therefore, we examined the locations of all predicted TF footprints in the genome. We observed that in all TFs, the footprints were identified in both the promoter and distal to the promoter regions of the nearest gene (Supplementary Fig. S17). In 72% (147 of 203) of the TFs, the majority of the footprints were observed at a distance of more than 3 kb from the nearest gene, whereas in the remaining 28% (56 of 203) of the TFs, the footprints were observed within the promoter region (i.e., TSS \pm 3 kb). Intriguingly, we found that all TFs associated with AR-NE- were preferentially localized in the promoter region. This finding was corroborated by analyzing independent publicly available ChIP-seq profiles of TFs associated with the AR-NE- subtype (Supplementary Fig. S18). Given that the promoters in humans are enriched for the CpG dinucleotide, TFs such as SP1, NRF1, and ETS and many C2H2 zinc finger proteins including KLF-family proteins and ZNF263 are known to preferentially bind proximal promoter DNA sequences (44, 57). Thus, ATAC-seq-based TF footprinting allows us to examine the preferred binding regions of a vast number of TFs.

Identification of the influential transcription factors in mCRPC transcriptional subtypes

To identify the functional impact of mCRPC subtype-specific TF-binding patterns, we constructed unbiased gene expression networks associated with each TF-binding event. We integrated ATAC-seq and matched RNA-seq data from our mCRPC cohort to comprehensively identify correlated or anticorrelated genome-wide cis-interacting peak-gene pairs (Materials and Methods; Supplementary Fig. S19). We identified a set of 37,865 robust peak-gene pairs (36,616 correlated and 1,249 anticorrelated pairs) consisting of 23,089 unique peaks and 7,710 unique genes (Supplementary Fig. S20A-C). On average, 1.64 (min = 1, max = 29) genes were associated with a peak (Supplementary Fig. S20D). Approximately 98% of peaks were correlated with five genes or less, and 66% of

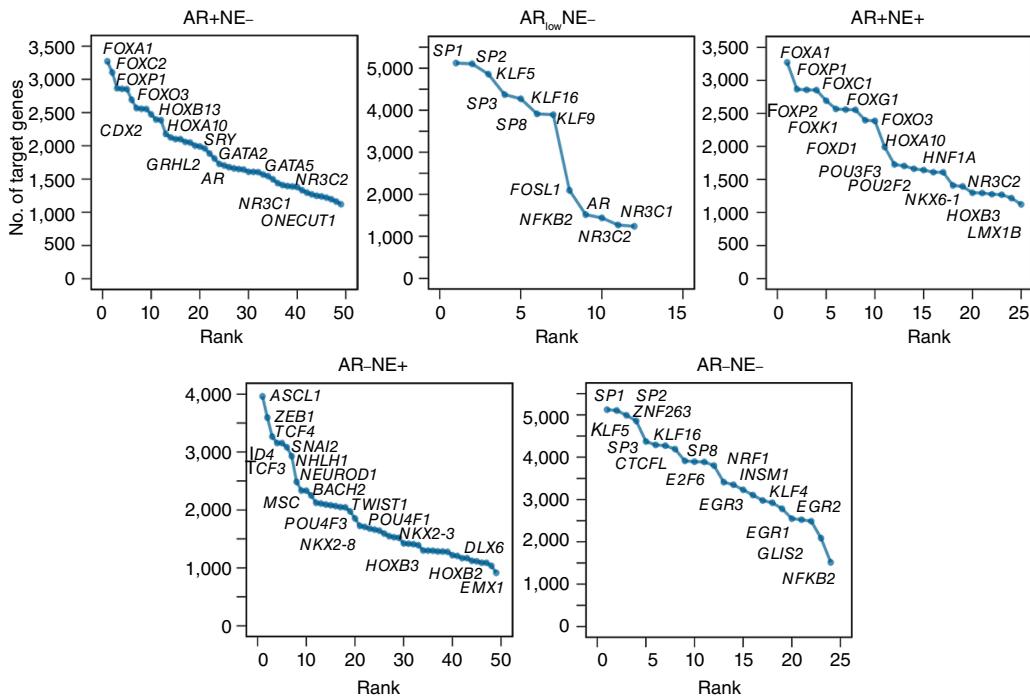
peaks were associated with only one gene. Similarly, on average, five (min = 1, max = 108) peaks were associated with a gene (Supplementary Fig. S20E). About 76% of genes were associated with at most five peaks, and 35% of genes were associated with only one peak. Most correlated peaks were proximal to the TSS region of a gene, as compared with its distal region (Supplementary Fig. S20F). These mCRPC TF network characteristics are consistent with the results obtained from TF networks derived from ATAC-seq in various types of cancer (21, 22). We further integrated the TF footprint sites identified in the accessible peak to the genes correlated with the peak to construct a TF-target gene regulatory network and identify the target genes of TFs. Restricting our analysis to the 203 mCRPC-associated TFs, the regulatory network represented 22,608 unique peaks and 7,632 unique genes. Based on the ATAC-seq peaks observed in individual mCRPC subtypes and the associated TFs, we derived the mCRPC subtype-specific regulatory network.

