The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis

Mark M Pomerantz1,6, Fugen Li2,3,6, David Y Takeda1,4,6, Romina Lenci1, Apurva Chonkar1, Matthew Chabot1, Paloma Cejas1,2, Francisca Vazquez4, Jennifer Cook1, Ramesh A Shivdasani1,2, Michaela Bowden5, Rosina Lis5, William C Hahn1,4, Philip W Kantoff1, Myles Brown1,2, Massimo Loda1,4,5, Henry W Long1,2 & Matthew L Freedman1,2,4

Master transcription factors interact with DNA to establish cell type identity and to regulate gene expression in mammalian cells1,2. The genome-wide map of these transcription factor binding sites has been termed the cistrome3. Here we show that the androgen receptor (AR) cistrome undergoes extensive reprogramming during prostate epithelial transformation in man. Using human prostate tissue, we observed a core set of AR binding sites that are consistently reprogrammed in tumors. FOXA1 and HOXB13 colocalized at the reprogrammed AR binding sites in human tumor tissue. Introduction of FOXA1 and HOXB13 into an immortalized prostate cell line reprogrammed the AR cistrome to resemble that of a prostate tumor, functionally linking these specific factors to AR cistrome reprogramming. These findings offer mechanistic insights into a key set of events that drive normal prostate epithelial transformation and establish the centrality of epigenetic reprogramming in human prostate tumorigenesis.

Activation of the AR is essential for cell growth and cancer progression in the prostate. Disruption of androgen metabolism influences prostate cancer incidence, and depletion of AR ligand has been the foundation of prostate cancer treatment for decades4,5. Yet, the role of the AR in transformation is unclear. There are, for instance, no recurrent AR genetic alterations in primary tumors6,7. Although many cofactors influence AR signaling in model systems8–11, it is unknown which factors are relevant for human prostate tumorigenesis. Several lines limit a comprehensive understanding of AR and cofactor binding during transformation. In contrast to luminal epithelial cells and prostate tumors, most cell line models of normal prostate epithelium do not express AR12. Moreover, all currently available prostate cancer cell lines are derived from metastatic disease and thus may not adequately model localized disease. Performing AR chromatin immunoprecipitation and high-throughput sequencing (ChIP-seq) directly in primary human tissue overcomes these impediments. To this end, we performed an AR cistrome-wide association study (CWAS) in a cohort of normal and tumor human prostate tissue samples.

We conducted the CWAS using chromatin extracted from 13 independent prostate cancers and seven histologically normal samples from areas of fresh-frozen radical prostatectomy specimens having at least 70% epithelial enrichment (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1). Six cases had matched pairs of tumor and normal tissue. Sequencing reads were aligned to the human genome (hg19), and AR binding sites were called using a standard algorithm13,14. A total of 76,553 unique AR binding sites were identified across the 20 samples at a false discovery rate (FDR) <0.01 (Fig. 1b). On the basis of the rate at which new peaks approached saturation, we estimated that our sampling captured the majority of common AR binding sites (Supplementary Figs. 3a and 4a). Although normal tissues showed fewer AR binding sites overall (median = 9,049), the distribution of personal and shared sites was similar to that in tumors (Supplementary Figs. 3b and 4b). To formally compare the AR peaks called in our cohort with those identified in a previous AR ChIP study15, we subjected the raw sequence data from that study to the exact analysis pipeline used here and found that 11 of the 12 samples from the previous study yielded fewer than 1,000 AR binding sites (Supplementary Table 2).

An unsupervised analysis of the AR cistromes clustered specimens distinctly into tumor and normal groups (Fig. 1c). These data show that AR binding is extensively and consistently reprogrammed during prostate tumorigenesis. AR ChIP-seq profiles from two prostate cancer cell lines, LNCaP and VCaP, clustered more closely with the primary tumor specimens, although they formed a distinct subset16–18 (Fig. 1c). The AR ChIP-seq profile in LHSAR, an immortalized prostate epithelial cell line with AR exogenously introduced19, clustered closest to normal human prostate samples (Fig. 1c).

1Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA. 2Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. 3Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. 4The Eli and Edythe L. Broad Institute, Cambridge, Massachusetts, USA. 5Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. 6These authors contributed equally to this work. Correspondence should be addressed to M.L.F. (freedman@broadinstitute.org).

