

GENOMIC AND EPIGENOMIC ABERRATIONS OF GENE REGULATION IN PROSTATE
CANCER AND LEUKEMIA

by

James Hawley

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Graduate Department of Medical Biophysics
University of Toronto

Contents

1 Noncoding mutations target <i>cis</i>-regulatory elements of the <i>FOXA1</i> plexus in prostate cancer	1
1.1 Abstract	2
1.2 Introduction	2
1.3 Results	4
1.3.1 <i>FOXA1</i> is essential for prostate cancer proliferation	4
1.3.2 Identifying putative <i>FOXA1</i> CREs	4
1.3.3 Putative <i>FOXA1</i> CREs harbour TF binding sites and SNVs	7
1.3.4 Disruption of CREs reduces <i>FOXA1</i> mRNA expression	7
1.3.5 <i>FOXA1</i> CREs collaborate to regulate its expression	11
1.3.6 Disruption of <i>FOXA1</i> CREs reduces prostate cancer (PCa) cell growth	12
1.3.7 SNVs mapping to <i>FOXA1</i> CREs can alter their activity	12
1.3.8 SNVs mapping to <i>FOXA1</i> CREs can modulate the binding of TFs	15
1.4 Discussion	15
1.5 Methods	17
1.5.1 Cell Culture	17
1.5.2 Prostate tumours and cancer cell lines expression	17
1.5.3 Prostate cancer cell line gene essentiality	17
1.5.4 siRNA knockdown and cell proliferation assay	18
1.5.5 Identifying putative <i>FOXA1</i> CREs	18
1.5.6 Hi-C and TADs in LNCaP cells	18
1.5.7 Clonal wild-type Cas9 and dCas9-KRAB mediated validation	19
1.5.8 Transient Cas9-mediated disruption of CREs	20
1.5.9 RT-PCR assessment of gene expression upon deletion of CREs	20
1.5.10 Confirmation of Cas9-mediated deletion of CREs	21
1.5.11 Cell proliferation upon deletion of <i>FOXA1</i> CREs	21

1.5.12 Luciferase reporter assays	21
1.5.13 Allele-specific ChIP-qPCR	22
1.6 Data availability	22
Glossary	24
References	26

Chapter 1

Noncoding mutations target *cis*-regulatory elements of the ***FOXA1*** plexus in prostate cancer

This chapter is a version of the paper published in *Nature Communications* as follows:

Zhou, S., Hawley, J. R., Soares, F., Grillo, G., Teng, M., Tonekaboni, S. A. M., Hua, J. T., Kron, K. J., Mazrooei, P., Ahmed, M., Arlide, C., Yun, H. Y., Livingstone, J., Huang, V., Yamaguchi, T. N., Espiritu, S. M. G., Zhu, Y., Severson, T. M., Murison, A., Cameron, S., Zwart, W., van der Kwast, T., Pugh, T. J., Fraser, M., Boutros, P. C., Bristow, R. G., He, H. H., and Lupien, M. Noncoding mutations target *cis*-regulatory elements of the *FOXA1* plexus in prostate cancer. ***Nature Communications***, 2020;11:1–13.

Contributions per the manuscript: S.Z. and M.L. conceptualized the study. S.Z. designed and conducted most of the experiments with help from F.S., G.G., M.T., K.J.K., J.T.H., C.A., H.Y.Y., Y.Z. and S.C. J.R.H. implemented most of the computational analyses and statistical approaches with help from S.A.M., P.M., M.A., A.M., V.H., T.N.Y., S.M.G.E., T.M.S. and J.L. under the supervision of W.Z., T.v.d.K., T.J.P., M.F., P.C.B., R.G.B., H.H.H., or M.L. Figures were designed by S.Z. with assistance from J.R.H. and S.A.M. The manuscript was written by S.Z., J.R.H. and M.L. with assistance from all authors. M.L. oversaw the study.

1.1 Abstract

PCa is the second most commonly diagnosed malignancy among men worldwide. Recurrently mutated in primary and metastatic prostate tumours, *FOXA1* encodes a pioneer transcription factor involved in disease onset and progression through both androgen receptor (AR)-dependent and AR-independent mechanisms. Despite its oncogenic properties however, the regulation of *FOXA1* expression remains unknown. Here, we identify a set of six *cis*-regulatory elements (CREs) in the *FOXA1* regulatory plexus harboring somatic single nucleotide variantss (SNVs) in primary prostate tumours. We find that deletion and repression of these CREs significantly decreases *FOXA1* expression and PCa cell growth. Six of the ten SNVs mapping to *FOXA1* regulatory plexus significantly alter the transactivation potential of CREs by modulating the binding of transcription factors (TFs). Collectively, our results identify CREs within the *FOXA1* plexus mutated in primary prostate tumours as potential targets for therapeutic intervention.

1.2 Introduction

PCa is the second most commonly diagnosed cancer among men with an estimated 1.3 million new cases worldwide in 2018 [1]. Although most men diagnosed with primary PCa are treated with curative intent through surgery or radiation therapy, treatments fail in 30% of patients within 10 years [2] resulting in a metastatic disease [3]. Patients with metastatic disease are typically treated with anti-androgen therapies, the staple of aggressive PCa treatment [4]. Despite the efficacy of these therapies, recurrence ultimately develops into lethal metastatic castration-resistant prostate cancer (mCRPC) [4]. As such, there remains a need to improve our biological understanding of PCa development and find novel strategies to treat patients. Sequencing efforts identified coding somatic SNVs mapping to *FOXA1* in up to 9% [5–10] and 13% [9–11] of primary and metastatic PCa patients, respectively. These coding somatic SNVs target the Forkhead and transactivation domains of *FOXA1* [12], altering its pioneering functions to promote PCa development [10, 13]. Outside of coding SNVs, whole genome sequencing (WGS) also identified somatic SNVs and indels in the 3' untranslated region (UTR) and C-terminus of *FOXA1* in textapprox 12% of mCRPC patients [14]. In addition to SNVs, the *FOXA1* locus is a target of structural rearrangements in both primary and metastatic PCa tumours, inclusive of duplications, amplifications, and translocations [9, 10]. Taken together, *FOXA1* is recurrently mutated taking into account both its coding and flanking noncoding sequences across various stages of PCa development.

FOXA1 serves as a pioneer TF that can bind to heterochromatin, promoting its remodelling to increase accessibility for the recruitment of other TFs [15]. *FOXA1* binds to chromatin at cell-type specific genomic coordinates facilitated by the presence of mono- and dimethylated lysine 4 of histone H3 (H3K4me1 and H3K4me2) histone modifications [16, 17]. In PCa, *FOXA1* is known to pioneer and reprogram the binding of AR alongside *HOXB13* [18]. Independent from its role in AR signalling, *FOXA1* also regulates the expression of genes involved in cell cycle regulation in PCa [19, 20]. For instance, *FOXA1* co-localizes with *CREB1* to regulate the transcription of genes involved in cell cycle processes, nuclear division and mitosis in mCRPC [19–25]. *FOXA1* has also been shown to promote feed-forward mechanisms to drive disease progression [26, 27]. Hence, *FOXA1* contributes to AR-dependent and AR-independent processes favouring PCa development.

Despite the oncogenic roles of *FOXA1*, therapeutic avenues to inhibit its activity in PCa are lacking. In the breast cancer setting for instance, the use of cyclin-dependent kinases inhibitors have been suggested based on their ability to block *FOXA1* activity on chromatin [28]. As such, understanding the governance of *FOXA1* messenger RNA (mRNA) expression offers an alternative strategy to find modulators of its activity. Gene expression relies on the interplay between distal CREs, such as enhancers and anchors of chromatin interaction, and their target gene promoter(s) [29]. These elements can lie tens to hundreds of kilobases (kbps) away from each other on the linear genome but physically engage in close proximity with each other in the three-dimensional space [30]. By measuring contact frequencies between loci through the use of chromatin conformation capture (3C)-based technologies, it enables the identification of regulatory plexuses corresponding to sets of CREs in contact with each other [31, 32]. By leveraging these technologies, we can begin to understand the three-dimensional organization of the PCa genome and delineate the *FOXA1* regulatory plexus.

Here, we integrate epigenetics and genetics from PCa patients and model systems to delineate CREs establishing the regulatory plexus of *FOXA1*. We functionally validate a set of six mutated CREs that regulate *FOXA1* mRNA expression. We further show that SNVs mapping to these CREs are capable of altering their transactivation potential, likely through modulating the binding of key PCa TFs.

1.3 Results

1.3.1 *FOXA1* is essential for prostate cancer proliferation

We interrogated *FOXA1* expression levels across cancer types. We find that *FOXA1* mRNA is consistently the most abundant in prostate tumours compared to 25 other cancer types across patients (Figure 1.1a), ranking in the 95th percentile for 492 of 497 prostate tumours profiled in The Cancer Genome Atlas (TCGA) (??a). Using the same dataset we also find that *FOXA1* is the most highly expressed out of 41 other forkhead box (FOX) factors in prostate tumours (??b). We next analyzed expression data from Cancer Dependency Map (DEPMAP) and observed *FOXA1* to be most highly expressed in PCa cell lines compared to cell lines of other cancer types (??a). Amongst the eight PCa cell lines in the dataset (22Rv1, DU145, LNCaP, MDA-PCa-2B, NCI-H660, PrECLH, PC3, and VCaP), *FOXA1* mRNA abundance is above the 90th percentile in all but one cell line (PrECLH) compared to the > 56,000 protein coding and non-protein coding genes profiled (??b). These new results gained from the TCGA and DEPMAP validate previous understanding that *FOXA1* is one of the highest expressed genes in PCa [33].

Following up on *FOXA1* mRNA expression levels, we interrogated the essentiality of *FOXA1* for PCa cell growth. RNAi-mediated essentiality screens compiled in DEPMAP show that *FOXA1* lies in the 94th percentile across 6 of the 8 available PCa cell lines: 22Rv1, LNCaP, MDA PCa 2B, NCI-H660, PC3, and VCaP cells (Figure 1.1b-c). The median RNAi-mediated essentiality score for all prostate cell lines is significantly lower than all other cell lines, suggesting that *FOXA1* is especially essential for PCa cell proliferation (permutation test, $p = 1 \cdot 10^{-6}$, see Methods) (??a). Growth assays in LNCaP and VCaP cells following *FOXA1* knockdown using two independent siRNAs (Figure 1.1d, ??b) show significant growth inhibition in LNCaP (siRNA #1: 4-fold, siRNA #2: 3.35-fold) and VCaP (siRNA #1: 8.7-fold, siRNA #2: 2-fold) cells five days post-transfection (Mann-Whitney U Test, $p < 0.05$; Figure 1.1e-f). In accordance with previous reports, our results using essentiality datasets followed by knockdown validation reveals that *FOXA1* is oncogenic and essential for PCa cell proliferation.

