



**QUEEN'S  
UNIVERSITY  
BELFAST**

## **Genome-wide analysis of AR binding and comparison with transcript expression in primary human fetal prostate fibroblasts and cancer associated fibroblasts**

Nash, C., Boufaied, N., Mills, I. G., Franco, O. E., Hayward, S. W., & Thomson, A. A. (2017). Genome-wide analysis of AR binding and comparison with transcript expression in primary human fetal prostate fibroblasts and cancer associated fibroblasts. *Molecular and Cellular Endocrinology*. Advance online publication. <https://doi.org/10.1016/j.mce.2017.05.006>

**Published in:**  
Molecular and Cellular Endocrinology

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

### **Publisher rights**

© 2017 Elsevier. This manuscript is distributed under a Creative Commons Attribution-NonCommercial-NoDerivs License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

### **General rights**

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

### **Open Access**

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>



**Genome-wide analysis of AR binding and comparison with transcript expression in primary human fetal prostate fibroblasts and cancer associated fibroblasts**

Claire Nash<sup>1</sup>, Nadia Boufaied<sup>1</sup>, Ian G. Mills<sup>2</sup>, Omar E. Franco<sup>3</sup>, Simon W. Hayward<sup>3</sup> and Axel A. Thomson<sup>1\*</sup>

<sup>1</sup> Department of Surgery, Division of Urology, McGill University and the Cancer Research Program of the McGill University Health Centre Research Institute, Montreal, Quebec, H4A 3J1, Canada.

<sup>2</sup> Movember/Prostate Cancer UK Centre of Excellence for Prostate Cancer Research, Centre for Cancer Research and Cell Biology (CCRCCB), Queen's University Belfast, 97 Lisburn Road, Belfast, BT9 7AE, UK.

<sup>3</sup> Department of Surgery, NorthShore University HealthSystem Research Institute, 1001 University Place, Evanston, IL 60201, USA.

\* Corresponding author; [axel.thomson@mcgill.ca](mailto:axel.thomson@mcgill.ca)

Keywords;

Chromatin immunoprecipitation, ChIPseq, Androgen receptor, RNAseq, prostate, primary cells, cancer associated fibroblasts, mesenchyme, stroma, prostate cancer, prostate development.

## Abstract

The androgen receptor (AR) is a transcription factor, and key regulator of prostate development and cancer, which has discrete functions in stromal versus epithelial cells. AR expressed in mesenchyme is necessary and sufficient for prostate development while loss of stromal AR is predictive of prostate cancer progression. Many studies have characterized genome-wide binding of AR in prostate tumour cells but none have used **primary** mesenchyme or stroma.

We applied ChIPseq to identify genomic AR binding sites in primary human fetal prostate fibroblasts and patient derived cancer associated fibroblasts, as well as the WPMY1 cell line overexpressing AR. We identified AR binding sites that were specific to fetal prostate fibroblasts (7534), cancer fibroblasts (629), WPMY1-AR (2561) as well as those common among all (783). Primary fibroblasts had a distinct AR binding profile versus prostate cancer cell lines and tissue, and showed a localisation to gene promoter binding sites 1kb upstream of the transcriptional start site, as well as non-classical AR binding sequence motifs.

We used RNAseq to define transcribed genes associated with AR binding sites and derived cistromes for embryonic and cancer fibroblasts as well as a cistrome common to both. These were compared to several in vivo ChIPseq and transcript expression datasets; which identified subsets of AR targets that were expressed in vivo and regulated by androgens. This analysis enabled us to deconvolute stromal AR targets active in stroma within tumour samples.

Taken together, our data suggest that the AR shows significantly different genomic binding site locations in primary prostate fibroblasts compared to that observed in tumour cells. Validation of our AR binding site data with transcript expression in vitro and in vivo suggests that the AR target genes we have identified in primary fibroblasts may contribute to clinically significant and biologically important AR-regulated changes in prostate tissue.

## Introduction

The androgen receptor (AR) is known to be essential for masculinisation of sex-accessory reproductive organs such as the prostate, seminal vesicle and penis, as well as playing a central role in prostate cancer.

Paracrine signalling mechanisms active in the developing prostate are often dysregulated in prostate tumours [1-3]. Stromal-specific secreted molecules identified in embryonic rat and human prostate tissues have been shown to be re-expressed in cancer-associated fibroblasts (CAFs) and can modulate tumour growth [4, 5]. In addition, stromal markers have been shown to correlate with disease progression and may be predictors of aggressive prostate tumour subtypes [6].

During development, AR expressed in mesenchymal cells is both necessary and sufficient for androgen mediated reproductive development; AR expressed in epithelial cells is not required [7, 8]. In prostate cancer, AR functions in both the tumour cells and via CAFs. CAFs are known to play an important role in prostate tumour growth [9] and AR action within CAFs is able to contribute to hormone sensitive tumour growth [10]. In these experiments, AR was absent from the tumour cells, demonstrating that hormonal effects upon growth are likely to be mediated by AR in CAFs [10]. The function of AR has not been extensively examined in mesenchymal or stromal cells, despite their physiological importance, and our current knowledge of AR activity is derived from studies in prostate cell lines and patient tumour samples. Thus, the current AR cistrome reflects AR activity in cells that do not reflect androgen function in vivo during development or androgen action in fibroblasts present in prostate cancer stroma.

Recent advances in ChIP sequencing (ChIPseq) and gene profiling technologies have enabled the characterization of genome-wide AR binding and the AR transcriptional network (reviewed in [11]). The majority of these studies have been conducted on prostate cancer cell lines and considerable effort has been made in defining AR targets in tumour epithelium [12-17]. This has revealed a transcriptional role for AR in tumour cells regulating cell proliferation, metabolism, survival and DNA repair [18, 19]. Prostate cancer cell lines have also been used to determine the changes in AR transcriptional networks in both a hormone-naïve- and castrate-resistant setting and numerous studies have revealed distinct binding profiles in the androgen-independent setting [13, 17, 18, 20, 21]. More recently, the genomic binding profile of AR variants have also been explored and have been shown to compensate for full length AR in an endocrine therapy-like setting [22]. These studies have provided valuable insights in to the molecular biology and function of the AR and have identified several co-activators and co-repressors that modulate AR transcriptional activity [23-28].

It has been shown that the genome-wide AR binding profiles identified in prostate cancer cell line models are not reflective of in vivo tissues and that AR target

gene signatures identified in prostate cancer tissues are not represented in prostate cancer cell line models [29]. This has highlighted the importance of studying AR genomic binding in primary cells or tissues. Therefore, more recently, the genomic behavior of AR has been investigated in prostate cancer tissues and has revealed that AR genomic binding and interactions are reprogrammed with different tumour stages and have prognostic value [30-33]. However, tissue samples typically contain multiple distinct cellular components and the majority of studies to date have neglected the role of the AR in CAFs present in the tumour microenvironment. It has been shown that human prostate myofibroblast cell lines have a different AR binding profile to prostate cancer epithelium and are not dependent on the classic AR pioneer factors such as FOXA1 [34]. The role of AR in stroma is likely to be important, since stromal expression of AR is prognostic [35, 36] and the genes regulated by AR in stroma may themselves be prognostic and control cell proliferation and differentiation. However, no study has addressed the genomic behavior of AR in primary prostate developmental or cancer-associated fibroblasts. In addition, the majority of studies to date investigating the AR transcriptional network have been limited by genome coverage on microarrays and thus may not adequately capture the full landscape of AR binding and transcriptional regulation in cells and tissues.

We have applied ChIPseq and RNA sequencing of the transcriptome (RNAseq) to provide the first AR cistrome in primary prostate embryonic and cancer-associated fibroblasts. In fibroblasts, the AR genomic distribution was distinct from prostate cancer cell lines, with a strong enrichment at proximal promoter regions of genes and with preference for non-classical AR binding motifs. Genes associated with these binding sites correlated well with the fibroblast transcriptomes and allowed us to define cancer-fibroblast specific and embryonic prostate fibroblast specific AR cistromes. The AR cistromes showed a high degree of overlap with transcripts expressed in embryonic prostate tissue and prostate cancer stroma. Overall, our study provides the first insight in to the AR genomic distribution in both primary developmental and primary cancer-associated prostate fibroblasts.

## Materials and Methods

### *Cell culture*

Immortalized human prostatic stromal cells overexpressing wildtype AR (WPMY1-AR cells) [37] were a gift from Drs M. Grabowska and R. Matusik (Vanderbilt University, TN, USA). Primary embryonic prostate fibroblasts (EPF) were derived from 12-16 week human embryonic tissue as per [38] in accordance with UK guidelines and following approval by the Lothian Research Ethics Committee. Primary prostate cancer-associated fibroblasts (CAF) were isolated from prostate specimens of patients undergoing radical prostatectomy and bioassayed in an in-vivo tissue recombination model to determine their pro-tumorigenic potential as previously described [9]. Primary fibroblast cells were assessed for the expression of known fibroblast markers from RNAseq data to confirm fibroblast identity (Supplementary Table 1). All cells were maintained in DMEM supplemented with 10% FBS and used at passages between 5 and 7 for all ChIPseq assays. All primary cells were used under ethical approval 15-631-MUHC.

For testosterone treatment, WPMY1-AR cells ( $1 \times 10^6$ ) were plated in 10cm<sup>2</sup> plates and grown overnight. Cells were washed 3 x PBS and medium was replaced with DMEM supplemented with 10% charcoal-stripped fetal bovine serum  $\pm$  10nM, 100nM testosterone (Sigma-Aldrich, Missouri, USA) or ethanol as vehicle control, and cells were cultured for 72 hours.

### *Western blotting*

This was performed using Mini-PROTEAN® TGX Stain-Free™ Precast Gels (4-15%) and proteins transferred to 0.2µm polyvinylidene difluoride membrane using the Trans-Blot® Turbo™ Blotting system (Bio-Rad Laboratories, Inc. California, USA). Membranes were incubated (4°C overnight) with AR H280 IgG (Rabbit Polyclonal, Santa Cruz Biotechnology, Santa Cruz, USA; dilution 1:1000), ZFX IgG (Rabbit Polyclonal, SAB3501008 Sigma-Aldrich, Missouri, USA; dilution 1:800) or  $\beta$ -actin IgG (Clone AC-15, Sigma-Aldrich, Missouri, USA; dilution 1:10000) and HRP conjugated secondary antibodies (Jackson ImmunoResearch Laboratories, Inc.; dilution 1:10000; 1 hr) before visualisation with Clarity™ Western ECL Substrate (Bio-Rad Laboratories, Inc. California, USA).

### *Immunofluorescence staining*

Cells in either full serum medium or cells previously serum starved for 48 hours were seeded into sterile multi-chamber slides (Sarstedt AG & Co, Germany) and allowed to recover overnight so that 60% confluency was reached. Cells were treated with 10nM testosterone in ethanol or ethanol (control) in full serum or serum-free media and incubated for 24 hours. Cells were rinsed in PBS and fixed/permeabilised for 20 mins in 1:1 ethanol/acetone solution at -20°C. Fixed cells were blocked with 1% bovine serum albumin before incubation with AR H280 diluted 1:200 in blocking solution for 1 hour at room temperature. Primary

antibody was omitted to serve as a negative control. Goat anti-rabbit IgG conjugated to Texas Red®-X fluorescent dye (Invitrogen, ThermoFisher Scientific, Massachusetts, USA) was applied and incubated in the dark for 40 mins. Slides were mounted using VECTASHIELD hard-set mountant containing 4',6-diamidino-2-phenylindole (DAPI) (Vector Laboratories Inc. California, USA) and visualized with a Zeiss LSM 780 laser scanning confocal microscope.