Received 17 May; accepted 17 September; published online 12 October 2015; doi:10.1038/ng.3419
To identify AR binding sites that distinguished normal from cancerous prostate tissue, we selected sites with significantly elevated binding intensities across multiple tumor specimens relative to normal tissue and vice versa (t test, P ≤ 0.001; Online Methods and Supplementary Figs. 1 and 5). A total of 9,179 sites had higher binding intensity in tumors (tumor AR binding sites, T-ARBSs), and 2,690 sites had higher binding intensity in normal samples (normal AR binding sites, N-ARBSs) (Fig. 2a). Differential sites demonstrated a fourfold average difference in binding intensity (Fig. 2b). Analysis of these 11,869 tissue-specific sites in prostate cell lines showed strong concordance with the observations in primary human tissue.

In LNCaP cells, AR binding sites coincided with T-ARBSs, whereas AR binding was largely absent at N-ARBSs (Supplementary Fig. 6). In LHSAR cells, by contrast, AR binding coincided with N-ARBSs and was notably diminished at T-ARBSs (Supplementary Fig. 6).

In gene set enrichment analysis (GSEA) of the transcripts nearest AR binding sites in normal tissue as well as the tissue-specific sites, we hypothesized that genes differentially expressed by normal prostate and prostate tumors would be enriched for T-ARBSs and N-ARBSs. Gene expression data (RNA sequencing, RNA-seq) were obtained from The Cancer Genome Atlas (TCGA) data set of 220 samples (44 normal and 176 tumor samples), and genes were rank-ordered by differential expression. A total of 3,743 genes met the criteria for differential expression (Online Methods). When these genes were rank-ordered by their differential expression values, those highly expressed in tumors relative to normal epithelium were significantly enriched for T-ARBSs (within a 50-kb window centered on the gene transcription start site), whereas genes downregulated in tumors were substantially enriched for N-ARBSs (Fig. 2c, Supplementary Fig. 7 and Supplementary Table 5).

Next, to identify transcription factors colocalizing with AR, we performed transcription factor binding motif analysis for AR binding sites in normal tissue, tumor tissue and the LNCaP cell line. The AR binding sites present in normal tissue as well as the N-ARBS subset, the most significantly enriched motif (z score 5−15) was the canonical AR binding motif (Fig. 3a). At the AR binding sites identified in tumor tissue, two motifs were enriched: the motifs for AR and the forkhead transcription factor FOXA1 (Fig. 3a). At the T-ARBS subset, however, a third motif—for HOXB13—emerged (Fig. 3a and
Supplementary Tables 7 and 8). In LNCaP cells, motif analysis identified only AR and FOXA1 motifs. However, when the LNCaP AR binding sites were intersected with the T-ARBSs, the HOXB13 motif clearly achieved significance (Fig. 3b). These findings emphasize the importance of interrogating (i) normal as well as tumor human tissues (to define tumor-specific sites) and (ii) the cistromes of multiple individuals (to determine consensus sites).

To validate the presence of these transcription factors at T-ARBSs, we performed ChIP-seq for HOXB13 and FOXA1 in both human prostate tissue and prostate cell lines. HOXB13 and FOXA1 binding sites in human tumors overlapped extensively with T-ARBSs but not with N-ARBSs (Fig. 3c and Supplementary Fig. 9). HOXB13 and FOXA1 binding sites in the LNCaP and VCaP cell lines showed a similar pattern at these sites (Supplementary Figs. 9 and 10). Multiplexed immunohistochemical analysis of FOXA1, HOXB13 and AR in 49 radical prostatectomy specimens confirmed that all three proteins were consistently coexpressed in human prostate cancer cells (Fig. 3d), with FOXA1 and HOXB13 coexpressed in 97% of nuclei (n > 73,000 nuclei examined). To quantify protein levels, we performed immunoblots for AR, HOXB13 and FOXA1 and found that expression was stronger in tumor tissue than in paired normal specimens (Supplementary Fig. 11).