1.3.2 Identifying putative *FOXA1* CREs

The interweaving of distal CREs with target gene promoters establishes regulatory plexuses with some to be ascribed to specific genes [31, 32]. Regulatory plexuses stem from chromatin interactions orchestrated by various factors including ZNF143, YY1, CTCF and the cohesin complex

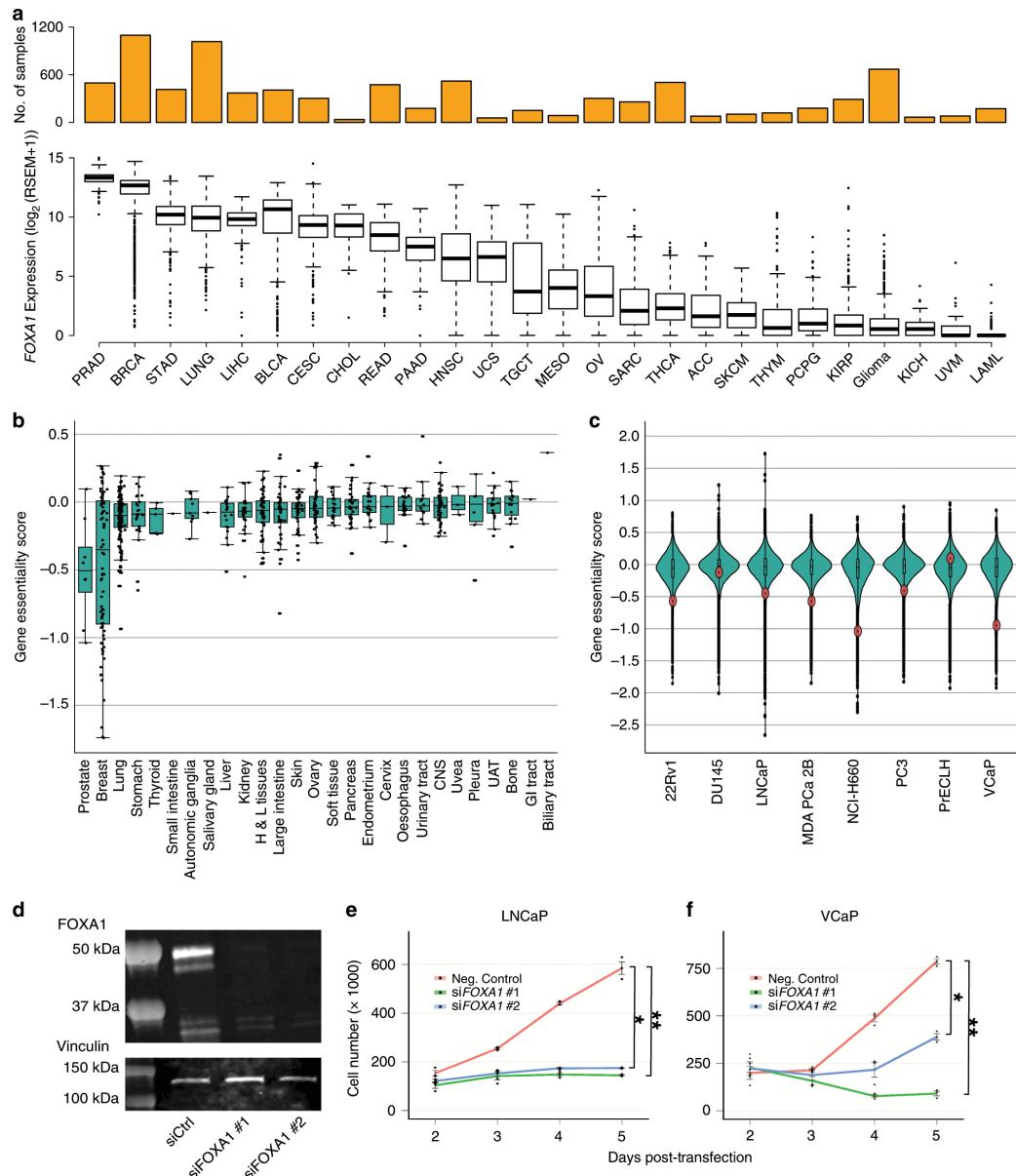


Figure 1.1: *FOXA1* is highly expressed in PCa and essential for PCa cell proliferation..

a. The mRNA expression of *FOXA1* across tumour types ($n = 26$) from RNA-seq data of TCGA.

b. *FOXA1* essentiality mediated through RNAi across various cell lines ($n = 707$) from DEPMAP. Gene essentiality scores are normalized z -scores. Higher scores indicate less essential, and lower scores indicate more essential for cell proliferation. x -axis indicate tissue of origin for each cell line tested. Each dot indicates one cell line.

c. Gene essentiality mediated through RNAi across PCa cell lines ($n = 8$) from DEPMAP. Each dot indicates one gene, red indicates *FOXA1*.

d. Representative Western blot against *FOXA1* in LNCaP cells 5 days post-transfection of non-targeting siRNA and two independent siRNA targeting *FOXA1*.

e. Cell proliferation assay conducted in LNCaP cells upon siRNA-mediated knockdown of *FOXA1* across 5 days.

f. Cell proliferation assay conducted in VCaP cells upon siRNA-mediated knockdown of *FOXA1* across 5 days. Error bars indicate \pm s.d. $n = 3$ independent experiments. Mann-Whitney U test, * $p < 0.05$, ** $p < 0.01$.

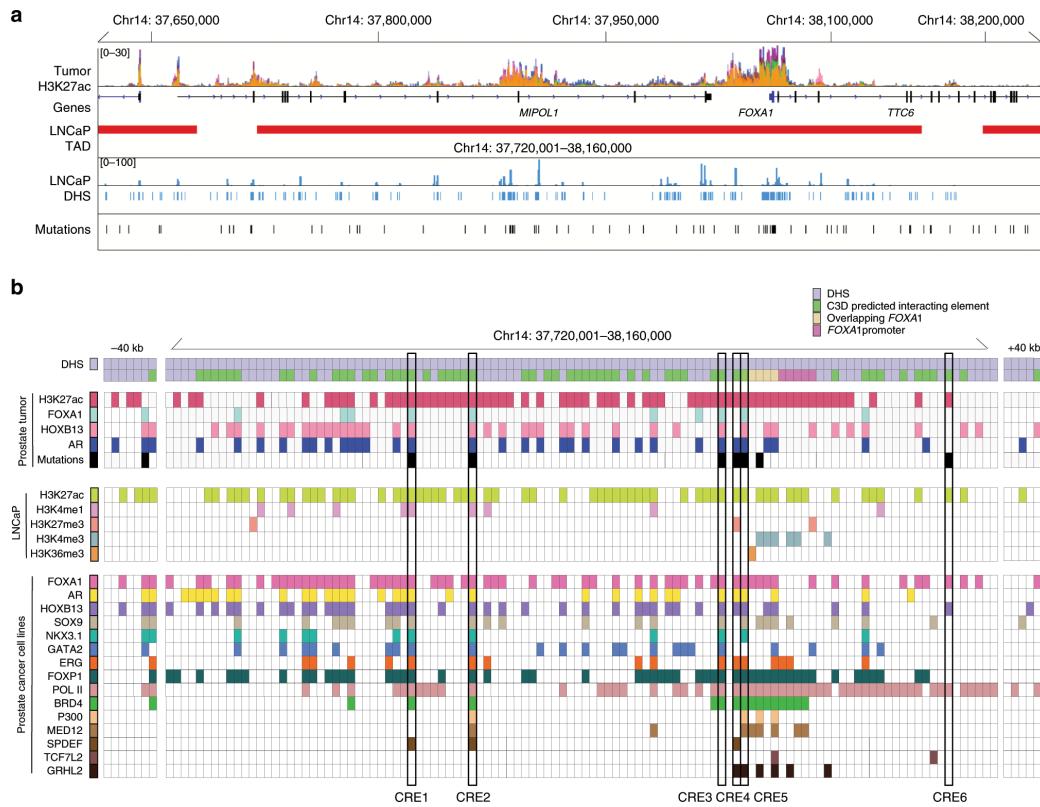


Figure 1.2: Epigenetic annotation of 14q21.1 locus and identification of *FOXA1* CREs.

a. Overview of cis-regulatory landscape surrounding *FOXA1* on the 14q21.1 locus. H3K27ac signal track is the ChIP-seq signal overlay of 19 primary prostate tumours. LNCaP Hi-C depicts the TAD structure around *FOXA1*. Mutations indicate SNVs identified in 200 primary prostate tumours.

b. Functional annotation of putative *FOXA1* CREs using TF and histone modification ChIP-seq conducted in primary tumours and PCa cell lines. Annotated in the matrix are all DHS within the TAD and ± 40 kbp resolution left and right of the TAD. Putative *FOXA1* CREs targeted by noncoding SNVs for downstream validation are boxed.

[34–36]. Motivated by the oncogenic role of *FOXA1* in PCa, we investigated its regulatory plexus controlling its expression. According to chromatin contact frequency maps generated from Hi-C assays performed in LNCaP PCa cells, *FOXA1* lies in a 440 kbp topologically associated domain (TAD) (chr14: 37720002-38160000 ± 40 kbp adjusting for resolution) (Figure 1.2a). By overlaying DNase-seq data from LNCaP PCa cells, there are a total of 123 putative CREs reported as DNase I hypersensitive sites (DHSs) that populate this TAD (Figure 1.2a). We next inferred the regulatory plexus of *FOXA1* using the C3D method [37]. C3D aggregates and draws correlation of DHS signal intensities between the cell line of choice and the DHS signal across all systems in the database [37]. Anchoring our analysis to the *FOXA1* promoter and using accessible chromatin regions defined in LNCaP PCa cells identified 55 putative CREs to the *FOXA1* regulatory plexus ($r > 0.7$) (Figure 1.2b).

1.3.3 Putative *FOXA1* CREs harbour TF binding sites and SNVs

To delineate the CREs that could be actively involved in the transcriptional regulation of *FOXA1*, we annotated the DHS with available ChIP-seq data for histone modifications and TFs conducted in LNCaP, 22Rv1, VCaP PCa cell lines and primary prostate tumours (Figure 1.2b) [18, 38]. Close to 60% (33/55) of the putative *FOXA1* plexus CREs are positively marked by H3K27ac profiled in primary prostate tumours [38], indicative of active CREs in tumours (Figure 1.2b) [39]. Next, considering that noncoding SNVs can target a set of CREs that converge on the same target gene in cancer [32], we overlapped the somatic SNVs called from the whole-genome sequencing across 200 primary prostate tumours to the 33 H3K27ac-marked DHS predicted to regulate *FOXA1* [6, 40]. This analysis identified 6 out of the 33 DHS marked with H3K27ac (18.2%) harboring one or more SNVs (10 total SNVs called from 9 tumours) (Figure 1.2b). We observe that these 6 CREs can be bound by multiple TFs in PCa cells, including *FOXA1*, *AR* and *HOXB13* (Figure 1.2b, ??). The Hi-C data from the LNCaP PCa cells corroborates the C3D predictions as demonstrated by the elevated contact frequency between the region harboring the *FOXA1* promoter and where the 6 CREs are located, compared to other loci in the same TAD (Figure 1.3a). The 6 CREs lie in intergenic or intronic regions (Figure 1.3b-h). Together, histone modifications, TF binding sites and noncoding SNVs support that these 6 putative CREs are active in primary PCa. The Hi-C and C3D predictions suggest that they regulate *FOXA1* expression.