*Chromatin Immunoprecipitation and sequencing library preparation (Illumina)*  
ChIP sequencing was carried out by Active Motif (Carlsbad, CA, USA) as follows. Cells were fixed with 1% formaldehyde for 15 min and quenched with 0.125 M glycine. Chromatin was isolated by the addition of lysis buffer, followed by disruption with a Dounce homogenizer. Lysates were sonicated and the DNA sheared to an average length of 300-500 bp. Genomic DNA (Input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by ethanol precipitation. Pellets were resuspended and the resulting DNA was quantified on a NanoDrop spectrophotometer. Extrapolation to the original chromatin volume allowed quantitation of the total chromatin yield. An aliquot of chromatin (30 µg) was precleared with protein A agarose beads (Invitrogen, ThermoFisher Scientific, Massachusetts, USA). Genomic DNA regions of interest were isolated using 4 µg of antibody against ARH280. Complexes were washed, eluted from the beads with SDS buffer, and subjected to RNase and proteinase K treatment. Crosslinks were reversed by incubation overnight at 65 C, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation.

Quantitative real-time PCR reactions were carried out in triplicate on specific genomic regions using SYBR Green Supermix (Bio-Rad Laboratories, Inc. California, USA). The resulting signals were normalized against Input DNA for primer efficiency according to the Active Motif ChIP-IT® qPCR analysis kit protocol.

Illumina sequencing libraries were prepared from the ChIP and Input DNAs by the standard consecutive enzymatic steps of end-polishing, dA-addition, and adaptor ligation. After a final PCR amplification step, the resulting DNA libraries were quantified and sequenced on Illumina's NextSeq 500 (75 nt reads, single end).

#### *ChIP sequence read and peak analysis*

Reads were aligned to the human genome (hg19) using the Bowtie2 algorithm (default settings) [39]. Duplicate reads were removed and only uniquely mapped reads (mapping quality  $\geq 25$ ) were used for further analysis. AR peak locations were determined by comparing with input samples using the MACS v2.0.10 algorithm [40] with a cutoff of p-value =  $1e-7$  and the HOMER findPeaks algorithm (default settings). Only peaks that were co-identified by both algorithms were used for further analysis. Overlap, feature annotation, peak characterization and enriched motif analysis (*de novo*) were performed using the HOMER suite of algorithms [41]. Known motif analysis was also performed using MEME [42] with the JASPAR 2016 database [43]. The deepTools suite of algorithms [44] were



used to determine sample similarity by hierarchical clustering and Spearman correlation and for ChIPseq read enrichment analysis around specific peak sites. All size proportional Venn diagrams were created using BioVenn [45]. **For prostate cancer in vitro cell line (PCCL) datasets, only cell lines which contained androgens in their culture media were selected for analysis.** PCCL and prostate cancer *in vivo* datasets of AR binding sites were generated by combining AR peak regions (if reads were aligned to hg19 and peaks called with MACS2 or HOMER) or by re-analyzing raw ChIP-seq read sequences as above with publicly available ChIP sequencing datasets detailed in Supplementary Tables 2 and 11. Visualization of ChIPseq read coverage was performed using IGB software [46].

#### *RNA extraction and quantitative real-time PCR*

Total RNA was extracted using Qiazol followed by the RNeasy™ Mini kit (Qiagen, Venlo, Netherlands) following manufacturers instructions. Complementary DNA synthesis was performed using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems- ThermoFisher Scientific, Massachusetts, USA) and qPCR was performed on an ABI 7500 Fast machine using SYBR Select Mastermix (ThermoFisher Scientific, Massachusetts, USA). Transcript abundance was normalized to two housekeeping genes; RPLP0 and EIF2B1. Primers used are provided in Supplementary Table 3.

#### *Immunohistochemistry*

Immunostaining of AR and ZFX on serial sections of male rat P0 urogenital tract tissue (isolated as per [38]) was performed using AR H280 IgG (Rabbit Polyclonal, Santa Cruz Biotechnology, Santa Cruz, USA; dilution 1:800) and ZFX IgG (Rabbit Polyclonal, HPA003877 Sigma-Aldrich, Missouri, USA; dilution 1:1500). Primary antibody was omitted to serve as a negative control. Tissue was de-paraffinised in xylene and hydrated in ethanol and rinse under water. Antigen retrieval was performed by pressure cooking in 0.2M citrate buffer, pH 6.0 (Poly Scientific R&D Corp, Bay Shore, NY, USA) for 10 min. Endogenous peroxidase activity was blocked for 30 min with 3% H<sub>2</sub>O<sub>2</sub>. After a 3 h block with IHC Select (Millipore, Billeria, MA, USA), the primary antibody was used overnight at 4°C. Histofine Simple Stain MAX PO multi (Nichirei BioSciences, Inc., Tokyo, Japan) was used as secondary antibody and chromogenic detection was carried out using Simple Stain AEC Solution (Nichirei BioSciences, Inc.). Images were taken with an Aperio Slide Scanner (Leica, Wetzlar, Germany).

#### *RNA sequencing library preparation*

RNA sequencing was carried out by Exiqon, Inc (Denmark) as follows. Libraries were prepared using a TruSeq® stranded total RNA sample preparation kit with rRNA depletion (Illumina, Inc, San Diego, USA).

100 ng of total RNA was used for library construction and rRNA was depleted using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal. The isolated mRNA was subsequently fragmented using enzymatic fragmentation. Then first strand synthesis and second strand synthesis were performed and the double stranded cDNA was purified (AMPure XP, Beckman Coulter). The cDNA was end repaired, 3' adenylated and Illumina sequencing adaptors ligated onto the fragments ends, and the library was purified (AMPure XP). The mRNA stranded libraries were pre-amplified with PCR and purified (AMPure XP). The libraries size distribution was validated and quality inspected on a Bioanalyzer high sensitivity DNA chip (Agilent Technologies). High quality libraries were quantified using qPCR, the concentration normalized and the samples pooled according to the project specification (number of reads). The library pool(s) were re-quantified with qPCR and optimal concentration of the library pool used to generate the clusters on the surface of a flowcell before sequencing on Nextseq500/ High Output sequencing kit (51 cycles according to the manufacturer instructions (Illumina Inc.).

#### *RNA sequence read analysis*

Sequencing reads were aligned to the human genome (hg19) using the Tophat 2.1.0 algorithm [47] and only uniquely mapped, non-redundant reads were used for further analysis. Read counts were quantified with the featureCounts algorithm [48] and reads with a transcripts per million (TPM) frequency greater than ten were retained.

#### *Tag sequence read analysis*

Tag sequencing expression data from human embryonic prostate tissue from a previously published study from our group were re-analysed [5]. Briefly, tag sequences were aligned to the human genome (hg19) using the Bowtie2 algorithm (default settings) and only uniquely mapped, non-redundant reads were used for further analysis. Read counts were quantified with the featureCounts algorithm [48] and reads with a reads per million (RPM) frequency greater than six were retained.

#### *Microarray analysis*

Microarray raw data for three publically available datasets (GSE20758, GSE8218 and GSE17951) were uploaded from GEO (NCBI) [49] and normalized using robust multi-array Averaging (RMA) [50] and batch corrected using Combat [51]. Differentially expressed genes between epithelial and stromal samples were identified using *limma* package [52, 53]. Only genes with adjusted p value < 0.05 were used for comparison to cistromes. GSE20758 contains microdissected stromal and epithelial tissue from cancer tissue. With dataset GSE8218 and GSE17951, the authors provide cell-type composition estimated by pathologists.

For dataset GSE8218, samples from cancer patients with stromal content greater than 90% have been compared to samples with tumour content greater than 70%. For data set GSE17951, samples from cancer patients with tumour content equal to 0 were compared to samples from cancer patients with tumour content higher than 65%. Expression matrix (normalised and log transformed data) for dataset GSE70758 and GSE72920 as provided by the authors in GEO were used to identify differentially expressed transcripts in cancer tissues or normal tissues treated with a GnRH antagonist (Degarelix) vs non-treated tissues. All bioinformatics and statistical analysis were conducted in R and Bioconductor suite (<http://www.r-project.org>) [54].

## Results

To establish the AR protein levels present within the primary cells used for ChIPseq studies, we performed western blotting with the AR H280 antibody (Figure 1A). This showed that AR levels in EPF and CAF were lower than tissue *in vivo* (rat prostate and epididymis) and over 100 fold lower than AR levels observed in LNCaP and WPMY1-AR cells. Immunofluorescence staining showed heterogeneous expression of AR in WPMY1-AR and CAFs and a shift from cytoplasmic to nuclear distribution in response to testosterone (T) in serum-free medium in CAFs (Supplementary Figure 1) using AR H280. Due to the response to T coupled with the fact we did not observe any isoforms or truncations of the AR on our blots using AR H280, we subsequently used this antibody for ChIP analysis of AR. ChIP was performed using EPF, CAF and WPMY1-AR cells, followed by qPCR analysis of selected AR targets to verify the success of the immunoprecipitation prior to full sequencing of the libraries. We saw enrichment of AR binding events associated with a known AR target gene ATAD2 [55] and the ARL8B gene in EPF and CAF cells with a ~8 fold and ~30-45 fold range of enrichment respectively. Enrichment of AR binding events associated with the known AR target gene ZBTB16 [28, 56] was also seen in WPMY1-AR cells with a 16 fold enrichment (Supplementary Figure 2). This suggested that, despite low AR levels in EPF and CAF, we were able to detect AR binding events to chromatin in fibroblasts. We were unable to use established AR target genes such as PSA since these were derived in epithelial cell types, and are not expressed in fibroblasts.

We observed differences in the numbers of peaks defined following ChIPseq and peak calling with both MACS (2.1.0) and HOMER; WPMY1-AR, EPF and CAF yielded 5495, 11120, and 2311 peaks, respectively (Figure 1B, Supplementary Table 4) and demonstrated enrichment at ATAD2 and ZBTB16 control gene regions (Supplementary Figure 3). WPMY1-AR cells were included as a positive control, since we were unsure that our primary cells expressed sufficient AR to yield reliable data. The peak number observed following ChIPseq showed no relationship with the level of AR expression within the samples used. Indeed, the numbers of peaks observed in EPF were double that observed in WPMY1-AR cells – yet EPF expressed approximately 100 fold less AR protein than WPMY1-AR (Figure 1A).

When peaks were mapped to the genome and annotated with genomic features, we observed that an unusually high proportion were associated with proximal gene promoter regions; defined as within +100bp and -1Kb of the transcriptional start site (TSS) (Figure 1C). EPF, CAF and WPMY1-AR showed 37% (4061), 41% (948), and 33% (1834) of total peaks associated with promoters. Given that the size defined as the proximal promoter was small (1.1Kb) such a high frequency of binding sites seems unlikely to occur by chance. Comparison of EPF, CAF and WPMY1-AR peaks defined 783 common to all 3 cell types (Figure 1D). Lists of EPF, CAF and WPMY1-AR peaks along with those that are specific

or intersect are provided in the Supplementary data file. We selected candidate AR target genes that were associated with an AR peak upstream from the transcriptional start site (TSS) in at least 2 cell types. We assessed if these (along with AQP3, a known stromal androgen-regulated transcript) were androgen-regulated by quantifying their expression by qPCR following treatment with T in WPMY1-AR as per [37] (Supplementary Figure 4). We saw that T induced a transcriptional response in all 5 candidates suggesting our ChIPseq datasets contain target genes that are androgen-regulated.