Taking advantage of these dense regulatory networks, we assessed if we could identify influential TFs potentially regulating the transcriptional programs and driving the mCRPC subtypes. We hypothesized that highly influential TFs regulate the transcriptional activity of a large fraction of the downstream target genes. Thus, we computed the node degree of the TFs in mCRPC subtype-specific regulatory networks (Fig. 5; Supplementary Table S6). We found that several of our top influential TF hits were well-established drivers of prostate cancer. FOXA1 and several FOX-family TFs, including FOXC2, AR, HOXB13, GRHL2, and SRY, were predicted as the top influential TFs in the AR+NE- subtype. FOXA1 and AR are well established as drivers of mCRPC (11, 12, 16, 17, 58, 59). The top influential TFs in the AR-NE+ subtype, such as ASCL1 (42, 55), NEUROD1 (42), ZEB1 (60), TCF4 (61), and SNAI2 (62), are known to promote neuroendocrine differentiation in prostate cancer. The AR+NE+ subtype was enriched with influential TFs found in both AR+ and NE+ subtypes. The Stripe family (63) of TFs, including SP1, SP2, and KLF5 (64), was predicted to drive both AR_{low}NE- and AR-NE- subtypes. This suggests that our analysis nominated consistent TF hits associated with prostate cancer. Furthermore, this analysis identified several influential TFs that were not previously linked to prostate cancer, including ZNF263 and RREB1 in the AR-NE- subtype, ZNF384 and CDX1 in the AR+NE- subtype, and BACH2 and ZBTB18 in the AR-NE+ subtype. To explore the regulatory impact of the hits identified through TF footprinting analysis on their target genes, we opted to investigate ZNF263 because it has not been studied in the context of mCRPC.

ZNF263 regulates the MYC signaling pathway in mCRPC

ZNF263 has been implicated in modulating oncogenic signaling in cancers. For example, ZNF263 is the most significant TF bound to the endoplasmic reticulum stress-specific super-enhancer and is highly expressed in hepatocellular carcinoma (65). ZNF263 knockdown in hepatocellular carcinoma cell lines leads to reduced proliferation, apoptosis resistance, and chemoresistance (65). ZNF263 enhances EGFR signaling and glioblastoma progression (66). Despite having a KRAB domain that typically facilitates transcriptional repression, ZNF263 can exert both positive and negative impacts on the transcriptional regulation of the genes it targets (44). However, ZNF263 has not been studied in the context of prostate cancer.

First, to establish the relevance of ZNF263 in prostate cancer, we interrogated its mRNA expression in several publicly available datasets. Our analysis of gene expression profiles of non-diseased tissues indicated that ZNF263 is highly expressed in prostate tissues

**Figure 5.**

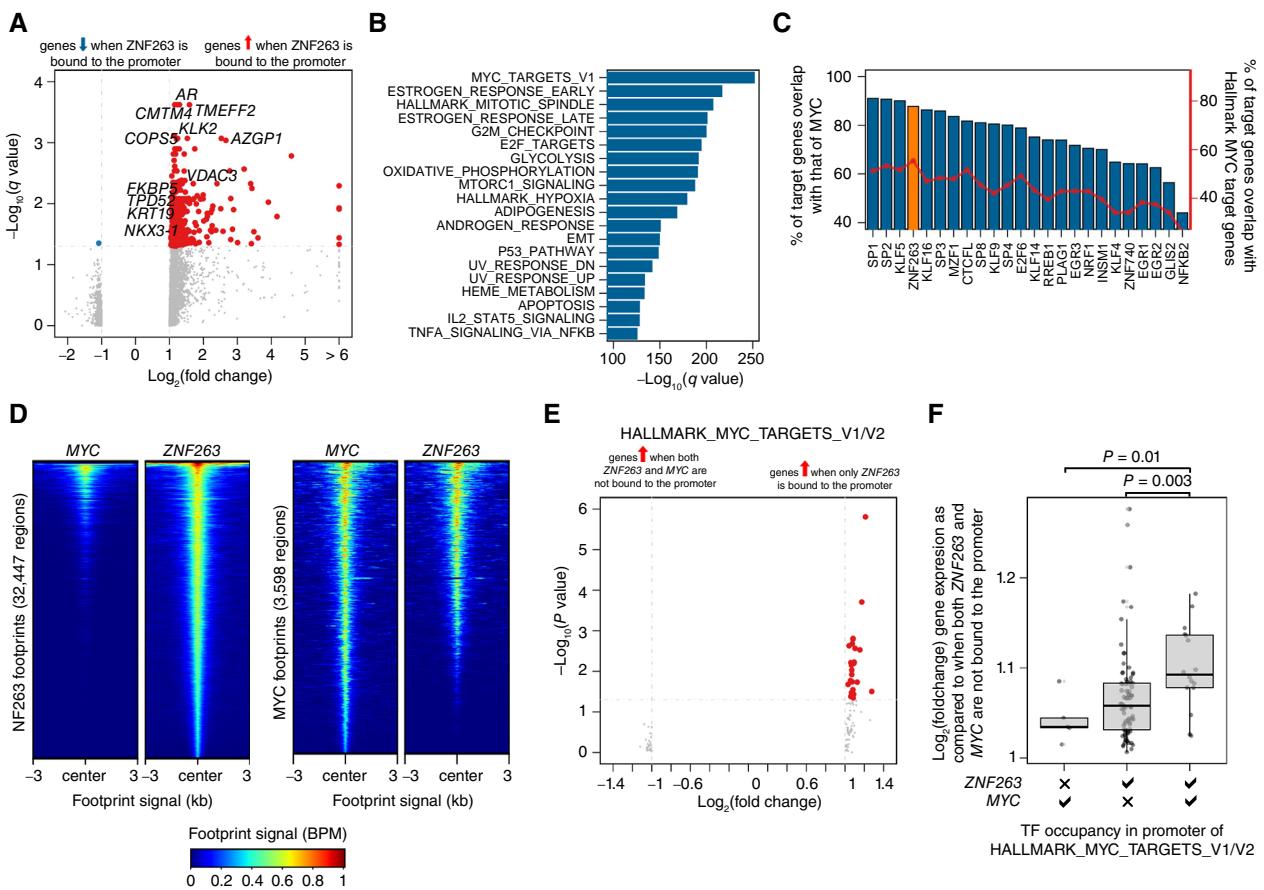
Identification of the influential TFs driving mCRPC transcriptional subtypes. We hypothesized that highly active TF regulate (or influence) gene expression activity of a large fraction of target genes. The plot indicates the top influential mCRPC transcriptional subtype-associated TFs ranked by the number of target genes (based on gene expression) they influence.