To investigate whether FOXA1 and HOXB13 were sufficient to reprogram the AR cistrome, we performed AR ChIP-seq in LHSAR cells jointly transduced with viruses encoding HOXB13 and FOXA1 (Supplementary Fig. 12). A 48-h time point for analysis was selected because expression of FOXA1 and HOXB13 proteins could not be stably maintained for longer intervals (Supplementary Fig. 13). In comparison to the AR cistrome in control cells transduced to express LacZ, the AR cistrome of LHSAR cells jointly expressing these transcription factors was strikingly reprogrammed, with a profound shift toward T-ARBSs and away from N-ARBSs (Fig. 4a and Supplementary Figs. 14–16). To evaluate whether HOXB13 or FOXA1 alone could reprogram the LHSAR AR cistrome to a similar extent, we independently transduced cells to express each one of these transcription factors. Neither transcription factor alone could recapitulate the reprogramming of T-ARBSs and N-ARBSs to the same degree as both transcription factors together (Fig. 4a,b and Supplementary Fig. 17). In a supervised analysis, LHSAR cells clustered with normal prostate tissue, whereas the addition of FOXA1 and HOXB13 reclassified the LHSAR cell line with prostate tumors (Fig. 4b). The introduction of FOXA1 or HOXB13 alone, in both instances, resulted in the LHSAR cell line being positioned outside of the normal and tumor clusters (Fig. 4b and Supplementary Fig. 17). Modulation of AR levels alone in LHSAR cells, in the absence of HOXB13 or FOXA1 addition, did not alter the AR binding pattern at tissue-specific sites (Supplementary Fig. 18). Taken together, these results show that increased AR levels are unable to explain the profound change in the cistrome observed in LHSAR cells and prostate tumors and support the hypothesis that expression of HOXB13 and FOXA1 in immortalized LHSAR cells reprograms the AR cistrome to resemble the pattern observed in prostate tumors.
Next, we performed transcriptome sequencing (RNA-seq) under these same four conditions in LHSAR cells, for cells transduced to express LacZ control, FOXA1 alone, HOXB13 alone, or both FOXA1 and HOXB13. The three LHSAR cell lines transduced to express transcription factors were compared to the control LHSAR-LacZ line to generate six lists of differentially expressed genes (three lists of upregulated genes and three lists of downregulated genes). We performed GSEA for each one of these six lists, comparing the genes to the predefined sets of T-ARBS and N-ARBS genes described above (n = 324 and 212, respectively). Strikingly, for both up- and downregulated transcripts, we observed greater overlap with T-ARBS and N-ARBS genes in the cells jointly expressing FOXA1 and HOXB13 in comparison to the cells transduced to express each transcription factor individually (Supplementary Fig. 19). Comparing the six gene lists to publicly available gene sets, GSEA identified more highly ranked prostate cancer terms in LHSAR cells coexpressing FOXA1 and HOXB13 than under the other conditions (Supplementary Tables 9–14).

We next addressed whether decreasing FOXA1 and HOXB13 levels in the LNCaP prostate cancer cell line would influence tissue-specific AR binding sites. Previous data demonstrated that knockdown of FOXA1 in LNCaP cells results in AR reprogramming, in which AR binding sites are lost, maintained or gained. We calculated the proportion of T-ARBSs across these three categories (Supplementary Fig. 20). T-ARBSs were disproportionately enriched at FOXA1-dependent AR binding sites, consistent with a pioneering effect of FOXA1 on T-ARBSs. Depletion of HOXB13 in the LNCaP cell line, however, substantially decreased AR protein levels (Supplementary Fig. 21). Therefore, any interrogation of the AR cistrome upon HOXB13 knockdown cannot unambiguously be attributed to decreased HOXB13 levels.

We assessed cell proliferation upon depletion of these transcription factors in LNCaP cells. LNCaP cells were dependent on both FOXA1 and HOXB13 (Fig. 4c). This dependence was confirmed in a large cell line database, Project Achilles (Fig. 4d). In comparison to 102 cell lines representing multiple tissue types, LNCaP cells scored second highest for HOXB13 dependency and were ranked fifth highest for FOXA1 dependency.
In summary, the marked redistribution of AR binding sites during tumorigenesis represents one of the most recurrent epigenetic or genetic alterations yet discovered in prostate cancer. Our analysis showed that two factors—FOXA1, a general pioneer factor \cite{10,23,24}, and HOXB13, a highly lineage-specific factor \cite{25}—colocalize at most T-ARBSs. Furthermore, these same two factors were sufficient to reprogram the AR cistrome. Consistent with these observations, a recently performed proteomic analysis in LNCaP cells showed that AR, HOXB13 and FOXA1 were present in the same complex \cite{10}.

In addition, a rare protein-coding polymorphism was shown to significantly elevate prostate cancer risk \cite{26,27}.

These findings offer the first mechanistic insights into a key set of events that drive normal prostate epithelium toward a clinically relevant neoplastic phenotype. More broadly, this work lays a foundation of events that drive normal prostate epithelium toward a clinically relevant neoplastic phenotype. More broadly, this work lays a foundation for characterizing the cistrome in primary human tissue samples, relevant neoplastic phenotype. More broadly, this work lays a foundation for characterizing the cistrome in primary human tissue samples, in addition to cell lines and other models, to advance insights into cellular transformation.