Since WPMY1-AR cells were both immortalised and expressed supra-physiological levels of AR, we chose to focus upon data derived from primary EPF and CAF. **Using correlational analysis, we compared normalized ChIPseq read counts in EPF and CAF to each other, as well as to a dataset comprised of several publicly available AR ChIPseq binding profiles from a panel of prostate cancer cell lines (PCCLs) stimulated with androgens (summarised in Supplementary Table 2).** Our goal was to define whether the ChIPseq peaks in our two primary fibroblasts showed a conserved fibroblast-specific pattern that was distinct from that in PCCLs. Figure 2A shows a higher correlation between EPF and CAF than either of these versus PCCLs. The mean ChIPseq read coverage of peaks common to PCCLs at EPF and CAF peaks was low (Figure 2B). Similarly, the mean ChIPseq read coverage of EPF and CAF was low at PCCL peaks (Supplementary Figure 5). This indicates that tumour cell AR exhibits a distinct binding profile from that of AR in fibroblasts. The genomic locations of peaks within the PCCL dataset are shown in Figure 2C. In PCCL there was a low (2.2%) proportion of peaks in proximal promoter regions, in contrast to the high frequency of proximal promoter peaks (30-40%) observed in fibroblasts (Figure 2C). To illustrate enrichment of ChIPseq reads relative to the transcriptional start site (TSS), read density per peak was plotted for EPF, CAF and PCCL datasets, and showed a 2-3 fold enrichment in both EPF and CAF relative to PCCL (Figure 2D). We compared peaks identified in EPF, CAF and PCCL to define those peaks common to fibroblasts (EPF and CAF) as well as specific to fibroblasts (those observed in both EPF and CAF but absent from PCCL). In general, there was a very small overlap between the fibroblasts and PCCL (131, 56, and 18 peaks among datasets of 11120, 37156 and 2311 peaks). We observed 1467 peaks that were specific to fibroblasts (and absent in PCCL) and common between EPF and CAF (Figure 2E). Lists of specific and intersecting EPF, CAF and PCCL peaks are provided in the Supplementary data file.

Having identified significant numbers of peaks in fibroblasts, we next examined whether these might be associated with actively transcribed genes. We performed RNAseq on the same samples used for our ChIPseq studies, to measure mRNA expression at genome-wide resolution. Comparison of the peaks in EPF and CAF (following removal of those common to PCCLs) identified 9466 EPF-specific peaks, 770 CAF-specific peaks and 1467 peaks shared between EPF and CAF (Figure 3A). Examples of peaks specific to EPF or CAF, and common to both (Fibroblast), are shown in Figure 3B. Construction and sequencing of libraries of EPF and CAF identified 8334 and 6177 mRNAs with a read count of > 10 transcripts per million (>10 TPM) in EPF and CAF respectively (Figure 3C). To derive cistromes for EPF, CAF and fibroblasts, we selected peaks mapped within a 100Kb genomic window 5' to the TSS, and compared these to transcript expression defined by RNAseq. For CAF, 770 CAF-specific peaks yielded 308 peaks within 100Kb of the TSS, while 9466 EPF-specific peaks yielded 3580 promoter peaks, and the 1467 peaks common to fibroblasts yielded 565 promoter peaks. These were then compared with the RNAseq defined transcriptomes to derive the EPF, CAF and fibroblast AR cistromes. This identified 1560 genes in EPF, 112 genes in CAF, and 208 genes in fibroblasts with an associated promoter AR peak (Figure 3D). Gene ontology analysis of the cistromes are presented in Supplementary Tables 5-10 and show enrichment of ubiquitin-transferase activity, RNA binding and glucocorticoid activity in EPF-specific, CAF-specific and fibroblast cistromes respectively. Lists of EPF-specific, CAF-specific and fibroblast peaks and cistromes along with lists of transcripts > 10 TPM from RNAseq are provided in the Supplementary data file.

We evaluated the relative position of peaks to the TSS to examine whether there was non-random distribution of peaks in promoters (100Kb 5' to TSS), and observed a striking localisation to sequences immediately adjacent to the TSS. Figure 4A shows the distribution of peaks across 100Kb located 5' of the TSS, divided into a series of 20Kb subregions. We included two subregions that were not 20Kb in size; one was a 1Kb region directly adjacent to the TSS (proximal promoter) and the other was the region furthest from the TSS which was defined as 100Kb and greater. Graphs of peaks validated by transcript expression are shown in Figure 4B. We plotted both peaks as well as those validated by transcript expression, in order to observe possible artefacts associated with gene expression or mapping. We determined that the number of peaks that are associated with expressed transcripts is not dependent on the distance 5' to the TSS (Supplementary Figure 6) since peak validation by transcript data was approximately equal in all regions. We observed that there was a very high association of peaks in the proximal promoter regions of EPF, CAF and fibroblasts. A similar pattern of proximal promoter peaks was observed in WPMY1-AR cells, though this fibroblast cell line expresses 100 fold more AR than primary EPF or CAF, and we suggest that AR level does not influence

promoter binding location. The proximal position of peaks observed in fibroblasts was in marked contrast to PCCL where few peaks were observed in the proximal promoter region, but which showed a greater number of peaks in the 100Kb+ group. A dataset comprised of several publicly available AR ChIPseq binding profiles from a panel of prostate cancer tissues (PCa in vivo, summarised in Supplementary Table 11) showed a distribution of peaks similar to PCCL, though with some increase in proximal promoter peaks (Figure 4A). We speculate that these proximal promoter peaks may arise from the stromal component present within the tissue used for ChIP analysis. Enrichment of peaks in the proximal promoter region of genes was conserved in our transcript valid peaks with 37.5%, 28.3% and 34.2% of 5' peaks in EPF, CAF and Fibroblasts respectively (Figure 4B). To investigate whether AR bound at proximal promoter sites might be involved in transcriptional activation or repression, we examined the transcript levels for genes with AR peaks in EPF and CAF. This illustrated the distribution of transcript levels and showed that a high proportion of AR sites were associated with low transcript levels, and we also observed a few AR sites associated with abundantly expressed transcripts (Supplementary Figure 7). This association is of value in distinguishing the general balance between activation and repression, however it does not provide definitive proof of regulation.

Next we addressed whether our AR targets are represented in other published datasets. First we compared to in vitro ChIPseq and gene expression microarray data of an immortalised prostate fibroblast cell line that overexpresses AR (PShTERT-AR) [34]. We observed little overlap between our EPF, CAF and WPMY1-AR fibroblast AR peaks with PShTERT-AR peaks. Less than 10% of our EPF and CAF promoter-bound cistromes were co-identified in the PShTERT-AR cistrome (Supplementary Figure 8).

We then evaluated whether EPF-specific, CAF-specific and fibroblast-specific peaks and cistromes were physiologically relevant by comparing to in vivo tissue datasets. Comparisons of EPF-specific, CAF-specific and fibroblast peaks to PCa in vivo AR ChIPseq peaks (Supplementary Table 10) co-identified 2021 (21%), 171 (22%) and 452 (30%) peaks with PCa in vivo tissue respectively (Figure 5A). We examined whether the primary fibroblast cistromes were associated with actively transcribed genes expressed in human tissue in vivo. We first compared EPF-specific, CAF-specific and fibroblast cistromes to the genome-wide transcriptome of human embryonic prostate tissue (EMB transcriptome, Figure 5B,[5]). This showed a high degree of overlap (83-86%) with EMB transcripts suggesting our primary fibroblast cistromes are well conserved in developing human prostate in vivo. We then compared our cistromes to prostate cancer patient transcripts, and particularly datasets in which the stromal proportion of samples were defined. We generated three datasets of stromal-specific or stromal-enriched transcripts that were significantly upregulated in prostate cancer tissue microarrays compared with tumour epithelium. The first dataset consisted of laser-capture microdissected prostate cancer stroma vs epithelium yielding 876 stroma-specific transcripts (GSE20758 [57]). Following removal of transcripts from our cistromes that were not present

on the microarrays, comparison to our EPF-specific, CAF-specific and fibroblast cistromes identified 111 (12.9%), 4 (7.8%) and 7 (6.1%) transcripts respectively which were stromal-specific in prostate cancer in vivo (Figure 5C). The second and third datasets consisted of microarray expression data of prostate cancer tissue with known stromal and epithelial content; defined by pathologist. Stromal-enriched transcripts were determined by comparing samples with high stromal-content and low epithelial content versus samples with high epithelial content and low stromal content. Transcripts that were significantly upregulated in high stroma samples were considered stromal-enriched. These yielded 557 and 5653 stromal-enriched transcripts (GSE8218 and GSE17951 [58]). Comparison of these with our cistromes (following removal of transcripts not present on the arrays) yielded 61 (4.8%) and 477 (32.1%) EPF-specific, 6 (7.0%) and 32 (31.4%) CAF-specific and 7 (4.0%) and 59 (30.1%) fibroblast transcripts which are stromal-enriched in prostate cancer in vivo (Figure 5C). Lists of all transcripts that intersected with EPF-specific, CAF-specific and fibroblast cistromes are provided in the Supplementary Data file.

In order to determine whether our cistromes represented androgen-regulated targets in vivo, we compiled a list of differentially regulated transcripts in response to testosterone in an embryonic mouse model of androgen treated urogenital sinus [3]. Comparisons of EPF-specific, CAF-specific and fibroblast cistromes to these transcripts co-identified 188 (12.1%), 14 (12.5%) and 20 (9.6%) with androgen-regulated transcripts respectively (Figure 6A). Next we compiled a list of differentially expressed transcripts between non-treated and androgen-deprivation treated (ADT) human prostate cancer samples and benign adjacent to prostate cancer tissue samples in order to define AR regulated transcripts in vivo (GSE72920 and GSE70768 [59]). Comparisons of EPF-specific, CAF-specific and fibroblast cistromes to these transcripts showed a larger overlap with ADT treated benign adjacent tissue transcripts (11-13%) than prostate cancer tissue transcripts (5-6%) (Figure 6B). Lists of all transcripts that intersected with EPF-specific, CAF-specific and fibroblast cistromes are provided in the Supplementary Data file.

Overall, comparison of our cistrome data with transcript expression data derived from tissues verified that a high proportion of our experimentally defined AR targets were genes expressed in vivo; especially in regard to prostate development and also in regard to prostate tumour stroma.

