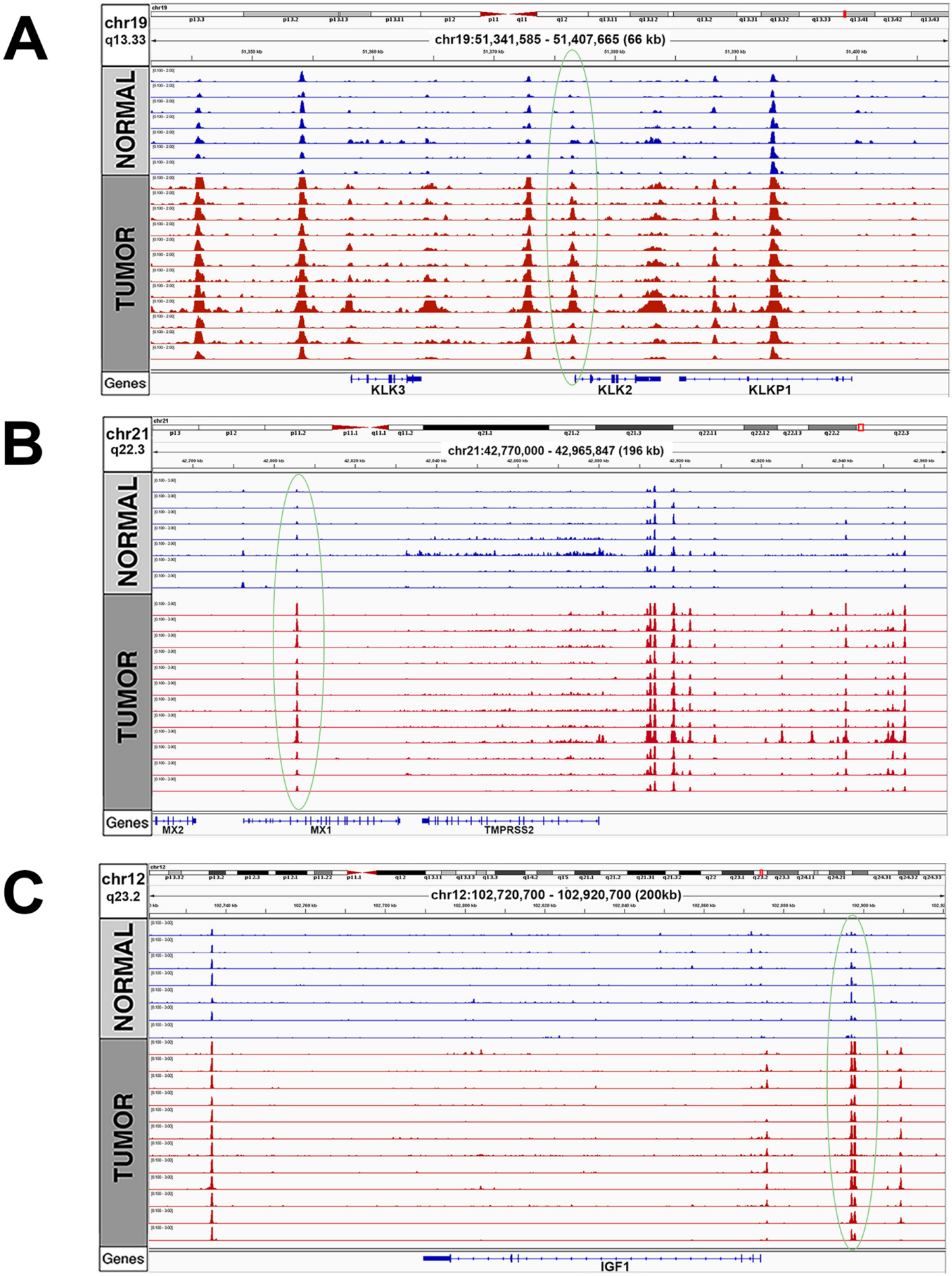
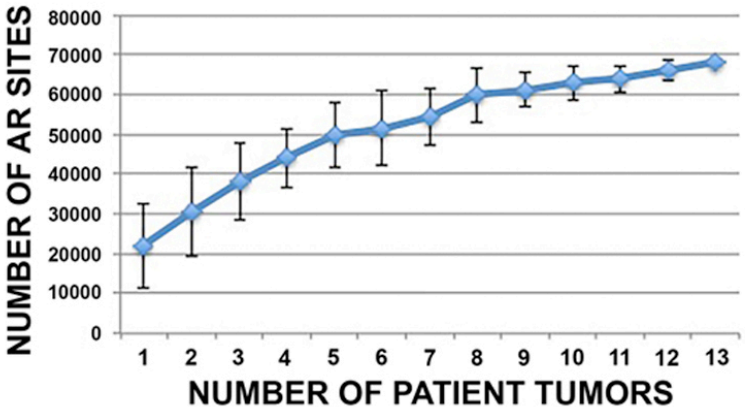
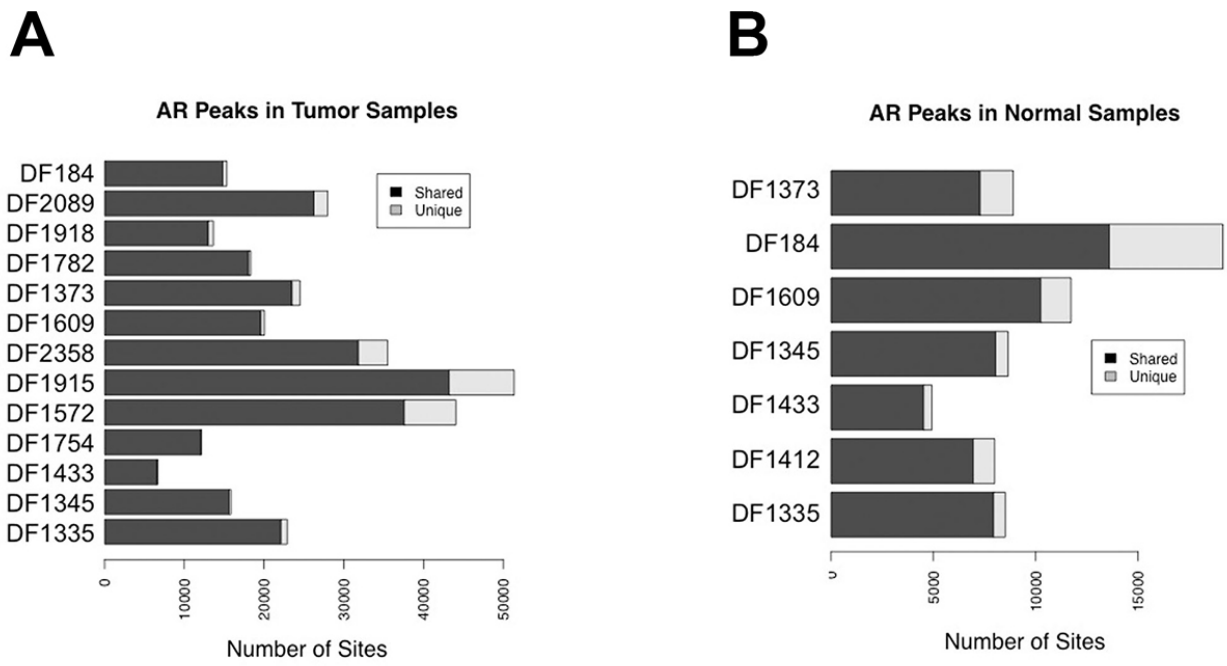


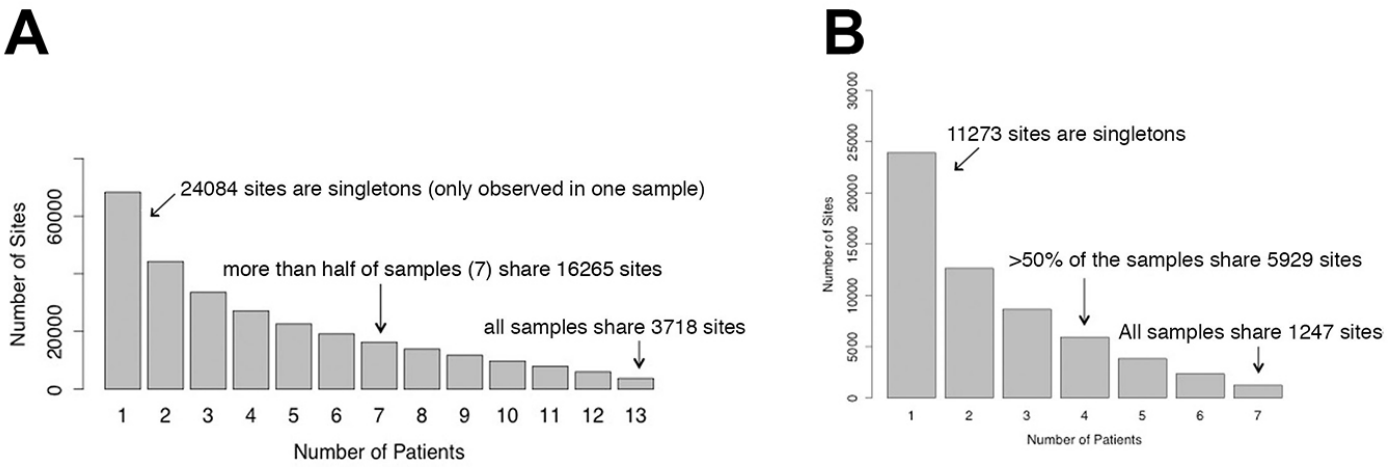
Supplementary Fig. 1. Androgen receptor (AR) binding in normal prostate epithelium and tumor tissue at known AR-regulated genes. Each track depicts normalized ChIP-seq AR binding intensity for a given normal or tumor tissue sample. **A.** Locus at chromosome 19, including the KLK3 gene. **B**. Locus at chromosome 21, including the TMPRSS2 gene. **C**. Locus at chromosome 12, including the IGF1 gene. In each image, examples of tissue-specific AR binding sites are circled in green.



