RESEARCH ARTICLE

Genome-wide analysis of HOXC4 and HOXC6 regulated genes and binding sites in prostate cancer cells

Zhifei Luo, Peggy J. Farnham [ID]*

Department of Biochemistry and Molecular Medicine and the Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States of America

* peggy.farnham@med.usc.edu

Abstract

Aberrant expression of HOXC6 and HOXC4 is commonly detected in prostate cancer. The high expression of these transcription factors is associated with aggressive prostate cancer and can predict cancer recurrence after treatment. Thus, HOXC4 and HOXC6 are clinically relevant biomarkers of aggressive prostate cancer. However, the molecular mechanisms by which these HOXC genes contribute to prostate cancer is not yet understood. To begin to address the role of HOXC4 and HOXC6 in prostate cancer, we performed RNA-seq analyses before and after siRNA-mediated knockdown of HOXC4 and/or HOXC6 and also performed ChIP-seq to identify genomic binding sites for both of these transcription factors. Our studies demonstrate that HOXC4 and HOXC6 co-localize with HOXB13, FOXA1 and AR, three transcription factors previously shown to contribute to the development of prostate cancer. We suggest that the aberrantly upregulated HOXC4 and HOXC6 proteins may compete with HOXB13 for binding sites, thus altering the prostate transcriptome. This competition model may be applicable to many different human cancers that display increased expression of a HOX transcription factor.

Introduction

Prostate cancer is estimated to be the most common cancer type for new cancer cases and the second ranked cause of death by cancer for men in the USA [1]. A better understanding of the mechanisms that drive prostate cancer could lead to more effective cancer prevention, earlier diagnosis, and increased treatment options. Previous studies have shown an association of HOX family members with prostate cancer [2]. For example, HOXB13 controls the normal embryological development of the prostate gland [3, 4]. Studies have shown HOXB13-mediated repression of Androgen Receptor (AR) signaling, suggesting that HOXB13 may function as a growth suppressor in prostate tumors [5, 6]. In contrast, others have linked HOXB13 expression to androgen-dependent proliferation and migration in prostate cancer cells and it has been proposed that HOXB13 contributes to the development of prostate cancer by reprogramming AR binding sites [7–10]. HOXC family members are not expressed in normal

prostate tissue but increased expression of HOXC genes is commonly detected in prostate cancers and multiple studies have identified HOXC4 and HOXC6 as important classifiers in panels of 3–8 genes that can be used for early diagnosis of prostate cancer, identify patients with aggressive prostate cancer, and predict recurrence of prostate cancer after treatment [11–13]. Using DNA methylation data from The Cancer Genome Atlas (TCGA), we have previously identified HOXC4 and HOXC6 in the set of top-ranked transcription factors (TFs) whose high expression correlates with the creation of prostate tumor-specific enhancers [14]. These previous findings, combined with the knowledge that decreased levels of HOXC proteins leads to decreased proliferation of prostate cancer cells [15], suggest that HOXC proteins are drivers of tumorigenesis in prostate cancer. However, there is a lack of knowledge concerning the mechanisms of HOXC-mediated gene regulation. Therefore, we have developed genome-wide binding profiles for both HOXC4 and HOXC6 and identified HOXC4- and HOXC6-regulated genes in 22Rv1 prostate cancer cells.

Results

Characterizing the HOXC4- and HOXC6-regulated transcriptomes in 22Rv1 prostate cancer cells

To investigate the relationship between HOXC family members and prostate cancer, we used 22Rv1 cells, which were derived from a castration-resistant human prostate carcinoma [16]. RNA-seq analysis reveals that, similar to our previous analysis of prostate tumors [2], *HOXC4* and *HOXC6* are robustly expressed in 22Rv1 cells [17]. Therefore, we have used 22Rv1 cells as our model system to study HOXC4 and HOXC6. HOXC4 and HOXC6 are very similar proteins, having 91% similarity in their C-terminal DNA binding homeodomains. Although the N-terminal regions of the two proteins are not as similar, both do have the YPWM motif thought to be critical for protein-protein interactions [18]. One important question that we wished to address is whether these two highly related HOX genes regulate the same genes. We treated 22Rv1 cells with control siRNAs or siRNAs targeting *HOXC4* and/or *HOXC6* mRNA and performed 5 independent replicates of each of 4 different treatments (5 siControl, 5 siHOXC4, 5 siHOXC6, and 5 double knockdowns); see S1 Table. With more replicates, more differentially expressed genes were identified (S1 Fig); therefore, we used all 5 replicates for each condition for our analyses.

The genes differentially expressed after treatment with siRNAs to *HOXC4* and/or *HOXC6* are displayed as Volcano plots that separate the transcriptome into up-regulated vs down-regulated genes (see S2 Table for a list of all genes affected by knockdown of *HOXC4* and/or *HOXC6*). This analysis revealed that although *HOXC4* and *HOXC6* were both reduced by approximately the same amount, the number of genes affected upon *HOXC6* knockdown was greater than the number of genes affected by *HOXC4* knockdown (Fig 1A). Because the DNA binding domains of the two TFs are very similar, it was possible that the genes we identified were the subset of genes that were uniquely responsive to reduction of *HOXC4* or *HOXC6*, but that we had missed a set of genes for which the two TFs play redundant regulatory functions. Therefore, we also performed a double knockdown of *HOXC4* and *HOXC6* and identified a larger set of genes that were deregulated upon simultaneous knockdown of the two HOXC TFs (Fig 1B), suggesting that these TFs have at least partially redundant functions. Pathway analysis revealed that the genes down-regulated upon siRNA knockdown of *HOXC4* and/or *HOXC6* (i.e. genes whose expression is positively regulated by the TFs in the control 22Rv1 cells) are related to cancer and cell cycle (Fig 1C); these categories were not observed for the set of up-regulated genes. These findings are consistent with a previous study showing that
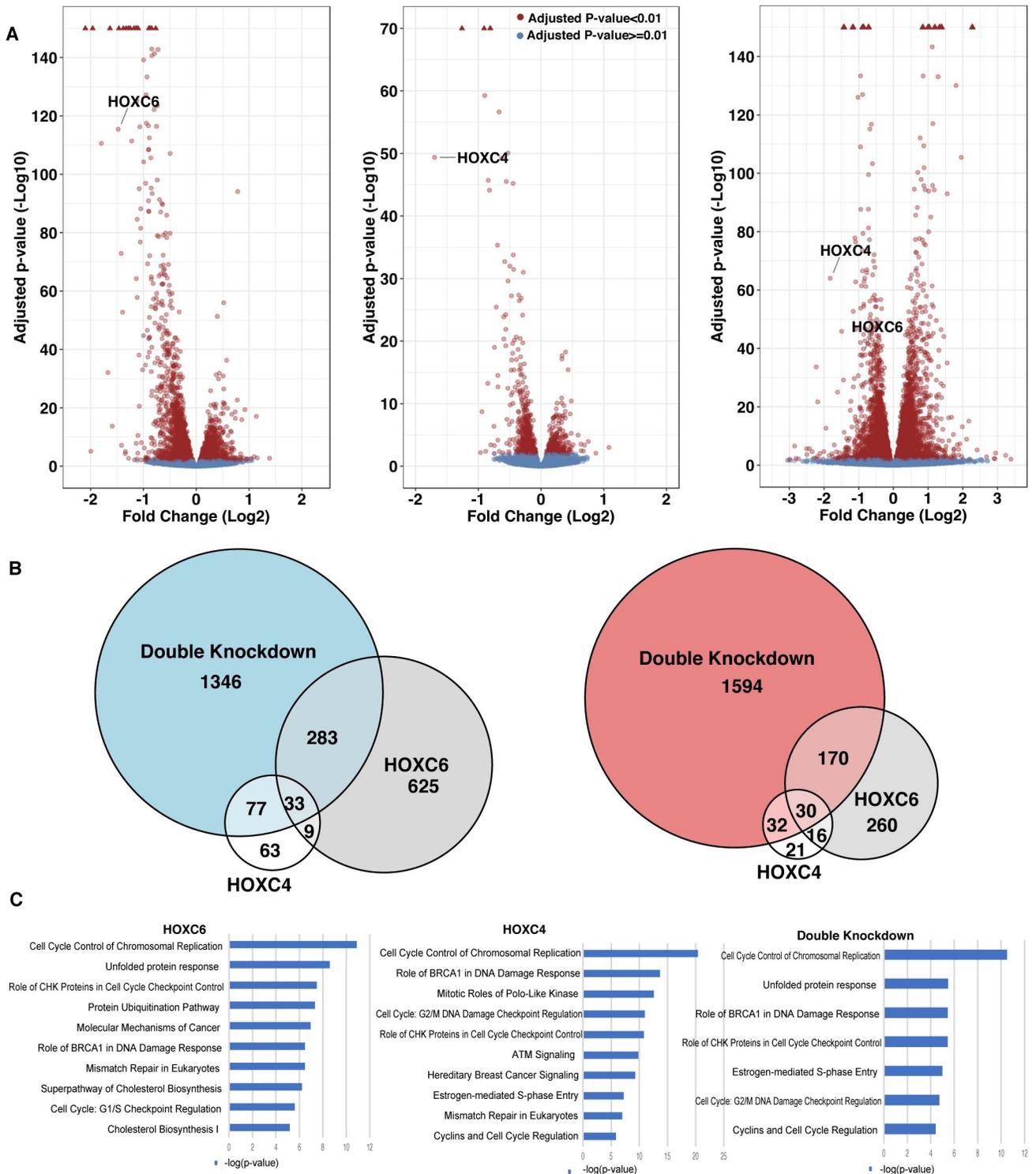
Fig 1. Identification of genes regulated by HOXC6 and/or HOXC4. (A) Volcano plots of genes up- or down-regulated after siRNA-mediated knockdown of *HOXC6*, *HOXC4*, or both. Red arrowheads; genes with extremely high adjusted p-values; the adjusted p-values for these genes are not shown to scale for ease of display; see S2 Table for exact adjusted p-values for all genes and S2 Fig for validation of several top-ranked genes by RT-qPCR and western blot. (B) Venn diagrams comparing the genes that are down-regulated (left panel) or up-regulated (right panel) upon knockdown of *HOXC6*, *HOXC4*, or both, using an adjusted p-value of <0.01 and a 1.2-fold cut-off. (C) Pathway analysis of down-regulated genes after knockdown of *HOXC6* and/ or *HOXC4*.