Having identified that AR has a different binding distribution across the genome compared to PCCLs, we then addressed whether the DNA sequences which stromal AR binds to also differ. We performed de novo motif analysis on EPF, CAF and WPMY1-AR peaks using HOMER. We identified DNA sequence motifs that were significantly enriched in EPF and WPMY1-AR and we observed that these differed from the classical androgen-response element (AREs) reported for epithelial cells (Figure 7A). HOMER provided a series of transcription factors with a DNA binding motif most similar to the enriched motif. For EPF and CAF, to



validate the identified transcription factor motifs, we chose the best matching transcription factor with an expression read count of > 10 transcripts per million RNAseq data. For WPMY1-AR, we chose the transcription factor with the highest match score from HOMER. We identified that ZFX was the most probable transcription factor associated with EPF peaks, SWI5 as the most probable transcription factor in CAF and ATF3 as the most probable transcription factor in WPMY1-AR (Figure 7B). Motif analysis was also performed using the MEME algorithm [42] which identified REST as the most probable transcription factor for both EPF and CAF which reached statistical significance in EPF only (Supplementary Table 12). We note that the MEME algorithm does not allow analysis of the whole peak dataset for each sample and only a 60bp fraction of each peak is considered in the analysis that could account for differences between MEME and HOMER.

We compared the number of ARE positive peaks in our stromal cells vs PCCLs and PCa in vivo tissues. We observed ~100 fold less ARE positive peaks in our primary stromal cells and ~10 fold less ARE positive peaks in our WPMY1-AR than PCCLs. PCa in vivo tissues contain ~4 fold less ARE positive peaks than PCCLs (Figure 7C).

In order to address if ZFX was a potential AR co-factor in developing prostate mesenchyme in vivo, we performed immunohistochemistry for AR and ZFX on serial sections of male P0 rat urogenital tract tissue (Figure 7D). We observed that AR expression is nuclear and largely restricted to mesenchymal cells of the developing ventral prostate (VPM) with little or no expression in the epithelial cells of the ventral prostatic buds (VPE). ZFX was detected in both the VPM and VPE but we observed more nuclear expression in the AR positive VPM and more cytoplasmic staining in the AR low/negative VPE. Western blot analysis of ZFX also confirmed expression in our in vitro stromal cells as well as LNCaP and PC3 PCCLs and suggested a mix of nuclear and cytoplasmic expression (determined as per [60]) that is not influenced by the levels of AR expression (Supplementary Figure 9).

Overall, we have determined that primary prostate stromal cells have a different AR binding profile to prostate cancer epithelial cells both in genomic distribution and in specific DNA binding sequences. This suggests that stromal AR may co-operate with different co-factors, with ZFX being a potential AR co-factor in developing prostate.

## Discussion

We have performed a genome wide analysis of AR binding in primary prostate embryonic fibroblasts and cancer associated fibroblasts, and derived AR cistromes using RNAseq-based transcript expression data. Fibroblasts and stroma play key roles in regulating prostate development and cancer, yet the role of AR in these cells has received little investigation, despite their physiological importance.

We observed that AR showed a unique genomic distribution in regard to the genes and promoter locations occupied in fibroblasts compared to tumour cells. This has been shown in a previous study using the PShTERT-AR myofibroblast cell line [34]. There was little overlap between our AR peaks and those of PShTERT-AR which could be attributed to differences between primary and immortalized cell cultures and differences in experimental approaches. In particular, the proportion of peaks located close to the TSS was very high in our fibroblasts and consisted of approximately 33-40% of all peaks observed. The high proportion of promoter bound sites in primary fibroblasts was in marked contrast to the location in PCCL or prostate cancer in vivo where approximately 2% of AR sites were located in gene promoters. The fibroblast promoter peaks that we observed were located within 1kb of the TSS, a very small and well defined genomic region, and we suggest that this is unlikely to arise by chance or random distribution. The location of the peaks in close proximity to the TSS suggests that we can be confident of the involvement of AR in the regulation of these genes, in contrast to distally located 'enhancer' type binding sites that may be over 100kb from the TSS which are difficult to assign to the regulation of any particular gene. AR binding sites within 1kb of the TSS are likely to play a role in gene expression, and we correlated these binding sites with transcript expression levels (Supplementary Figure 7). This identified genes with proximal promoter AR binding sites and high or low mRNA levels; enabling a general comparison of activation versus repression. We observed that there were many AR sites associated with low transcript levels and a few with high transcript levels in both EPF and CAF. Although we provide some evidence of androgen regulation of some of our candidate AR target genes in WPMY1-AR cells (Supplementary Figure 4), and through comparisons to published data for androgen-stimulated PShTERT-AR cells (Supplementary Figure 8, [34]), we are yet to experimentally address transcript response to androgens of our target genes in EPF and CAF cells and to determine if our genes with proximal promoter AR binding sites are activated or repressed by AR.

The AR protein level in primary fibroblasts was substantially lower than in LNCaP or WPMY1-AR cells as well as prostate tissue, and also showed heterogeneity among cells in culture (Figure 1 and Supplementary Figure 1). At the inception of this work, we were concerned that such low AR levels might not yield sufficient peak numbers, which led us to include WPMY1-AR cells engineered to express high levels of AR. Interestingly, the peak distribution among EPF, CAF and

WPMY1-AR showed considerable overlap (Figure 1), and also showed enrichment at proximal promoter sites (Figure 4), despite the considerable difference in AR levels between the primary fibroblasts (low) and WPMY1-AR cells (100 fold higher). We speculate that the fibroblast signature of AR is independent of AR expression level. It is also interesting to note that CAF and WPMY1-AR cells show a heterogeneous pattern of AR expression, similar to the heterogeneous expression of AR in human prostate stroma – in both development and cancer. The ChIPseq peaks derived from these cells must be considered to be an ‘average’ of a population made up of cells with different AR levels. This issue may also be relevant to AR studies in tumour cells, however most published ChIPseq studies have not addressed the expression level or heterogeneity of AR. It has been shown that supra-physiological changes in AR protein level can affect transcriptional programs in tumour epithelial cell lines [61]. We suggest that analysis of AR binding in cells with uniform high or low AR expression may be important to precisely define the role of AR level and genomic distribution in fibroblasts.

We observed varying proportions of overlap when we compared the AR cistromes with transcript datasets derived from tissues in vivo. In general, we have used transcript expression to validate and refine our AR peak data, with an emphasis on primary cells and tissue expression. The strongest overlap was observed when our AR cistromes were compared to fetal prostate tissue [5], which may reflect the fact that fetal human prostate is comprised of a high proportion of mesenchyme (relative to epithelia), is actively growing, and likely shows less cellular heterogeneity than adult tissue. A low proportion of overlap was observed between AR cistromes and prostate cancer tissue. In such comparisons we predict that the stromal AR cistrome would match transcripts expressed in the stroma that comprises a small proportion of the samples. However, stroma is of heterogeneous cellular composition, and the extent of this heterogeneity is not fully defined. We would expect little overlap between the fibroblasts cistromes and tumour cell specific transcripts, since tumour epithelial AR peaks were removed from our AR cistromes (Figure 2). To further refine the AR cistromes, we compared them with tissue-based datasets derived from samples with known stromal content. This showed relatively low overlap with one stromal dataset (GSE20758, [57]) but stronger overlap with another (GSE8218/GSE17951 [58]), and may show low similarity due to the cellular heterogeneity inherent within tumour stroma. In addition, it has been shown that stromal AR expression is decreased with increasing grade of prostate tumour [35, 36] and therefore a larger dataset comprised of prostate tumours with microdissected stroma and sufficient numbers of lower grade tumours would be required to formally address this. **In all comparisons, we observed a higher degree of overlap of EPF-specific targets with PCa stromal-specific/enriched transcripts than CAF-specific targets. This may reflect the larger peak number in EPF compared to CAF, and perhaps lower cellular heterogeneity in developmental versus adult fibroblasts. Additionally, our CAFs represent one prostate cancer patient subtype with appreciable levels of AR expression.**

Patients with an AR positive stroma have a better outcome than those with loss of AR [35,62,63]. The PCa data may be enriched for AR negative stroma, due to inclusion of higher stage and grade, which would contribute to a lack of overlap with our CAF-specific (AR positive) cistrome. Nevertheless, our comparisons with EPF-specific and fibroblast cistromes indicate a subset of our stromal AR targets are expressed in tumour stroma in vivo and such genes may have prognostic potential in prostate cancer.

To define which genes may be AR targets in vivo, we compared our fibroblast AR cistromes to datasets addressing androgen regulation in vivo - in both development and cancer. We observed an overlap of 9-12% between fibroblast cistromes and transcripts regulated by testosterone in urogenital sinus development in vivo [3]. There was a similar proportion of overlap when fibroblast cistromes were compared with benign-adjacent human prostate undergoing androgen deprivation (11-13%); though in this dataset there will be varying proportions of stroma and changes in cellular composition in response to androgen deprivation. It is also uncertain whether developmental growth that is stimulated by androgens bears any similarity to the regression of adult prostate, and in either case true gene regulation must be distinguished from changes in cell proportions. When stromal cistromes were compared to patient tumour tissue undergoing androgen deprivation, there was a lower overlap (6-8%), which may reflect the lower proportion of stroma present in cancer samples. In general, comparisons of our cistromes with in vivo data will be limited by the fact that the proportion of stroma within the tissue can be low, and will change in response to hormonal conditions. This limitation notwithstanding, such comparisons provide a method to parse large datasets in order to identify those genes that may be of physiological importance. It remains possible that such genes will be missed, if they are expressed in small subsets of cells present at low proportions within the sample (eg stem cells). Comparison with cancer-based datasets is most subject to high levels of cellular heterogeneity and this will lead to considerable random associations as well as low overlap. We suggest that developmental systems are less subject to these particular artefacts.

Analysis of the binding sequences present in the AR-derived peaks showed that classical palindromic ARE sequences were very poorly represented in fibroblasts, which suggests that the AR is likely to bind to chromatin via co-factors or other transcription factors [20]. AP1 was identified as a co-factor for AR in the PShTERT-AR cell line [34], however we did not observe many AP1 sites in our data. In this study, we have used primary fibroblasts which could account for this difference as it has been shown that the AR binding profiles can differ between cell lines and primary cells and tissues [29]. The most commonly identified site in our study was ZFX, in EPF; it was also observed in CAF, however the lower peak number in CAF meant that there was no statistically significant binding site identification in these cells.

ZFX is a krueppel C<sub>2</sub>H<sub>2</sub>-type zinc finger protein located on the X chromosome, with poorly defined physiological functions [64]. Studies have shown that ZFX

functions as a transcription factor [65] with the capacity to regulate self-renewal and cell survival in embryonic and haematopoietic stem cells [66, 67]. ZFX has been recognised as an oncogene in a number of cancers (reviewed in [68]). In prostate cancer cell lines, ZFX has been shown to drive cell proliferation and promote cell survival [69] and expression of ZFX has been shown to be elevated in prostate tumours [69, 70]. There is no known link between the functions of ZFX and AR. We have shown that ZFX mRNA and protein is expressed in AR positive prostate stromal cells and is co-expressed with AR in developing prostate mesenchyme. This suggests ZFX may associate with AR; however further study is required to address whether ZFX binds to and serves as a co-factor for AR.

In conclusion, we have identified genome wide AR DNA binding in primary prostate fibroblasts and observed considerable differences with AR binding in tumour cells. We suggest that further verification of stromal AR binding patterns will eventually lead to a more complete and physiologically relevant understanding of AR action.