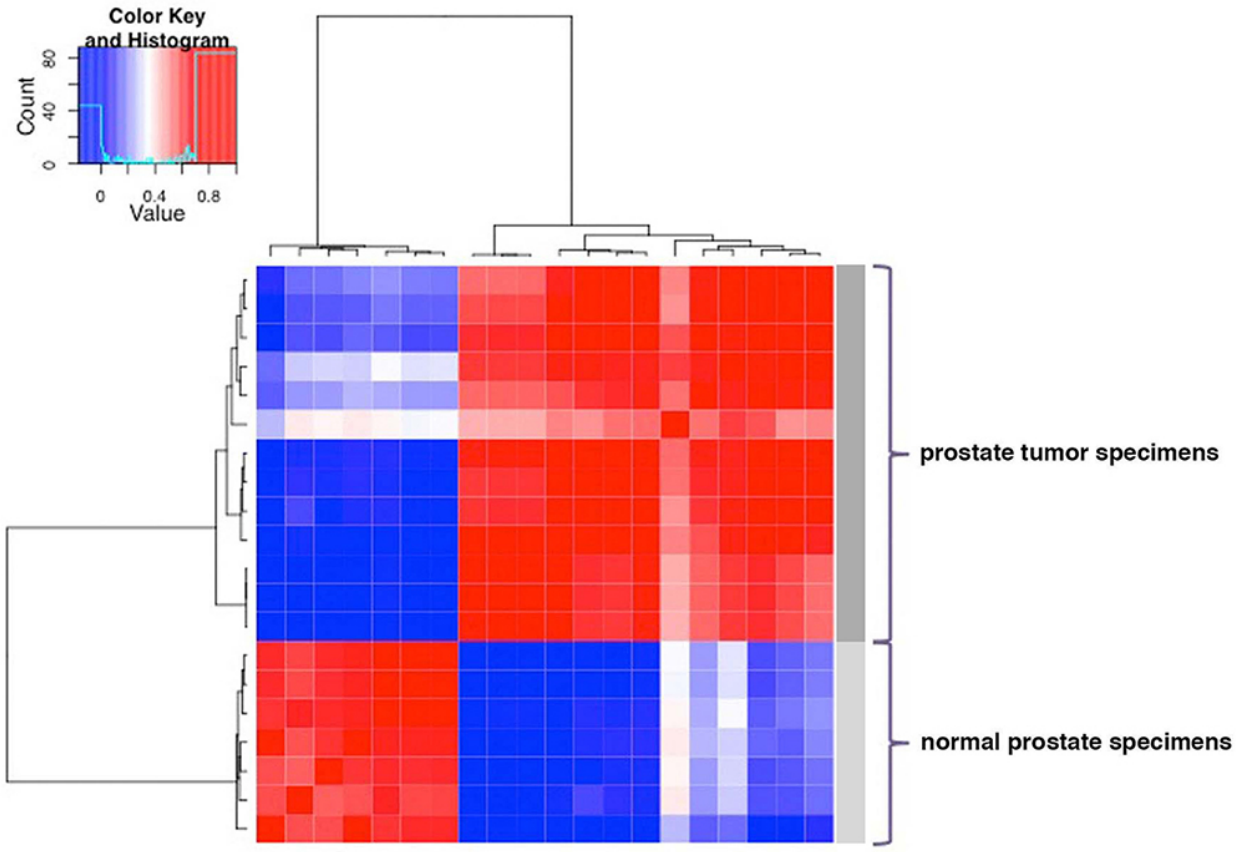
Supplementary Fig. 2. The rate of increase in AR binding sites identified as a function of the number of individuals undergoing ChIP-seq analysis.



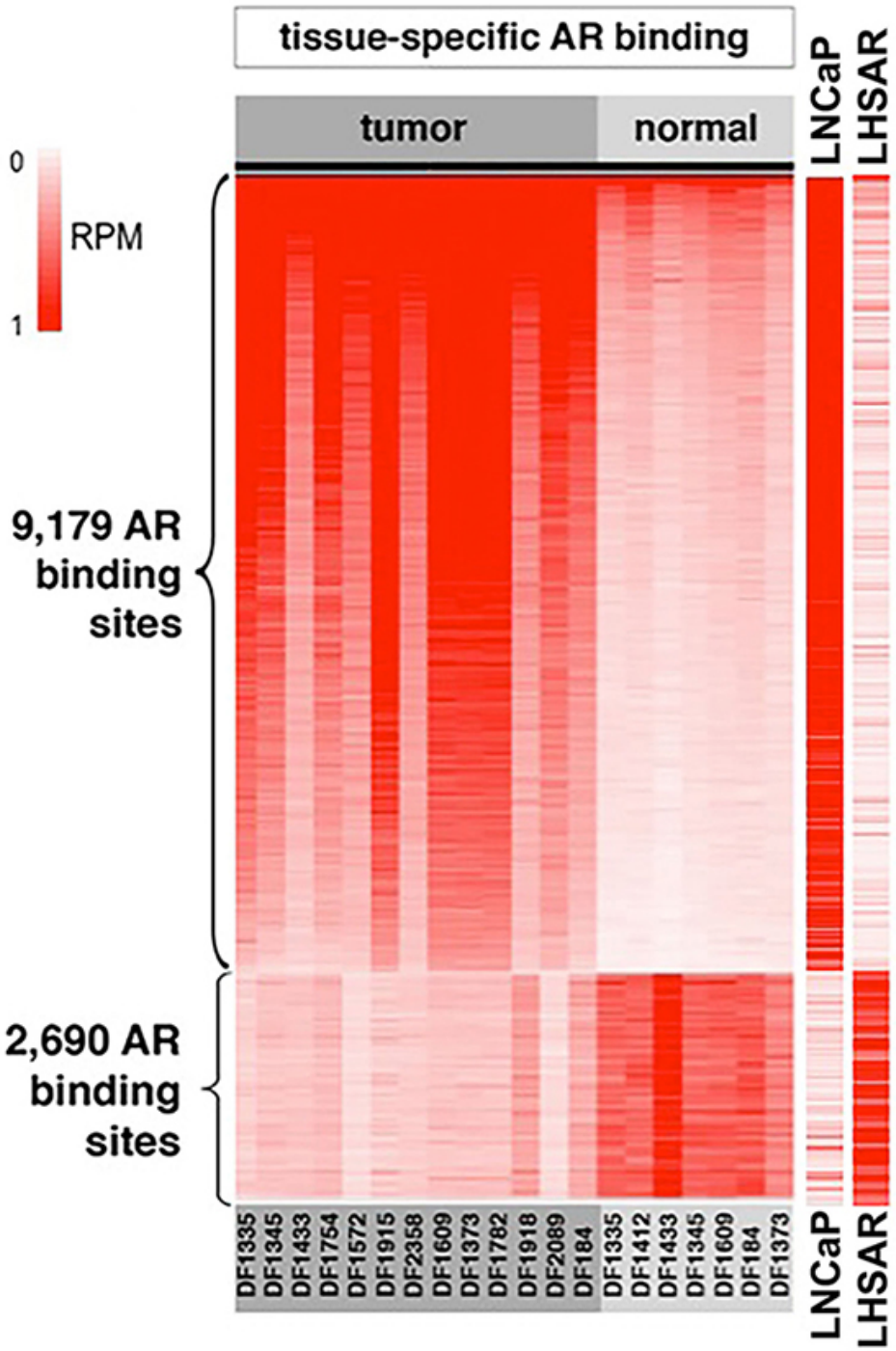
Supplementary Fig. 3. The number of androgen receptor (AR) binding sites genome-wide discovered by ChIP-seq in each individual in the cohort. **A**. Number of sites in tumor specimens. The dark part of each bar depicts AR sites shared with at least one other individual and the light gray depicts AR sites unique to that individual specimen. **B.** AR binding in the normal prostate specimens.



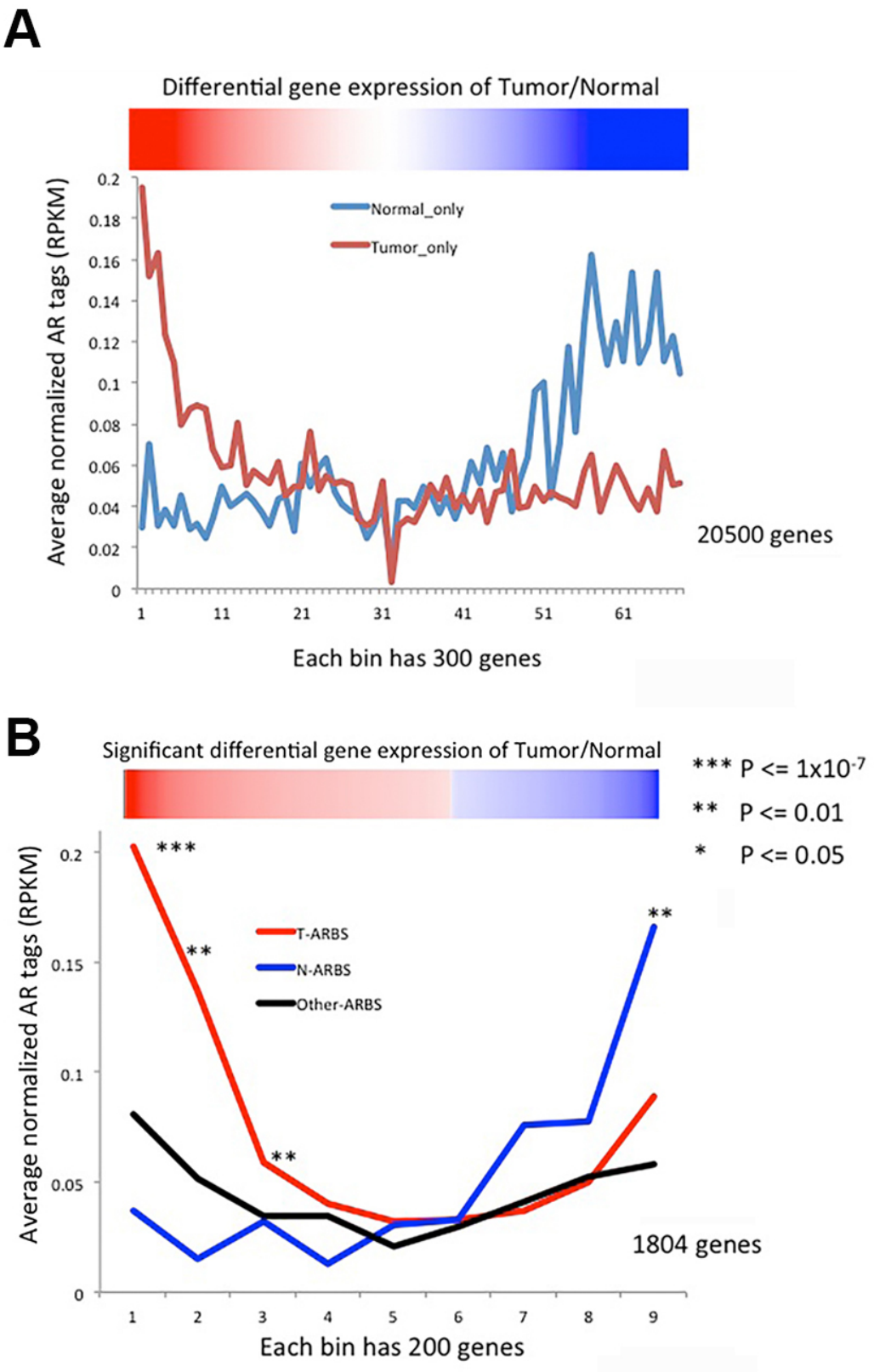
Supplementary Fig. 4. Androgen receptor (AR) binding sites shared by multiple individuals.
A. Cumulative AR binding site overlap in tumor specimens. **B.** Cumulative AR binding site overlap in normal prostate specimens.