https://doi.org/10.1371/journal.pone.0228590.g001

decreasing the levels of HOXC proteins can decrease proliferation rates of prostate cancer cell lines [15].

Identification of HOXC4 and HOXC6 genomic binding sites in 22Rv1 prostate cancer cells

Because the RNA-seq analyses suggested that HOXC4 and HOXC6 may have redundant functions, we next wanted to determine their genome-wide binding profiles. We note that a genome-wide analysis of binding sites has not yet been published for either HOXC4 or HOXC6 in any human cell type and that there are very few published binding site datasets for any of the human HOX proteins. In fact, there are no ChIP-seq validated antibodies and no ChIP-seq datasets for any of the 39 human HOX TFs in the most recent ENCODE database (encodeproject.org). We obtained numerous HOXC4 and HOXC6 antibodies from a variety of commercial sources and performed ChIP-seq in 22Rv1 cells. However, no peaks were identified by ChIP-seq using any of the commercial antibodies, suggesting that it may be very difficult to develop ChIP-grade antibodies to the HOXC TFs (perhaps due to high sequence conservation). Therefore, we used Flag-tagged HOXC6 and HOXC4 proteins for our ChIP-seq experiments. We transfected 22Rv1 cells with a plasmid expressing tagged HOXC6 or tagged HOXC4, harvested the cells 96 hours after transfection, and then performed ChIP-seq using an antibody against the Flag epitope. We first performed two independent biological replicates for HOXC6; peaks were called using MACS2 and IDR was used to determine the number of reproducible peaks (see S3 Fig for our experimental and analytical ChIP-seq pipeline). We identified ~14,000 reproducible peaks for HOXC6 (see S3 Table); de novo motif analysis showed that ~73% of the HOXC6 peaks contain the canonical AT-rich HOX binding motif which is located centrally in the peaks (Fig 2A). We then annotated the HOXC6 binding sites relative to their location in a specific genomic feature. We found that the HOXC6 peaks are generally located in distal intergenic regions or in introns. We performed a similar annotation of the HOXC6 peaks with respect to the distance of the peak from the nearest transcription start site (TSS), finding that HOXC6 binds mainly to regulatory elements located farther than 10 kb upstream or downstream from a TSS. We next performed similar ChIP-seq experiments using a Flag-tagged HOXC4, identifying an AT-rich motif located centrally in the peaks and observing a similar distribution of peaks as shown for HOXC6 regarding distance from the TSS and gene structure (Fig 2B).