### **Acknowledgements**

We would like to thank Drs. Magdalena Grabowska and Robert Matusik (Vanderbilt University, TN, USA) for providing WPMY1-AR cells and discussion. Supported by Canadian Cancer Research Society Grant; INNOV14-1 #702423. Additionally, Dr Nadia Boufaied was funded by Prostate Cancer Canada and the Movember Foundation – Grant #T2014-01. Drs. Omar Franco and Simon Hayward were supported by NIH grant 5U01 CA151924.

## Figure Legends

**Figure 1** AR protein expression and genome-wide AR binding in prostate stromal fibroblasts. **(A)** Western blot analysis of AR (110 kDa) and  $\beta$ -actin (42 kDa); male rat brain tissue at P0 and PC3 cells (negative controls), male rat epididymis and whole prostate tissue at P0 (positive controls), LNCaP, WPMY1-AR prostate fibroblast cell line, primary embryonic prostate fibroblasts (EPF) and primary prostate cancer-associated fibroblasts (CAF).  $\beta$ -actin was used as the loading control. \* lysates diluted 1:100 with male rat brain tissue lysate at P0. Representative blot from 2 biological replicates and 2 technical replicates (n=4). **(B)** An overview of AR ChIP-seq peaks of WPMY1-AR, EPF and CAF cells detailing number of peaks called by MACS2.0 and Homer algorithms and final co-identified peak libraries used in subsequent analyses. **(C)** Genomic location analysis of AR peaks in EPF, CAF and WPMY1-AR cells in comparison to the whole genome (hg19). AR peaks were enriched (33-41%) within the promoter regions of genes (-1000bp to +100bp from the TSS) in all 3 cell types. **(D)** Venn diagram illustrating the overlap of AR peaks between WPMY1-AR, EPF and CAF. 783 peaks were co-identified in all 3 peak datasets. CAF had the lowest number of unique AR peaks (629) representing 27% of the total CAF peaks. 67% (1558) of CAF peaks were co-identified in EPF. EPF had the highest number of unique AR peaks (7534) representing 68% of the total EPF peaks. Abbreviations: 5'UTR = 5' untranslated region, 3' UTR = 3' untranslated region, TTS = translational termination site.

**Figure 2** Genome-wide AR binding differs between primary prostate fibroblasts and prostate cancer epithelium. **(A)** Genome-wide correlation of ChIPseq read coverage between primary embryonic prostate fibroblasts (EPF), prostate cancer-associated fibroblasts (CAF) and multiple prostate cancer epithelial cell lines (PCCLs). The heatmap shows the pairwise Spearman correlation coefficient based on read coverage of 10kb bins excluding signal artifact regions. **(B)** ChIPseq read coverage of PCCLs at primary fibroblast peak sites. Heatmaps show the mean ChIPseq read coverage of PCCLs between -2Kb and +2Kb from the centre (0Kb) of called EPF, CAF and PCCL called AR peaks. Heatmaps illustrate a paucity of ChIPseq read coverage of PCCLs at primary fibroblast AR binding sites **(C)** Genomic location analysis of AR peaks in cancer cells in comparison to the whole genome (hg19). Only 2.2% of AR peaks were enriched in the promoter regions of genes (-1000bp to +100bp from the transcriptional start site (TSS)) in PCCLs in comparison to 36.5 and 41.0% in EPF and CAF respectively. **(D)** Average ChIPseq read coverage of EPF, CAF and PCCLs centred around promoter regions of genes (-1000bp to +1000bp from the TSS) expressed as mapped reads per base pair per peak normalized per  $1 \times 10^7$  reads. The histogram shows approximately 3-fold enrichment of EPF and 2-fold enrichment of CAF cells at TSSs compared to PCCLs. **(E)** Venn diagram illustrating the overlap of AR peaks between EPF, CAF and PCCLs. Only 0.35%, 0.05% and 0.15% of epithelial AR peaks are co-identified in EPF, CAF and both

EPF and CAF respectively. Abbreviations: 5'UTR = 5' untranslated region, 3'UTR = 3' untranslated region, TTS = translational termination site.

**Figure 3** Identification of primary embryonic and cancer-associated prostate fibroblast AR cistromes. **(A)** Venn diagram illustrating the overlap of non-epithelial embryonic prostate fibroblast (EPF) and prostate cancer-associated fibroblast (CAF) AR peaks. 9466 AR peaks were identified as EPF-specific, 770 AR peaks were identified as CAF specific and 1467 were co-identified in both fibroblast cell types (Fibroblast peaks) **(B)** Gene track examples of EPF-specific, CAF-specific and Fibroblast peaks at loci on chromosomes 2, 7 and 10 respectively (hg19). **(C)** An overview of RNA-seq transcriptomes of EPF and CAF detailing the total number of reads sequenced, number of paired reads, number of reads mapped the hg19 genome and the total number of genes with read counts > 10 transcripts per million (TPM). **(D)** Transcript validation of genes associated with upstream AR peaks (between 0 to -100000bp upstream of the transcriptional start site (TSS)) from EPF-specific, CAF-specific and Fibroblast AR peak datasets. Venn diagrams illustrate 1560 unique genes with an EPF-specific upstream AR peak have read counts of > 10 TPM in the EPF RNA-seq transcriptome (43.0%). 112 genes with a CAF-specific upstream AR peak have read counts of > 10 TPM in the CAF RNA-seq transcriptome (33.1%). 208 genes with a Fibroblast upstream AR peak have reads counts of >10 TPM in both CAF and EPF RNA-seq transcriptomes (35.6%). Final EPF-specific cistrome contains 1560 genes, CAF-specific cistrome contains 112 genes and Fibroblast cistrome contains 208 genes.

**Figure 4** AR peak distribution differs within the promoter regions of genes between primary prostate fibroblasts and prostate cancer in vitro epithelium and in vivo tissues. **(A)** Genomic location analysis of EPF-specific, CAF-specific, fibroblast, WPMY1-AR, PCCLs and prostate cancer in vivo tissue (PCa in vivo) AR peaks upstream of gene transcriptional start sites (TSS). 28%, 38% and 34% of upstream AR peaks were located in proximal promoter regions of genes (+100bp to -1kb from the transcriptional start sites of genes (TSS)) in EPF-specific, CAF-specific and fibroblast respectively. This AR binding distribution was conserved in WPMY1-AR control fibroblast cell lines with 63% of upstream AR peaks in proximal promoter regions. Only 4% and 19% of upstream AR peaks were located in proximal promoter regions of genes in PCCLs and PCa in vivo tissues respectively. **(B)** Genomic location analysis of transcript-validated EPF-specific, CAF-specific and fibroblast AR peaks upstream of gene TSSs. Enrichment of AR binding at proximal promoter regions of genes was conserved in AR peaks which have been validated with RNAseq transcript expression.

**Figure 5** Identification of AR cistrome targets within in vivo prostate tissue datasets. **(A)** Venn diagrams illustrating the overlap of primary embryonic prostate fibroblast-specific (EPF-specific), primary prostate cancer-associated fibroblast-specific (CAF-specific) and Fibroblast AR peaks with multiple in vivo

prostate cancer (PCa in vivo) AR peaks. Between ~22-30% of the cell AR peaks were co-identified in PCa in vivo tissue AR peaks with an overlap of 2021, 171 and 452 EPF-specific peaks, CAF-specific peaks and Fibroblast peaks respectively. **(B)** Venn diagrams illustrating the overlap of EPF-specific, CAF-specific and Fibroblast AR cistromes with the human embryonic prostate tissue transcriptome (EMB transcriptome). Between ~82-86% of the AR cistromes were co-identified in the EMB transcriptome with an overlap of 1350, 94 and 172 with EPF-specific, CAF-specific and Fibroblast cistromes respectively. **(C)** Venn diagrams illustrating the overlap of EPF-specific, CAF-specific and Fibroblast AR cistromes with microdissected PCa in vivo stroma-specific transcripts (PCa stromal-specific, GSE20758, [57]) and PCa stromal-enriched transcripts (PCa stromal-enriched, GSE8218 and GSE17951,[58]) following removal of transcripts not present on the arrays. For the stromal-specific transcripts, 61 and 477 were co-identified with the EPF-specific cistrome, 6 and 32 were co-identified with the CAF-specific cistrome and 7 and 59 were co-identified with the Fibroblast cistrome.

**Figure 6** Comparison of AR cistrome targets to androgen regulated transcripts within in vivo prostate tissue datasets. **(A)** Venn diagrams illustrating the overlap of EPF-specific, CAF-specific and Fibroblast AR cistromes with androgen-responsive transcripts from the embryonic mouse urogenital sinus (UGS DHT-regulated, [3]). Between ~9-12% of the AR cistromes were co-identified with androgen-responsive UGS transcripts with an overlap of 188, 14 and 20 with EPF-specific, CAF-specific and Fibroblast cistromes respectively. **(B)** Venn diagrams illustrating the overlap of EPF-specific, CAF-specific and Fibroblast AR cistromes with transcripts that are differentially expressed following androgen-deprivation treatment in human prostate cancer tissue samples (PCa ADT responsive) and in human normal adjacent tissues (Benign ADT responsive). Between ~6-8% of the AR cistromes were co-identified with PCa ADT responsive transcripts with an overlap of 88, 6 and 16 with EPF-specific, CAF-specific and Fibroblast cistromes respectively. Between ~11-13% of the AR cistromes were co-identified with Benign ADT responsive transcripts with an overlap of 159, 13 and 23 with EPF-specific, CAF-specific and Fibroblast cistromes respectively.

**Figure 7** Identification of ZFX as a potential AR co-factor in primary prostate fibroblasts. **(A)** Consensus sequence logos of the top sequence motifs and their significance identified in primary prostate embryonic fibroblast (EPF), primary prostate cancer-associated fibroblast (CAF) and WPMY1-AR using *de novo* motif analysis with HOMER. \*  $p < 1E-50$ . Classical androgen-response element sequence (ARE) is provided for comparison. **(B)** An overview of the number of AR ChIPseq peaks containing the top sequence motif, the best match transcription factor (TF) based on positive mRNA expression from RNAseq analysis and expression value in transcripts per million (TPM). **(C)** Histogram showing the percentage (bold) of ChIPseq peaks positive for the classical ARE in EPF, CAF, WPMY1-AR, prostate cancer cell lines (PCCL) and prostate cancer tissues (PCa). EPF and CAF had 100-fold less and WPMY1-AR 10-fold less ARE



positive peaks than PCCLs. **(D)** Localization of AR and ZFX in serial sections of P0 male urogenital tract by immunohistochemistry. Ur denotes urethra, UrSM denotes urethral smooth muscle, VP denotes ventral prostate, DP denotes dorsal prostate, VPE denotes epithelial buds of the ventral prostate and VPM denotes ventral prostate mesenchyme. (i-ii) P0 male UGT showed AR expression localized to VP mesenchymal cells and little or no expression in epithelial cells of the VP prostatic buds (low magnification i, high magnification ii). (iii-iv) ZFX expression was localized to both VP mesenchyme and in the epithelium of prostatic buds but appeared more nuclear in VP mesenchyme. (low magnification iii, high magnification iv). Negative control omitting primary antibody inset (iii-iv). Left panels x 8 magnification, scale bars = 300 $\mu$ m and right panels at x 40 magnification, scale bars = 60 $\mu$ m.