URLs. The Cancer Genome Atlas (TCGA) data, https://tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp; Broad Institute Project Achilles, http://www.broadinstitute.org/achilles; R package, http://www.r-project.org/; consensus signal artifact (Velcro) regions, http://hgdownload.cse.ucsc.edu/hg19/encodeDCC/wgEncodeMapability/wgEncodeDacMapabilityConsensusExcludable.bed.gz; Genomic Regions Enrichment of Annotations Tool (GREAT), http://bejerano.stanford.edu/great/public/html/.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. All ChIP-seq and RNA-seq data sets have been deposited in the Gene Expression Omnibus (GEO) under accession GSE70079.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
We thank Z. Herbert and the Molecular Biology Core Facility at the Dana-Farber Cancer Institute for their sequencing expertise. We acknowledge funding from the Prostate Cancer Foundation (Challenge Award), the H.L. Snyder Medical Foundation (M.L.F.), US National Institutes of Health (NIH) grant R01CA199310, the Anna Fuller Fund (PS#6241901; Inseed 2013-0659), the NIH GAME-ON U19 (grant U19CA148537), the Dana-Farber Prostate Cancer Specialized Program of Research Excellence (SPORE; NIH P50CA093931-11) and the NIH grant U01CA176658 (F.W. and W.C.H.).

AUTHOR CONTRIBUTIONS
M.M.P. designed the study, performed ChIP-seq assays, analyzed data and wrote the manuscript. P.L. analyzed ChIP-seq data and performed biostatistical analysis. Y.T. performed shRNA and gene transduction experiments. A.C. and M.C. performed ChIP-seq experiments. R. Lenci performed cell line ChIP experiments. P.C. and J.C. assisted with ChIP-seq assays. R.A.S. participated in analyzing the manuscript. F.L. analyzed ChIP-seq data and performed biostatistical analysis.

Figure 4 HoxB13 and FoxA1 are sufficient for reprogramming of the AR cistrome in LHSAR cells and are essential for prostate cancer cell survival. (a) AR binding intensity in LHSAR cells virally transduced to express HoxB13 and FoxA1, HoxB13 alone, FoxA1 alone or LacZ control at T-ARBSs and N-ARBSs. The joint effects of FoxA1 and HoxB13 recapitulate AR binding patterns in tumor. (b) Cluster analysis of all human specimens and the LHSAR modified cell lines supervised by T-ARBSs and N-ARBSs. The LNCaP cell line is shown in red. Data are part of the Broad Institute’s Project Achilles.
COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Whyte, W.A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319 (2013).
2. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
3. Lupien, M. & Brown, M. Cistromics of hormone-dependent cancer. Endocr. Relat. Cancer 16, 381–389 (2009).
4. Thompson, I.M. et al. Androgen receptor function in promoting metastatic prostate cancer. Cancer Res. 64, 8867–8875 (2004).
5. Williams, S.E. et al. Integrative genomic profiling of human prostate cancer. Cancer Cell 18, 11–22 (2010).
6. Wang, Q. et al. Androgen receptor regulates a distinct transcription program in prostate cancer. Cancer Cell 18, 11–22 (2010).
7. Wang, D. et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature 474, 390–394 (2011).
8. Jin, H.J., Zhao, J.C., Wu, L., Kim, J. & Yu, J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. Cell 138, 245–256 (2009).
9. Economicides, K.D. & Capecci, M.R. HOXB13 is required for normal differentiation and secretory function of the ventral prostate. Development 130, 2061–2069 (2003).
10. Wang, Q. et al. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. Cell 138, 245–256 (2009).
11. Economides, K.D. & Capecci, M.R. HOXB13 is required for normal differentiation and secretory function of the ventral prostate. Development 130, 2061–2069 (2003).