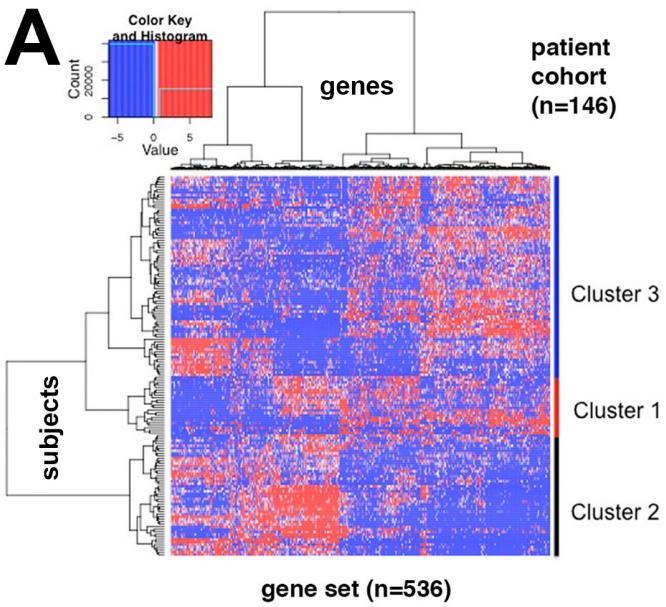
Supplementary Fig. 5. Supervised analysis of all specimens in the cohort (13 tumor and 7 normal prostate tissues), using only tissue-specific ARBS to cluster the samples: 9,179 T-ARBS and 2,690 N-ARBS.

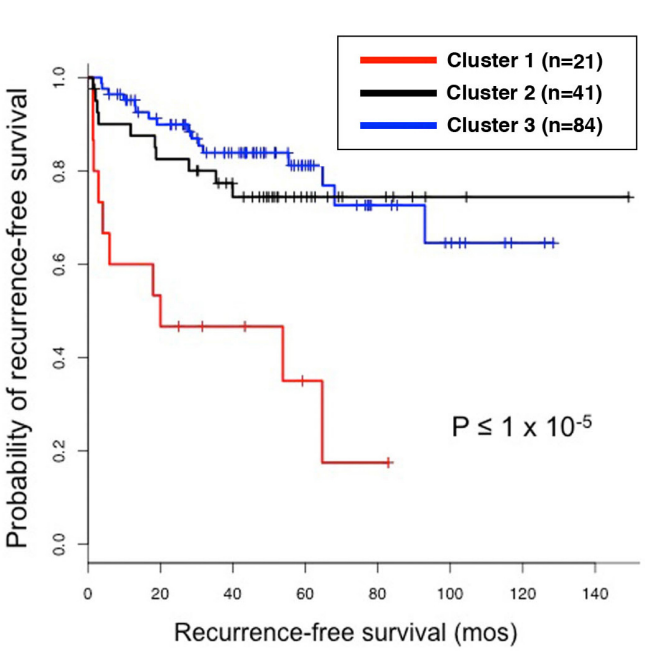


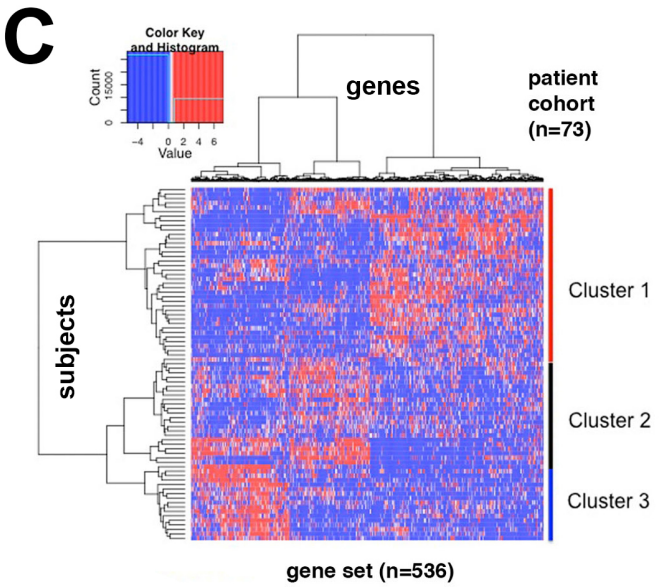
Supplementary Fig. 6. AR binding at human T-ARBS and N-ARBS in the LNCaP and LHSAR prostate cell lines. At left, AR binding intensity at tissue-specific AR sites in 13 human tumor specimens and 7 normal prostate epithelial specimens. At right, AR binding intensity at these sites in the LNCaP and LHSAR cell lines.

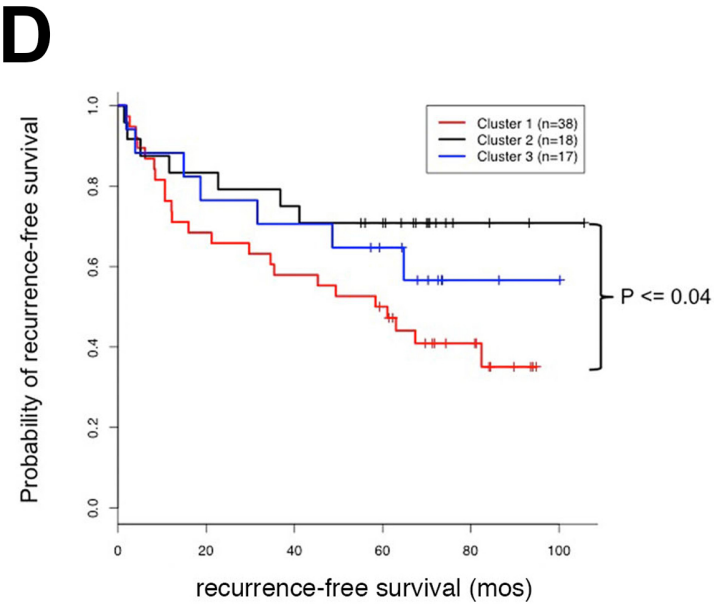


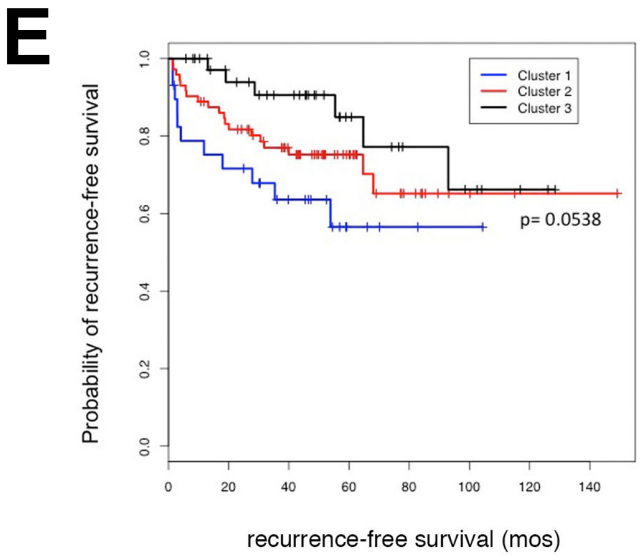
Supplementary Fig. 7. Association between tumor/normal differential gene expression and proximity of tissue-specific ARBS. **A.** All genes annotated in the TCGA data set. The bar atop the graph depicts the genes most overexpressed in tumor relative to normal (red) and normal relative to tumor (blue). The genes from this expression dataset were binned into groups of 300, displayed along the x-axis in order of expression level. The graph depicts the type of AR binding sites in human specimens most proximal to these 20,500 genes (T-ARBS in red and N-ARBS in blue). The y-axis shows the average of AR tags within 50kb of these genes. **B.** The most widely differentially expressed genes identified from a dataset of tumor/normal pairs from 103 human radical prostatectomy specimens (103 tumor and 63 matched normal prostate epithelial specimens). The 1804 differentially expressed genes from this expression dataset were binned into groups of 200, displayed along the x-axis in order of expression level. The graph depicts the type of AR binding sites in human specimens most proximal to these 1804 genes (T-ARBS, N-ARBS, or AR sites shared by tumor and normal tissue). Asterisks along the curves indicate the p-value (upper right) for enrichment of tissue-specific ARBS at these genes relative to non-tissue-specific ARBS.