HOXC6 and HOXC4 can bind to both nucleosome-free and nucleosome-occupied regions

The distal binding patterns of HOXC4 and HOXC6 suggested that they may be enhancer-binding TFs. Enhancers are generally identified as regions of open chromatin that are marked by H3K27Ac. Therefore, to further characterize the HOXC binding sites, we determined if HOXC4 or HOXC6 peaks are in regions of open chromatin, using publicly available Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) data. Surprisingly, the results revealed that the majority of the HOXC6 and HOXC4 binding sites are not in open chromatin (Fig 3A). The unexpected finding that the majority of HOXC4 and HOXC6 binding sites are not in regions identified by ATAC-seq led us to further investigate the characteristics of the peaks. Similar to the comparison to ATAC-seq, we found that the majority of the HOXC4 and HOXC6 binding sites are not marked by H3K27Ac. However, we note that the strength of binding of HOXC6 and HOXC4 is approximately equal at peaks that are H3K27Ac + or H3K27Ac- and at peaks that are ATAC-seq+ or ATAC-seq, as noted by the similar peak heights in the tag density plots (Fig 3B), indicating that the binding of HOXC4 and HOXC6 in
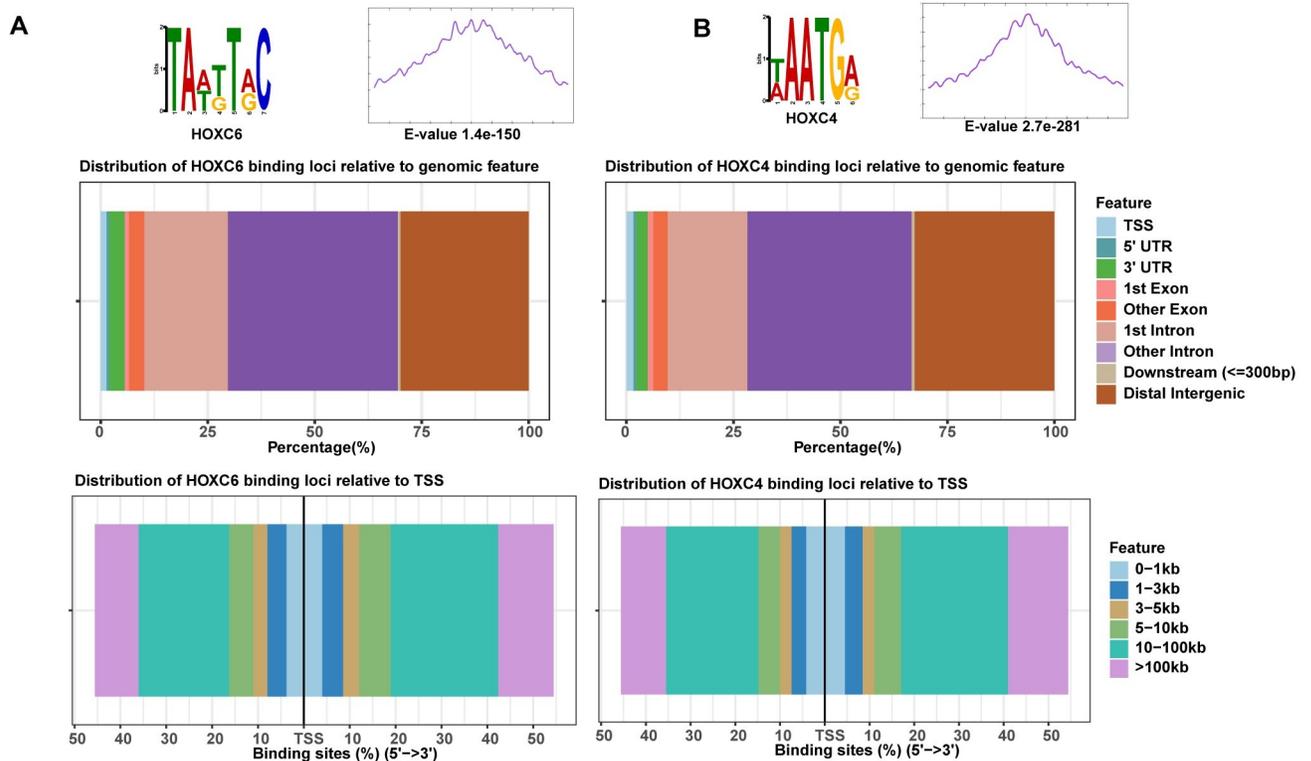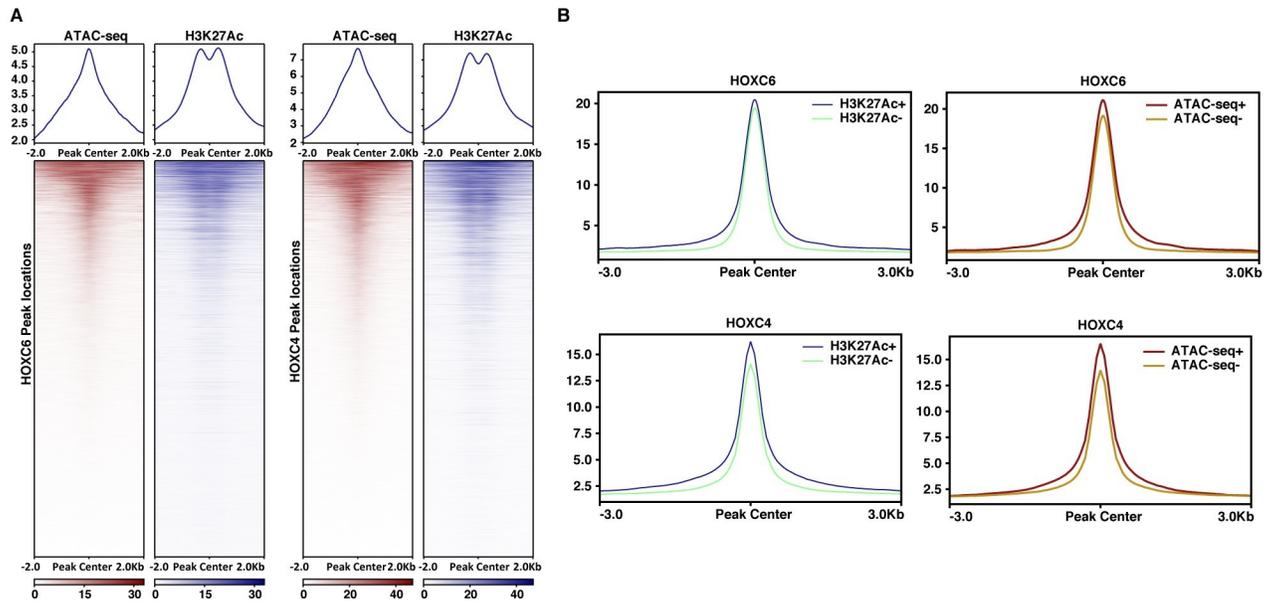
Fig 2. The majority of HOXC4 and HOXC6 binding sites are far from a TSS. (A) Top: Shown is the highest ranked de novo motif identified in the set of all HOXC6 binding sites; enrichment of the motif is shown within a -/+ 250bp window centered on the HOXC6 peaks. Middle: Shown is the location of HOXC6 peaks relative to gene structure. Bottom: shown is the location of HOXC6 peaks relative to distance from the nearest TSS. (B) Panels are the same as in (A) but for HOXC4 binding sites.

regions of nucleosome-occupied chromatin is as robust as the binding of the TFs at nucleosome-free, active enhancers.

HOXC4, HOXC6, and HOXB13 have a common set of binding sites

There are 39 human HOX genes located at 4 chromosomal loci [2]. These HOX genes are believed to have evolved by a series of duplication events and they have similar DNA binding domains. Previous studies have shown that different HOX genes from Drosophila can bind to similar genomic loci [18–20]. To test if this is also true for the human HOX genes that are upregulated in prostate cancer, we compared the binding sites for HOXC6 and HOXC4. The DNA binding domains of the HOXC4 and HOXC6 proteins are 81% identical and have a 91% similarity and, perhaps not surprisingly, we found that the majority of the HOXC4 binding sites overlap with the HOXC6 binding sites (**Fig 4A**). As described above, HOXB13 has also been associated with prostate cancer. Therefore, we next included HOXB13 in our comparisons (the HOXB13 DNA binding domain is 73.7% similar to the HOXC6 DNA binding domain). We first validated the specificity of a commercially available HOXB13 antibody (S4 Fig) and then performed HOXB13 ChIP-seq in 22Rv1 cells. We found that 37% of the HOXB13 binding sites are also bound by HOXC4 and HOXC6; there are a total of 2321 HOXB13 peaks shared with either HOXC4 and/or HOXC6 (Fig 4A). Characterization of the peaks co-bound by HOXC4, HOXC6 and HOXB13 revealed that the majority of the co-bound sites are within distal intergenic or intronic regions that are more than 10kb away from a TSS (Fig 4B).

Fig 3. The majority of HOXC6 and HOXC4 binding sites are in nucleosome-occupied regions. (A) Shown are tag density plots and heatmaps of ATAC-seq and H3K27Ac ChIP-seq data centered on the genomic locations of the HOXC6 or HOXC4 peaks. (B) Shown are tag density plots of HOXC6 and HOXC4 peaks divided into sets that are or are not marked by H3K27Ac and sites that are or are not identified as open chromatin by ATAC-seq.