12. Sampson, N., Neuwirt, H., Puhm, M., Klocker, H. & Eder, I.E. In vitro model systems to study androgen receptor signaling in prostate cancer. Endocr. Relat. Cancer 20, R49–R64 (2013).
13. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X.S. Identifying ChIP-seq enrichment using MACS. Nat. Protoc. 7, 1728–1740 (2012).
14. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
15. Sharma, N.L. et al. The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. Cancer Cell 23, 35–47 (2013).
16. He, H.H. et al. Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. Genome Res. 22, 1015–1025 (2012).
17. Massie, C.E. et al. The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. EMBO J. 30, 2719–2733 (2011).
18. Yu, J. et al. An integrated network of androgen receptor, Polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. Cancer Cell 17, 443–454 (2010).
19. Berger, R. et al. Androgen-induced differentiation and tumorigenicity of human prostate epithelial cells. Cancer Res. 64, 8867–8875 (2004).
20. Taylor, B.S. et al. Integrative genomic profiling of human prostate cancer. Cancer Cell 18, 11–22 (2010).
21. Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M. & Gerald, W.L. Gene expression profiling predicts clinical outcome of prostate cancer. J. Clin. Invest. 113, 913–923 (2004).
22. Wang, D. et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature 474, 390–394 (2011).
23. Jin, H.J., Zhao, J.C., Wu, L., Kim, J. & Yu, J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. Nat. Commun. 5, 3972 (2014).
24. Wang, Q. et al. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. Cell 138, 245–256 (2009).
25. Economicides, K.D. & Capecci, M.R. HOXB13 is required for normal differentiation and secretory function of the ventral prostate. Development 130, 2061–2069 (2003).
26. Ewing, C.M. et al. Germline mutations in HOXB13 and prostate-cancer risk. N. Engl. J. Med. 366, 141–149 (2012).
27. Xu, J. et al. HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). Hum. Genet. 132, 5–14 (2013).
ONLINE METHODS

Tissue cohort. Fresh-frozen radical prostatectomy specimens were selected from the Dana-Farber Cancer Institute/Harvard Cancer Center SPORE biobank and database, as part of DFCI Protocol 01-045, approved by the Dana-Farber Cancer Institute/Harvard Cancer Center Institutional Review Board. Informed consent was obtained from all subjects whose samples were included in the study. A genitourinary pathologist (M.L.) reviewed slides stained with hematoxylin and eosin from each case and isolated areas estimated to be enriched by >70% for prostate tumor tissue or normal prostate epithelium at least 1 cm away from any tumor focus. Fourteen subjects were selected for ChIP analysis (Supplementary Table 1). An additional 89 subjects were selected for expression analysis. Of the 103 subjects in total, there was matched normal prostate epithelium for 63 individuals. A tissue microarray (TMA) comprised of a separate cohort of 49 subjects was used for analysis by immunohistochemistry.

ChIP-seq in human tissue specimens. Using a 2-mm² core needle, approximately four cores were extracted from areas circled on each slide. Frozen cores were pulverized using the Covaris CryoPrep system. Tissue was then fixed using 1% formaldehyde for 18 min, and fixation was quenched with glycine. Chromatin was sheared to 300–500 bp in size using the Covaris E220 ultrasonicator. The resulting chromatin was incubated overnight with 6 µg of antibody—to AR (N-20, Santa Cruz Biotechnology), HOXB13 (H-80, Santa Cruz Biotechnology) or FOXA1 (ab23738, Abcam)—bound to protein A or protein G beads (Life Technologies). A fraction of the sample was not exposed to antibody to be used as control (input). Samples underwent cross-linking reversal and were treated with RNase and proteinase K, and DNA was extracted. The samples were then rehydrated to 100–300 bp in size using the Covaris ultrasonicator, and the concentrations of ChIP DNA were quantified by Qubit fluorometer (Life Technologies). DNA sequencing libraries were prepared using the ThruPLEX-FD Prep kit (Rubicon Genomics). Libraries were sequenced using 50-bp reads on the Illumina platform at the Dana-Farber Cancer Institute.

ChIP-seq in cell lines. LHSAR cells were obtained from W.G.H., and LNCaP cells were obtained from the American Type Culture Collection. ChIP of cultured LNCaP and LHSAR cells was performed as described previously. Briefly, chromatin from approximately 1 × 10⁶ fixed cells was sonicated to a size range of 200–300 bp. Solubilized chromatin was subjected to immunoprecipitation with the antibodies described above. Reversal of cross-linking, DNA extraction, library preparation and genetic sequencing were performed as they were for human tissue. Mycoplasma contamination was checked for at least once per month (LookOut Mycoplasma PCR Detection kit, Sigma-Aldrich).