## References

1. Hayward, S.W., G.R. Cunha, and R. Dahiya, *Normal development and carcinogenesis of the prostate. A unifying hypothesis*. Ann N Y Acad Sci, 1996. **784**: p. 50-62.
2. Vanpoucke, G., et al., *Transcriptional profiling of inductive mesenchyme to identify molecules involved in prostate development and disease*. Genome Biol, 2007. **8**(10): p. R213.
3. Schaeffer, E.M., et al., *Androgen-induced programs for prostate epithelial growth and invasion arise in embryogenesis and are reactivated in cancer*. Oncogene, 2008. **27**(57): p. 7180-91.
4. Orr, B., et al., *Reduction of pro-tumorigenic activity of human prostate cancer-associated fibroblasts using Dlk1 or SCUBE1*. Dis Model Mech, 2013. **6**(2): p. 530-6.
5. Orr, B., et al., *Identification of stromally expressed molecules in the prostate by tag-profiling of cancer-associated fibroblasts, normal fibroblasts and fetal prostate*. Oncogene, 2012. **31**(9): p. 1130-42.
6. Rochette, A., et al., *Asporin is a stromally expressed marker associated with prostate cancer progression*. Br J Cancer, 2017. **116**(6): p. 775-784.
7. Cunha, G.R. and L.W. Chung, *Stromal-epithelial interactions--I. Induction of prostatic phenotype in urothelium of testicular feminized (Tfm/y) mice*. J Steroid Biochem, 1981. **14**(12): p. 1317-24.
8. Murashima, A., et al., *Androgens and mammalian male reproductive tract development*. Biochim Biophys Acta, 2015. **1849**(2): p. 163-70.
9. Olumi, A.F., et al., *Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium*. Cancer Res, 1999. **59**(19): p. 5002-11.
10. Ricke, E.A., et al., *Androgen hormone action in prostatic carcinogenesis: stromal androgen receptors mediate prostate cancer progression, malignant transformation and metastasis*. Carcinogenesis, 2012. **33**(7): p. 1391-8.
11. Lamb, A.D., C.E. Massie, and D.E. Neal, *The transcriptional programme of the androgen receptor (AR) in prostate cancer*. BJU Int, 2014. **113**(3): p. 358-66.
12. Andreu-Vieyra, C., et al., *Dynamic nucleosome-depleted regions at androgen receptor enhancers in the absence of ligand in prostate cancer cells*. Mol Cell Biol, 2011. **31**(23): p. 4648-62.
13. Decker, K.F., et al., *Persistent androgen receptor-mediated transcription in castration-resistant prostate cancer under androgen-deprived conditions*. Nucleic Acids Research, 2012.
14. Jin, H.J., J. Kim, and J. Yu, *Androgen receptor genomic regulation*. Transl Androl Urol, 2013. **2**(3): p. 157-177.
15. Tao, S., H. He, and Q. Chen, *ChIP-seq analysis of androgen receptor in LNCaP cell line*. Mol Biol Rep, 2014. **41**(9): p. 6291-6.

16. Tewari, A.K., et al., *Chromatin accessibility reveals insights into androgen receptor activation and transcriptional specificity*. *Genome Biol*, 2012. **13**(10): p. R88.
17. Zhu, Z., et al., *Dose-dependent effects of small-molecule antagonists on the genomic landscape of androgen receptor binding*. *BMC Genomics*, 2012. **13**: p. 355.
18. Massie, C.E., et al., *The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis*. *Embo j*, 2011. **30**(13): p. 2719-33.
19. Polkinghorn, W.R., et al., *Androgen receptor signaling regulates DNA repair in prostate cancers*. *Cancer Discov*, 2013. **3**(11): p. 1245-53.
20. Wilson, S., J. Qi, and F.V. Filipp, *Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines*. *Sci Rep*, 2016. **6**: p. 32611.
21. Wang, Q., et al., *Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer*. *Cell*, 2009. **138**(2): p. 245-56.
22. Chan, S.C., et al., *Targeting chromatin binding regulation of constitutively active AR variants to overcome prostate cancer resistance to endocrine-based therapies*. *Nucleic Acids Res*, 2015. **43**(12): p. 5880-97.
23. Chng, K.R., et al., *A transcriptional repressor co-regulatory network governing androgen response in prostate cancers*. *Embo j*, 2012. **31**(12): p. 2810-23.
24. Jehle, K., et al., *Coregulator control of androgen receptor action by a novel nuclear receptor-binding motif*. *J Biol Chem*, 2014. **289**(13): p. 8839-51.
25. Jia, L., et al., *Genomic androgen receptor-occupied regions with different functions, defined by histone acetylation, coregulators and transcriptional capacity*. *PLoS One*, 2008. **3**(11): p. e3645.
26. Jin, H.J., et al., *Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program*. *Nat Commun*, 2014. **5**: p. 3972.
27. Pihlajamaa, P., et al., *Tissue-specific pioneer factors associate with androgen receptor cisomes and transcription programs*. *Embo j*, 2014. **33**(4): p. 312-26.
28. Yu, J., et al., *An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression*. *Cancer Cell*, 2010. **17**(5): p. 443-54.
29. Sharma, N.L., et al., *The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man*. *Cancer Cell*, 2013. **23**(1): p. 35-47.
30. Chen, Z., et al., *Agonist and antagonist switch DNA motifs recognized by human androgen receptor in prostate cancer*. *Embo j*, 2015. **34**(4): p. 502-16.
31. Nevedomskaya, E., et al., *Androgen receptor DNA binding and chromatin accessibility profiling in prostate cancer*. *Genom Data*, 2016. **7**: p. 124-6.
32. Pomerantz, M.M., et al., *The androgen receptor cisome is extensively reprogrammed in human prostate tumorigenesis*. *Nat Genet*, 2015. **47**(11): p. 1346-51.
33. Stelloo, S., et al., *Androgen receptor profiling predicts prostate cancer outcome*. *EMBO Mol Med*, 2015. **7**(11): p. 1450-64.

34. Leach, D.A., et al., *Cell-lineage specificity and role of AP-1 in the prostate fibroblast androgen receptor cistrome*. Mol Cell Endocrinol, 2016.
35. Leach, D.A., et al., *Stromal androgen receptor regulates the composition of the microenvironment to influence prostate cancer outcome*. Oncotarget, 2015. **6**(18): p. 16135-50.
36. Huber, F., et al., *Comprehensive validation of published immunohistochemical prognostic biomarkers of prostate cancer -what has gone wrong? A blueprint for the way forward in biomarker studies*. Br J Cancer, 2015. **112**(1): p.140-8.
37. Tanner, M.J., et al., *Effects of androgen receptor and androgen on gene expression in prostate stromal fibroblasts and paracrine signaling to prostate cancer cells*. PLoS One, 2011. **6**(1): p. e16027.
38. Orr, B., et al., *Expression of pleiotrophin in the prostate is androgen regulated and it functions as an autocrine regulator of mesenchyme and cancer associated fibroblasts and as a paracrine regulator of epithelia*. Prostate, 2011. **71**(3): p. 305-17.
39. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie2*. Nat Methods, 2012. **9**(4): p. 357-9.
40. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS)*. Genome Biol, 2008. **9**(9): p. R137.
41. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities*. Mol Cell, 2010. **38**(4): p. 576-89.
42. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching*. Nucleic Acids Res, 2009. **37**(Web Server issue): p. W202-8.
43. Mathelier, A., et al., *JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles*. Nucleic Acids Res, 2016. **44**(D1): p. D110-5.
44. Ramirez, F., et al., *deepTools: a flexible platform for exploring deep-sequencing data*. Nucleic Acids Res, 2014. **42**(Web Server issue): p. W187-91.
45. Hulsen, T., J. de Vlieg, and W. Alkema, *BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams*. BMC Genomics, 2008. **9**: p. 488.
46. Nicol, J.W., et al., *The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets*. Bioinformatics, 2009. **25**(20): p. 2730-1.
47. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions*. Genome Biol, 2013. **14**(4): p. R36.
48. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features*. Bioinformatics, 2014. **30**(7): p. 923-30.
49. Davis, S. and P.S. Meltzer, *GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor*. Bioinformatics, 2007. **23**(14): p. 1846-7.
50. Gautier, L., et al., *affy--analysis of Affymetrix GeneChip data at the probe level*. Bioinformatics, 2004. **20**(3): p. 307-15.

51. Jaffe, A.E., et al., *Practical impacts of genomic data "cleaning" on biological discovery using surrogate variable analysis*. BMC Bioinformatics, 2015. **16**: p. 372.
52. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. Nucleic Acids Res, 2015. **43**(7): p. e47.
53. Phipson, B., Lee, S, Majewski, IJ, Alexander, WS, and Smyth, GK *Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression*. Annals of Applied Statistics 2016. **10**(2): p. 946-963.
54. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics*. Genome Biol, 2004. **5**(10): p. R80.
55. Zou, J.X., et al., *Androgen-induced coactivator ANCCA mediates specific androgen receptor signaling in prostate cancer*. Cancer Res, 2009. **69**(8): p. 3339-46.
56. Chen, H., et al., *Genome-wide analysis of androgen receptor binding and gene regulation in two CWR22-derived prostate cancer cell lines*. Endocr Relat Cancer, 2010. **17**(4): p. 857-73.
57. Gregg, J.L., et al., *Analysis of gene expression in prostate cancer epithelial and interstitial stromal cells using laser capture microdissection*. BMC Cancer, 2010. **10**: p. 165.
58. Wang, Y., et al., *In silico estimates of tissue components in surgical samples based on expression profiling data*. Cancer Res, 2010. **70**(16): p. 6448-55.
59. Ross-Adams, H., et al., *Integration of copy number and transcriptomics provides risk stratification in prostate cancer: A discovery and validation cohort study*. EBioMedicine, 2015. **2**(9): p. 1133-44.
60. Zhu, Z., et al., *ZFX regulates glioma cell proliferation and survival in vitro and in vivo*. J Neurooncol, 2013. **112**(1): p. 17-25.
61. Waltering, K.K., et al., *Increased expression of androgen receptor sensitizes prostate cancer cells to low levels of androgens*. Cancer Res, 2009. **69**(20): p. 8141-9.
62. Li, Y., et al., *Decrease in stromal androgen receptor associates with androgen-independent disease and promotes prostate cancer cell proliferation and invasion*. J Cell Mol Med, 2008. **12**(6b): p. 2790-8.
63. Wikstrom, P., et al., *Low stroma androgen receptor level in normal and tumor prostate tissue is related to poor outcome in prostate cancer patients*. Prostate, 2009. **69**(8): p. 799-809.
64. Iwase, M., et al., *The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species*. Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5258-63.
65. Xu, S., et al., *Zinc Finger and X-Linked Factor (ZFX) Binds to Human SET Transcript 2 Promoter and Transactivates SET Expression*. Int J Mol Sci, 2016. **17**(10).
66. Arenzana, T.L., M.R. Smith-Raska, and B. Reizis, *Transcription factor Zfx controls BCR-induced proliferation and survival of B lymphocytes*. Blood, 2009. **113**(23): p. 5857-67.