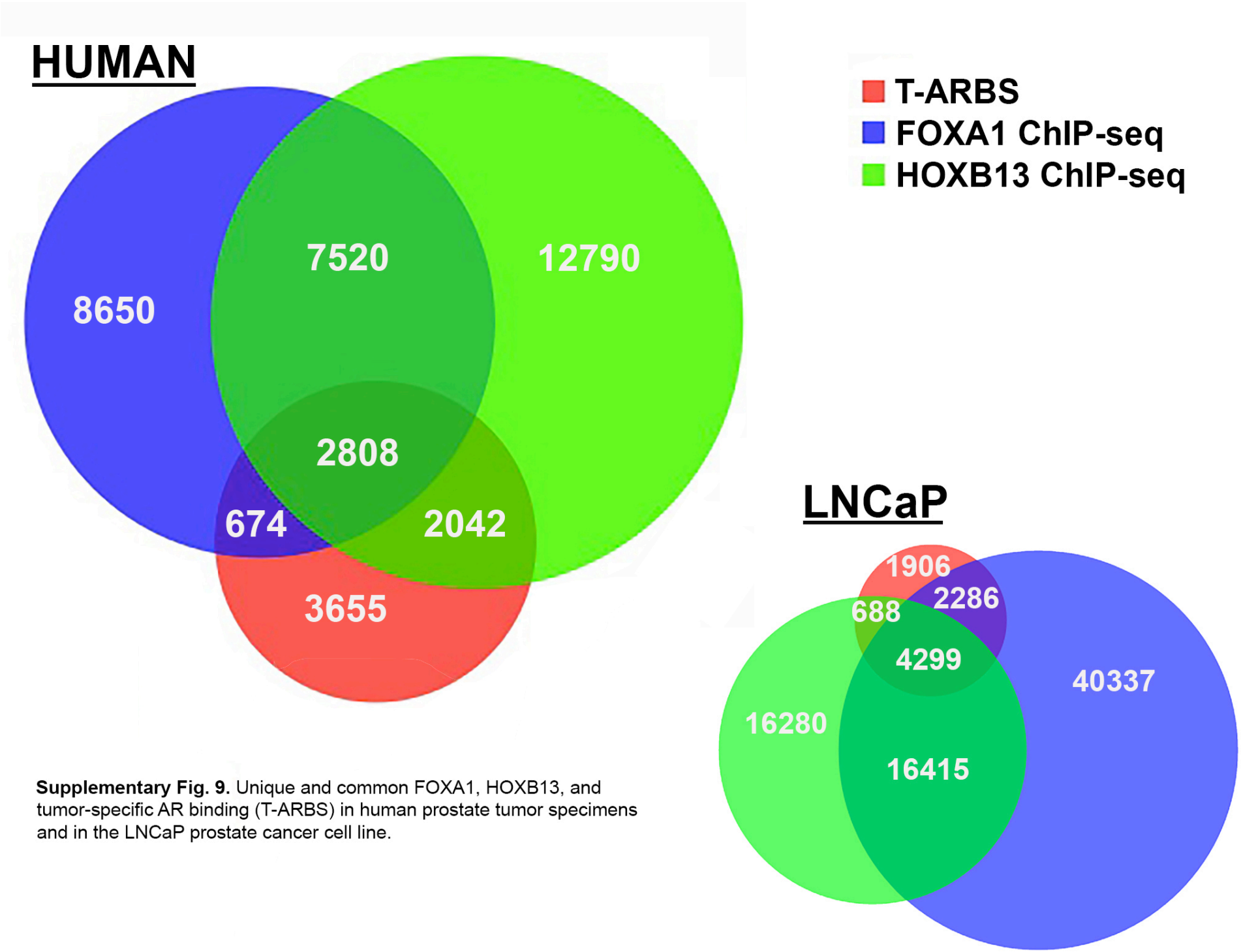




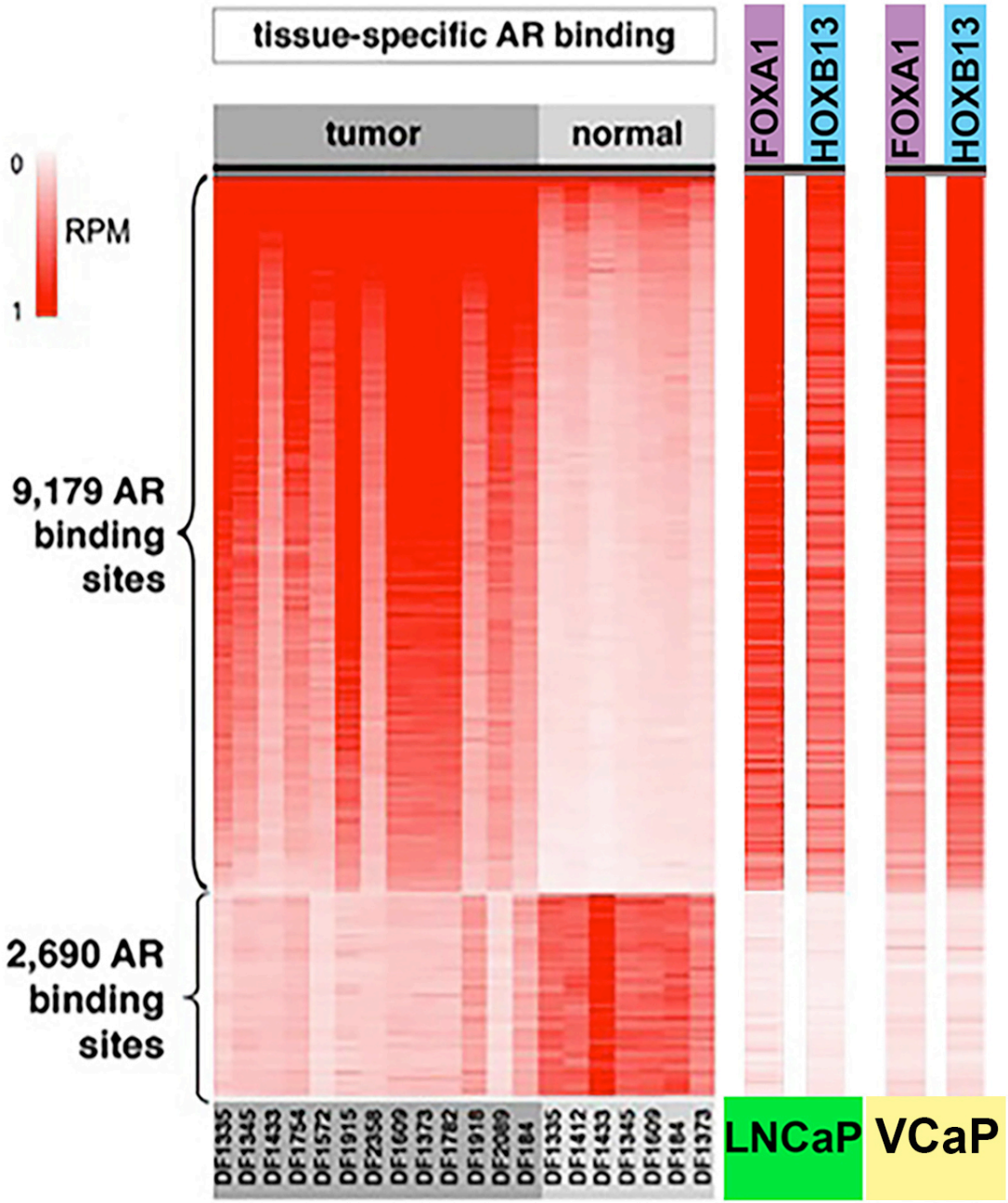




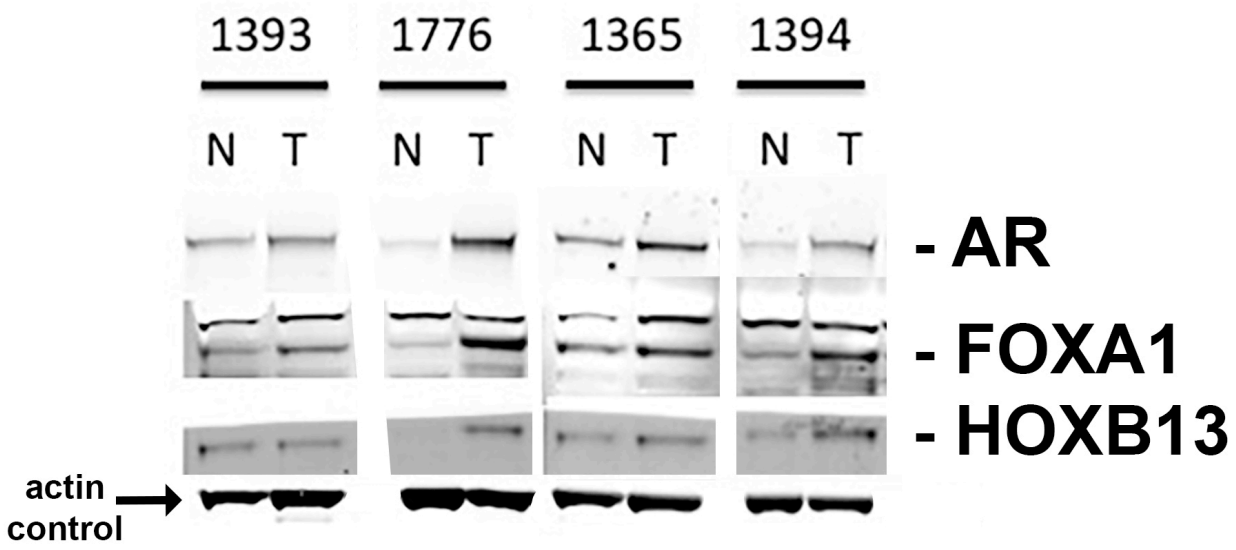
Supplementary Fig. 8. Clinical outcome is associated with gene expression at tissue-specific AR binding sites. A gene expression signature comprising 536 genes was derived based on proximity to tissue-specific AR binding sites (Methods). Unsupervised hierarchical clustering of 146 prostate cancer subjects was performed based on expression of these genes. **A.** Unsupervised hierarchical clustering in Taylor et al. cohort based on expression of genes in the AR target gene set. Patients clustered into three major classes. **B.** Kaplan-Meier analysis of recurrence-free survival after radical prostatectomy, based on gene expression cluster. Patients assigned to Cluster 1 (red) demonstrated significantly more aggressive disease. **C.** Unsupervised cluster analysis in the Glinsky et al. cohort. **D.** Kaplan Meier analysis of the expression clusters derived from the Glinsky et al cohort. **E.** Kaplan-Meier survival analysis of Taylor et al. subjects clustered based on expression of the top 200 differentially expressed genes in TCGA, NOT proximal to tissue-specific ARBS.



Supplementary Fig. 9. Unique and common FOXA1, HOXB13, and tumor-specific AR binding (T-ARBS) in human prostate tumor specimens and in the LNCaP prostate cancer cell line.