1.3.4 Disruption of CREs reduces *FOXA1* mRNA expression

We next assessed the role of CREs toward *FOXA1* expression using LNCaP and 22Rv1 clones stably expressing the wild-type (WT) Cas9 protein (Figure 1.4a-b). Guide RNA (gRNA) designed against the *FOXA1* gene (exon 1 and intron 1) served as positive controls while an outside-TAD region (i.e termed Chr14 (-)), a region on a different chromosome (the human *AAVS1* safe-harbor site at the *PPP1R12C* locus [38, 41]), and three regions within the TAD predicted to be excluded from the *FOXA1* plexus served as negative controls. Individual deletion of the *FOXA1* plexus CREs through transient transfection of gRNAs into the LNCaP cells (See Methods) led to significantly decreased *FOXA1* mRNA expression (Δ CRE1 $\sim 29.3 \pm 8.3\%$, Δ CRE2 $\sim 40.1 \pm 11.8\%$, Δ CRE3 $\sim 30.6 \pm 9.1\%$, Δ CRE4 $\sim 23.6 \pm 8.2\%$, Δ CRE5 $\sim 25.3 \pm 6.6\%$, Δ CRE6 $\sim 24.5 \pm 10.2\%$ and Δ *FOXA1* (exon 1 and intron 1) $\sim 87.4 \pm 8.8\%$ reduction relative to basal levels) (Figure 1.4c, ??a-f). In contrast, deletion of several negative control regions within the same TAD did not significantly reduce *FOXA1* mRNA level (Figure 1.4c, ??g-i). Similar results were observed in 22Rv1

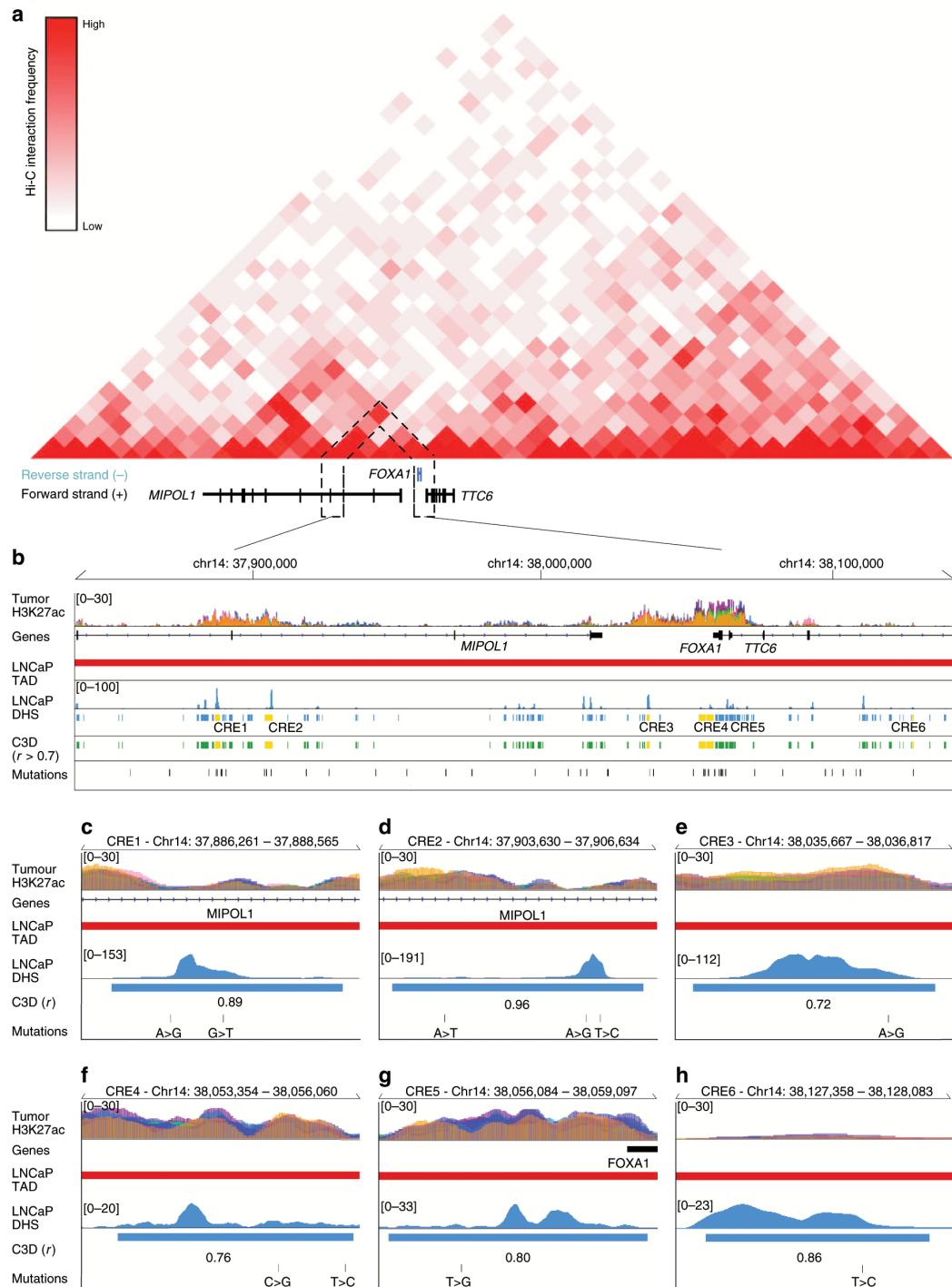


Figure 1.3: Putative CREs predicted to interact with *FOXA1* promoter. **a.** Hi-C conducted in LNCaP cells indicating physical interactions between putative *FOXA1* CREs and the *FOXA1* promoter. Hi-C resolution is 40 kbp. **b.** The six putative *FOXA1* CREs are coloured in yellow. **c-h.** Zoom-in of each individual putative *FOXA1* CRE. C3D value is the Pearson correlation of DHS signal between LNCaP and the DHS reference matrix.

PCa cells (Figure 1.4d). As each clone expressed Cas9 protein at different levels, there may be a difference between genome editing efficiencies between the clones. We compared the CRISPR/Cas9 on-target genome editing efficiency across the five LNCaP cell line-derived clones with the relative *FOXA1* mRNA levels, and indeed observe a significant inverse correlation across all CREs (Pearson's correlation $r = 0.49, p < 0.005$) (??a) and agreeing trends for each individual CRE (??b).

Complementary to our findings using the WT CRISPR/Cas9 system, we next generated four LNCaP and four 22Rv1 cell line-derived dCas9-KRAB fusion protein expressing clones (Figure 1.4e-f). Transient transfection of the same gRNAs used in the WT Cas9 experiments, targeting the six *FOXA1* plexus CREs into our dCas9-KRAB LNCaP clones significantly decreased *FOXA1* expression relative to basal levels (iCRE1 $\sim 24.6 \pm 6.2\%$, iCRE2 $\sim 42.2 \pm 10.8\%$, iCRE3 $\sim 25.3 \pm 9.2\%$, iCRE4 $\sim 23.3 \pm 4.3\%$, iCRE5 $\sim 30.2 \pm 3.4\%$ and iCRE6 $\sim 23.1 \pm 8.1\%$ reduction). Similarly, gRNAs targeting the dCas9-KRAB fusion protein to *FOXA1* decreased its expression (i*FOXA1* $\sim 81.6 \pm 11.8\%$ reduction; Student's *t*-test, $p < 0.05$, Figure 1.4g). Analogous results were also observed in our four clonal 22Rv1 dCas9-KRAB cell lines (Student's *t*-test, $p < 0.05$, Figure 1.4h). Collectively, our results suggest that the six CREs control *FOXA1* expression.

We further assessed the regulatory activity of the six *FOXA1* plexus CREs by testing the consequent mRNA expression on other genes within the same TAD, namely *MIPOL1* and *MIPOL1*. Δ CRE1 and Δ CRE2 significantly reduced *MIPOL1* mRNA expression by $\sim 38.4 \pm 6.4\%$ and $\sim 48.4 \pm 9\%$, respectively relative to basal levels, whereas deletion of the other four CREs did not result in any significant *MIPOL1* expression changes (Student's *t*-test, $p < 0.05$, ??a). On the other hand, deletion of CREs each significantly reduced *MIPOL1* mRNA expression relative to its basal levels (Δ CRE1 $\sim 52.9\% \pm 6.4\%$, Δ CRE2 $\sim 66 \pm 11.3\%$, Δ CRE3 $\sim 55.5 \pm 12.8\%$, Δ CRE4 $44.9 \pm 10.6\%$, Δ CRE5 $43.1 \pm 11.9\%$ and Δ CRE6 $52.2 \pm 7.3\%$ reduction (Student's *t*-test, $p < 0.05$, ??b), in agreement with the fact that *MIPOL1* shares its promoter with *FOXA1* as both genes are transcribed on opposing strands (??c).

Reduction in *FOXA1* mRNA expression resulting from the deletion of *FOXA1* plexus CREs may also impact gene expression downstream of *FOXA1*, we assessed the mRNA expression of several *FOXA1* target genes, namely *SNAI2*, *ACPP*, and *GRIN3A*. Deletion of CREs resulted in significant change in *SNAI2* (up-regulation; Δ CRE1 $\sim 190\%$, Δ CRE2 $\sim 162.8\%$, Δ CRE3 $\sim 147.5\%$, Δ CRE4 $\sim 133.3\%$, Δ CRE5 $\sim 137.3\%$, Δ CRE6 $\sim 120.8\%$, Δ *FOXA1* $\sim 266.7\%$), *ACPP* (down-regulation; Δ CRE1 $\sim 73.5\%$, Δ CRE2 $\sim 62.5\%$, Δ CRE3 $\sim 69.6\%$, Δ CRE4 $\sim 75.6\%$, Δ CRE5 $\sim 70.9\%$, Δ CRE6 $\sim 74.6\%$, Δ *FOXA1* $\sim 52.2\%$) and *GRIN3A* expression (up-regulation; Δ CRE1 $\sim 138.2\%$, Δ CRE2 $\sim 168.8\%$, Δ CRE3 $\sim 144.6\%$, Δ CRE4 $\sim 132.1\%$, Δ CRE5 $\sim 131.4\%$, Δ CRE6 $\sim 127\%$, Δ *FOXA1*