(Supplementary Fig. S21A). Moreover, we found elevated expression levels of ZNF263 in mCRPC as compared with both benign prostate and localized prostate cancer tissues (Supplementary Fig. S21B and S1C). Notably, ZNF263 was expressed in all mCRPC subtypes, and there was no difference in expression levels between the subtypes (Supplementary Fig. S21D). Next, to inspect the reproducibility of our predicted ZNF263 footprint sites, we compared the predicted binding sites in mCRPC to previously published ZNF263 ChIP-seq profiles measured on human embryonic stem cells and erythroblast cells. We hypothesized that although some ZNF263 binding sites would be unique to mCRPC, many binding sites would be invariant among cell types, and demonstrating nonrandom enrichment for experimentally identified binding sites would support the validity of our computational analysis. Indeed, we found that 6% to 16% of ZNF263-bound footprint sites were also observed in the ChIP-seq peaks (hypergeometric test P value $< 2.2 \times 10^{-16}$; Supplementary Fig. S22A). Importantly, sites where we predicted TF binding in mCRPC had a greater degree of overlap with the ChIP-seq peaks than sites predicted to be unbound in mCRPC, supporting the robustness of our footprint predictions.

To identify targets of ZNF263 transcriptional regulation in mCRPC, we assessed variation in gene expression levels in the presence or absence of ZNF263 binding. As described in the previous section, because ZNF263 has a strong affinity for binding to the promoter region (Supplementary Figs. S17 and S18), we focused our analysis on ZNF263 footprints identified in this region. For each gene, we grouped the mCRPC samples based on whether ZNF263 was present or absent in the promoter region of the gene. Then we measured the fold change in the gene

expression levels between the two groups. A significant increase in gene expression levels was observed for most genes when ZNF263 was bound to their promoter region (Fig. 6A). We conducted an unbiased enrichment analysis to identify common functions of genes in which expression increased when ZNF263 was bound and identified enrichment in the MYC signaling pathways (Supplementary Fig. S22B). Conducting the same analysis on individual mCRPC transcriptional subtypes produced comparable findings to the analysis that included all samples (Supplementary Fig. S22C and S22D). We observed that ZNF263 footprints were prevalent in all mCRPC subtypes. However, when compared with other subtypes, the AR-NE- subtype exhibited a higher occupancy of ZNF263 footprints (Supplementary Fig. S22E). Interestingly, the target genes of ZNF263 including MYC targets were expressed in the AR-NE- subtype as well as many other samples from different other mCRPC subtypes (Supplementary Fig. S23). These results suggest that ZNF263 potentially acts as an activator of gene

expression levels between the two groups. A significant increase in gene expression levels was observed for most genes when ZNF263 was bound to their promoter region (Fig. 6A). We conducted an unbiased enrichment analysis to identify common functions of genes in which expression increased when ZNF263 was bound and identified enrichment in the MYC signaling pathways (Supplementary Fig. S22B). Conducting the same analysis on individual mCRPC transcriptional subtypes produced comparable findings to the analysis that included all samples (Supplementary Fig. S22C and S22D). We observed that ZNF263 footprints were prevalent in all mCRPC subtypes. However, when compared with other subtypes, the AR-NE- subtype exhibited a higher occupancy of ZNF263 footprints (Supplementary Fig. S22E). Interestingly, the target genes of ZNF263 including MYC targets were expressed in the AR-NE- subtype as well as many other samples from different other mCRPC subtypes (Supplementary Fig. S23). These results suggest that ZNF263 potentially acts as an activator of gene