https://doi.org/10.1371/journal.pone.0228590.g003
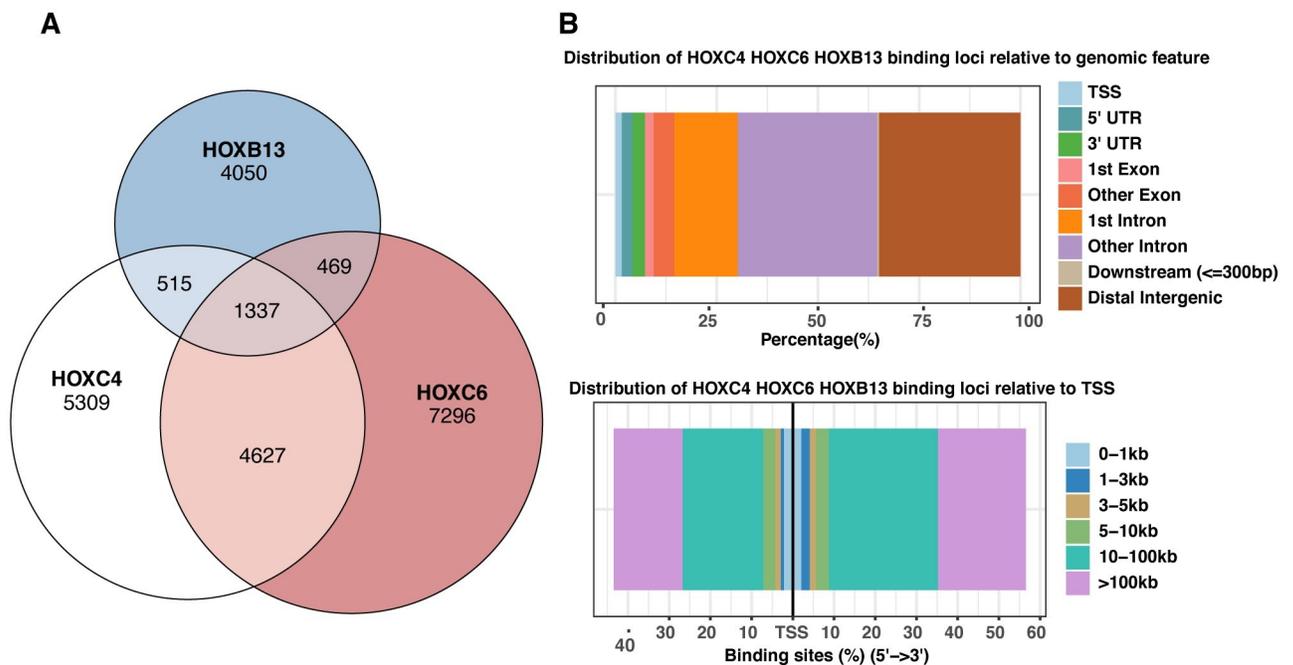
Fig 4. HOXC6, HOXC4 and HOXB13 bind to a common set of regulatory elements. (A). Shown is a 3-way Venn diagram comparing HOXC6, HOXC4 and HOXB13 binding sites; see also S5 Fig for additional measurements of peak overlap (B) TOP: Shown is the location of HOXC6, HOXC4 and HOXB13 co-bound sites relative to gene structure. Bottom: shown is the location of HOXC6, HOXC4 and HOXB13 co-bound sites relative to distance from the nearest TSS.

https://doi.org/10.1371/journal.pone.0228590.g004

Because it has been previously reported that HOXB13, FOXA1 and AR can colocalize at common binding sites, our next step was to compare HOXC4 and HOXC6 binding sites to FOXA1 and AR binding sites. We performed ChIP-seq to identify FOXA1 sites in 22Rv1 cells. To ensure high quality of our FOXA1 ChIP-seq data, we used two different FOXA1 antibodies and confirmed, using IDR, that similar sets of binding sites were detected by both antibodies. We used published ChIP-seq datasets from 22Rv1 cells for AR [21]. We found that there is a striking correlation between the strength of the HOXC4 or HOXC6 binding sites and the strength of the binding sites for the key prostate cancer transcription factors HOXB13, FOXA1, and AR (Fig 5A). However, only a subset of the binding sites is bound by all 5 TFs. Interestingly, this subset of sites is specifically the sites that are located in open chromatin marked by H3K27A (Fig 5B).

Discussion

There is much effort directed towards the identification of markers for detection of prostate cancer. A standard screening approach is to use serum prostate specific antigen (PSA) as a bio-marker for prostate cancer and, fortunately, PSA screening has reduced prostate cancer mortality rate by 20% [22–24]. However, this method falls short in accuracy and does not distinguish aggressive prostate cancer from non-life-threating prostate cancer, which can lead to overtreatment of men with localized tumors. HOXC6 and HOXC4 have been previously identified as promising, easy to assay (urinary-based) biomarkers of aggressive prostate cancer that can enhance early diagnosis and predict cancer recurrence after treatment [11–13]. The incorporation of these two genes into the screening process may lead to better clinical regimens. In addition to being good biomarkers, HOXC4 and HOXC6 may also be critical drivers of prostate neoplasia due to their ability to alter the prostate transcriptome [14]. In support of this hypothesis, studies have shown that targeted reduction in levels of HOXC6 can increase apoptosis while reducing proliferation of prostate cancer cells [15]. However, the mechanisms by which HOXC TFs influence the prostate tumor transcriptome and contribute to tumorigenesis have not been elucidated. A better understanding of the molecular function of HOXC4 and HOXC6 could provide a more comprehensive understanding of prostate cancer. Towards this goal, we identified genes that are affected by knockdown of *HOXC4* and/or *HOXC6* in 22Rv1 prostate cancer cells and performed ChIP-seq to characterize their genome-wide binding profiles.

Pathway analysis revealed that genes whose expression is positively regulated by HOXC4 and HOXC6 in prostate tumor cells are related to cancer and the cell cycle. In accordance with the hypothesis that HOX proteins are master regulators of TF pathways [25], we identified ~300 TFs that are affected in cells treated with siRNAs to *HOXC4* and *HOXC6* (S2 Table). For example, members of the E2F family, which are known regulators of the G1 to S phase transition [26–29], showed reduced expression upon knockdown of *HOXC4* and *HOXC6*. The reduction of these E2Fs supports previous studies linking HOXC6 expression to prostate cancer cell growth [15]. The oncogenic transcription regulator *YAP1* is the most significantly down-regulated gene upon the knockdown of *HOXC4*. YAP1 has been shown to promote cell proliferation and suppresses apoptosis. Thus, increased levels of HOXC4 in prostate cancer may promote tumorigenesis via upregulation of YAP1. Several of the top-ranked genes that are down-regulated upon knockdown of HOXC6, such as *EIF4G1* (which regulates a rate-limiting step in translation initiation) and *ALY/REF* (which plays a key function in mRNA export) are known to influence protein abundance. OS9, which binds to HIF-1 (a key regulator of the hypoxic response and angiogenesis) and promotes the degradation of one of the HIF-1 subunits, is also down-regulated in the *HOXC6* knockdown cells. Thus, increased levels of
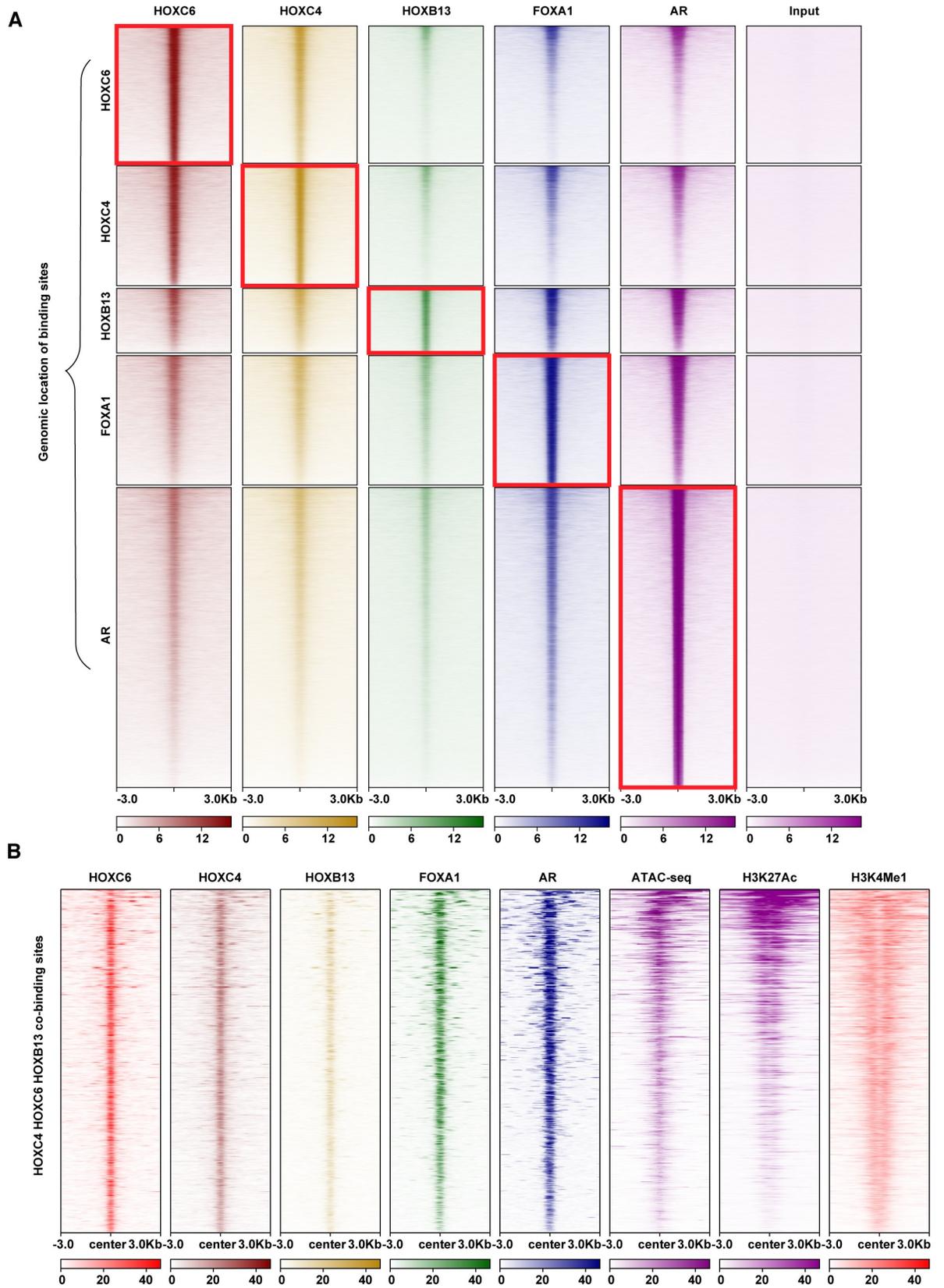
Fig 5. HOXC6, HOXC4, HOXB13, FOXA1, and AR co-bind to active enhancers. (A). Shown are HOXC4, HOXC6, HOXB13, FOXA1, and AR heatmaps centered on the genomic locations of HOXC4, HOXC6, HOXB13, FOXA1 or AR binding sites. Each row represents the genomic locations of the binding sites for the TF shown next to the row. The ChIP-seq tags for the 5 different factors (shown above each column) are then plotted at the genomic locations of each of the 5 factors. In each column, the red box outlines the binding sites for a given factor mapped at the binding locations for that same factor. (B) Shown are tag density plots and heatmaps of HOXC6, HOXC4, HOXB13, FOXA1, AR, ATAC-seq, H3K27Ac, and H3K4me1 data (indicated by the column heading) centered on the genomic location of the sites co-bound by HOXC6, HOXC4 and HOXB13.