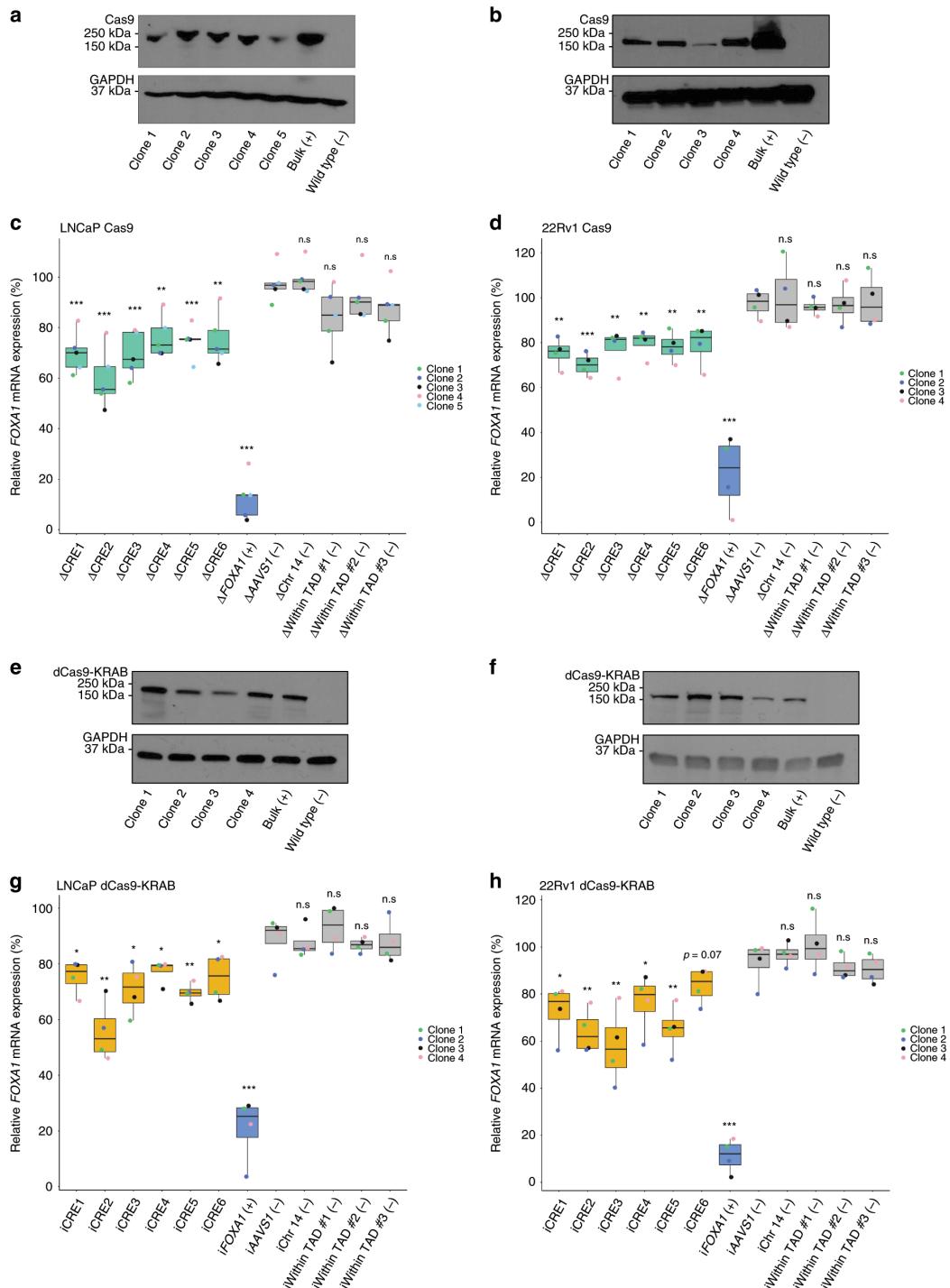


Figure 1.4: Functional dissection of putative *FOXA1* CREs. (Continued on the following page)

Figure 1.4: **a.** Representative western blot probed against Cas9 in LNCaP clones ($n = 5$ clones) derived to stably express Cas9 protein upon blasticidin selection. **b.** Representative western blot probed against Cas9 in 22Rv1 clones ($n = 4$ clones) derived to stably express Cas9 protein upon blasticidin selection. **c.** *FOXA1* mRNA expression normalized to housekeeping *TBP* mRNA expression upon CRISPR/Cas9-mediated deletion of each CRE using LNCaP clones ($n = 5$ independent experiments, each dot represents an independent clone). **d.** *FOXA1* mRNA expression normalized to housekeeping *TBP* mRNA expression upon CRISPR/Cas9-mediated deletion of each CRE using 22Rv1 clones ($n = 4$ independent experiments, each dot represents an independent clone). **e.** Representative western blot probed against Cas9 in LNCaP clones ($n = 4$ clones) derived to stably express the dCas9-KRAB fusion protein upon blasticidin selection. **f.** Representative western blot probed against Cas9 in 22Rv1 clones ($n = 4$ clones) derived to stably express dCas9-KRAB fusion protein upon blasticidin selection. **g.** *FOXA1* mRNA expression normalized to housekeeping *TBP* mRNA expression upon dCas9-KRAB-mediated repression of each CRE using LNCaP clones ($n = 4$ independent experiments, each dot represents an independent clone). **h.** *FOXA1* mRNA expression normalized to housekeeping *TBP* mRNA expression upon dCas9-KRAB-mediated repression of each CRE using 22Rv1 clones ($n = 4$ independent experiments, each dot represents an independent clone). *FOXA1* mRNA expression was normalized to basal *FOXA1* expression prior to statistical testing. Δ indicates CRISPR/Cas9-mediated deletion, i indicates dCas9-KRAB-mediated repression. Error bars indicate \pm s.d. Student's *t*-test, n.s. not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

$\sim 228\%$) (Student's *t*-test, $p < 0.05$, ??d-f). Collectively, our results support the restriction of most *FOXA1* plexus CREs towards *FOXA1* and its target genes.

1.3.5 *FOXA1* CREs collaborate to regulate its expression

Expanding on the idea that multiple CREs can converge to regulate the expression of a single target gene [31, 32, 42], we asked whether the CREs we identified collaboratively regulate *FOXA1* mRNA expression. Here, we applied a transient approach that delivers Cas9 protein:gRNA as a ribonucleoprotein (RNP) complex formed prior to transfection that would avoid the heterogeneity of Cas9 protein expression across the PCa cell clones (See Methods). We first validated this system through single CRE deletions, where we transiently transfected a set of gRNA targeting the CRE of interest. In accordance with data from our PCa cell clones stably expressing WT Cas9 and dCas9-KRAB, individual CRE deletion resulted in a significant reduction in *FOXA1* mRNA expression: (Δ CRE1 $\sim 29.3 \pm 7.3\%$, Δ CRE2 $\sim 36 \pm 11.8\%$, Δ CRE3 $\sim 30.6 \pm 12.7\%$, Δ CRE4 $\sim 24.5 \pm 6.1\%$, Δ CRE5 $\sim 23.7 \pm 13.2\%$, Δ CRE6 $\sim 26.8 \pm 14.2\%$ and Δ *FOXA1* $\sim 96.2 \pm 1.4\%$ reduction (Student's *t*-test, $p < 0.05$, Figure 1.5a, ??a-f). Next for combinatorial deletions, we prioritized the CREs that harbor more than 1 SNV (i.e CRE1, CRE2, CRE4), and transiently transfected RNP complexes that target both CREs in various combinations (i.e CRE1 + CRE2, CRE1 + CRE4, CRE2 + CRE4), and assessed *FOXA1* mRNA expression. Compared to negative control regions, the combinatorial deletion of Δ CRE1 + Δ CRE2, Δ CRE1 + Δ CRE4, and Δ CRE2 + Δ CRE4 resulted in a

significant $\sim 48.5 \pm 4.5\%$, $\sim 50.4 \pm 2.9\%$ and $\sim 45.2 \pm 5.5\%$ reduction in *FOXA1* mRNA expression, respectively (Student's *t*-test, $p < 0.05$, Figure 1.5b, ??a-f) a fold reduction greater than single CRE deletions (Student's *t*-test, ??, $p < 0.05$). These results together demonstrate that these CREs collaboratively contribute to the establishment and regulation of *FOXA1* expression in PCa.

1.3.6 Disruption of *FOXA1* CREs reduces PCa cell growth

As *FOXA1* is essential for PCa growth (Figure 1.1b-e), we next sought to assess the importance of the six *FOXA1* plexus CREs towards PCa cell growth. We adapted a lentiviral-based approach that expressed both the Cas9 protein and two gRNA that target each CRE for deletion (See Methods). Upon lentiviral transduction with subsequent selection, we separated LNCaP PCa cells for RNA, DNA and for cell proliferation. We first tested the system by measuring *FOXA1* mRNA expression, and independently observed significant reductions of *FOXA1* mRNA expression (Δ CRE1 $\sim 18\%$, Δ CRE2 $\sim 30\%$, Δ CRE3 $\sim 15\%$, Δ CRE4 $\sim 12\%$, Δ CRE5 $\sim 35\%$, Δ CRE6 $\sim 46\%$ and Δ *FOXA1* (exon 1 and intron 1) $\sim 48\%$ reduction (Student's *t*-test, $p < 0.05$, Figure 1.5c, ??a-f). We then seeded these cells at equal density. Six days post-seeding, we harvested the cells and observed a significant reduction in cell growth upon deleting any of the six *FOXA1* plexus CREs (Δ CRE1 $\sim 42\%$, Δ CRE2 $\sim 28\%$, Δ CRE3 $\sim 33\%$, Δ CRE4 $\sim 27\%$, Δ CRE5 $\sim 42\%$, Δ CRE6 $\sim 44\%$ and Δ *FOXA1* (exon 1 and intron 1) $\sim 50\%$ reduction (Student's *t*-test, $p < 0.05$, Fig 5d). These results suggest that the six *FOXA1* plexus contribute to PCa etiology, in agreement with their ability to regulate *FOXA1* expression and the essentiality of this gene in PCa cell growth.

1.3.7 SNVs mapping to *FOXA1* CREs can alter their activity

SNVs can alter the transactivation potential of CREs [32, 43–51]. In total, we found 10 SNVs called from 9 out of the 200 tumours that map to the six *FOXA1* plexus CREs (Figure 1.6a). To assess the impact of these noncoding SNVs, we conducted luciferase assays comparing differential reporter activity between the variant and the WT allele of each CRE (Figure 1.6b-k). We found that the variant alleles of 6 of the 10 SNVs displayed significantly greater luciferase reporter activity when compared to the WT alleles (Mann-Whitney U test, $p < 0.05$). Specifically, we observed the following fold-changes: chr14:37,887,005 A > G (1.65-fold), chr14:37,904,343 A > T (1.35-fold), chr14:37,905,854 A > G (1.28-fold), chr14:37,906,009 T > C (1.71-fold), chr14:38,036,543 A > G (1.44-fold), chr14:38,055,269 C > G (1.39-fold) (Figure 1.6b, d-h). These results indicate that these SNVs can alter the transactivation potential of *FOXA1* plexus CREs in PCa cells.