Sequencing and data analysis for ChIP and peak calling. All samples were processed through the exact same computational pipeline developed at the Center for Functional Cancer Epigenetics (CFCE) using primarily open source programs. Sequence tags were aligned with Bowtie to build hg19 of the human genome, and uniquely mapped, non-redundant reads were retained. These reads (termed ‘unique mappable locations’) are tabulated for each sample in Supplementary Table 1 and were used to generate AR binding sites with Model-Based Analysis of ChIP-seq 2 (MACS v2.0.10), with a q-value (FDR) threshold of 0.01 (refs. 13,28). We evaluated multiple quality control criteria based on alignment information and peak quality: (i) sequence quality score; (ii) uniquely mappable reads (reads that can only map to one location in the genome); (iii) uniquely mappable locations (locations that can only be mapped by at least one read); (iv) peak overlap with Velcro regions, a comprehensive set of locations—also called consensus signal artifact regions—in the human genome that have anomalous, unstructured high signal or read counts in next-generation sequencing experiments independent of cell line and of type of experiment; (v) number of total peaks (the minimum required was 1,000); (vi) high-confidence peaks (the number of peaks that are tenfold enriched over background); (vii) percentage overlap with known DHSs derived from the ENCODE Project (the minimum required to meet the threshold was 80%); and (viii) peak conservation (a measure of sequence similarity across species based on the hypothesis that conserved sequences are more likely to be functional). Typically, if a sample fails one of these criteria, it will fail many (locations with low mappability will likely have low peak numbers, many of which will likely be in high-mappability regions, etc.). All analyses were based on ChIP peaks located outside of promoter regions (≥2 kb upstream or ≥2 kb downstream of any transcription start site).

Differential AR peak and DNA binding motif analyses. Peaks from all study samples were merged to create a union set of AR binding sites (n = 76,553). Read densities were calculated for each peak for each sample in RPM, which were used for comparison of cistromes across samples. Sample similarity was determined by hierarchical clustering using the Spearman correlation between samples. Tissue-specific peaks (T-ARBSs and N-ARBSs) were identified by t test using limma with adjusted P ≤ 0.001. Equal numbers of peaks were randomly selected from each group of sites with differential binding (T-ARBSs and N-ARBSs) and were used for motif analysis by the motif search algorithm (MDSeqPos) with cutoff z score ≤ 0.05. For confirmation of motif analysis with HOMER (v3.0.0) 31, we used sequences from all sites with differential AR binding. The motif rankings in HOMER were slightly different as MDSeqPos employs a penalty if the matching motif is farther from the summit than from the AR peak, whereas HOMER does not have a distance penalty.

Determining the saturation level of common AR binding sites. In the set of tumor specimens, a number of subjects were randomly sampled and a number of union AR peaks within that set was determined. For each sample size, we repeated sampling of a cohort ten times and calculated the average number of union AR peaks and standard error. The rate of increase in union AR binding sites as the sample size increased was used to estimate the saturation of AR peak discovery.

Gene set enrichment analysis. For the 9,179 T-ARBSs and 2,690 N-ARBSs, the GREAT tool v3.0.0 was used to map a peak to a gene and perform GSEA using a 50-kb window around each state-specific AR binding site.

Gene expression and tissue-specific AR binding sites. mRNA expression data from the TCGA data set of 220 samples (44 normal and 176 tumor samples) were downloaded. Transcripts showing significantly different expression in tumor and normal samples (adjusted P ≤ 1 × 10−7) were selected for analysis (n = 3,743). These transcripts were rank-ordered by differential expression and grouped into bins of 300. The gene bins were plotted against the average read density at each T-ARBS and N-ARBS using a distance of 50 kb. A similar analysis was performed using the 103 radical prostatectomy specimens described above. RNA was extracted from fresh-frozen tissue using the Qiagen AllPrep DNA/RNA kit and prepared for expression analysis using the Affymetrix HT2.0 array. Expression data, in the form of .CEL files, were normalized using robust multichip average (RMA) and were log transformed. In total, 1,804 (adjusted P ≤ 0.001) genes that were significantly differentially expressed were binned into groups of 200 and analyzed as above.

HOXB13 motif enrichment analysis. The goal of this analysis was to evaluate the enrichment of the HOXB13 motif in certain sets of AR binding sites in LNCaP cells (Fig. 3b). Data for HOXB13 ChIP analysis in VCaP cells were downloaded from Huang et al. 33. The HOXB13 motif was mapped to each AR peak region with the position-specific scoring matrix (PSSM), using the HOMER program with default parameters. To evaluate the background HOXB13 genomic distribution, the union DHSs derived for 80 cell types from the ENCODE Project were used. Motif distribution across all conditions was compared to the DNase I hypersensitivity peak background, and enrichment was calculated by Fisher’s exact test.

AR targeted gene set and prostate cancer recurrence-free survival analysis. Each site with differential AR binding was assigned to the nearest gene (transcription start site ≤50 kb from the AR summit). Differential gene expression was calculated on the basis of the TCGA prostate data set, and RNA-seq data were downloaded from the Broad Institute. All expression values were normalized according to the TCGA expression pipeline (accessed 15 January 2014). Differential expression between prostate tumor and normal tissue was identified using limma with t test with adjusted P ≤ 1 × 10−7. If the closest gene was differentially expressed in the appropriate direction (for example, upregulated
in tumor for a T-ARBS and downregulated in tumor for an N-ARBS), it was selected. A total of 536 genes (324 upregulated genes associated with T-ARBSs and 212 downregulated genes associated with N-ARBSs) met the criteria.