HOXC6 in prostate cancer may promote tumorigenesis either by enhancing overall protein levels (which would enable accelerated cell proliferation) or by supporting prostate cancer metastasis to the bone environment. The genes that show altered regulation upon knockdown of HOXC4 or HOXC6 are a combination of direct and indirect target genes. Because only a very small percentage of HOXC6 or HOXC4 binding sites are within 1 kb of a TSS, it is not possible to link the individual HOXC4 or HOXC6 binding sites to regulation of specific genes. Therefore, we cannot distinguish genes directly regulated by these HOX proteins from genes whose expression is indirectly altered (due to changes in signaling pathways regulated by the direct targets).

The HOX genes are critical regulators of normal development, with different HOX genes being associated with the development of different organs. [2] Interestingly, although those studies suggest very specific functions of different HOX genes, other research has suggested that different HOX proteins can bind, with low affinity, to a similar set of binding sites. The ability to bind to the same places but produce different phenotypes is termed the HOX specificity paradox. It has been proposed that post-DNA binding actions (such as interaction of different HOX family members with distinct protein partners) are the primary determinant of normal HOX function. If so, this suggests a model for how the inappropriate expression of certain HOX genes could contribute to cancer. We hypothesize that dysregulated HOXC genes could affect tumor progression not only by creating new enhancers but also by hijacking distal regulatory elements where the normally expressed HOX genes bind in that tissue (Fig 6). In our study, we show that HOXC4 and HOXC6 colocalize with HOXB13, which is present at high levels in both normal and tumor prostate tissue. It is possible that HOXB13 (and its protein partners) regulate the normal prostate transcriptome (in the absence of HOXC4 and HOXC6, which are not expressed in the normal prostate), but when levels of HOXC4 and HOXC6 are inappropriately elevated in tumors, HOXB13 is kicked off the common binding sites causing alterations in the transcriptome. The outcome of this HOX substitution could be a combination of inappropriate activation and inappropriate repression of different genes, depending on the protein partners that are recruited by the different HOX TFs. Several co-regulators which have been shown to interact with other HOX proteins are expressed in the 22Rv1 prostate cancer cells, including PBX1-3, MEIS1 and MEIS3, and PKNOX1. However, there is no protein-protein interaction data for HOXC4 or HOXC6 in any human cell type. Future mass spectrometry studies could be useful in furthering our understanding of the mechanisms by which HOX proteins achieve specificity in the regulation of their target genes. We note that there are an increasing number of reports that associate elevated levels of HOX TFs with various cancers. In fact, every HOX gene has now been shown to have increased expression in at least one type of solid tumor [2, 30, 31]. Thus, this competition model may be relevant not only for prostate cancer but for many different cancer types.

In summary, we have shown that HOXC4 and HOXC6 regulate critical genes affecting prostate cancer proliferation. In addition, we showed colocalization of HOXC4 and HOXC6 at HOXB13 sites that are also bound by FOXA1 and AR, which have been previously linked to regulation of prostate cancer cell proliferation [9, 15, 21, 32, 33]. We suggest that the
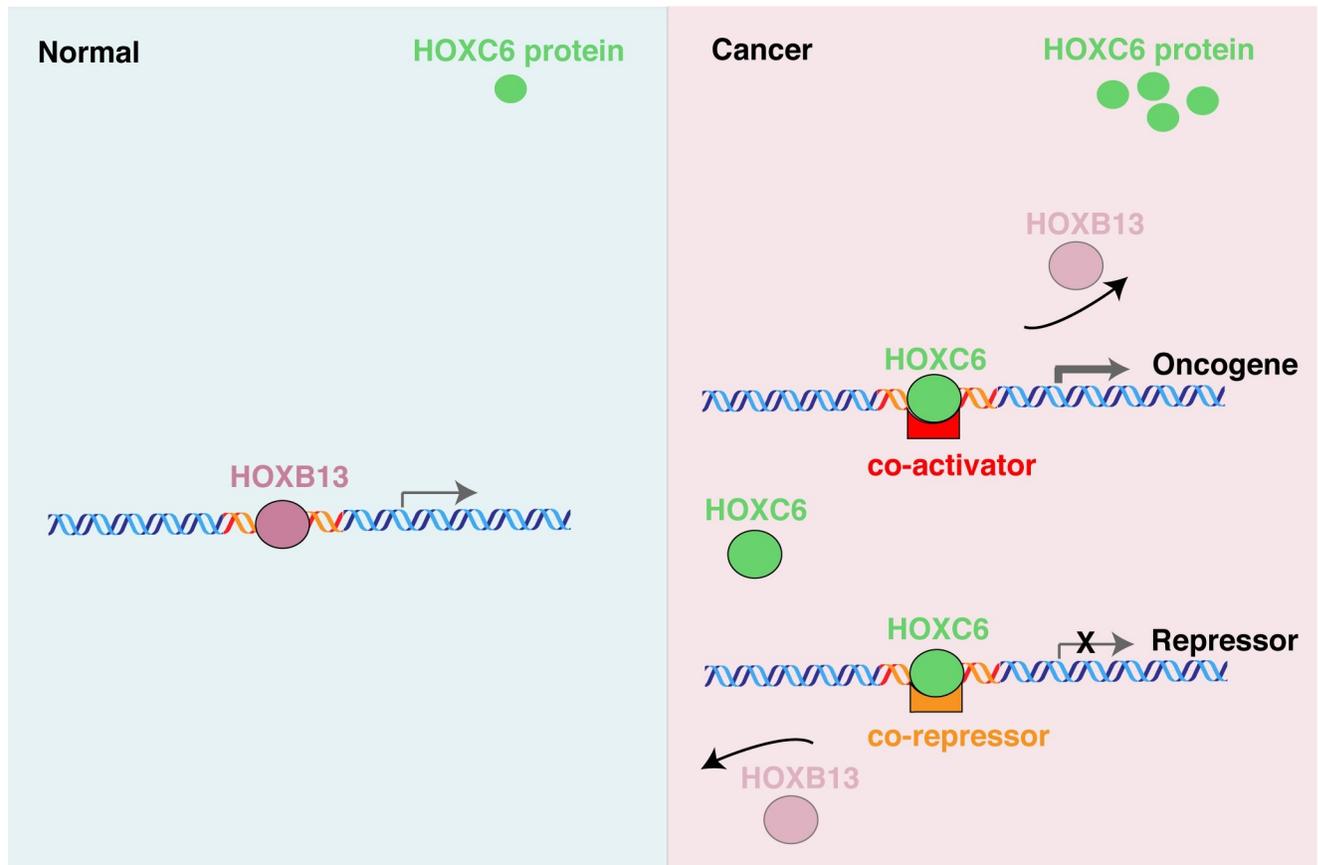
Fig 6. Competition model for HOX protein-mediated regulation of the prostate cancer transcriptome. Left panel: HOXB13 is expressed in normal prostate tissue, binds to its target sites, and regulates the prostate transcriptome. There are very low levels of HOXC6 in normal prostate cells and therefore this TF does not contribute to gene regulation. Right panel: HOXC6 is aberrantly expressed in prostate tumors and the high levels of HOXC6 protein compete with HOXB13 for occupancy of common binding sites. The bound HOXC6 could increase expression of an oncogene (if it recruits a co-activator) or reduce expression of a tumor suppressor (if it recruits a co-repressor).