67. Luoh, S.W., et al., *Zfx mutation results in small animal size and reduced germ cell number in male and female mice*. *Development*, 1997. **124**(11): p. 2275-84.
68. Jen, J. and Y.C. Wang, *Zinc finger proteins in cancer progression*. *J Biomed Sci*, 2016. **23**(1): p. 53.
69. Jiang, H., et al., *Knockdown of zinc finger protein X-linked inhibits prostate cancer cell proliferation and induces apoptosis by activating caspase-3 and caspase-9*. *Cancer Gene Ther*, 2012. **19**(10): p. 684-9.
70. Tricoli, J.V. and R.B. Bracken, *ZFY gene expression and retention in human prostate adenocarcinoma*. *Genes Chromosomes Cancer*, 1993. **6**(2): p. 65-72.

**Highlights for Nash et al**

**Highlights for Nash et al**

- 1. Genome-wide distribution of AR defined in primary prostate stromal cells using ChIPseq
- 2. Stromal AR shows non-ARE binding sites with a unique distribution in stroma compared to tumour cells
- 3. Frequent localisation of AR proximal to transcriptional start site in stroma
- 4. Deconvolution of stromal AR binding sites in prostate tumour tissue ChIPseq datasets

Figure 1

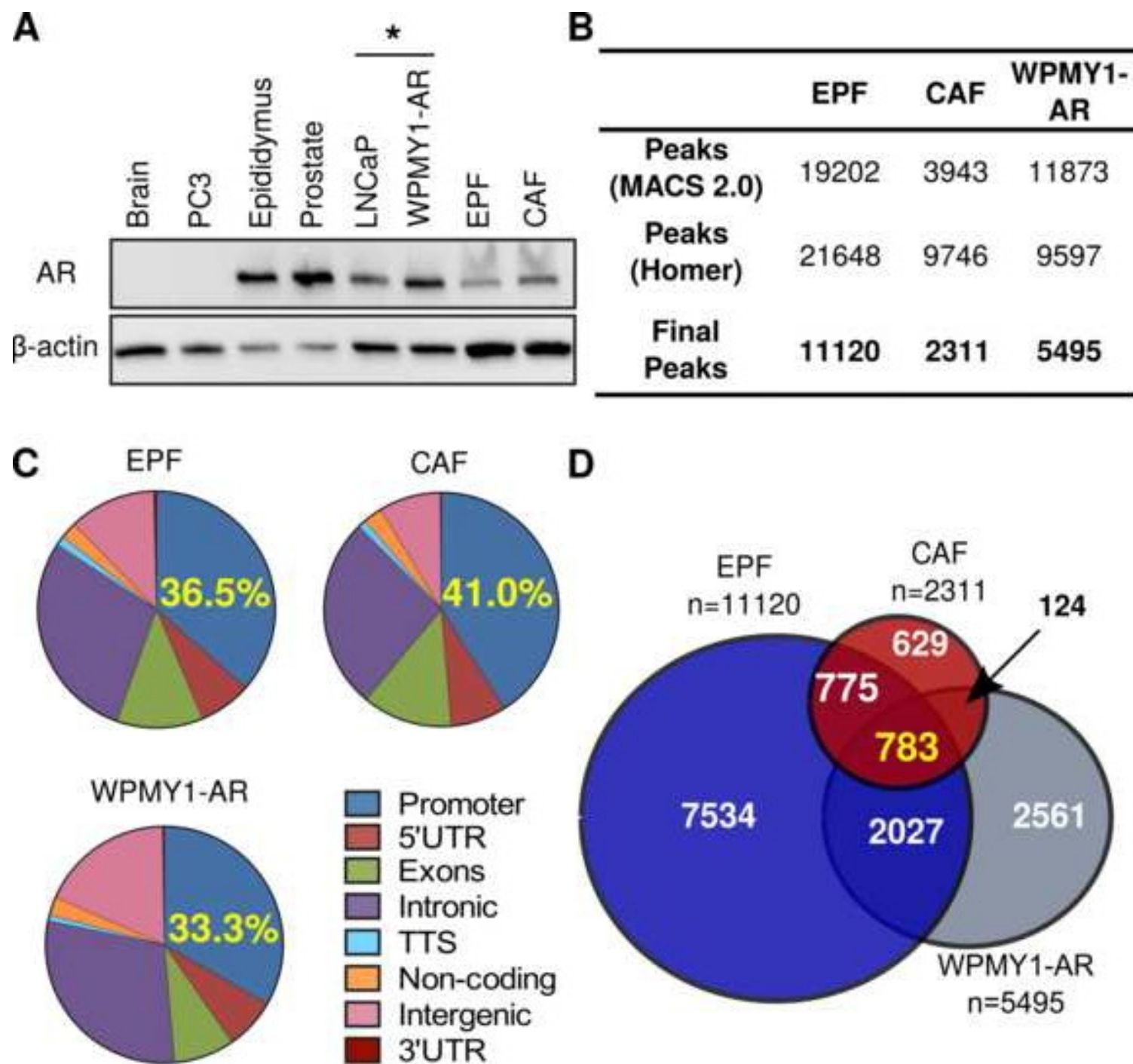
[Click here to download high resolution image](#)



Figure 2

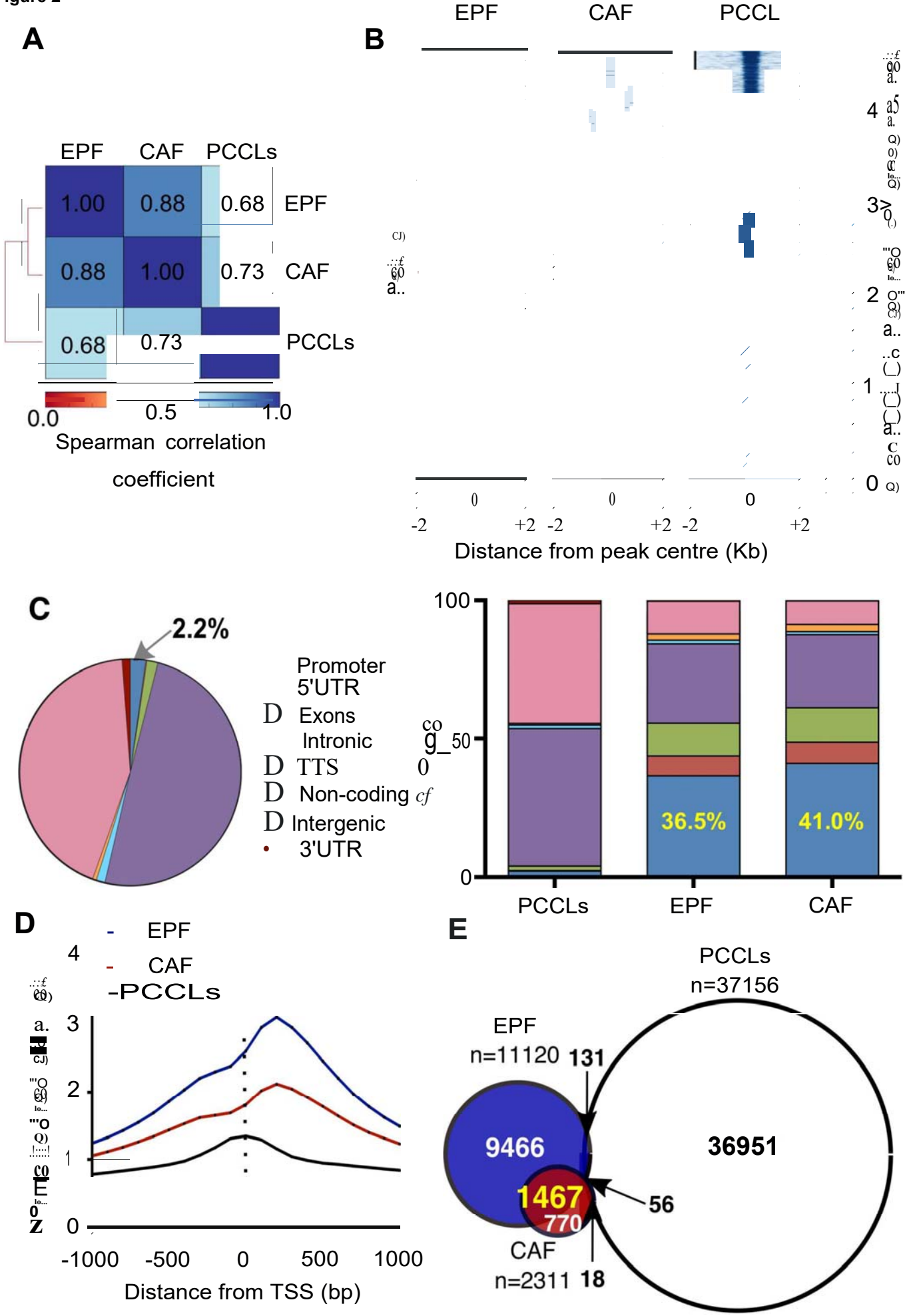
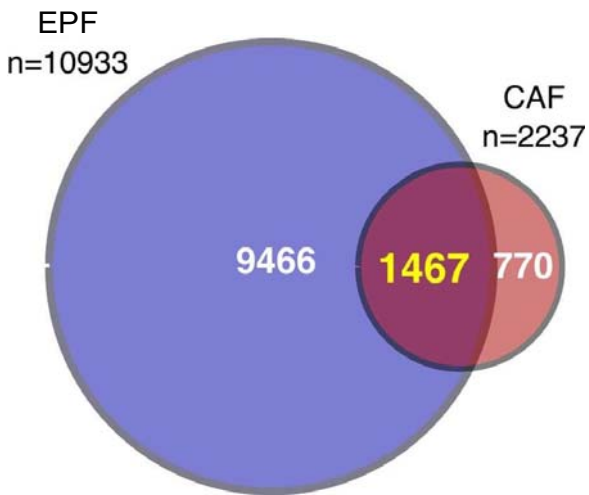
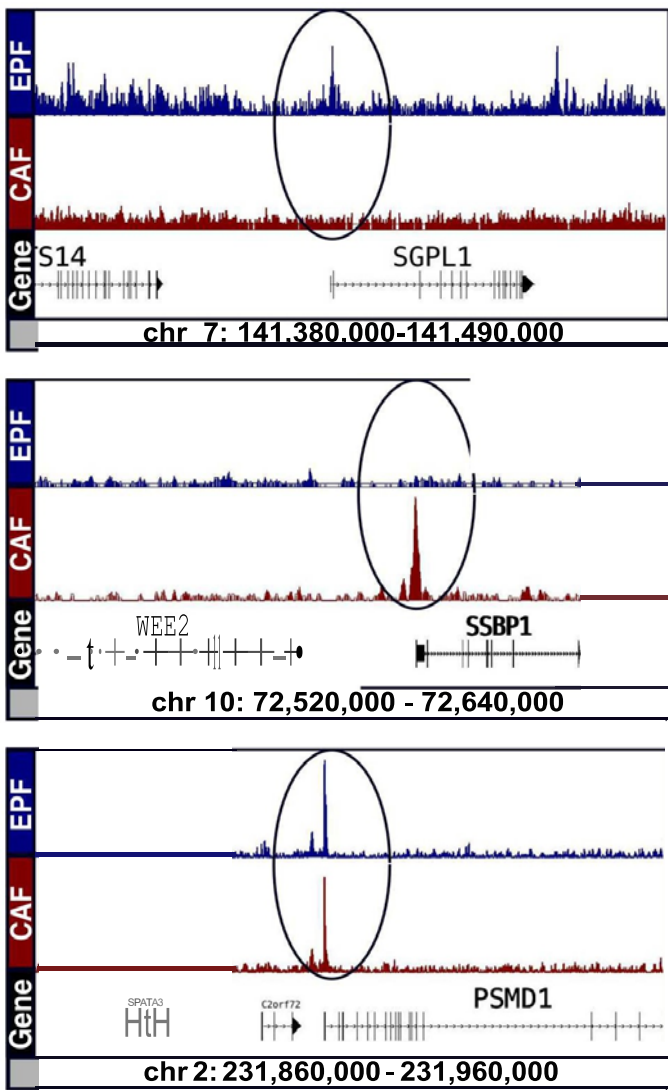