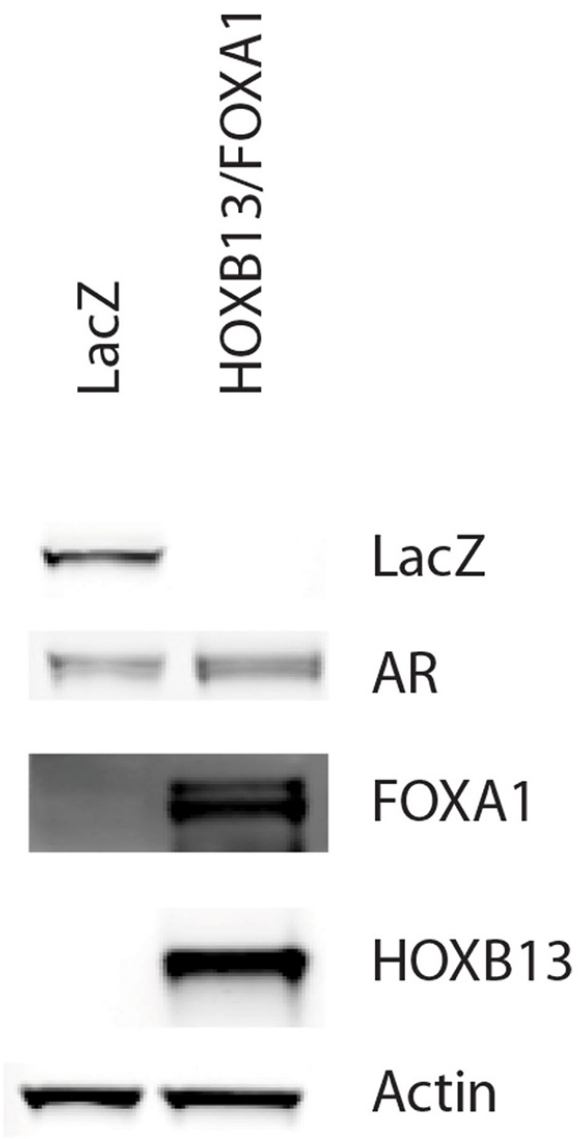
Supplementary Fig. 10. FOXA1 and HOXB13 binding in prostate cancer cell lines at human T-ARBS and N-ARBS. At left, AR binding intensity at tissue-specific AR sites in 13 human tumor specimens and 7 normal prostate epithelial specimens. At right, binding intensity at these sites in the LNCaP and VCaP cell lines upon ChIP-seq for the transcription factors FOXA1 and HOXB13. VCaP transcription factor cistrome data derived from publicly available data33.



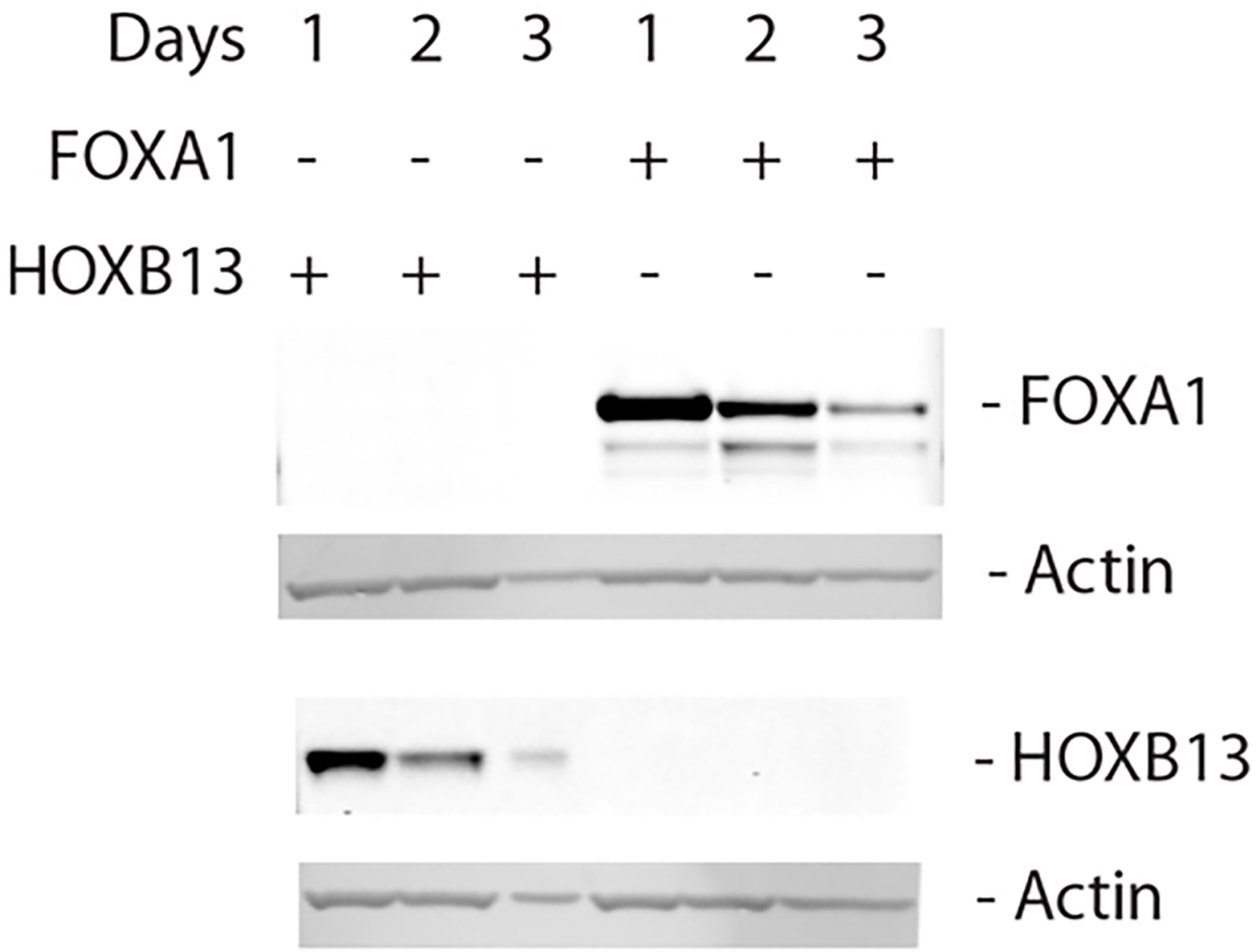
tumor / normal expression

	AR	FOXA1	HOXB13
1393	0.86	1.00	0.46
1776	12.63	12.96	13.59
1365	1.69	1.40	1.41
1394	4.06	3.66	3.17

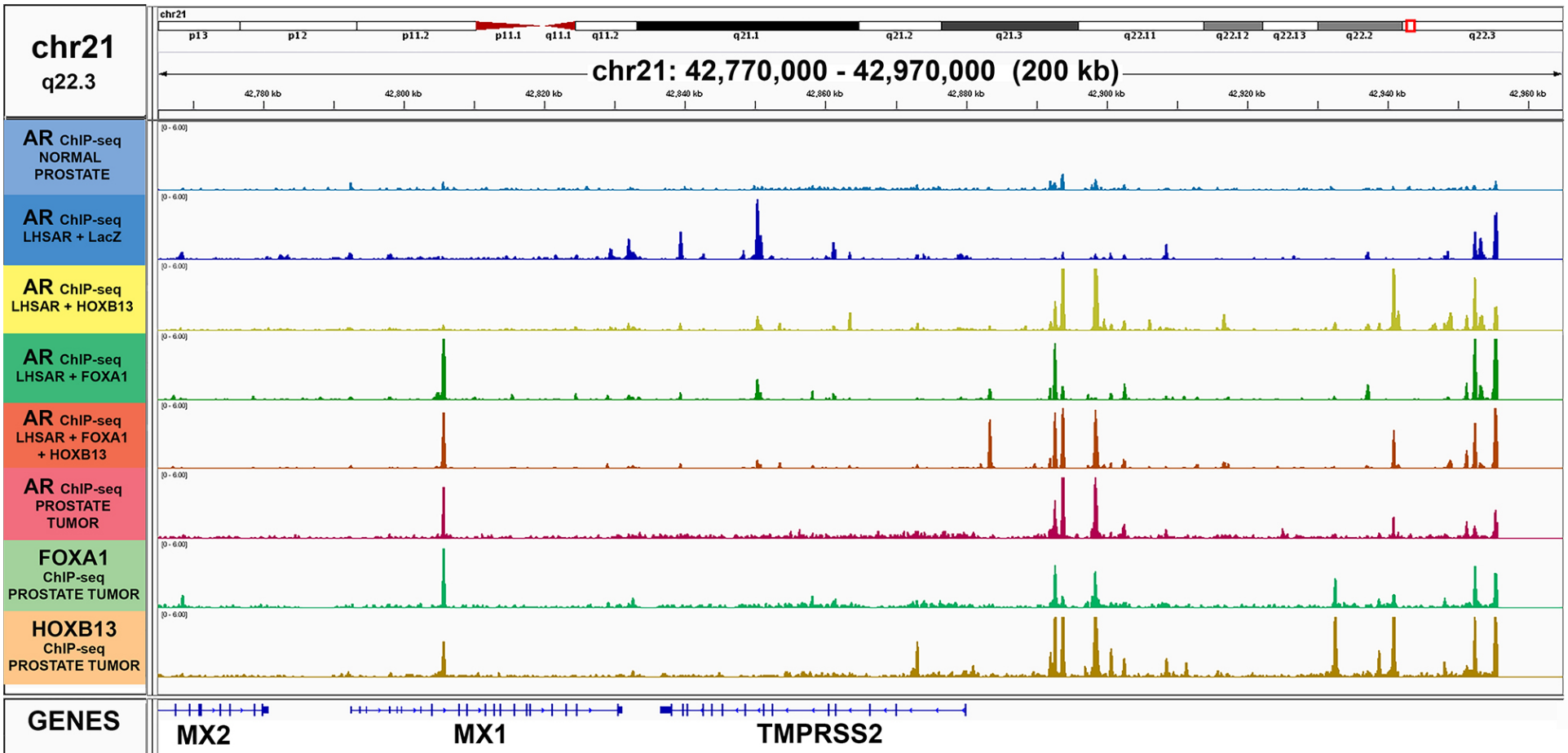
Supplementary Fig. 11. **Western immunoblot for AR, FOXA1, and HOXB13 in normal and tumor pairs from four radical prostatectomy specimens**. Immunoblots for each protein assayed are demonstrated at the top of the figure. At the bottom, tumor-to-normal expression ratios after normalizing to actin expression.