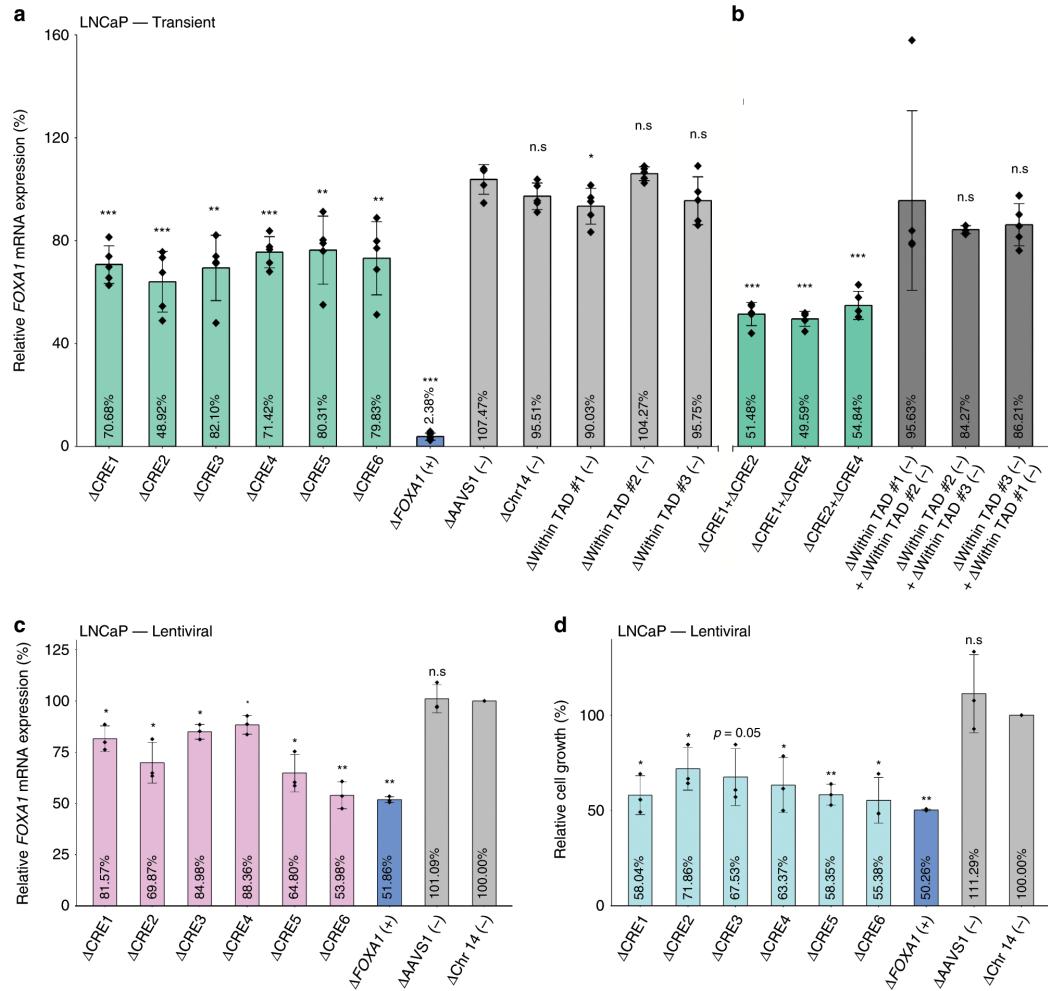


Figure 1.5: *FOXA1* CREs collaborate to regulate its expression and are critical for PCa cell proliferation.. **a.** *FOXA1* mRNA expression normalized to housekeeping *TBP* mRNA expression upon transient transfection-based CRISPR/Cas9-mediated deletion of CRE1, CRE2, CRE4, and sequential deletion combinations ($n = 5$ independent experiments). **b.** *FOXA1* mRNA expression normalized to housekeeping *TBP* mRNA expression upon bulk lentiviral-based CRISPR/Cas9-mediated deletion of each CRE in LNCaP cells ($n = 3$ independent experiments). **c.** Cell proliferation assay conducted after puromycin and blasticidin selection for LNCaP cells carrying deleted regions of interest. Data was based on cell counting 6 days after seeding post-selection ($n = 3$, representative of three independent experiments). *FOXA1* mRNA expression upon deletion was normalized to basal *FOXA1* expression prior to statistical testing. *FOXA1* mRNA expression was normalized to the basal LNCaP *FOXA1* expression prior to statistical testing. Δ indicates CRISPR/Cas9-mediated deletion. Error bars indicate \pm s.d. Student's *t*-test, n.s. not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

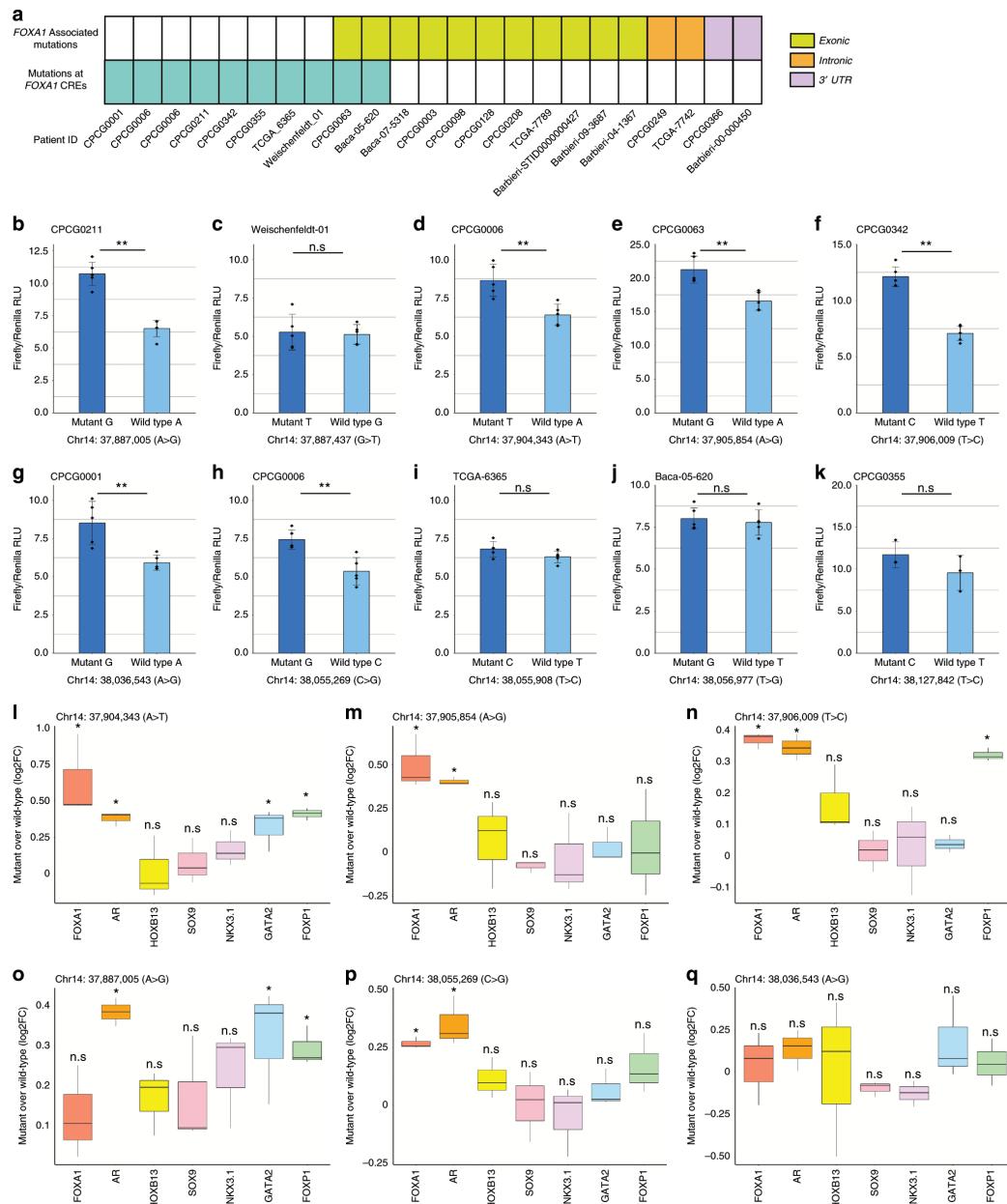


Figure 1.6: A subset of noncoding SNVs mapping to the *FOXA1* CREs are gain-of-function. **a.** Matrix showcasing the patients from the CPC-GENE dataset that harbour SNVs at the *FOXA1* CREs, exons, introns, and the 3' UTR of *FOXA1*. **b-k.** Luciferase assays are conducted in LNCaP cells. Bar plot showcases the mean firefly luciferase activity normalized by *renilla* luciferase activity. Error bars indicate \pm s.d. $n = 5$ independent experiments for all CREs except for chr14:38,127,842 T > C where $n = 3$. Each diamond represents an independent experiment. Hypothesis testing done with Mann-Whitney U test. **l-q.** Allele-specific ChIP-qPCR conducted on plasmids carrying the WT or variant sequence upon transient transfection in PCa cells. Data is presented as \log_2 fold-change of variant sequence upon comparison to WT sequence ($n = 3$ independent experiments per ChIP). Hypothesis testing done with Student's *t*-test, n.s. not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

1.3.8 SNVs mapping to *FOXA1* CREs can modulate the binding of TFs

We next assessed if the changes in transactivation potential induced by noncoding SNVs related to changes in TF binding to CREs by allele-specific ChIP-qPCR [32, 44, 51] in LNCaP PCa cells. We observed differential binding of *FOXA1*, *AR*, *HOXB13*, *GATA2* and *FOXP1* for the chr14:37887005 (A > G) SNV found in CRE1; the chr14:37904343 (A > T), chr14:37905854 (A > G) and chr14:37906009 (T > C) SNVs found in CRE2; and the chr14:38055269 (C > G) SNV found in CRE4 (Student's *t*-test, $p < 0.05$, Figure 1.6l-p). In contrast, *SOX9* and *NKX3.1* binding was unaffected by these SNVs (Figure 1.6l-q). Compared to the WT sequence, chr14:37,887,005 A > G significantly increased *AR* binding (1.31-fold increase), *GATA2* binding (1.25-fold increase) and *FOXP1* binding (1.23-fold increase); chr14:37,904,343 A > T significant increased *AR* binding (1.30-fold increase), *GATA2* (1.25-fold increase) and *FOXP1* (1.33-fold increase); chr14:37,905,854 A > G significantly increased *FOXA1* binding (1.41-fold increase) and *AR* binding (1.33-fold increase); chr14:37,906,009 T > C significantly increased the binding of *FOXA1* (1.29-fold increase), *AR* (1.31-fold increase), *HOXB13* (1.13-fold increase) and *FOXP1* (1.25-fold increase); and chr14:38,055,269 C > G significantly increased *FOXA1* binding (1.20-fold increase). Notably all six SNVs increased the binding of the TFs known to bind at these CREs. In contrast, none of the SNVs significantly decreased the binding of these TFs. Our observations suggest that gain-of-function SNVs populate the *FOXA1* plexus CREs.

1.4 Discussion

Modern technologies and understanding of the epigenome allow the possibility of probing CRE(s) involved in regulating genes implicated in disease. Despite *FOXA1* being recurrently mutated [5–8, 11] and playing potent oncogenic roles in prostate cancer etiology [9, 10, 13], the CREs involved in its transcriptional regulation are poorly understood. Understanding how *FOXA1* is expressed can provide a complementary strategy to antagonize *FOXA1* in prostate cancer.

We used the DHS profiled in prostate cancer cells to identify putative *FOXA1* CREs through annotating these regions with five different histone modifications, TF binding sites and noncoding SNVs profiled in prostate cancer cells and primary prostate tumours. Our efforts identified and validated a set of six active CREs involved in *FOXA1* regulation, agreeing with a recent report where a subset of our CREs map to loci suggested to be in contact with the *FOXA1* promoter [52]. The disruption of these six distal CREs each significantly reduced *FOXA1* mRNA levels, similar to

what has been demonstrated for ESR1 in luminal breast cancer [32], MLH1 in Lynch syndrome [53], MYC in lung adenocarcinoma and endometrial cancer [54], and AR in mCRPC [55, 56]. Through combinatorial deletion of two CREs, *FOXA1* mRNA levels were further reduced in comparison with single CRE deletions, raising the possibility of CRE additivity [57]. The deletion of the *FOXA1* plexus CREs also significantly reduced prostate cancer cell proliferation at levels comparable to what has been reported upon deletion of the amplified CRE upstream of the AR gene in mCRPC [55], suggestive of onco-CREs as reported in lung [54] and prostate [55] cancer.

More than 90% of SNVs found in cancer map to the noncoding genome [58, 59] with a portion of these SNVs mapping to CREs altering their transactivation potential [32, 44–46] and/or downstream target gene expression [48, 58, 60]. We extended this concept with SNVs identified from primary prostate tumours mapping to *FOXA1* plexus CREs. We observed that a subset of these SNVs can alter transactivation potential by modulating the binding of specific TFs whose cistromes are preferentially burdened by SNVs in primary prostate cancer [59]. Our findings complement recent reports of SNVs found in the noncoding space of *FOXA1* that could affect its expression [14, 61]. The *FOXA1* plexus CREs we identified here are also reported to be target of structural variants in both the primary and metastatic settings [9, 62], including tandem duplication in ~14% (14/101) mCRPC tumours over CRE2 [62], amplification, duplication and translocation over CRE3, CRE4, and CRE5 [9]. Notably, the translocation and duplication defining the FOXMIND enhancer driving *FOXA1* expression reported in primary and metastatic settings harbors the CRE3 element we characterized [9]. Collectively, these studies combined with our discoveries reveal the fundamental contribution of the *FOXA1* plexus in prostate cancer etiology. As a whole, our findings in conjunction with recent reports suggest that CREs involved in the transcriptional regulation of *FOXA1* may be hijacked in prostate tumours through various types of genetic alterations.