https://doi.org/10.1371/journal.pone.0228590.g006

inappropriately high levels of HOXC4 and HOXC6 that occurs upon neoplastic transformation of prostate cells may contribute to tumor progression by driving oncogene expression or repressing tumor suppressor genes via binding at distal regulatory elements and altering the levels of chromatin-bound HOXB13. To accurately test this model, what is needed is a comparison of the binding of the endogenous HOXC4, HOXC6, and HOXB13 in normal and tumor prostate cells. However, there is a lack of functional antibodies for HOXC4 and HOXC6; in fact, our studies using the tagged HOXC proteins provide the first genome-wide binding profiles of HOXC4 or HOXC6 in any human cell type. Therefore, testing the proposed model will require tagging of the endogenous HOX proteins, which will be a laborious task, but will allow further insights into how HOX proteins regulate the transcriptome in normal and cancer cells.

Materials and methods

Cell culture

22Rv1 (CRL-2505) cells were purchased from American Type Culture Collection (ATCC) and cultured in RPMI 1640 supplemented with 10% fetal bovine serum (Gibco by Thermo Fisher,

#10437036) plus 1% penicillin and 1% streptomycin at 37˚C with 5% CO_2 for less than 20 passages. Cell lines were authenticated via the Short Tandem Repeat method by the Norris Comprehensive Cancer Center Media Core and tested to be mycoplasma free using a universal mycoplasma detection kit (ATCC 30-1012K).

siRNA knockdown

22Rv1 cells were transfected with DharmaFECT1 reagent (Dharmacon, #T-2001-01) plus siR-NAs against human HOXC4 (ON-TARGETplus Human HOXC4 (3221) siRNA-SMARTpool, L-013462-00-0005), human HOXC6 (ON-TARGETplus Human HOXC6 (3223) siRNA-S-MARTpool, L-011871-00-0005), or a non-targeting control (ON-TARGETplus Non-targeting Pool). The final concentration of siRNAs in the cell culture media was 50nmol. Cell were incubated for 24 hours and then transfected again using same amount of the same siRNAs and incubated for another 24 hours when they were harvested for analysis by RNA-seq.

RNA-seq

RNA integrity was checked using a 2100 Bioanalyzer (Agilent technologies, G2939AA) and an RNA 6000 nano kit (Agilent technologies, 50671511). Libraries were made using the Kapa mRNA HyperPrep Kit (KAPA, KKK8581) with 500ng total RNA. Samples were sequenced on an Illumina HiSeq3000 with single-end 50 bp read length to achieve at least 30 million reads. Quality control of the sequenced datasets was performed using FastQC and reads were trimmed using Trimmomatic [34]; reads were mapped to the human genome GRCh38 reference using Gencode V29 annotation and STAR[35]. Differentially expressed genes were determined using DESeq2 after normalizing raw read counts [36]. Transcripts per million (TPM) for each gene were calculated as following with R script:

$$TPM_i = \frac{x_i}{l_i} \times \left\{ \frac{1}{\sum_j \frac{x_j}{l_j}} \right\} \times 10^6$$

where x_i is counts for a particular gene and l_i denotes its length. If a gene had less than 1 TPM in both the deletion and control datasets, then it was excluded from further analysis. The genes that showed a significant change in expression were evaluated using the Ingenuity Pathway Analysis tool (QIAGEN) for functional annotations.

Transient transfection

15cm dishes of 22Rv1 cells were transfected using lipofectamine 3000 (Life Technologies, L3000015) with 40μg of a Flag-tagged expression vector encoding HOXC4 (GeneCopoeia, EX-A3619-M11) or HOXC6 (GeneCopoeia, EX-E1530-M14), or a corresponding control plasmid expressing eGFP (GeneCopoeia, EX-EGFP-M14); the cells were grown for 96 hours, and then harvested for ChIP-seq.

ChIP-seq

Cells expressing the Flag-tagged HOXC4 or HOXC6 proteins (described above) were cross-linked with disuccinimidyl glutarate (DSG) (Thermo Fisher Scientific, 20593) for 45min and formaldehyde (VWR, EMD-FX0410-5) for 15min, followed by the addition of glycine to stop the crosslinking. After washing, chromatin was prepared using a cell lysis buffer (Cell Signaling, 81804), nuclear extraction, and sonication. 10μg anti-Flag antibody (Sigma, F1804-200UG) was used to immunoprecipitate 200μg of chromatin. Chromatin was pulled down using protein G beads (Thermo Fisher Scientific, 10003D), eluted from the beads using elution

buffer (1% SDS, 50nM NaHCO$_3$), then reverse-crosslinked overnight. ChIP-seq libraries were prepared using the Kapa Hyper prep kit (Kapa, KK8503) according to the manufacturer's protocol. FOXA1 and HOXB13 ChIP was performed similarly as described above but without DSG crosslinking and only treated with formaldehyde for 10 minutes. 5μl anti-FOXA1 antibody (Abcam, ab5089; Cell Signaling, 58613S) or anti-HOXB13 antibody (D7N8O; Rabbit mAb, Cell Signaling, 90944) was used to pulldown chromatin. Samples were sequenced on an Illumina HiSeq3000 machine using 50bp single end reads. All ChIP-seq data were quality checked using FASTQC and MultiQC[37] and mapped to GRCh38 using BWA [38]. Low quality, duplicated, not primary aligned, and not properly paired reads were filtered out using SAMtools, Sambamb and Samblaster[39–41]. Peaks were called using MACS2[42] on each ChIP-seq replicate and common peaks were identified using IDR (for transcription factors) or the overlap method (for histones). Motifs analysis was performed using MEME tools[43]. Other ChIP-seq datasets were downloaded from the ENCODE portal (encodeproject.org) or from GEO (see S1 Table for accession numbers and S3 Fig for a schematic of the ChIP-seq pipeline).