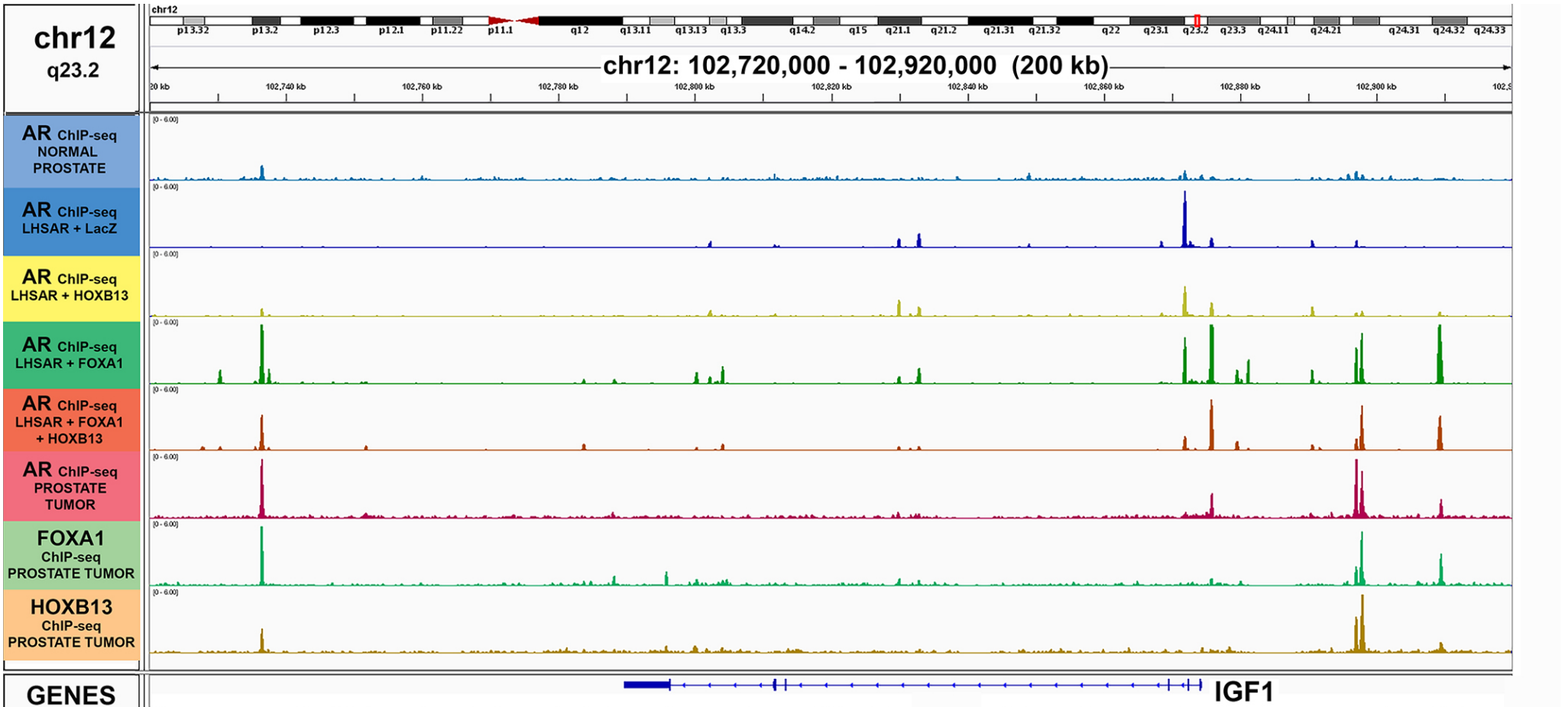
Supplementary Fig. 12. Western blot analysis of the cells in which AR ChIP-seq was performed in LHSAR. The figure demonstrates lack of HOXB13 or FOXA1 protein expression in control LHSAR cells transduced with LacZ. HOXB13 and FOXA1 proteins are clearly expressed in the LHSAR cells in which these genes were introduced.



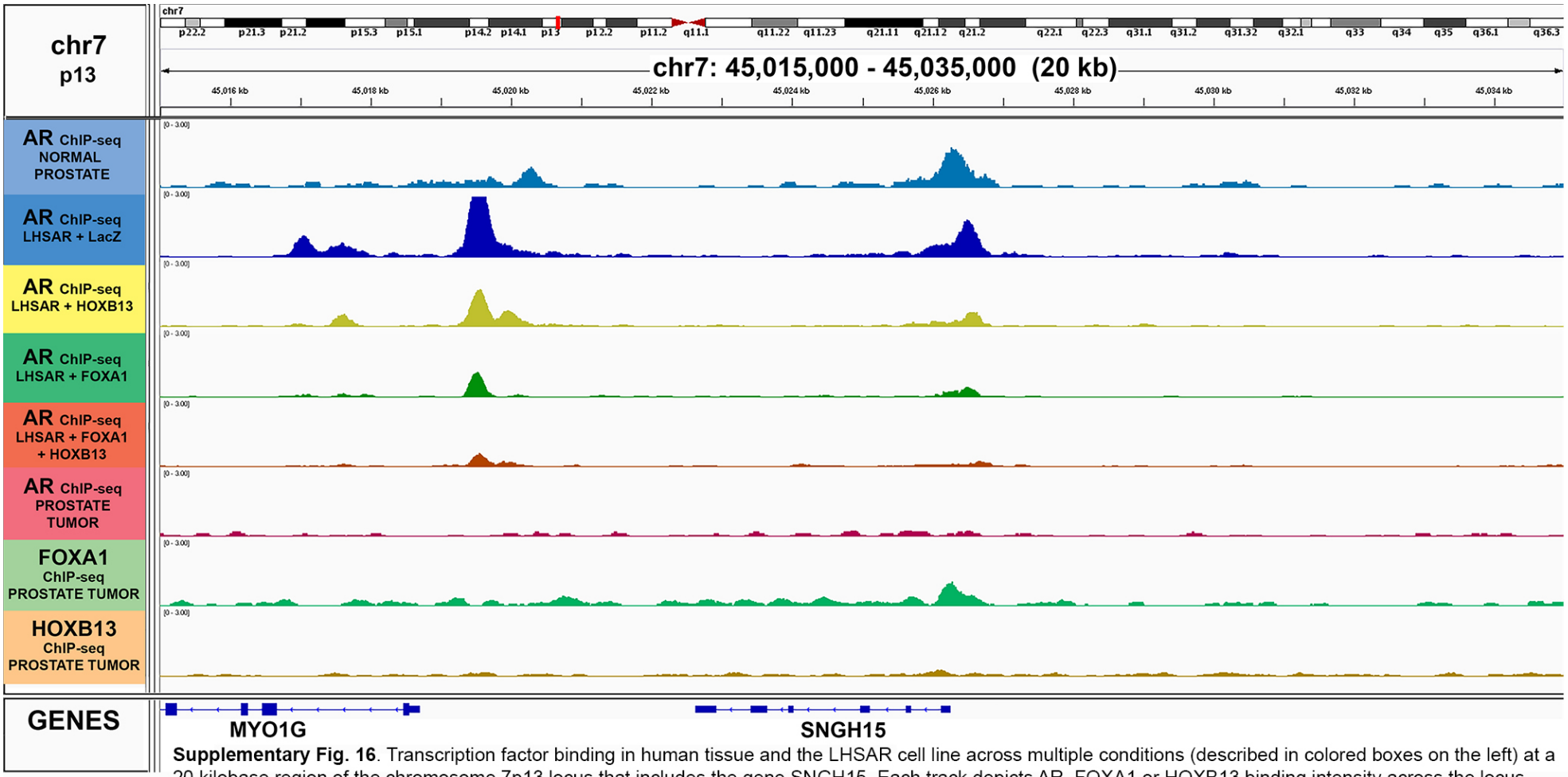
Supplementary Fig. 13. Time course of FOXA1 and HOXB13 expression in LHSAR cells transduced with these transcription factors.



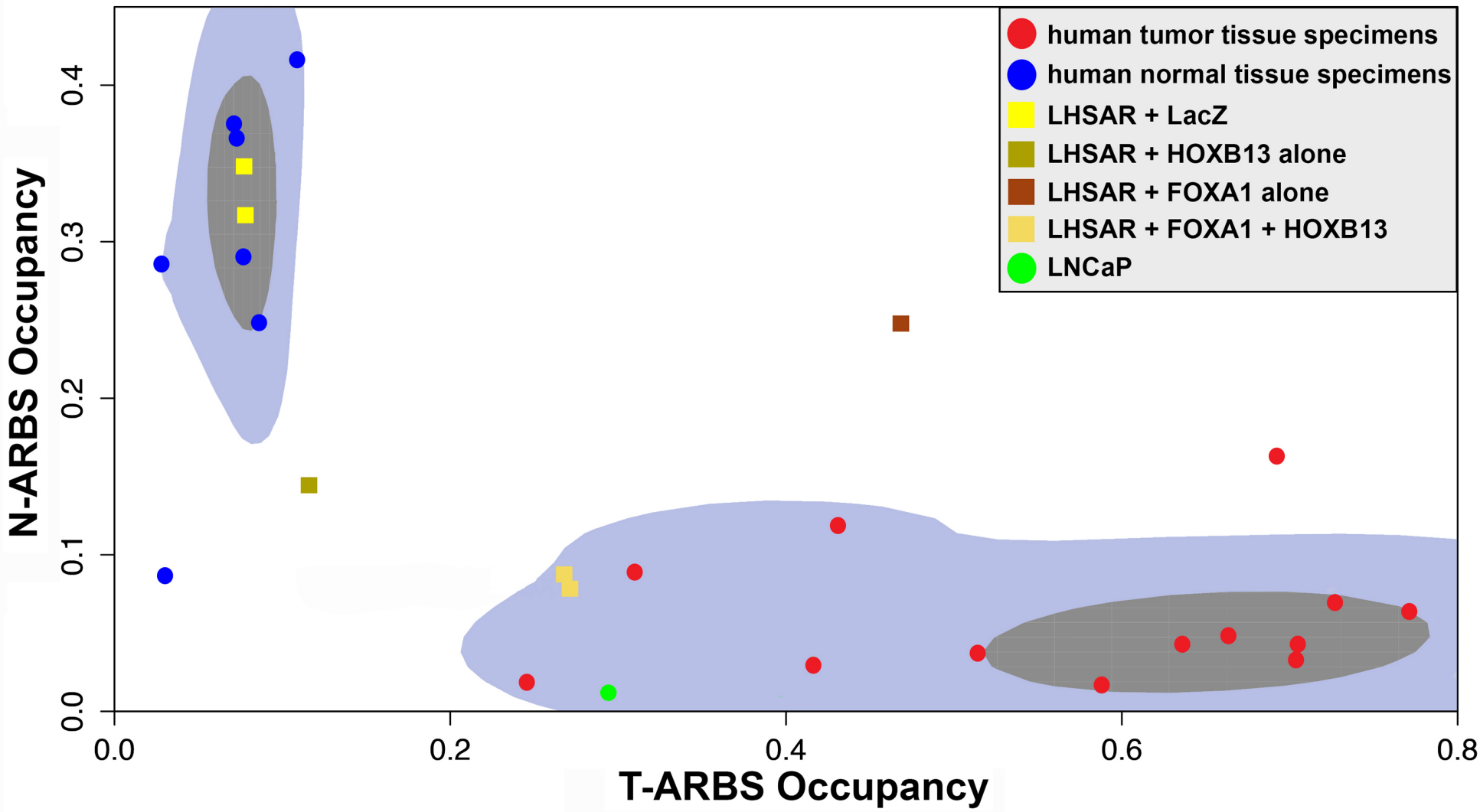
Supplementary Fig. 14. Transcription factor binding in human tissue and the LHSAR cell line across multiple conditions (described in colored boxes on the left) at a 200 kilobase region of the chromosome 21q22 locus that includes the AR-regulated gene TMPRSS2. Each track depicts AR, FOXA1 or HOXB13 binding intensity across the locus.