Kaplan-Meier survival analysis was carried out using the R package survival analysis (version 2.37-7). The end point for survival analysis was biochemical recurrence after radical prostatectomy. Patient samples underwent unsupervised clustering based on the 536-gene target gene set. The statistical significance of differences between the survival curves was assessed using z² and log-rank tests.

TSA-plus fluorescence immunohistochemistry. A total of 49 formalin-fixed, paraffin-embedded prostate cancer tissues were obtained from consented patients at Brigham and Women's Hospital and used to construct the TMA; each sample was represented by three TMA cores for tumor.

A multiplexed tyramide signal amplification (TSA) method was performed on 4-μm sections of the TMA for detection of the colocalization of AR, FOXA1 and HOXB13 proteins. The staining approach consisted of a multistep protocol of sequential TSA-amplified immunofluorescence labels for AR, FOXA1 and HOXB13 and counterstaining with DAPI. Briefly, sections were deparaffinized and hydrated; before each immunofluorescence labeling, AR, FOXA1 and HOXB13 antigens were retrieved with a single microwave step. Each labeling cycle consisted of the application of a primary antibody, followed by a secondary antibody conjugated to horseradish peroxidase (HRP) and TSA conjugated to a fluorophore. The slides were incubated with antibody against AR (AR441, mouse monoclonal, Dako) at a dilution of 1:500, FOXA1 (HNF-3α/b (C-20), goat polyclonal, sc-6553, Santa Cruz Biotechnology) at a dilution of 1:1,000 and HOXB13 (F-9, mouse monoclonal, sc-28333, Santa Cruz Biotechnology) at a dilution of 1:500 for 30 min. TSA reagents were obtained from PerkinElmer. TSA conjugated to fluorescein was used for AR, to Cy3 was used for FOXA1 and to Cy5 was used for HOXB13.

Prostate cancer tissue from radical prostatectomy specimens was used as positive controls for AR, FOXA1 and HOXB13. Omission of the primary antibody was used as a negative control. Slides with single staining for each antibody and with counterstaining for DAPI were used for spectral library construction.

The TMA was scanned on a PerkinElmer Vectra 2 imaging workstation. Appropriate filter cubes and 20x objective exposure times were set for each target-dye combination (specifically, nuclei-DAPI, AR-FITC, FOXA1-Cy3 and HOXB13-Cy5). Each single-stained control slide was imaged with the established exposure time to create spectral controls for generating the spectral library. A 4x magnification spectral scan was run with the correct number of TMA rows and columns, as determined from the TMA map. Three focal points were manually selected. The remainder of the scan was fully automated.

Using the ‘scale to max’ spectra viewer, DAPI peak brightness was verified. The constancy of the shape of the curve (dye spectra) was confirmed by overlaying spectra from individual pixel readouts using the ‘normalized’ spectra viewer. Three unique regions of the TMA were sampled. The same process was replicated for each of the individual target-dye combinations. We ran an algorithm learning tool using the InForm Analysis software package to train for the appropriate tumor, benign and prostatic intraepithelial neoplasia (PIN) regions of interest and subsequently completed cell segmentation. Normalized total expression intensity was recorded on a per-pixel basis. The reported mean for a given cell is the average intensity of all the normalized total pixel values in each nucleus. Nucleus finding was enabled by detecting circular objects in the DAPI channel. A two-pixel radius around the nucleus was defined as the cytoplasm. The algorithm was then applied to all the images contained in the TMA.

Positive staining thresholds were determined initially upon pathology review of the images generated and subsequently statistically measured by ranking nuclear expression levels for each marker independently and calculating the limit of detection (mean + 3 s.d.). Signal intensity cutoffs of 0.09, 1.05 and 0.12 for AR, FOXA1 and HOXB13, respectively, were determined. Statistical analysis was carried out using SPSS v.13.0. Spearman rank correlation was used to investigate the correlation of protein expression among AR, FOXA1 and HOXB13.