Despite initial treatment responses from treating aggressive primary and metastatic prostate cancer through castration to suppress AR signalling [4], resistance ensues as 80% of mCRPC tumours harbor either AR gene amplification, amplification of a CRE upstream of AR, or activating AR coding mutations [11, 55, 62]. Given the AR-dependent [15, 18] and AR-independent [25] oncogenic activity of *FOXA1* in prostate cancer, its inhibition is an appealing alternative therapeutic strategy. Our dissection of the *FOXA1* cis-regulatory landscape complement recent findings through revealing loci that are important for the regulation of *FOXA1*. Theoretically, direct targeting of the CREs regulating *FOXA1* would down-regulate *FOXA1* levels and could therefore serve as a valid alternative to antagonize its function.

Taken together, we identified *FOXA1* CREs targeted by SNVs that are capable of altering

transactivation potential through the modulation of key prostate cancer TFs. The study supports the importance of considering CREs not only as lone occurrences but as a team that works together to regulate their target genes, particularly when considering the impact of genetic alterations. As such, our work builds a bridge between the understanding of FOXA1 transcriptional regulation and new routes to FOXA1 inhibition. Aligning with recent reports [9, 10, 13], our findings support the oncogenic nature of FOXA1 in prostate cancer. Gaining insight on the cis-regulatory plexuses of important genes such as FOXA1 in prostate cancer may provide new avenues to inhibit other drivers across various cancer types to halt disease progression.

1.5 Methods

1.5.1 Cell Culture

LNCaP and 22Rv1 cells were cultured in RPMI medium, and VCaP cells were cultured in DMEM medium, both supplemented with 10% FBS, and 1% penicillin-streptomycin at 37 °C in a humidified incubator with 5% CO₂. These prostate cancer cells originated from ATCC. 293FT cells were purchased from ThermoFisherScientific (Cat No. R70007) maintained in complete DMEM medium (DMEM with 10% FBS (080150, Wisent), L-glutamine (25030-081, ThermoFisher) and non-essential amino acids (11140-050, ThermoFisher) supplemented with 50mg/mL Geneticin (4727894001, Sigma-Aldrich). The cells are regularly tested for Mycoplasma contamination. The authenticity of these cells was confirmed through Short Tandem Repeat profiling.

1.5.2 Prostate tumours and cancer cell lines expression

Cancer cell line mRNA abundance data were collected from the Cancer Dependency Map Project (DEPMAP; <https://depmap.org/portal/>; RNA-seq TPM values from 2018q4 version with all 5 non-cancer cell lines were removed) [63] projects. Prostate tumour mRNA abundance data was collected from The Cancer Genome Atlas (TCGA) prostate cancer (TCGA-PRAD) project via the Xena Browser (<https://xenabrowser.net/>; dataset description: TCGA prostate adenocarcinoma (PRAD) gene expression by RNA-seq (polyA+ Illumina HiSeq; RSEM)).

1.5.3 Prostate cancer cell line gene essentiality

Essentiality scores were collected from the Cancer Dependency Map Project [64]. To compare gene essentiality between prostate cancer cell lines and others, essentiality scores for FOXA1 were

collected from all available cell lines ($n = 707$). To perform a permutation test, the median of 8 randomly selected cell lines was calculated one million times to generate a background distribution of essentiality scores across all cell types available. The median essentiality score from the 8 prostate cancer cell lines was calculated and its percentile within the background distribution is reported.

1.5.4 siRNA knockdown and cell proliferation assay

300,000 LNCaP cells (Day 0) were reverse transfected with siRNA (siFOXA1 using Lipofectamine®RNAimax reagent (ThermoFisher Scientific, Cat No. 13778150)). Cells were counted using Countess™ automated cell counter (Invitrogen). Whole cell lysates LNCaP cells after siRNA-mediated FOXA1 knockdown was collected at 96-hours post-transfection in RIPA buffer. Protein concentrations were determined through the bicinchoninic acid method (ThermoFisher Scientific, Cat No. 23225). Then 25 μ g of lysate was subjected to SDS-PAGE. Upon completion of SDS-PAGE, protein was transferred onto PVDF membrane (Bio-Rad, Cat No. 1704156). The membrane was blocked with 5% non-fat milk for one hour at room temperature with shaking. After blocking, anti-FOXA1 (Abcam Cat No. 23737) in 2.5% non-fat milk was added, and was incubated at 4 °C overnight. Next day, the blot was washed and incubated with IRDye®800CW Goat Anti-Rabbit IgG secondary antibody (LI-COR, Cat No. 925-32211) at room temperature for 1 hour. The blot was then washed and assessed with the Odyssey®CLX imaging system (LI-COR).

1.5.5 Identifying putative FOXA1 CREs

Putative FOXA1 CREs were identified through the use of Cross Cell-Type Correlation based on DNase I Hypersensitivity (C3D) (<https://github.com/tahmidmehdi/C3D>) [37]. Predicted interacting DNase I Hypersensitivity Sites (DHS) with a Pearson’s correlation above 0.7 [65] were kept for downstream analysis.

1.5.6 Hi-C and TADs in LNCaP cells

Hi-C and TADs conducted and called, respectively, in LNCaP cells are publicly available off ENCODE portal (ENCSR346DCU). Visualization of the Hi-C dataset is available on the Hi-C Browser (<http://promoter.bx.psu.edu/hi-c/>) [66].

1.5.7 Clonal wild-type Cas9 and dCas9-KRAB mediated validation

Lentiviral particles were generated in 293FT cells (ThermoFisher) using the pMDG.2 and psPAX2 packaging plasmids (Addgene; #12259 and #12260, a gift from Didier Trono) alongside the Lenti-Cas9-2A-Blast plasmid (Addgene #73310, a gift from Jason Moffat) and collected 72 hrs post transfection. LNCaP and 22Rv1 cells were then transduced for 24-48 hours with equal amounts of virus followed by selection with media containing blasticidin (7.5 μ g/mL for LNCaP cells, 6 μ g/mL for 22Rv1 cells). Upon selection, clones were derived by serial dilution with subsequent single cell seeding into 96-well plates containing selection media. Cas9 protein expression for each clone was then assessed through Western blotting (1 $^{\circ}$ Ms-Cas9 (Cell Signalling Technology, Cat No. #14697) 1:1000, Ms-GAPDH 1:5000 (Santa Cruz Biotechnology, Cat No. #sc47724) in 5% non-fat milk; 2 $^{\circ}$ HRP-linked Anti-Mouse IgG (Cell Signalling Technology, Cat No. #7076S) 1:10 000 in 2.5% non-fat milk. The full unprocessed blot is in the Source Data File.

Lentiviral particles were generated in 293FT cells (ThermoFisher) using the pMDG.2 and psPAX2 packaging plasmids (Addgene; #12259 and #12260, a gift from Didier Trono) alongside the Lenti-dCas9-KRAB-blast plasmid (Addgene #89567, a gift from Gary Hon) and collected 72 hrs post transfection. LNCaP and 22Rv1 cells were then transduced for 24-48 hours with equal amounts of virus followed by selection with media containing blasticidin (7.5 μ g/mL for LNCaP cells, 6 μ g/mL for 22Rv1 cells). Upon selection, clones were derived by serial dilution with subsequent single cell seeding into 96-well plates containing selection media. dCas9-KRAB protein expression for each clone was then assessed through Western blotting (1 $^{\circ}$ Ms-Cas9 (Cell Signalling Technology, Cat No. #14697) 1:1000, Ms-GAPDH 1:5000 (Santa Cruz Biotechnology, Cat No. #sc47724) in 5% non-fat milk; 2 $^{\circ}$ HRP-linked Anti-Mouse IgG (Cell Signalling Technology, Cat No. #7076S) 1:10 000 in 2.5% non-fat milk. The full unprocessed blot is in the Source Data File.

For gRNA design, five to six unique crRNA molecules (Integrated DNA Technologies) were designed to tile across the region of interest using the CRISPR (<http://crispor.tefor.net/>) [67] and the Zhang lab CRISPR Design tools (<http://crispr.mit.edu/>) [68]. See published manuscript for gRNA. Each CRISPR RNA (crRNA) and tracrRNA (Integrated DNA Technologies) were duplexed according to company supplier protocol to a concentration of 50 μ M. Upon generation of the clones, six guides (crRNA-tracrRNA duplexes) for each region of interest were pooled into a single tube (1 μ L each guide, 6 μ L per reaction) (Integrated DNA Technologies). Lastly, 1 μ L (100 μ M) of electroporation enhancer (Integrated DNA Technologies) was added to the mix (7 μ L total) prior to transfection. The entire transfection reaction was transfected into 350 000 cells through Nu-

cleofection (SF Solution EN120 - 4D Nucleofector, Lonza). Cells were then harvested 24 hours post-transfection for RNA and DNA for RT-PCR and confirmation of deletion, respectively.

1.5.8 Transient Cas9-mediated disruption of CREs

Deletion of elements through this method were achieved through the transfection of Cas9 nucleic acid protein complexed with the crRNA (Integrated DNA Technologies). Briefly, five to six unique crRNA molecules (Integrated DNA Technologies) were designed to tile across the region of interest using the CRISPR (<http://crispor.tefor.net/>) [67] and the Zhang lab CRISPR Design tools (<http://crispr.mit.edu/>) [68]. Each crRNA and tracrRNA (Integrated DNA Technologies) were duplexed according to company supplier protocol to a concentration of 50 μ M. The six crRNA-tracrRNA duplexes were pooled into a single tube (6 μ L per reaction), prior to adding 1 μ L (5 μ g) of Alt-R \textcircled{R} S.p HiFi Cas9 Nuclease 3NLS (Integrated DNA Technologies). The reaction was incubated at room temperature for 10 minutes for ribonucleoprotein (RNP) complex formation. Lastly, 1 μ L (100 μ M) of electroporation enhancer (Integrated DNA Technologies) was added to the mix prior to transfection. The entire transfection reaction was transfected into 350 000 cells through Nucleofection (SF Solution EN120 - 4D Nucleofector, Lonza). Cells were then harvested 24 hours post-transfection for RNA and DNA for RT-PCR and confirmation of deletion, respectively. For double deletions, two sets of guide RNA-RNP complex (10 μ g of Alt-R \textcircled{R} S.p HiFi Cas9 Nuclease 3NLS) were transfected and harvested 24 hours post-transfection for RNA and DNA for RT-PCR and confirmation of deletion, respectively. To control for double deletions, two negative control regions within the TAD were also compounded. Due to size, see published manuscript for primers.

1.5.9 RT-PCR assessment of gene expression upon deletion of CREs

DNA and RNA were harvested with Qiagen AllPrep RNA/DNA Kit (Qiagen, Cat No. 80204). Next, cDNA was synthesized from 300 ng of RNA using SensiFast cDNA Synthesis kit (Bioline, Cat No. BIO-65054), and mRNA expression levels for various genes of interest were assessed. Due to size, see published manuscript for the primer sequences used for expression evaluation. Differential gene expression was calculated upon normalization with TBP (housekeeping gene). Statistical significance was calculated using Student's t-test in R.