**Figure 6.**

ZNF263 activates MYC signaling targets. **A**, The volcano plot depicts the genes that undergo activation or repression upon ZNF263 binding to their specific promoter region. Each gene is represented by a dot, and the difference in gene expression between samples with and without ZNF263 in the promoter region was measured as fold change. Additionally, the statistical significance of the difference was evaluated using the Wilcoxon rank-sum test to calculate the P value between the two groups. **B**, Overrepresentation analysis of the predicted ZNF263 target genes against the Hallmark pathways. **C**, Percentage of the predicted target genes of MYC that overlap with those of TFs associated with the AR–NE– subtype. The overlap of the respective TF target genes with genes in the Hallmark MYC targets geneset is illustrated as the red line. **D**, Heatmap of overlapping ZNF263 and MYC footprint sites. Red highlights the direct overlap between ZNF263 and MYC footprints. **E**, Volcano plot showing the genes that are activated when ZNF263 binds to the promoter region compared with genes that are activated when both ZNF263 and MYC are not bound to the promoter region. **F**, Box plot of gene expression fold change when different combinations of ZNF263 and MYC bind to the promoter as compared with when both ZNF263 and MYC are simultaneously absent in the promoter. Each dot represents a gene in Hallmark MYC targets geneset.

suggests that ZNF263 potentially influences the activity of MYC target genes.

Further, we inspected the ZNF263 footprints and their association with that of MYC. Despite having a different motif sequence than that of MYC (Supplementary Fig. S24C), approximately 0.35% of all ZNF263 footprint sites overlap with the footprints of MYC, and 2.75% of all MYC footprints overlap with that of ZNF263 (Fig. 6D). We further excluded these overlapping footprint regions and measured the distance between the remaining ZNF263 and its nearest MYC binding site. We found MYC-occupied regions near ZNF263 binding sites more often in the AR–NE– subtype as compared with the rest of the mCRPC transcriptional subtypes (Supplementary Fig. S24D). Furthermore, we evaluated if ZNF263 footprints predicted using ATAC-seq overlap with MYC ChIP-seq predicted binding sites. For this, we leveraged publicly available MYC ChIP-seq profiles

measured on different prostate cancer and osteosarcoma cell lines. We found that about 15% to 20% of ZNF263-bound footprint sites overlap with the MYC ChIP-seq-predicted binding sites (hypergeometric test P value $< 2.2 \times 10^{-16}$; Supplementary Fig. S24E), thus implying that MYC binds near the binding sites of ZNF263.

We next tested the hypothesis that ZNF263 is a coactivator of MYC transcriptional targets in mCRPC. ZNF263 was predicted to bind more gene promoter sites than MYC in AR+ and AR– subtypes (Supplementary Fig. S24F). We tested whether MYC targets were differentially expressed when both ZNF263 and MYC were bound in the promoter region, in comparison with ZNF263 alone or neither protein. Supporting our hypothesis, the presence of ZNF263 binding increased the expression level of these genes (Fig. 6E), and the concomitant binding of ZNF263 and MYC further increased the expression level of MYC target genes (Fig. 6F). These observations

are consistent with a model that ZNF263 collaborates with MYC to activate MYC targets in mCRPC.

Discussion

The prolonged usage of ADT and/or ARSI in advanced prostate cancer leads to the emergence of a diverse spectrum of mCRPC subtypes. Although there are emerging genomic and transcriptomic distinctions between the mCRPC subtypes, information about the variations in the epigenomic regulatory landscape between the subtypes is scarce. In this study, we present a comprehensive characterization of the chromatin accessibility of mCRPC using integrated analysis of ATAC-seq and RNA-seq from matched samples. Earlier studies (15, 21, 22) on mCRPC predominantly depended on cell lines, organoids, PDXs, or a limited number of mCRPC tissue biopsies. To date, our study represents the largest group of mCRPC tissue biopsies that have been characterized using ATAC-seq. Here, we show that chromatin accessibility increases during prostate cancer progression to mCRPC. Importantly, we found mCRPC to have unique chromatin accessibility profiles compared with localized prostate cancer and benign prostate. AR signaling is the major driver of mCRPC (11–13, 16, 17), and our investigation indicated that the functional activity of AR predominantly governs the chromatin accessibility patterns in mCRPC. The ATAC-seq profile of our mCRPC cohort closely recapitulates the chromatin accessibility heterogeneity in the patient population with advanced prostate cancer. Furthermore, the subtypes Tang and colleagues (22) identified through the analysis of ATAC-seq profiles in CRPC cell lines, organoid, and PDX models were similar to the mCRPC subtypes reported in our study. Particularly noteworthy was the finding that the CRPC-WNT and CRPC-SCL subtypes predominantly represented AR– subtypes.