Western blot

Whole cell lysate was prepared using RIPA buffer (20mM Tris pH 7.4, 150mM NaCl, 1% NP40, 0.25% deoxycholic acid, 0.1% SDS) with protease inhibitors (cOmplete™ Protease Inhibitor Cocktail, Sigma Aldrich) and sonicated briefly using a Bioruptor Pico (Diagenode) sonicator to reduce sample viscosity. Total cellular protein was denatured at 95˚C for 5min, separated on 7.5% or 12.5% SDS-PAGE gels (BioRAD, #4561025, #4561045), and then transferred to a nitrocellulose membrane. The membrane was blocked for 1 hour with 5% nonfat milk (BioRAD, 1706404XTU) in TBST (20mM Tris pH 7.4, 150mM NaCl, 0.1% Tween 20) then washed 3 times (5 min each time) with TBST. Primary antibody was added at 1:1000 dilution in TBST with 5% BSA (Sigma Aldrich, #A9647-50G) and incubated overnight at 4˚C. The primary antibodies used were raised against Beta Actin (proteintech, 60008-1-Ig), OS-9 (D8P4G; Rabbit mAb, Cell Signaling, 12497), YAP (D8H1X; XP® Rabbit mAb, Cell Signaling, 14074), THOC4/ ALY (D3R4R; Rabbit mAb, Cell Signaling, 12655), eIF4GI (D6A6; Rabbit mAb, Cell Signaling, 8701) and HOXB13 (D7N8O; Rabbit mAb, Cell Signaling, 90944). The next day, secondary antibodies (Thermo Scientific, Rabbit anti-Goat IgG (H+L) Cross Adsorbed Secondary Antibody, DyLight 800 conjugate, and Goat Anti-Mouse IgG (H+L), DyLight 680 Conjugated) were incubated for 1 hour at room temperature and washed as described above. Fluorophore signal was detected using the Li-COR Odyssey system and quantified using ImageJ.

RT-qPCR

RNA was extracted using the DirectZol RNA kit (Zymo, R2062). The iScript Reverse Transcription SuperMix (Bio-RAD, 1708841BUN) was used to synthesize cDNA from 500ng RNA. The PerfecTa SYBR Green SuperMix (Quantabio, 95054-02K) was used in quantitative PCR reactions. Primers used in RT-qPCR were designed using primer-blast and listed in S4 Table.

Supporting information

S1 Fig. Increasing the number of RNA-seq replicates can identify a larger number of differentially expressed genes. (A) Shown is the number of differentially expressed genes (adjusted p-value<0.01) identified by RNA-seq using 2 to 5 replicates of control and HOXC6 siRNA-treated 22 Rv1 cells. (B) Shown is a four-way Venn diagram displaying the overlap of differentially expressed genes (adjusted p-value <0.01) identified by RNA-seq using 2 to 5 replicates of

control and HOXC6 siRNA-treated cells.
(PDF)

S2 Fig. Confirmation of a subset of HOXC4 and HOXC6 target genes. (A) RT-PCR analysis of HOXC4- and HOXC6-regulated genes. A second siRNA knockdown of HOXC4 or HOXC6 was performed using 22Rv1 cells. In each panel, the genes identified to be regulated by the specific HOXC protein are indicated; the genes identified as targets for the other HOXC protein are used as controls. (B) Western blot confirmation of HOXC4- and HOXC6-regulated genes; OS9, EIF4G1, and ALY/REF were identified by RNA-seq as HOXC6-regulated genes, whereas YAP1 was identified by RNA-seq as a HOXC4-regulated gene.
(PDF)

S3 Fig. ChIP-seq experimental and analytical flowchart. Shown are the steps used to perform and analyze the HOXC6 ChIP-seq experiments; see Methods for details.
(PDF)

S4 Fig. Validation of the specificity of the HOXB13 antibody. Shown is a Western blot demonstrating the specificity of the HOXB13 antibody; siRNA-mediated knockdown of HOXB13 mRNA eliminates the signal detected by the HOXB13 antibody.
(PDF)

S5 Fig. Quantitative measures of co-binding of transcription factors. Shown are 3 tests that measure the overlap between the binding sites of HOXC6, HOXC4, HOXB13, FOXA1 and AR. The yellow number is the P-value for a two tail fisher exact test obtained using the bedtools fisher function, the red number is the Jaccard value generated using the bedtools jaccard function, the blue value is the number of overlapped peaks called using the MACS2 peak caller.
(PDF)

S1 Table. Genomic datasets.
(XLSX)

S2 Table. HOXC6- and HOXC4-regulated genes.
(XLSX)

S3 Table. HOXC6- and HOXC4 ChIP-seq Peaks.
(XLSX)

S4 Table. Primers used in RT-qPCR and qPCR.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Zhifei Luo, Peggy J. Farnham.

Formal analysis: Zhifei Luo.

Funding acquisition: Peggy J. Farnham.

Investigation: Zhifei Luo.

Methodology: Zhifei Luo.

Supervision: Peggy J. Farnham.

Visualization: Zhifei Luo.

Writing – original draft: Peggy J. Farnham.

Writing – review & editing: Zhifei Luo, Peggy J. Farnham.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019; 69(1):7–34. Epub 2019/01/09. https://doi.org/10.3322/caac.21551 PMID: 30620402.

2. Luo Z, Rhie SK, Farnham PJ. The enigmatic HOX genes: can we crack their code? Cancers. 2019; 11 (3):323. https://doi.org/doi.org/10.3390/cancers11030323

3. Huang L, Pu Y, Hepps D, Danielpour D, Prins GS. Posterior Hox gene expression and differential androgen regulation in the developing and adult rat prostate lobes. Endocrinology. 2007; 148(3):1235–45. Epub 2006/12/02. https://doi.org/10.1210/en.2006-1250 PMID: 17138648; PubMed Central PMCID: PMC2276874.

4. Economides KD, Capecchi MR. Hoxb13 is required for normal differentiation and secretory function of the ventral prostate. Development. 2003; 130(10):2061–9. Epub 2003/04/02. https://doi.org/10.1242/dev.00432 PMID: 12668621.

5. Jung C, Kim RS, Zhang HJ, Lee SJ, Jeng MH. HOXB13 induces growth suppression of prostate cancer cells as a repressor of hormone-activated androgen receptor signaling. Cancer Res. 2004; 64 (24):9185–92. Epub 2004/12/18. https://doi.org/10.1158/0008-5472.CAN-04-1330 PMID: 15604291.

6. Kim SD, Park RY, Kim YR, Kim IJ, Kang TW, Nam KI, et al. HOXB13 is co-localized with androgen receptor to suppress androgen-stimulated prostate-specific antigen expression. Anatomy & cell biology. 2010; 43(4):284–93. Epub 2011/01/27. https://doi.org/10.5115/acb.2010.43.4.284 PMID: 21267402; PubMed Central PMCID: PMC3026180.

7. Norris JD, Chang CY, Wittmann BM, Kunder RS, Cui H, Fan D, et al. The homeodomain protein HOXB13 regulates the cellular response to androgens. Mol Cell. 2009; 36(3):405–16. Epub 2009/11/18. https://doi.org/10.1016/j.molcel.2009.10.020 PMID: 19917249; PubMed Central PMCID: PMC2788777.

8. Stelloo S, Nevedomskaya E, Kim Y, Hoekman L, Bleijerveld OB, Mirza T, et al. Endogenous androgen receptor proteomic profiling reveals genomic subcomplex involved in prostate tumorigenesis. Oncogene. 2018; 37(3):313–22. Epub 2017/09/20. https://doi.org/10.1038/onc.2017.330 PMID: 28925401.

9. Pomerantz MM, Li F, Takeda DY, Lenci R, Chonkar A, Chabot M, et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. Nat Genet. 2015; 47(11):1346–51. Epub 2015/10/13. https://doi.org/10.1038/ng.3419 PMID: 26457646; PubMed Central PMCID: PMC4707683.

10. Chen J, Zhu S, Jiang N, Shang Z, Quan C, Niu Y. HoxB3 promotes prostate cancer cell progression by transactivating CDCA3. Cancer Lett. 2013; 330(2):217–24. Epub 2012/12/12. https://doi.org/10.1016/j.canlet.2012.11.051 PMID: 23219899.