1.5.10 Confirmation of Cas9-mediated deletion of CREs

Deletion of CREs were confirmed through PCR amplification of the intended region for deletion, followed by the T7 Endonuclease Assay (Integrated DNA Technology). Due to size, see published manuscript for primer sequences used for PCR amplification. PCR products were then loaded onto a 1% agarose gel for visualization. The agarose gel to assess the on-target genome editing efficiency was done through densitometry using ImageJ. The correlation between on-target genome editing efficiency and *FOXA1* mRNA expression reduction was drawn through Pearson's correlation in R.

1.5.11 Cell proliferation upon deletion of *FOXA1* CREs

Pairs of gRNAs flanking the CREs of interest, *FOXA1* promoter and control regions were designed using CRISPOR (<http://crispor.tefor.net/>) and Zhang lab CRISPR Design tool (<http://crispr.mit.edu/>) (due to size, see published manuscript). Each pair of gRNAs were cloned into the lentiCRISPRv2 (Addgene; a gift from Feng Zhang #52961) and the lentiCRISPRv2-Blast (Addgene; a gift from Feng Zhang #83480) plasmid as previously described [69]. Lentiviral particles were generated in 293FT cells (ThermoFisher) using the pMDG.2 and psPAX2 packaging plasmids (Addgene; #12259 and #12260, a gift from Didier Trono), and collected 72 hrs post transfection. LNCaP cells were transduced for 24-48hrs with equal amounts of virus, followed by selection with media containing puromycin (3.5 µg/mL, ThermoFisher) and blasticidin (7 µg/mL, Wisent). Cells were harvested upon selection for RNA and DNA for RT-PCR and confirmation of DNA cleavage, respectively. For cell proliferation, cells were seeded at equal density per well (on a 96-well plate; Day 1) upon puromycin and blasticidin selection. Growth of the cells were monitored through cell counting using Countess™ automated cell counter (Invitrogen). Cell numbers were calculated as a percentage compared to negative control. Statistical significance was calculated using Student's t-test.

1.5.12 Luciferase reporter assays

Each region of interest was ordered as gBlocks from Integrated DNA Technologies. The regions were cloned into the BamHI restriction enzyme digest site of the pGL3 promoter plasmid (Promega). On Day 0, 90 000 LNCaP cells were seeded in 24-well plates. Next day (Day 1), pGL3 plasmids harboring the wild-type and variant sequences were co-transfected with the pRL Renilla plasmid (Promega) using Lipofectamine 2000. 48-hours later, the cells were harvested, and dual luciferase reporter assays were conducted (Promega). Notably, inserts of both forward and reverse directions were tested using this assay as enhancer elements are known to be direction-independent. Final

luminescence readings are reported as firefly luciferase normalized to renilla luciferase activity. The assessment of each mutation was conducted in five biological replicates. Statistical significance was assessed by Mann-Whitney U test in R. See published manuscript for gBlock sequences.

1.5.13 Allele-specific ChIP-qPCR

Briefly, pGL3 plasmids containing the wild-type sequence and the mutant sequence used in the luciferase reporter assay were transfected into 7 million cells (2 μ g per allele, per 1 million cells) using Lipofectamine 2000 (ThermoFisher Scientific), per manufacturer's instructions. Next day, each antibody (FOXA1 5 μ g, Abcam, ab23738; AR 5 μ g, Abcam, ab1083241; HOXB13 5 μ g, Abcam, ab201682; SOX9 5 μ g, Abcam, ab3697; GATA2 5 μ g, Abcam, ab22849; FOXP1 5 μ g, Abcam, ab16645; NKX3.1 10 μ l, Cell Signalling Technology, #83700) was conjugated with 10 uL of each Dynabeads A and G (Thermo Fisher Scientific) for each ChIP for 6 hours with rotation at 4 °C. When antibody-beads conjugates were ready for use, cells were lifted using trypsin and fixed by re-suspending with 300 uL of 1% formaldehyde in PBS for 10 minutes at room temperature. 2.5M Glycine was added to quench excess formaldehyde (final concentration 0.125 M). Cells were then washed with cold PBS and lysed using 300 uL of Modified RIPA buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 140 mM NaCl; 1% Triton X-100; 0.1% SDS; 0.1% sodium deoxycholate) supplemented with protease inhibitor. The lysate was subject to 25 cycles of sonication (30s ON 30s OFF) using Diagenode Bioruptor Pico (Diagenode). 15 uL of sonicated lysate was set aside as input with the rest used for chromatin pulldown through addition of antibody-beads conjugates and overnight incubation at 4 °C with rotation. Next day, the beads were washed once with Modified RIPA buffer, washed once with Modified RIPA buffer + 500 mM NaCl, once with LiCl buffer (10 mM TrisHCl, pH 8.0; 1 mM EDTA; 250 mM LiCl; 0.5% NP-40; 0.5% sodium deoxycholate) and twice with Tris-ETDA buffer (pH 8). After washes, beads and input were de-crosslinked by addition of 100 μ L De-crosslinking buffer and incubation at 65 °C for 6 hours. Samples were then purified and eluted. ChIP and input DNA were then used for allele-specific ChIP-qPCR using MAMA primers as described previously. Fold-change significance was calculated using Student's t-test in R.

All analyses were done using hg19 reference genome coordinates.

1.6 Data availability

Genomic and Epigenomic data sets used to support this study can be found from the following accession codes: primary tumors—H3K27ac ChIP-seq (GSE96652), SNVs called from primary tu-

mors (<https://dcc.icgc.org/projects/PRAD-CA>), FOXA1, AR, and HOXB13 ChIP-seq in primary prostate tumors is available under the following accession code: GSE137527 and EGAS00001003928, TF ChIP-seq data were from public databases of ReMap and ChIP-Atlas. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request.

Glossary

3C chromatin conformation capture

AR androgen receptor

ChIP-seq chromatin immunoprecipitation sequencing

CPC-GENE Canadian Prostate Cancer Genome Network

CRE *cis*-regulatory element

DEPMAP Cancer Dependency Map

DHS DNase I hypersensitive sites

FOX forkhead box

gRNA guide RNA

kbp kilobase

mCRPC metastatic castration-resistant prostate cancer

mRNA messenger RNA

RNAi RNA interference

RNA-seq RNA sequencing

SNV single nucleotide variants

PCa prostate cancer

TAD topologically associated domain

TCGA The Cancer Genome Atlas

TF transcription factor

UTR untranslated region

WGS whole genome sequencing

WT wild-type

References

1. Bray, F. *et al.* Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. en. *CA: A Cancer Journal for Clinicians* **68**, 394–424. ISSN: 00079235 (Nov. 2018).
2. Boorjian, S. A. *et al.* Long-Term Outcome After Radical Prostatectomy for Patients With Lymph Node Positive Prostate Cancer in the Prostate Specific Antigen Era. en. *Journal of Urology* **178**, 864–871. ISSN: 0022-5347, 1527-3792 (Sept. 2007).
3. Litwin, M. S. & Tan, H.-J. The Diagnosis and Treatment of Prostate Cancer: A Review. en. *JAMA* **317**, 2532. ISSN: 0098-7484 (June 2017).
4. Attard, G. *et al.* Prostate Cancer. en. *The Lancet* **387**, 70–82. ISSN: 01406736 (Jan. 2016).
5. Abeshouse, A. *et al.* The Molecular Taxonomy of Primary Prostate Cancer. en. *Cell* **163**, 1011–1025. ISSN: 00928674 (Nov. 2015).
6. Fraser, M. *et al.* Genomic Hallmarks of Localized, Non-Indolent Prostate Cancer. en. *Nature* **541**, 359–364. ISSN: 1476-4687 (Jan. 2017).
7. Barbieri, C. E. *et al.* Exome Sequencing Identifies Recurrent SPOP , FOXA1 and MED12 Mutations in Prostate Cancer. en. *Nature Genetics* **44**, 685–689. ISSN: 1546-1718 (June 2012).
8. Grasso, C. S. *et al.* The Mutational Landscape of Lethal Castration-Resistant Prostate Cancer. en. *Nature* **487**, 239–243. ISSN: 0028-0836, 1476-4687 (July 2012).
9. Parolia, A. *et al.* Distinct Structural Classes of Activating FOXA1 Alterations in Advanced Prostate Cancer. en. *Nature* **571**, 413–418. ISSN: 1476-4687 (July 2019).
10. Adams, E. J. *et al.* FOXA1 Mutations Alter Pioneering Activity, Differentiation and Prostate Cancer Phenotypes. en. *Nature* **571**, 408–412. ISSN: 0028-0836, 1476-4687 (July 2019).
11. Robinson, D. *et al.* Integrative Clinical Genomics of Advanced Prostate Cancer. en. *Cell* **161**, 1215–1228. ISSN: 00928674 (May 2015).

12. Robinson, J. L. L., Holmes, K. A. & Carroll, J. S. FOXA1 Mutations in Hormone-Dependent Cancers. *Frontiers in Oncology* **3**. ISSN: 2234-943X (2013).
13. Gao, S. *et al.* Forkhead Domain Mutations in FOXA1 Drive Prostate Cancer Progression. en. *Cell Research* **29**, 770–772. ISSN: 1001-0602, 1748-7838 (Sept. 2019).
14. Annala, M. *et al.* Frequent Mutation of the FOXA1 Untranslated Region in Prostate Cancer. en. *Communications Biology* **1**, 122. ISSN: 2399-3642 (Aug. 2018).
15. Yang, Y. A. & Yu, J. Current Perspectives on FOXA1 Regulation of Androgen Receptor Signaling and Prostate Cancer. en. *Genes & Diseases* **2**, 144–151. ISSN: 23523042 (June 2015).
16. Lupien, M. *et al.* FoxA1 Translates Epigenetic Signatures into Enhancer-Driven Lineage-Specific Transcription. *Cell* **132**, 958–970. ISSN: 0092-8674 (Mar. 2008).
17. Eeckhoute, J. *et al.* Cell-Type Selective Chromatin Remodeling Defines the Active Subset of FOXA1-Bound Enhancers. en. *Genome Research* **19**, 372–380. ISSN: 1088-9051 (Dec. 2008).
18. Pomerantz, M. M. *et al.* The Androgen Receptor Cistrome Is Extensively Reprogrammed in Human Prostate Tumorigenesis. en. *Nature Genetics* **47**, 1346–1351. ISSN: 1061-4036, 1546-1718 (Nov. 2015).
19. Imamura, Y. *et al.* FOXA1 Promotes Tumor Progression in Prostate Cancer via the Insulin-Like Growth Factor Binding Protein 3 Pathway. en. *PLoS ONE* **7** (ed Agoulnik, I.) e42456. ISSN: 1932-6203 (Aug. 2012).
20. Xu, Y., Chen, S.-Y., Ross, K. N. & Balk, S. P. Androgens Induce Prostate Cancer Cell Proliferation through Mammalian Target of Rapamycin Activation and Post-Transcriptional Increases in Cyclin D Proteins. en. *Cancer Research* **66**, 7783–7792. ISSN: 0008-5472, 1538-7445 (Aug. 2006).
21. Jin, H.-J., Zhao, J. C., Ogden, I., Bergan, R. C. & Yu, J. Androgen Receptor-Independent Function of FoxA1 in Prostate Cancer Metastasis. en. *Cancer Research* **73**, 3725–3736. ISSN: 0008-5472, 1538-7445 (June 2013).
22. Yang, Y. A. *et al.* FOXA1 Potentiates Lineage-Specific Enhancer Activation through Modulating TET1 Expression and Function. en. *Nucleic Acids Research* **44**, 8153–8164. ISSN: 0305-1048, 1362-4962 (Sept. 2016).
23. Zhang, G. *et al.* FOXA1 Defines Cancer Cell Specificity. en. *Science Advances* **2**, e1501473. ISSN: 2375-2548 (Mar. 2016).