Immunoblotting. Whole-cell extracts were prepared by lysis in RIPA buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate and 0.1% SDS) supplemented with protease inhibitor cocktail (Roche). Nuclear extracts were prepared by hypotonic lysis (10 mM HEPES, pH 8.0, 5 mM MgCl₂, 250 mM sucrose and protease inhibitors) followed by extraction with RIPA buffer. Immunoblotting was performed according to standard procedures using the following antibodies: D6F11 for AR (Cell Signaling Technology), F-9 for HOXB13 (Santa Cruz Biotechnology), aβ23738 for FOXA1 (Abcam) and C4 for β-actin (Santa Cruz Biotechnology). All antibodies were used at a dilution of 1:1,000. The Odyssey imaging system was used to quantify the fluorescence intensity of infrared dye-conjugated secondary antibodies (LI-COR).

Transduction of HOXB13 and FOXA1 in the LHSAR cell line and knockdown in the LNCaP cell line. The ORF of HOXB13 or FOXA1 was cloned into the pLX302 lentiviral expression vector using the Gateway recombination system. Lentiviral vectors expressing shRNAs targeting HOXB13 or FOXA1 were obtained from the RNA interference (RNAi) consortium at the Broad Institute (clone numbers TRCN0000020845, TRCN0000020847, TRCN0000358367 and TRCN000014879). shRNAs targeting GFP or luciferase were used as negative controls, and shRNA targeting RP56 was used as a positive control for a gene required for cell proliferation. Viruses were generated by transfection of vectors into 293T cells along with the packaging vectors pVsVg and pDelta9.9, and transduction was performed in the presence of 4 μg/ml polybrene. For expression experiments in LHSAR cells, the medium was replaced 24 h after transduction with medium containing 2 mM R1881, and cells were incubated 48 h before collection for ChIP-seq analysis. For proliferation experiments, transduced cells were counted and replated in 12-well plates after selection with puromycin. Cells were counted at the indicated time points using a Vi-Cell analyzer. Results represent the average and 95% confidence interval of three biological replicates. Expression and knockdown were confirmed by immunoblotting.

RNA-seq in LHSAR cells and data analysis. LHSAR cells were virally transduced to express transcription factors as described above. For RNA-seq analysis, 75-bp reads were aligned to hg19 with the STAR aligner34. Cufflinks was used to generate expression values (in RPMK) for each gene. Differentially expressed genes were identified by limma for downstream analysis. Sample similarity was determined by hierarchical clustering in R Bioconductor using the Spearman correlation between samples based on the top 1,000 most variable genes.

Computational quantification of the gene suppression phenotypes from multi-sample RNAi screens. Project Achilles is a systematic effort to create a catalog of vulnerabilities across hundreds of cell lines by performing genome-wide pooled shRNA screens to identify essential genes. Each cell line is propagated for 16 population doublings or 40 d (whichever comes first), and the relative abundance of each shRNA is calculated with respect to the initial DNA plasmid pool. Data are submitted to a processing pipeline, and the data on shRNA levels are converted to gene-level scores using the ATARIS algorithm35. Here we used Achilles 2.9, a data set of 102 cell lines interrogated with a library containing ~98,000 shRNAs targeting ~17,000 genes, to extract the data for HOXB13 and FOXA1. Data for HOXB13_1_1100 and FOXA1_1_110011001 gene summarization scores (ATARIS solutions) were used to make the graphs. A complete data set will be available from the Project Achilles online portal. Dependency of the prostate cancer cell lines from the data set (LNCaP and PC3) on HOXB13 was determined using four separate shRNAs against HOXB13 (Supplementary Table 15). Cells were infected with lentivirus encoding shRNAs of interest in the presence of polybrene (4 μg/ml) followed by centrifugation. After 72 h of selection with puromycin, cells were trypsinized and replated in 12-well plates in duplicate. The number of viable cells was counted every 48 h using a Beckman Coulter Vi-Cell viability analyzer. Successful knockdown was confirmed by quantitative PCR using TaqMan assays.

28. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).
29. Smyth, G.K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3, Article3 (2004).
30. Liu, T. et al. Cistrome: an integrative platform for transcriptional regulation studies. 
*Genome Biol.* **12**, R83 (2011).

31. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. 
*Mol. Cell* **38**, 576–589 (2010).

32. Bolstad, B.M., Irizarry, R.A., Astrand, M. & Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. 
*Bioinformatics* **19**, 185–193 (2003).

33. Huang, Q. et al. A prostate cancer susceptibility allele at 6q22 increases *RFX6* expression by modulating HOXB13 chromatin binding. 
*Nat. Genet.* **46**, 126–135 (2014).

34. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. 
*Bioinformatics* **29**, 15–21 (2013).

35. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. 
*Nat. Protoc.* **7**, 562–578 (2012).

36. Shao, D.D. et al. ATARIS: computational quantification of gene suppression phenotypes from multisample RNAi screens. 
*Genome Res.* **23**, 665–678 (2013).