This study evaluated the differences in chromatin accessibility across five mCRPC subtypes. Among these subtypes, AR+NE– and AR–NE+ (t-SCNC/NEPC) subtypes have been extensively studied and well characterized, whereas AR_{low}NE–, AR+NE+, and AR–NE– subtypes remain relatively less explored. By integrating ATAC-seq and RNA-seq of matched tumors, we established a correlation between the accessibility of a regulatory element to the expression levels of predicted target genes. We then conducted a thorough analysis of TF occupancy signals across the entire accessible genomic regions, leading us to identify 203 TFs that are linked to specific mCRPC subtypes. Some TFs are uniquely enriched in a certain mCRPC subtype, whereas others are common to multiple subtypes. We found that a variety of TFs were associated with the t-SCNC phenotype. Furthermore, we found that AR_{low}NE– and AR–NE– have many common TFs, which further supports the previous (5) findings suggesting that these subtypes share common signaling pathways. We observed the presence of numerous TFs associated with AR+NE– in both AR_{low}NE– and AR+NE+ subtypes. Additionally, the AR+NE+ subtype displayed enrichment of several NE-related TFs. Thus, our analysis suggests that AR_{low}NE–, AR+NE+, and AR–NE– subtypes might potentially be an intermediate phenotype between the more extreme AR+NE– and AR–NE+ subtypes. A recent study (5) has reported the existence of multiple mCRPC subtypes within the same metastatic site of a patient supporting mCRPC disease continuum hypothesis. These diverse spectra of mCRPC subtypes are believed to emerge from intertumoral heterogeneity and treatment-induced selective pressures that can change the phenotypic and molecular landscapes of

mCRPC (5). This mandates a detailed study interrogating the molecular mechanisms driving these rare intermediate mCRPC subtypes and their clinical outcome for better management of mCRPC.

Our analysis identified both established mCRPC-associated TFs as well as TFs that are relatively understudied in the context of prostate cancer. Our interest in investigating ZNF263 was driven by several studies linking the altered activity of ZNF263 to oncogenic processes and chemotherapy resistance in different cancers (65, 66). To the best of our knowledge, the role of ZNF263 in prostate cancer has not been studied to date. Here, we demonstrate that ZNF263 has a considerable influence on modulating the gene expression in mCRPC and may collaborate with MYC in these tumors. Thus, the integration of ATAC-seq and RNA-seq data in our work demonstrates the ability to investigate the effects of TF binding on the activity of their downstream target genes.

Our comprehension of gene expression regulation relies heavily on understanding TF binding to regulatory elements (21). Traditionally, ChIP-seq has been the standard method for identifying TF-binding sites, but more recently, ATAC-seq has emerged as a promising alternative (20, 67). Unlike ChIP-seq, which requires specific antibodies targeting individual TF, ATAC-seq enables comprehensive genome-wide profiling of footprints of all known TFs within the accessible chromatin regions. Thus, ATAC-seq TF footprint prediction holds promise as a method to screen the genome-wide binding of an extensive range of TFs in a single analysis framework to gain a comprehensive understanding of gene regulation and further reconstruct subtype-specific regulatory networks.

TFs are alluring as therapeutic targets because they are master regulators of large gene networks that affect disease outcomes. Although TFs are conventionally considered difficult-to-drug proteins, promising technologies such as PROTAC (68) have enabled the targeted degradation of desired proteins, including TFs. Direct (PROTAC) or indirect inhibition of several TFs is currently being investigated in clinical trials (69). ATAC-seq TF footprinting can aid in the identification of potential therapeutic targets by providing information on the regulatory regions of genes that are accessible to TF binding.

In summary, this study characterizes the changes in chromatin accessibility in advanced prostate cancer. Our results illustrate the importance of studying chromatin shifts at regulatory regions to determine TFs actively occupied in the region and how they modulate transcriptional programs associated with oncogenic and tumor-suppressive functions. Overall, our findings provide valuable insights into the epigenetic changes that occur during mCRPC progression.

Authors' Disclosures

Dr. M. Sjöström reports grants from the Prostate Cancer Foundation during the conduct of the study and personal fees from Astellas outside the submitted work. X. Zhu reports grants from Conquer Cancer, the ASCO Foundation, Prostate Cancer Foundation, and Department of Defense, and grants from The Cancer League outside the submitted work. J.J. Alumkal reports other support from Fortis Therapeutics, Beactica, and Zenith Epigenetics, grants from NCCN/Astellas Pharma Global Development, Inc./Pfizer, Inc.; and research award outside the submitted work. R. Aggarwal reports personal fees from Johnson and Johnson, EcoR1, Dendreon, Novartis, Boxer Capital, Flare Therapeutics, and MJH Health; grants and personal fees from Merck, Amgen, and AstraZeneca; and personal fees from Harpoon Therapeutics outside the submitted work. E.J. Small reports personal fees and other support from Fortis, other support from Harpoon and Teon, personal fees from Janssen/J&J and Deciphera; and personal fees from Fibroten outside the submitted work. F.Y. Feng reports personal fees from Janssen Oncology, Bayer, PFS Genomics, Exact Sciences, Genentech, Myovant Sciences, Roivant Sciences, Astellas Pharma, Foundation Medicine, Varian, Bristol Meyers Squibb, Novartis, and Tempus; nonfinancial support from Serimmune, Clearnote