11. Leyten GH, Hessels D, Smit FP, Jannink SA, de Jong H, Melchers WJ, et al. Identification of a Candidate Gene Panel for the Early Diagnosis of Prostate Cancer. Clin Cancer Res. 2015; 21(13):3061–70. https://doi.org/10.1158/1078-0432.CCR-14-3334 PMID: 25788493.

12. Bibikova M, Chudin E, Arsanjani A, Zhou L, Garcia EW, Modder J, et al. Expression signatures that correlated with Gleason score and relapse in prostate cancer. Genomics. 2007; 89(6):666–72. Epub 2007/04/27. https://doi.org/10.1016/j.ygeno.2007.02.005 PMID: 17459658.

13. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, et al. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell. 2002; 1(2):203–9. Epub 2002/06/28. https://doi.org/10.1016/s1535-6108(02)00030-2 PMID: 12086878.

14. Rhie SK, Guo Y, Tak YG, Yao L, Shen H, Coetzee GA, et al. Identification of activated enhancers and linked transcription factors in breast, prostate, and kidney tumors by tracing enhancer networks using epigenetic traits. Epigenetics Chromatin. 2016; 9:50. https://doi.org/10.1186/s13072-016-0102-4 PMID: 27833659; PubMed Central PMCID: PMC5103450.

15. Ramachandran S, Liu P, Young AN, Yin-Goen Q, Lim SD, Laycock N, et al. Loss of HOXC6 expression induces apoptosis in prostate cancer cells. Oncogene. 2005; 24(1):188–98. Epub 2005/01/08. https://doi.org/10.1038/sj.onc.1207906 PMID: 15637592.

16. Sramkoski RM, Pretlow TG 2nd, Giaconia JM, Pretlow TP, Schwartz S, Sy MS, et al. A new human prostate carcinoma cell line, 22Rv1. In Vitro Cell Dev Biol Anim. 1999; 35(7):403–9. https://doi.org/10.1007/s11626-999-0115-4 PMID: 10462204.

17. Guo Y, Perez AA, Hazelett DJ, Coetzee GA, Rhie SK, Farnham PJ. CRISPR-mediated deletion of prostate cancer risk-associated CTCF loop anchors identifies repressive chromatin loops. Genome Biol. 2018; 19(1):160. Epub 2018/10/10. https://doi.org/10.1186/s13059-018-1531-0 PMID: 30296942; PubMed Central PMCID: PMC6176514.

18. Mann RS, Lelli KM, Joshi R. Hox specificity unique roles for cofactors and collaborators. Current topics in developmental biology. 2009; 88:63–101. Epub 2009/08/05. https://doi.org/10.1016/S0070-2153(09)88003-4 PMID: 19651302; PubMed Central PMCID: PMC2810641.

19. Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell. 2008; 133 (7):1277–89. Epub 2008/07/01. https://doi.org/10.1016/j.cell.2008.05.023 PMID: 18585360; PubMed Central PMCID: PMC2478728.

20. Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Pena-Castillo L, et al. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell. 2008; 133 (7):1266–76. Epub 2008/07/01. https://doi.org/10.1016/j.cell.2008.05.024 PMID: 18585359; PubMed Central PMCID: PMC2531161.

21. Chen Z, Wu D, Thomas-Ahner JM, Lu C, Zhao P, Zhang Q, et al. Diverse AR-V7 cistromes in castration-resistant prostate cancer are governed by HoxB13. Proc Natl Acad Sci U S A. 2018; 115(26):6810–5. Epub 2018/05/31. https://doi.org/10.1073/pnas.1718811115 PMID: 29844167; PubMed Central PMCID: PMC6042123.

22. Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009; 360(13):1320–8. Epub 2009/03/20. https://doi.org/10.1056/NEJMoa0810084 PMID: 19297566.

23. Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med. 2012; 366(11):981–90. Epub 2012/03/16. https://doi.org/10.1056/NEJMoa1113135 PMID: 22417251; PubMed Central PMCID: PMC6027585.

24. Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet. 2014; 384(9959):2027–35. Epub 2014/08/12. https://doi.org/10.1016/S0140-6736(14)60525-0 PMID: 25108889; PubMed Central PMCID: PMC4427906.

25. Carnesecchi J, Pinto PB, Lohmann I. Hox transcription factors: an overview of multi-step regulators of gene expression. Int J Dev Biol. 2018; 62(11–12):723–32. Epub 2019/01/04. https://doi.org/10.1387/ijdb.180294il PMID: 30604842.

26. Farnham PJ, Slansky JE, Kollmar R. The role of E2F in the mammalian cell cycle. Biochim Biophys Acta. 1993; 1155:125–31. https://doi.org/10.1016/0304-419x(93)90001-s PMID: 8357823

27. Slansky JE, Farnham PJ. Introduction to the E2F family: protein structure and gene regulation. In: Farnham PJ, editor. Transcriptional control of cell growth: the E2F gene family. 208. New York, New York: Springer-Verlag; 1996. p. 1–30.

28. Adams PD, Kaelin WG. Transcriptional Control of Cell Growth: The E2F Gene Family. Farnham PJ, editor. New York: Springer-Verlag; 1995. 79–93 p.

29. Nevins JR. The Rb/E2F pathway and cancer. Hum Mol Genet. 2001; 10:699–703. https://doi.org/10.1093/hmg/10.7.699 PMID: 11257102

30. Sun Y, Zhou B, Mao F, Xu J, Miao H, Zou Z, et al. HOXA9 Reprograms the Enhancer Landscape to Promote Leukemogenesis. Cancer Cell. 2018; 34(4):643–58 e5. Epub 2018/10/03. https://doi.org/10.1016/j.ccell.2018.08.018 PMID: 30270123; PubMed Central PMCID: PMC6179449.

31. Bhatlekar S, Fields JZ, Boman BM. HOX genes and their role in the development of human cancers. J Mol Med (Berl). 2014; 92(8):811–23. Epub 2014/07/06. https://doi.org/10.1007/s00109-014-1181-y PMID: 24996520.

32. Shukla S, Cyrta J, Murphy DA, Walczak EG, Ran L, Agrawal P, et al. Aberrant Activation of a Gastrointestinal Transcriptional Circuit in Prostate Cancer Mediates Castration Resistance. Cancer Cell. 2017; 32(6):792–806.e7. Epub 2017/11/21. https://doi.org/10.1016/j.ccell.2017.10.008 PMID: 29153843; PubMed Central PMCID: PMC5728174.

33. Kron KJ, Murison A, Zhou S, Huang V, Yamaguchi TN, Shiah YJ, et al. TMPRSS2-ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. Nat Genet. 2017; 49(9):1336–45. Epub 2017/08/08. https://doi.org/10.1038/ng.3930 PMID: 28783165.

34. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404; PubMed Central PMCID: PMC4103590.

35. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886; PubMed Central PMCID: PMC3530905.

36. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30(7):923–30. https://doi.org/10.1093/bioinformatics/btt656 PMID: 24227677.

37. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016; 32(19):3047–8. Epub 2016/06/18. https://doi.org/10.1093/bioinformatics/btw354 PMID: 27312411; PubMed Central PMCID: PMC5039924.

38. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. Epub 2009/05/20. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168; PubMed Central PMCID: PMC2705234.

39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943; PubMed Central PMCID: PMC2723002.

40. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics. 2015; 31(12):2032–4. Epub 2015/02/24. https://doi.org/10.1093/bioinformatics/btv098 PMID: 25697820; PubMed Central PMCID: PMC4765878.

41. Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics. 2014; 30(17):2503–5. Epub 2014/05/09. https://doi.org/10.1093/bioinformatics/btu314 PMID: 24812344; PubMed Central PMCID: PMC4147885.

42. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biology. 2008; 9:R137. https://doi.org/10.1186/gb-2008-9-9-r137 PMID: 18798982

43. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009; 37(Web Server issue):W202–8. Epub 2009/05/22. https://doi.org/10.1093/nar/gkp335 PMID: 19458158; PubMed Central PMCID: PMC2703892.