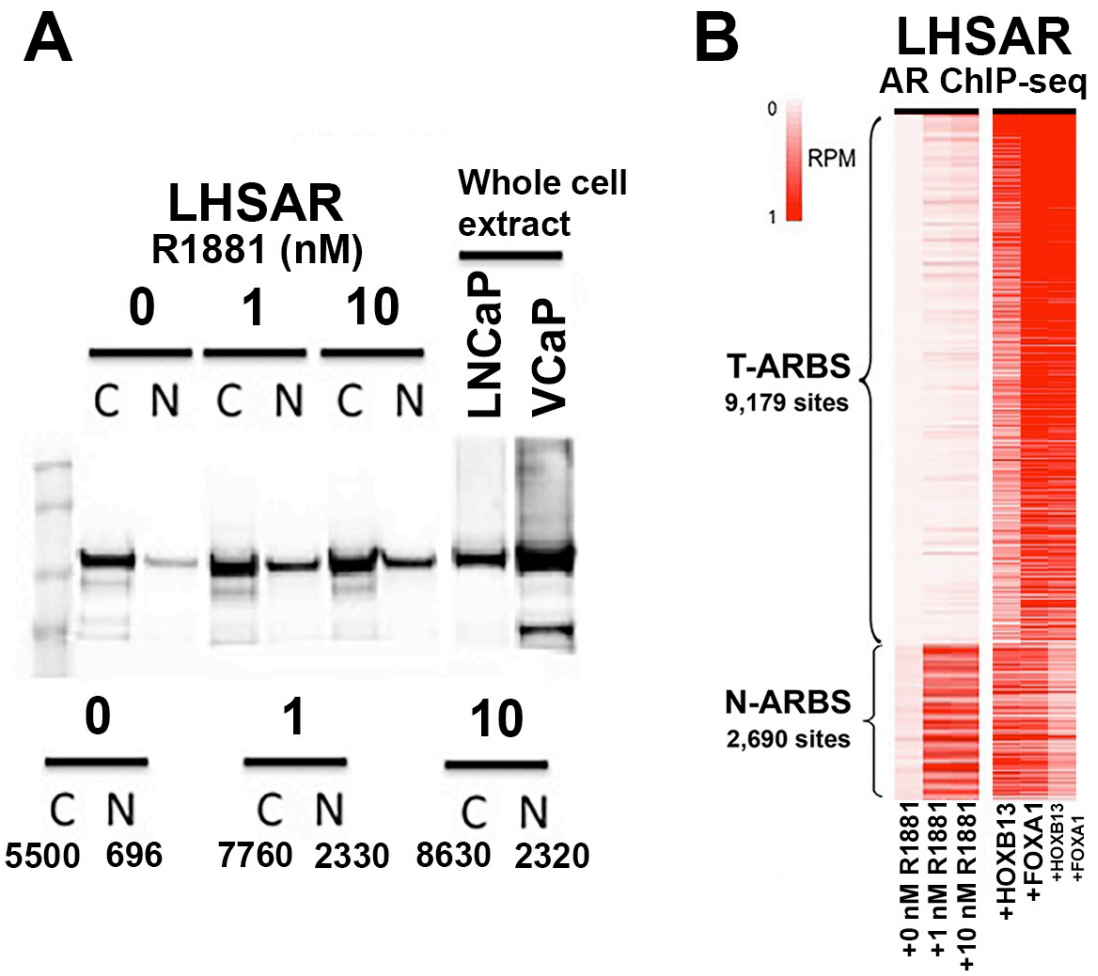
Supplementary Fig. 15. Transcription factor binding in human tissue and the LHSAR cell line across multiple conditions (described in colored boxes on the left) at a 200 kilobase region of the chromosome 12q23 locus that includes the AR-regulated gene IGF1. Each track depicts AR, FOXA1 or HOXB13 binding intensity across the locus.



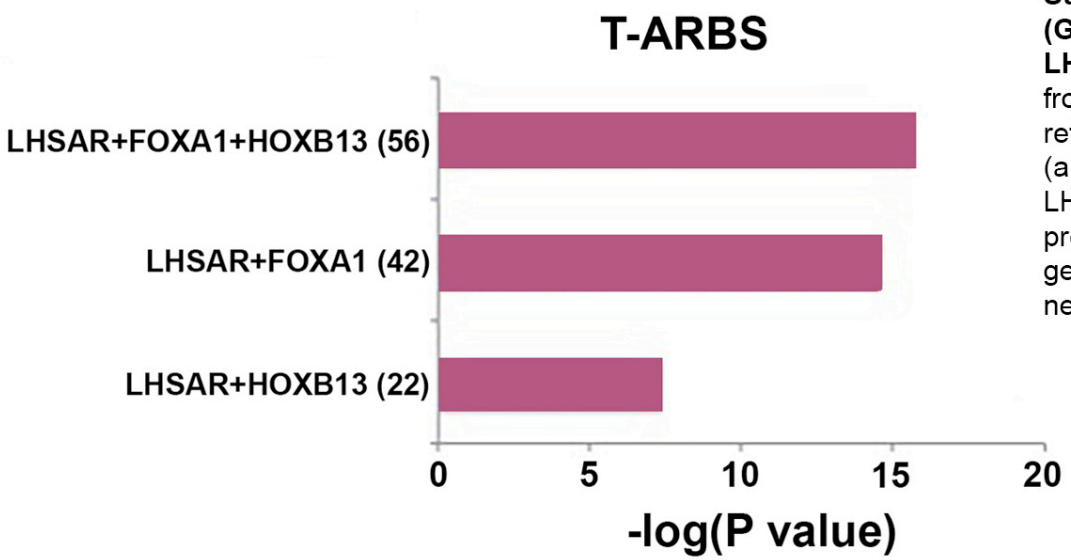
Supplementary Fig. 16. Transcription factor binding in human tissue and the LHSAR cell line across multiple conditions (described in colored boxes on the left) at a 20 kilobase region of the chromosome 7p13 locus that includes the gene SNGH15. Each track depicts AR, FOXA1 or HOXB13 binding intensity across the locus.

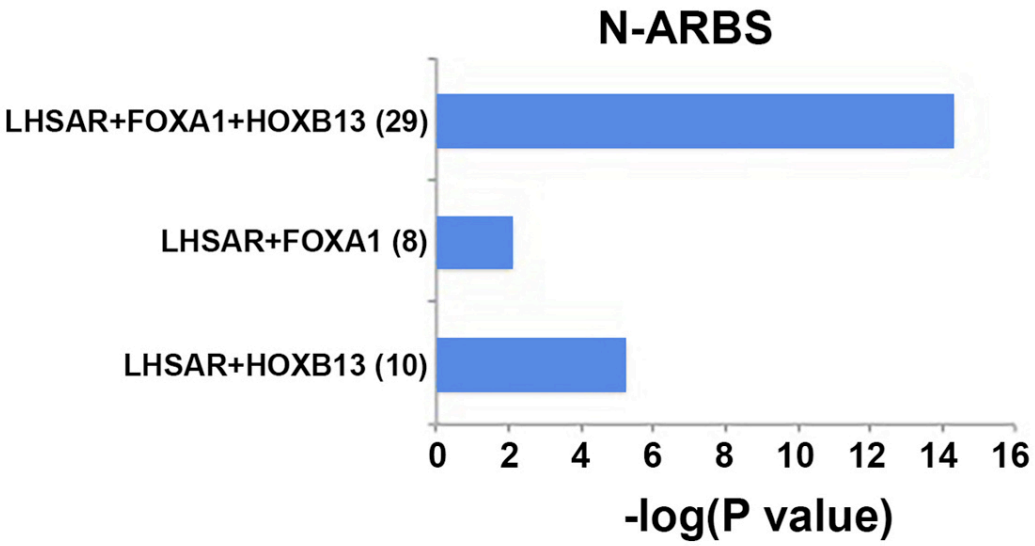


Supplementary Fig. 17. **AR occupancy at tissues-specific sites across various cistromes**. Each human sample from this study (round points), the LNCaP AR cistrome, and the LHSAR cell line AR cistrome after introduction of LacZ, HOXB13, FOXA1, or both HOXB13 and FOXA1 (square points) were plotted based on the percentage of N-ARBS and T-ARBS contained within their respective AR cistromes (all AR biding sites have FDR <0.01). To represent the regions on this plot most enriched in the primary human tumor and normal samples we calculated the high density regions for each sample type using the R-package hdrcde (version 3.1). The blue clouds in the graph represent 50% probability coverage. The dark gray clouds represent 95%. (Hyndman, R.J. (1996) Computing and graphing highest density regions, *American Statistician*, 50, 120-126.)