Health, and Artera; and personal fees from Amgen outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

R. Shrestha: Conceptualization, data curation, software, formal analysis, investigation, visualization, methodology, writing—original draft, writing—review and editing. **L.N. Chesner:** Data curation, methodology, writing—review and editing. **M. Zhang:** Data curation, software, methodology, writing—review and editing. **S. Zhou:** Methodology, writing—review and editing. **A. Foye:** Data curation, methodology, project administration, writing—review and editing. **A. Lundberg:** Data curation, software, methodology, writing—review and editing. **A.S. Weinstein:** Data curation, software, methodology, writing—review and editing. **M. Sjöström:** Data curation, software, methodology, writing—review and editing. **X. Zhu:** Data curation, software, methodology, writing—review and editing. **T. Moreno-Rodriguez:** Data curation, software, methodology, writing—review and editing. **H. Li:** Data curation, methodology, writing—review and editing. **SU2C/PCF West Coast Prostate Cancer Dream Team:** Resources, writing—review and editing. **J.J. Alumkal:** Resources, writing—review and editing. **R. Aggarwal:** Resources, writing—review and editing. **E.J. Small:** Resources, funding acquisition, writing—review and editing. **M. Lupien:** Supervision, funding acquisition, investigation, methodology, writing—review and editing. **D.A. Quigley:** Conceptualization, data curation, supervision, funding acquisition, investigation, methodology, project administration, writing—review and editing. **F.Y. Feng:** Conceptualization, supervision, funding acquisition, investigation, methodology, project administration, writing—review and editing.

Acknowledgments

The authors thank the patients who selflessly contributed samples to this study, without whom this research would not have been possible. This research was

supported by a Stand Up To Cancer – Prostate Cancer Foundation Prostate Cancer Dream Team Award (SU2C-AACR-DT0812 to E.J. Small) and by the Movember Foundation. This research grant was administered by the American Association for Cancer Research, the scientific partner of SU2C. R. Shrestha, M. Sjöström, X. Zhu, and H. Li were supported by the Prostate Cancer Foundation (PCF) Young Investigator Award. L.N. Chesner was supported by the PCF Young Investigator Award and the Department of Defense Early Investigator Award (W81XWH2110046). M. Lupien was funded by the Canadian Institute of Health Research (CIHR; FRN-153234 and 168933), the Ontario Institute for Cancer Research (OICR) Investigator Award through funding provided by the Government of Ontario, and the Princess Margaret Cancer Foundation. D.A. Quigley was funded by a Young Investigator and Challenge awards from the PCF and by the UCSF Benioff Initiative for Prostate Cancer Research. F.Y. Feng was funded by PCF Challenge Awards. Additional funding was provided by a UCSF Benioff Initiative for Prostate Cancer Research award. F.Y. Feng was supported by the National Institutes of Health (NIH)/National Cancer Institute (NCI) 1R01CA230516-01, by NIH/NCI 1R01CA227025 and PCF 17CHAL06, and by NIH P50CA186786. J.J. Alumkal was supported by NCI R01 CA251245, NCI P50 CA097186, NCI P50 CA186786, NCI P50 CA186786-07S1, Joint Institute for Translational and Clinical Research, and a PCF Challenge Award. The list of members in the SU2C/PCF West Coast Prostate Cancer Dream Team is provided in the Supplementary Appendix.