24. Augello, M. A., Hickey, T. E. & Knudsen, K. E. FOXA1: Master of Steroid Receptor Function in Cancer: FOXA1: Master of Steroid Receptor Function in Cancer. en. *The EMBO Journal* **30**, 3885–3894. ISSN: 02614189 (Oct. 2011).
25. Sunkel, B. *et al.* Integrative Analysis Identifies Targetable CREB1/FoxA1 Transcriptional Co-Regulation as a Predictor of Prostate Cancer Recurrence. en. *Nucleic Acids Research* **45**, 6993–6993. ISSN: 0305-1048, 1362-4962 (June 2017).
26. Ni, M. *et al.* Amplitude Modulation of Androgen Signaling by C-MYC. en. *Genes & Development* **27**, 734–748. ISSN: 0890-9369 (Apr. 2013).
27. Sasse, S. K. & Gerber, A. N. Feed-Forward Transcriptional Programming by Nuclear Receptors: Regulatory Principles and Therapeutic Implications. en. *Pharmacology & Therapeutics* **145**, 85–91. ISSN: 01637258 (Jan. 2015).
28. Wang, S., Singh, S., Katika, M., Lopez-Aviles, S. & Hurtado, A. High Throughput Chemical Screening Reveals Multiple Regulatory Proteins on FOXA1 in Breast Cancer Cell Lines. en. *International Journal of Molecular Sciences* **19**, 4123. ISSN: 1422-0067 (Dec. 2018).
29. Rowley, M. J. & Corces, V. G. Organizational Principles of 3D Genome Architecture. en. *Nature Reviews Genetics* **19**, 789–800. ISSN: 1471-0056, 1471-0064 (Dec. 2018).
30. Vernimmen, D. & Bickmore, W. A. The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. en. *Trends in Genetics* **31**, 696–708. ISSN: 01689525 (Dec. 2015).
31. Sallari, R. C. *et al.* Convergence of Dispersed Regulatory Mutations Predicts Driver Genes in Prostate Cancer. *bioRxiv*, 38–38 (2016).
32. Bailey, S. D. *et al.* Noncoding Somatic and Inherited Single-Nucleotide Variants Converge to Promote ESR1 Expression in Breast Cancer. *Nature Genetics* **48**, 1260–1269 (2016).
33. Tsourlakis, M. C. *et al.* FOXA1 Expression Is a Strong Independent Predictor of Early PSA Recurrence in ERG Negative Prostate Cancers Treated by Radical Prostatectomy. en. *Carcinogenesis* **38**, 1180–1187. ISSN: 0143-3334, 1460-2180 (Dec. 2017).
34. Phillips, J. E. & Corces, V. G. CTCF: Master Weaver of the Genome. en. *Cell* **137**, 1194–1211. ISSN: 00928674 (June 2009).
35. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. en. *Cell* **171**, 1573–1588.e28. ISSN: 00928674 (Dec. 2017).
36. Bailey, S. D. *et al.* ZNF143 Provides Sequence Specificity to Secure Chromatin Interactions at Gene Promoters. en. *Nature Communications* **6**, 6186. ISSN: 2041-1723 (May 2015).

37. Mehdi, T., Bailey, S. D., Guilhamon, P. & Lupien, M. C3D: A Tool to Predict 3D Genomic Interactions between Cis-Regulatory Elements. en. *Bioinformatics* **35** (ed Kelso, J.) 877–879. ISSN: 1367-4803, 1460-2059 (Mar. 2019).
38. Kron, K. J. *et al.* TMPRSS2–ERG Fusion Co-Opts Master Transcription Factors and Activates NOTCH Signaling in Primary Prostate Cancer. en. *Nature Genetics* **49**, 1336–1345. ISSN: 1546-1718 (Sept. 2017).
39. Creyghton, M. P. *et al.* Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State. en. *Proceedings of the National Academy of Sciences* **107**, 21931–21936. ISSN: 0027-8424, 1091-6490 (Dec. 2010).
40. Espiritu, S. M. G. *et al.* The Evolutionary Landscape of Localized Prostate Cancers Drives Clinical Aggression. English. *Cell* **173**, 1003–1013.e15. ISSN: 0092-8674, 1097-4172 (May 2018).
41. DeKelver, R. C. *et al.* Functional Genomics, Proteomics, and Regulatory DNA Analysis in Isogenic Settings Using Zinc Finger Nuclease-Driven Transgenesis into a Safe Harbor Locus in the Human Genome. en. *Genome Research* **20**, 1133–1142. ISSN: 1088-9051 (Aug. 2010).
42. Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. Enhancers: Five Essential Questions. en. *Nature Reviews Genetics* **14**, 288–295. ISSN: 1471-0056, 1471-0064 (Apr. 2013).
43. Rheinbay, E. *et al.* Recurrent and Functional Regulatory Mutations in Breast Cancer. *Nature* **547**, 55–60. ISSN: 1476-4687 (Electronic) 0028-0836 (Linking) (June 2017).
44. Zhang, X., Cowper-Sal{middle dot}lari, R., Bailey, S. D., Moore, J. H. & Lupien, M. Integrative Functional Genomics Identifies an Enhancer Looping to the SOX9 Gene Disrupted by the 17q24.3 Prostate Cancer Risk Locus. en. *Genome Research* **22**, 1437–1446. ISSN: 1088-9051 (Aug. 2012).
45. Huang, F. W. *et al.* Highly Recurrent TERT Promoter Mutations in Human Melanoma. en. *Science* **339**, 957–959. ISSN: 0036-8075, 1095-9203 (Feb. 2013).
46. Horn, S. *et al.* TERT Promoter Mutations in Familial and Sporadic Melanoma. en. *Science* **339**, 959–961. ISSN: 0036-8075, 1095-9203 (Feb. 2013).
47. Fuxman Bass, J. I. *et al.* Human Gene-Centered Transcription Factor Networks for Enhancers and Disease Variants. en. *Cell* **161**, 661–673. ISSN: 00928674 (Apr. 2015).
48. Zhou, S., Treloar, A. E. & Lupien, M. Emergence of the Noncoding Cancer Genome: A Target of Genetic and Epigenetic Alterations. *Cancer Discovery* **6**, 1215–1229 (Nov. 2016).

49. Feigin, M. E. *et al.* Recurrent Noncoding Regulatory Mutations in Pancreatic Ductal Adeno-carcinoma. en. *Nature Genetics* **49**, 825–833. ISSN: 1061-4036, 1546-1718 (June 2017).
50. Khurana, E. *et al.* Role of Non-Coding Sequence Variants in Cancer. en. *Nature Reviews Genetics* **17**, 93–108. ISSN: 1471-0056, 1471-0064 (Feb. 2016).
51. Cowper-Sal-lari, R. *et al.* Breast Cancer Risk–Associated SNPs Modulate the Affinity of Chromatin for FOXA1 and Alter Gene Expression. en. *Nature Genetics* **44**, 1191–1198. ISSN: 1061-4036, 1546-1718 (Nov. 2012).
52. Rhie, S. K. *et al.* A High-Resolution 3D Epigenomic Map Reveals Insights into the Creation of the Prostate Cancer Transcriptome. en. *Nature Communications* **10**, 1–12. ISSN: 2041-1723 (Sept. 2019).
53. Liu, Q. *et al.* Disruption of a -35 Kb Enhancer Impairs CTCF Binding and *MLH1* Expression in Colorectal Cells. en. *Clinical Cancer Research* **24**, 4602–4611. ISSN: 1078-0432, 1557-3265 (Sept. 2018).
54. Zhang, X. *et al.* Identification of Focally Amplified Lineage-Specific Super-Enhancers in Human Epithelial Cancers. en. *Nature Genetics* **48**, 176–182. ISSN: 1546-1718 (Feb. 2016).
55. Takeda, D. Y. *et al.* A Somatically Acquired Enhancer of the Androgen Receptor Is a Noncoding Driver in Advanced Prostate Cancer. English. *Cell* **174**, 422–432.e13. ISSN: 0092-8674, 1097-4172 (July 2018).
56. Viswanathan, S. R. *et al.* Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. en. *Cell* **174**, 433–447.e19. ISSN: 00928674 (July 2018).
57. Osterwalder, M. *et al.* Enhancer Redundancy Provides Phenotypic Robustness in Mammalian Development. en. *Nature* **554**, 239–243. ISSN: 0028-0836, 1476-4687 (Feb. 2018).
58. Melton, C., Reuter, J. A., Spacek, D. V. & Snyder, M. Recurrent Somatic Mutations in Regulatory Regions of Human Cancer Genomes. en. *Nature Genetics* **47**, 710–716. ISSN: 1061-4036, 1546-1718 (July 2015).
59. Mazrooei, P. *et al.* Cistrome Partitioning Reveals Convergence of Somatic Mutations and Risk Variants on Master Transcription Regulators in Primary Prostate Tumors. English. *Cancer Cell* **36**, 674–689.e6. ISSN: 1535-6108, 1878-3686 (Dec. 2019).

60. Weinhold, N., Jacobsen, A., Schultz, N., Sander, C. & Lee, W. Genome-Wide Analysis of Noncoding Regulatory Mutations in Cancer. en. *Nature Genetics* **46**, 1160–1165. ISSN: 1061-4036, 1546-1718 (Nov. 2014).
61. CAMCAP Study Group *et al.* Sequencing of Prostate Cancers Identifies New Cancer Genes, Routes of Progression and Drug Targets. en. *Nature Genetics* **50**, 682–692. ISSN: 1061-4036, 1546-1718 (May 2018).
62. Quigley, D. A. *et al.* Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. English. *Cell* **174**, 758–769.e9. ISSN: 0092-8674, 1097-4172 (July 2018).
63. The Cancer Cell Line Encyclopedia Consortium & The Genomics of Drug Sensitivity in Cancer Consortium. Pharmacogenomic Agreement between Two Cancer Cell Line Data Sets. en. *Nature* **528**, 84–87. ISSN: 0028-0836, 1476-4687 (Dec. 2015).
64. McFarland, J. M. *et al.* Improved Estimation of Cancer Dependencies from Large-Scale RNAi Screens Using Model-Based Normalization and Data Integration. en. *Nature Communications* **9**, 4610. ISSN: 2041-1723 (Dec. 2018).
65. Thurman, R. E. *et al.* The Accessible Chromatin Landscape of the Human Genome. en. *Nature* **489**, 75–82. ISSN: 0028-0836, 1476-4687 (Sept. 2012).
66. Wang, Y. *et al.* The 3D Genome Browser: A Web-Based Browser for Visualizing 3D Genome Organization and Long-Range Chromatin Interactions. en. *Genome Biology* **19**, 151. ISSN: 1474-760X (Dec. 2018).
67. Haeussler, M. *et al.* Evaluation of Off-Target and on-Target Scoring Algorithms and Integration into the Guide RNA Selection Tool CRISPOR. en. *Genome Biology* **17**, 148. ISSN: 1474-760X (Dec. 2016).
68. Hsu, P. D. *et al.* DNA Targeting Specificity of RNA-Guided Cas9 Nucleases. en. *Nature Biotechnology* **31**, 827–832. ISSN: 1087-0156, 1546-1696 (Sept. 2013).
69. Sanjana, N. E., Shalem, O. & Zhang, F. Improved Vectors and Genome-Wide Libraries for CRISPR Screening. en. *Nature Methods* **11**, 783–784. ISSN: 1548-7091, 1548-7105 (Aug. 2014).