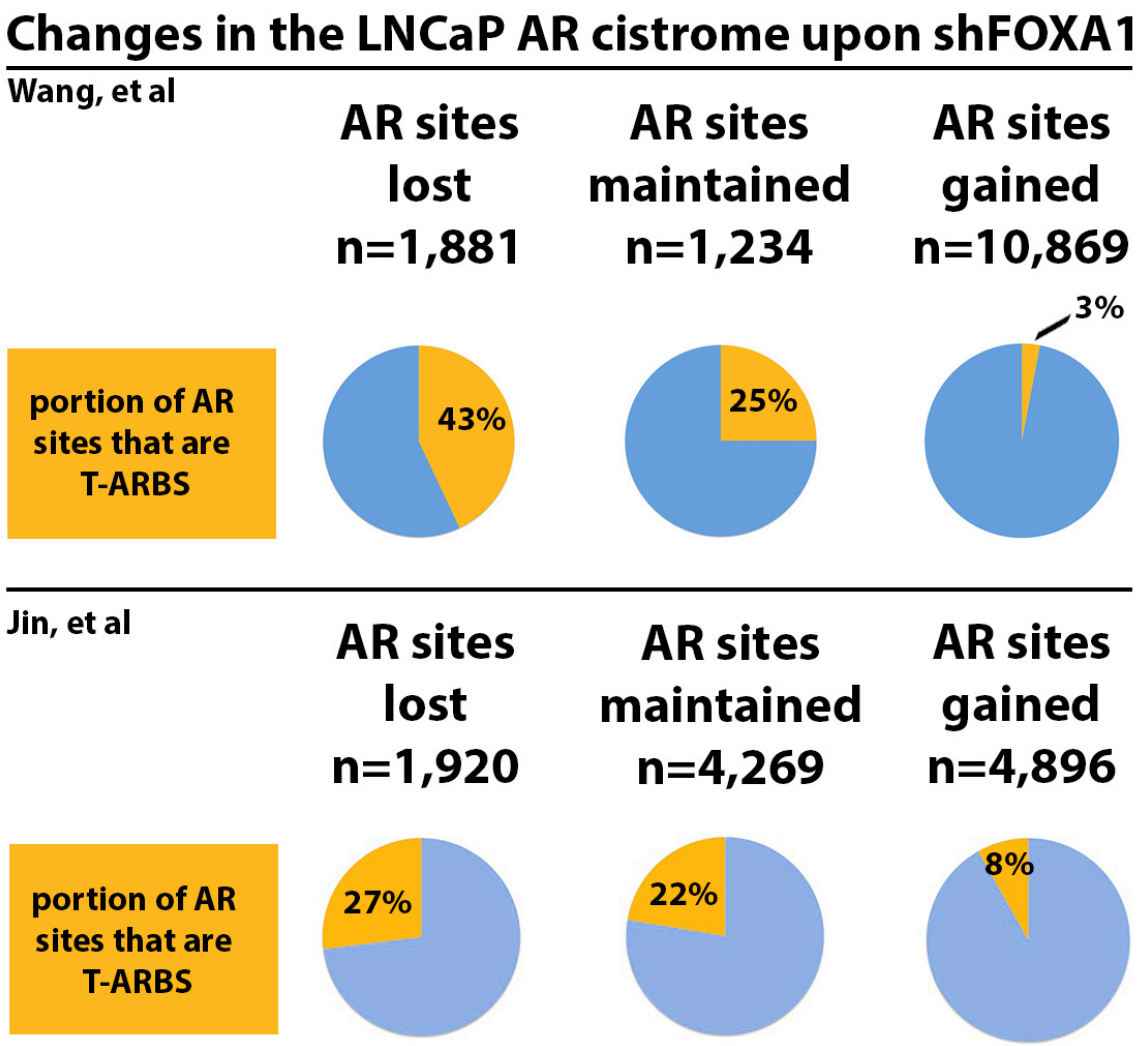


Supplementary Fig. 18. T**he LHSAR AR cistrome in the setting of increasing nuclear AR. A**. Western immunoblot for cytosolic (C) and nuclear (N) AR upon exposure to increasing amounts of androgen (R1881). Below, AR expression quantitation levels for each R1881 dose. **B**. AR ChIP-seq results at tissue-specific ARBS for the LHSAR cells exposed to varying R1881. Increased nuclear AR increases AR binding at N-ARBS, though not T-ARBS.

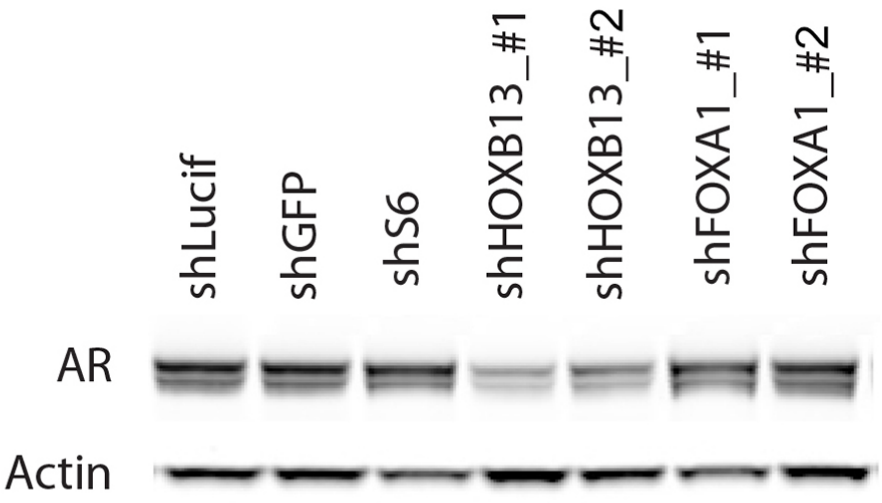




Supplementary Fig. 19. Gene set enrichment analysis (GSEA) of the transcription factor transduced LHSAR cell lines. The x-axis is the –log(P value) derived from a hypergeometric test. The numbers in parentheses refer to the number of genes in common between two lists: (a) the list of differentially expressed genes between the LHSAR modified lines and the LacZ control and (b) the previously defined list derived from the 324 up-regulated genes near T-ARBS or the 212 down-regulated genes near N-ARBS.



Supplementary Fig. 20. Androgen receptor binding in the LNCaP cell line upon shFOXA1 in two publicly available datasets. Shown are the numbers of AR sites lost, maintained or gained after FOXA1 knockdown22,23. The pie charts depict the proportion of AR sites within each category that are T-ARBS.



Supplementary Fig. 21. Knockdown of HOXB13 affects AR protein expression. Western immunoblots for LNCaP cell treated with shRNAs targeting Luciferase (shLucif) and GFP (shGFP) used as negative controls; ribosomal protein S6 (shS6), an essential gene for cell viability; HOXB13, using two different shHOXB13s; and FOXA1, using two different shFOXA1s. AR expression is depicted for each cell line condition, along with actin control.

Supplementary Table 1. Radical prostatectomy specimens analyzed by AR ChIP-seq

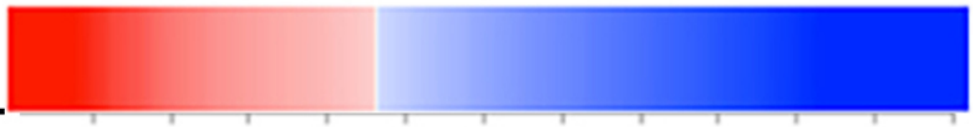
Subject	RP tissue analyzed	AR binding sites	Mappable loci	Year RP	Age at dx	PSA at dx	pT stage	RP Gleason score	Clinical f/u (yrs)	BCR?
DF1335	tumor and normal	23321(T) / 8336(N)	10347734/11548596	2006	63	7.2	T3a	3+4	0	unknown
DF1345	tumor and normal	16431(T) / 8842(N)	11767774/11726305	2006	62	unknown	T2c	3+3	6.8	no
DF1373	tumor and normal	24996(T) / 9049(N)	13699456/11671529	2006	53	6.6	T3a	4+5	7.3	yes
DF1412	normal only	9569	9376853	2006	59	4.1	T2c	3+3	5.1	no
DF1433	tumor and normal	6952(T) / 7776(N)	12510878/5701655	2005	51	2.6	T2c	3+3	5.0	no
DF1572	tumor only	44220	44490058	2006	60	16.3	T2c	3+3	2.6	no
DF1609	tumor and normal	20756(T) / 11160(N)	10764381/12581853	2009	53	14.1	T3b	4+4	4.1	yes
DF1754	tumor only	12714	9145753	2007	53	5.6	T3a	3+3	6.5	no
DF1782	tumor only	19256	9825325	2007	58	15.6	T2c	4+4	5.9	yes
DF184	tumor and normal	16124(T) / 20434(N)	12991150/15219395	unknown	unknown	unknown	unknown	3+3	0	unknown
DF1915	tumor only	51413	13637713	2008	41	4.9	T2c	3+3	2.9	no
DF1918	tumor only	10397	10844673	2008	67	5.7	T3b	4+3	5.6	yes
DF2089	tumor only	28431	37771661	2007	56	6.0	T2c	5+5	5.8	yes
DF2358	tumor only	30565	45494777	2008	64	6.0	T2c	3+3	4.5	no

RP, radical prostatectomy; dx, diagnosis; pT stage, pathologic tumor stage; f/u, follow-up; BCR, biochemical recurrence
Mappable loci: unique genomic locations identified via high-throughput sequencing

Supplementary Table 2. Genome-wide AR binding in Sharma et al. prostate specimens using the data analysis pipeline from the present study.

Patients	AR Peaks
ARM002_untreated	910
ARM003_Castrate-resistant	873
ARM009_untreated	526
NA_Treatment-responsive	597
NS012_Castrate-resistant	194
NS013_Treatment-responsive	607
NS018_untreated	720
NS019_Castrate-resistant	772
NS020_Castrate-resistant	4587
NS022_Castrate-resistant	739
Benign prostatic hyperplasia	<200

Supplementary Table 5. Number of T-ARBS and N-ARBS proximal to genes up- and down-regulated in tumor relative to normal tissue. Transcripts are rank ordered according to prostate tumor/normal differential expression.



Gene bins	1	2	3	4	5	6	7	8	9	10	11	12
No. of Genes	300	300	300	300	300	300	300	300	300	300	300	443
No. of T-ARBS	112	87	89	36	0	0	0	0	0	0	0	0
% T-ARBS	37.3	29.0	29.7	12.0	0	0	0	0	0	0	0	0
No. of N-ARBS	0	0	0	0	0	0	1	12	31	57	45	66
% N-ARBS	0	0	0	0	0	0	0.3	4	10.3	19.0	15.0	14.9

Supplementary Table 15. Four shRNAs used to determine dependency of prostate cancer cell lines on HOXB13.

shHOXB13_1	CCCGTGCCTTATGGTTACTTT
shHOXB13_2	CGCCAGATTACCATCTGGTTT
shHOXB13_3	GTTTGCCTTCTATCCGGGATA
shHOXB13_4	CTGTGGACAGTTACCAGTCTT
shHOXB13_GFP	GCAAGCTGACCCTGAAGTTCA