Note

Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Received March 15, 2024; revised April 16, 2024; accepted July 3, 2024; published first July 11, 2024.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* 2022;72:7–33.
2. de Bono JS, Logothetis CJ, Molina A, Fizazi K, North S, Chu L, et al. Abiraterone and increased survival in metastatic prostate cancer. *N Engl J Med* 2011;364:1995–2005.
3. Scher HI, Fizazi K, Saad F, Taplin M-E, Sternberg CN, Miller K, et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. *N Engl J Med* 2012;367:1187–97.
4. Ryan CJ, Smith MR, de Bono JS, Molina A, Logothetis CJ, de Souza P, et al. Abiraterone in metastatic prostate cancer without previous chemotherapy. *N Engl J Med* 2013;368:138–48.
5. Labrecque MP, Coleman IM, Brown LG, True LD, Kollath L, Lakely B, et al. Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. *J Clin Invest* 2019;129:4492–505.
6. Vlachostergios PJ, Puca L, Beltran H. Emerging variants of castration-resistant prostate cancer. *Curr Oncol Rep* 2017;19:32.
7. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrtà J, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med* 2016;22:298–305.
8. Aggarwal R, Huang J, Alumkal JJ, Zhang L, Feng FY, Thomas GV, et al. Clinical and genomic characterization of treatment-emergent small-cell neuroendocrine prostate cancer: a multi-institutional prospective study. *J Clin Oncol* 2018;36:2492–503.
9. Bluemn EG, Coleman IM, Lucas JM, Coleman RT, Hernandez-Lopez S, Tharakan R, et al. Androgen receptor pathway-independent prostate cancer is sustained through FGF signaling. *Cancer Cell* 2017;32:474–89.e6.
10. Westbrook TC, Guan X, Rodansky E, Flores D, Liu CJ, Udager AM, et al. Transcriptional profiling of matched patient biopsies clarifies molecular determinants of enzalutamide-induced lineage plasticity. *Nat Commun* 2022;13:5345.
11. Robinson D, Van Allen EM, Wu Y-M, Schultz N, Lonigro RJ, Mosquera J-M, et al. Integrative clinical genomics of advanced prostate cancer. *Cell* 2015;161:1215–28.
12. Quigley DA, Dang HX, Zhao SG, Lloyd P, Aggarwal R, Alumkal JJ, et al. Genomic hallmarks and structural variation in metastatic prostate cancer. *Cell* 2018;174:758–69.e9.
13. Viswanathan SR, Ha G, Hoff AM, Wala JA, Carrot-Zhang J, Whelan CW, et al. Structural alterations driving castration-resistant prostate cancer revealed by linked-read genome sequencing. *Cell* 2018;174:433–47.e19.
14. Park JW, Lee JK, Sheu KM, Wang L, Balanis NG, Nguyen K, et al. Reprogramming normal human epithelial tissues to a common, lethal neuroendocrine cancer lineage. *Science* 2018;362:91–95.
15. Pomerantz MM, Qiu X, Zhu Y, Takeda DY, Pan W, Baca SC, et al. Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nat Genet* 2020;52:790–9.
16. Zhao SG, Chen WS, Li H, Foye A, Zhang M, Sjöström M, et al. The DNA methylation landscape of advanced prostate cancer. *Nat Genet* 2020;52:778–89.
17. Sjöström M, Zhao SG, Levy S, Zhang M, Ning Y, Shrestha R, et al. The 5-hydroxymethylcytosine landscape of prostate cancer. *Cancer Res* 2022;82:3888–902.
18. Rubin MA, Bristow RG, Thienger PD, Dive C, Imielinski M. Impact of lineage plasticity to and from a neuroendocrine phenotype on progression and response in prostate and lung cancers. *Mol Cell* 2020;80:562–77.
19. Steloo S, Nevedomskaya E, van der Poel HG, de Jong J, van Leenders GJLH, Jenster G, et al. Androgen receptor profiling predicts prostate cancer outcome. *EMBO Mol Med* 2015;7:1450–64.
20. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 2013;10:1213–8.
21. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. *Science* 2018;362:eav1898.
22. Tang F, Xu D, Wang S, Wong CK, Martinez-Fundichely A, Lee CJ, et al. Chromatin profiles classify castration-resistant prostate cancers suggesting therapeutic targets. *Science* 2022;376:eabe1505.
23. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* 2017;14:959–62.
24. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011;17:10–12.

25. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.
27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25: 2078–9.
28. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016;44:W160–5.
29. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol* 2008;9:R137.
30. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015;31: 2382–3.
31. Ou J, Liu H, Yu J, Kelliher MA, Castilla LH, Lawson ND, et al. ATACseqQC: a bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genomics* 2018;19:169.
32. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 2019;47:e47.
33. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
34. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010;28:495–501.
35. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1:417–25.
36. Bhuvu DD, Cursons J, Davis MJ. Stable gene expression for normalization and single-sample scoring. *Nucleic Acids Res* 2020;48:e113.
37. Bentsen M, Goymann P, Schultheis H, Klek K, Petrova A, Wiegandt R, et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* 2020;11:4267.
38. Castro-Mondragon JA, Riudavets-Puig R, Raulusevičiute I, Lemma RB, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2022;50: D165–73.
39. Wilson S, Qi J, Filipp FV. Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines. *Sci Rep* 2016;6:32611.
40. Pinskaya M, Saci Z, Gallopin M, Gabriel M, Nguyen HT, Firlej V, et al. Reference-free transcriptome exploration reveals novel RNAs for prostate cancer diagnosis. *Life Sci Alliance* 2019;2:e201900449.
41. Chen S, Huang V, Xu X, Livingstone J, Soares F, Jeon J, et al. Widespread and functional RNA circularization in localized prostate cancer. *Cell* 2019;176: 831–43.e22.
42. Cejas P, Xie Y, Font-Tello A, Lim K, Syamala S, Qiu X, et al. Subtype heterogeneity and epigenetic convergence in neuroendocrine prostate cancer. *Nat Commun* 2021;12:5775.
43. Imbeault M, Helleboid PY, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* 2017;543:550–4.
44. Frietze S, Lan X, Jin VX, Farnham PJ. Genomic targets of the KRAB and SCAN domain-containing zinc finger protein 263. *J Biol Chem* 2010;285:1393–403.
45. Pope BD, Ryba T, Dilleep V, Yue F, Wu W, Denas O, et al. Topologically associating domains are stable units of replication-timing regulation. *Nature* 2014;515:402–5.
46. Barfield SJ, Urbanucci A, Itkonen HM, Fazli L, Hicks JL, Thiede B, et al. c-Myc antagonises the transcriptional activity of the androgen receptor in prostate cancer affecting key gene networks. *EBioMedicine* 2017;18:83–93.
47. See YX, Chen K, Fullwood MJ. MYC overexpression leads to increased chromatin interactions at super-enhancers and MYC binding sites. *Genome Res* 2022;32:629–42.
48. Guo H, Wu Y, Nouri M, Spisak S, Russo JW, Sowalsky AG, et al. Androgen receptor and MYC equilibration centralizes on developmental super-enhancer. *Nat Commun* 2021;12:7308.
49. Oki S, Ohta T, Shioi G, Hatanaka H, Ogasawara O, Okuda Y, et al. ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep* 2018;19:e46255.
50. Urbanucci A, Barfield SJ, Kyölä V, Itkonen HM, Coleman IM, Vodák D, et al. Androgen receptor deregulation drives bromodomain-mediated chromatin alterations in prostate cancer. *Cell Rep* 2017;19:2045–59.
51. Lundberg A, Zhang M, Aggarwal R, Li H, Zhang L, Foye A, et al. The genomic and epigenomic landscape of double-negative metastatic prostate cancer. *Cancer Res* 2023;83:2763–74.
52. Meyer RC, Giddens MM, Schaefer SA, Hall RA. GPR37 and GPR37L1 are receptors for the neuroprotective and glioprotective factors prosaptide and prosaposin. *Proc Natl Acad Sci U S A* 2013;110:9529–34.
53. Vierstra J, Stamatoyannopoulos JA. Genomic footprinting. *Nat Methods* 2016; 13:213–21.
54. Castro-Mondragon JA, Riudavets-Puig R, Raulusevičiute I, Berhanu Lemma R, Turchi L, Blanc-Mathieu R, et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2022;50:D165–73.
55. Nouruzi S, Ganguli D, Tabrizian N, Kobelev M, Sivak O, Namekawa T, et al. ASCL1 activates neuronal stem cell-like lineage programming through remodeling of the chromatin landscape in prostate cancer. *Nat Commun* 2022; 13:2282.
56. Bishop JL, Thaper D, Vahid S, Davies A, Ketola K, Kuruma H, et al. The master neural transcription factor BRN2 is an androgen receptor-suppressed driver of neuroendocrine differentiation in prostate cancer. *Cancer Discov* 2017;7:54–71.
57. FitsGerald PC, Shlyakhtenko A, Mir AA, Vinson C. Clustering of DNA sequences in human promoters. *Genome Res* 2004;14:1562–74.
58. Sahu B, Laakso M, Ovaska K, Mirtti T, Lundin J, Rannikko A, et al. Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. *EMBO J* 2011;30:3962–76.
59. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* 2012;487:239–43.
60. Bery F, Cancel M, Guéguinou M, Potier-Cartereau M, Vandier C, Chantôme A, et al. Zeb1 and SK3 channel are up-regulated in castration-resistant prostate cancer and promote neuroendocrine differentiation. *Cancers (Basel)* 2021;13: 2947.
61. Lee GT, Rosenfeld JA, Kim WT, Kwon YS, Palapattu G, Mehra R, et al. TCF4 induces enzalutamide resistance via neuroendocrine differentiation in prostate cancer. *PLoS One* 2019;14:e0213488.
62. Cmero M, Kurganova NJ, Stuchberry R, McCoy P, Grima C, Nguyen A, et al. Loss of SNAI2 in prostate cancer correlates with clinical response to androgen deprivation therapy. *JCO Precis Oncol* 2021;5:1048–59.
63. Zhao Y, Vartak SV, Conte A, Wang X, Garcia DA, Stevens E, et al. “Stripe” transcription factors provide accessibility to co-binding partners in mammalian genomes. *Mol Cell* 2022;82:3398–411.e11.
64. Che M, Chaturvedi A, Munro SA, Pitzen SP, Ling A, Zhang W, et al. Opposing transcriptional programs of KLF5 and AR emerge during therapy for advanced prostate cancer. *Nat Commun* 2021;12:6377.
65. Cui J, Liu J, Fan L, Zhu Y, Zhou B, Wang Y, et al. A zinc finger family protein, ZNF263, promotes hepatocellular carcinoma resistance to apoptosis via activation of ER stress-dependent autophagy. *Transl Oncol* 2020;13: 100851.
66. Yu Z, Feng J, Wang W, Deng Z, Zhang Y, Xiao L, et al. The EGFR-ZNF263 signaling axis silences SIX3 in glioblastoma epigenetically. *Oncogene* 2020;39: 3163–78.
67. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. *Nature* 2012;489: 75–82.
68. Samarasinghe KTG, Jaime-Figueroa S, Burgess M, Nalawansha DA, Dai K, Hu Z, et al. Targeted degradation of transcription factors by TRAFTACs: TRAnscription Factor TArgeting Chimeras. *Cell Chem Biol* 2021;28:648–61.e5.
69. Henley MJ, Koehler AN. Advances in targeting “undruggable” transcription factors with small molecules. *Nat Rev Drug Discov* 2021;20:669–88.