Figure 3

**A**



**B**



**C**

	EPF	CAF
Total mapped paired reads	69355917	55730323
Genes with mapped reads	13867	14800
Genes > 10TPM	8334	6177

**D**

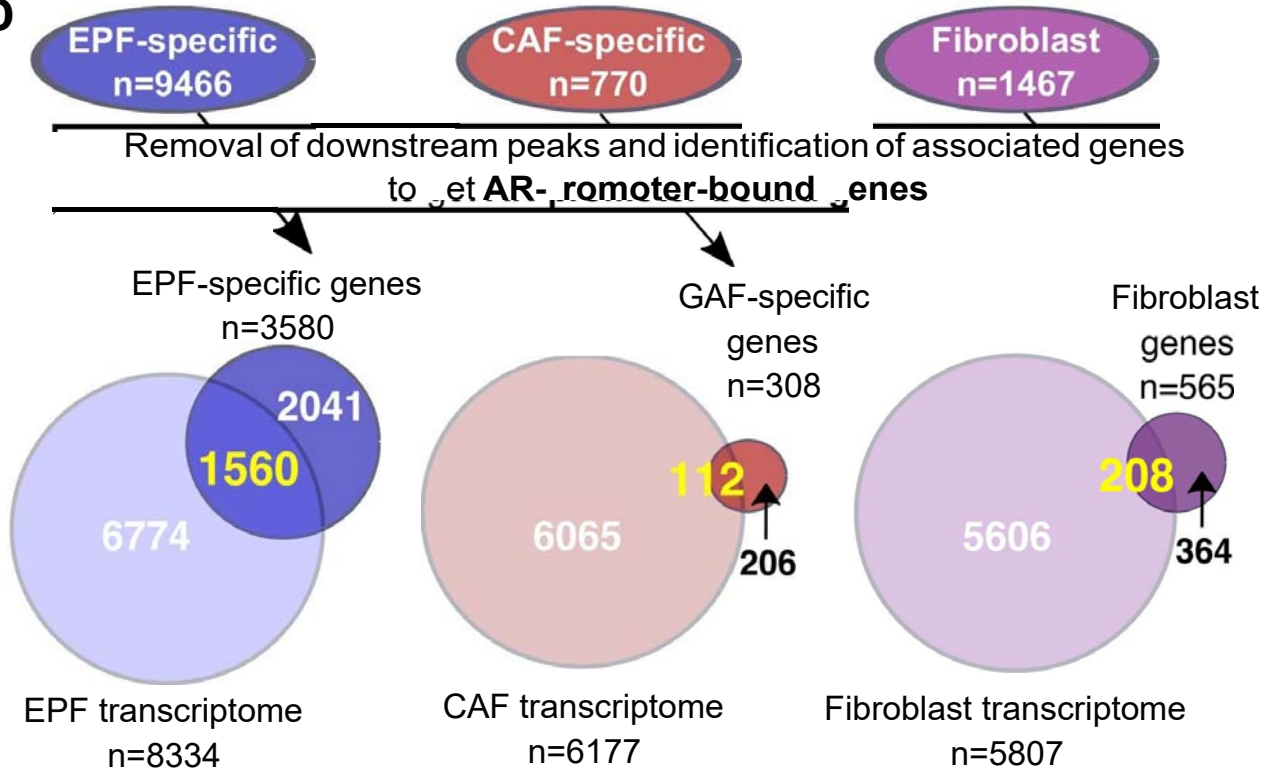


Figure 4

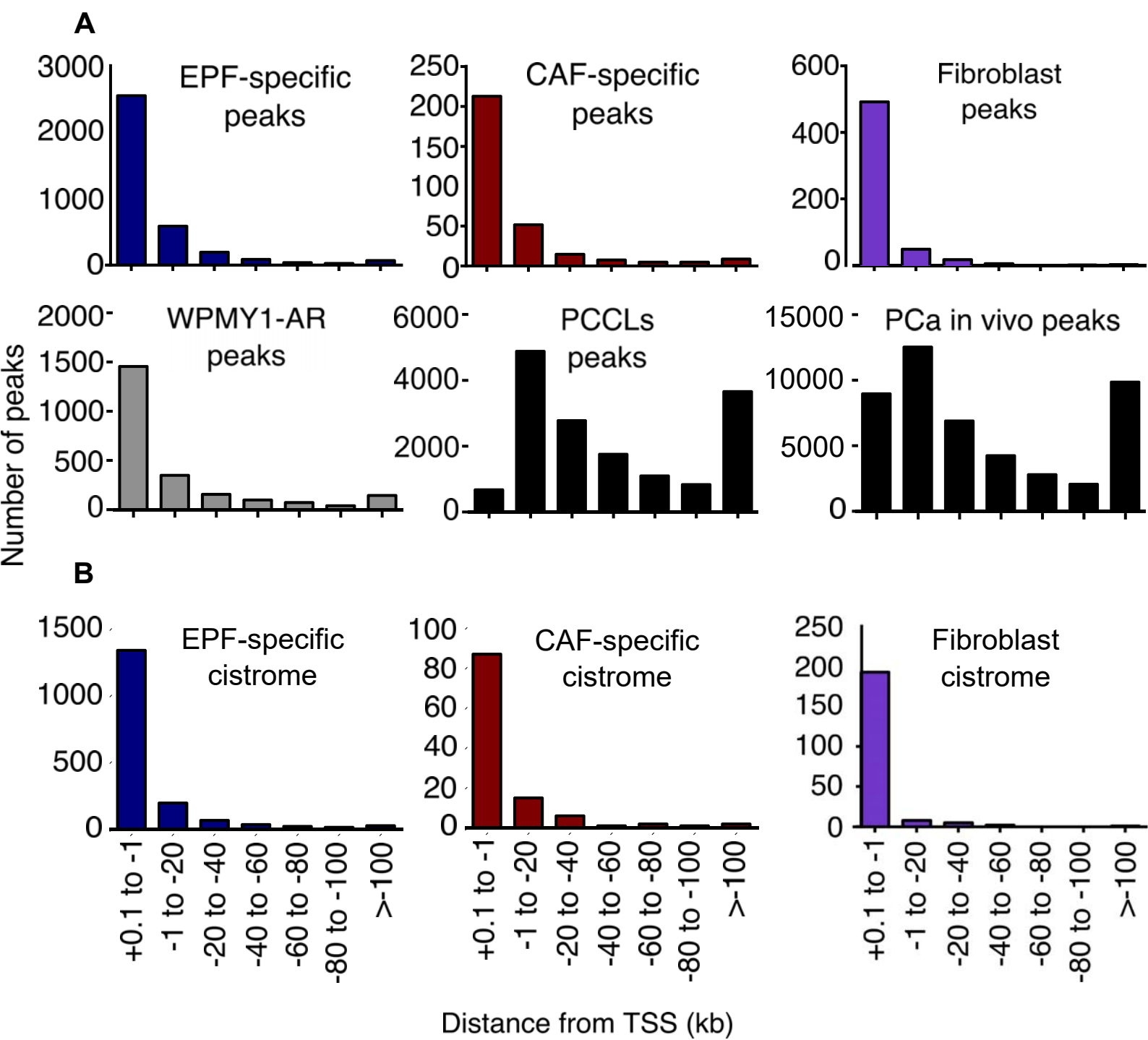
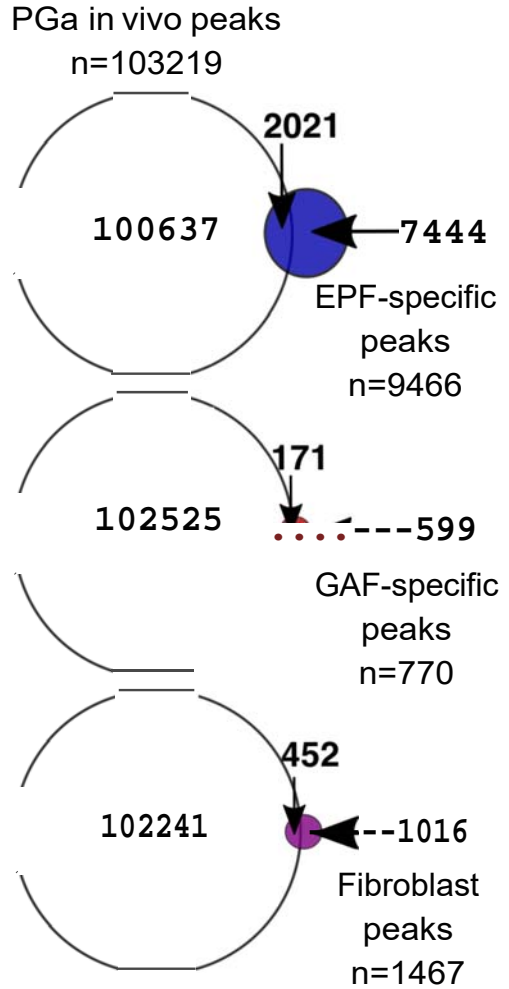
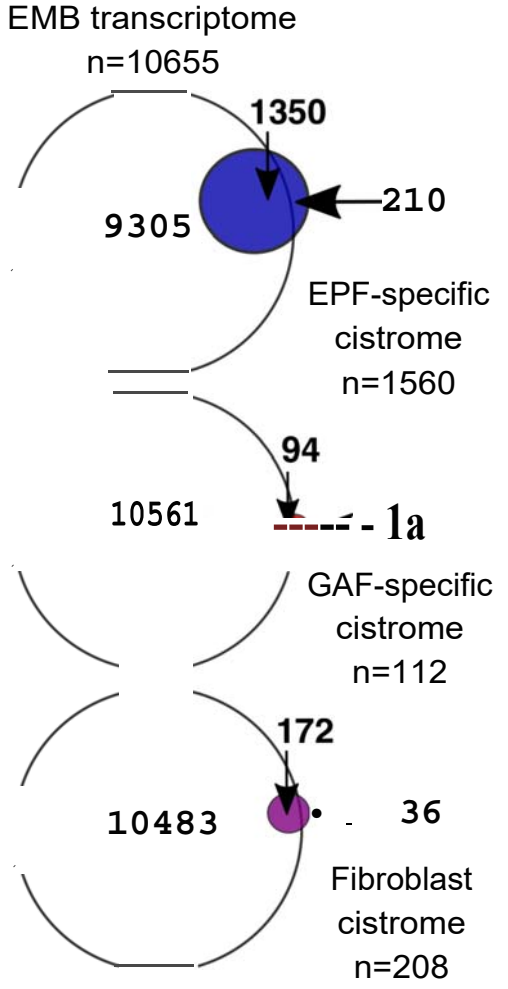


Figure 5

**A**



**8**



**C**

PGa stromal-specific

PGa stromal-enriched

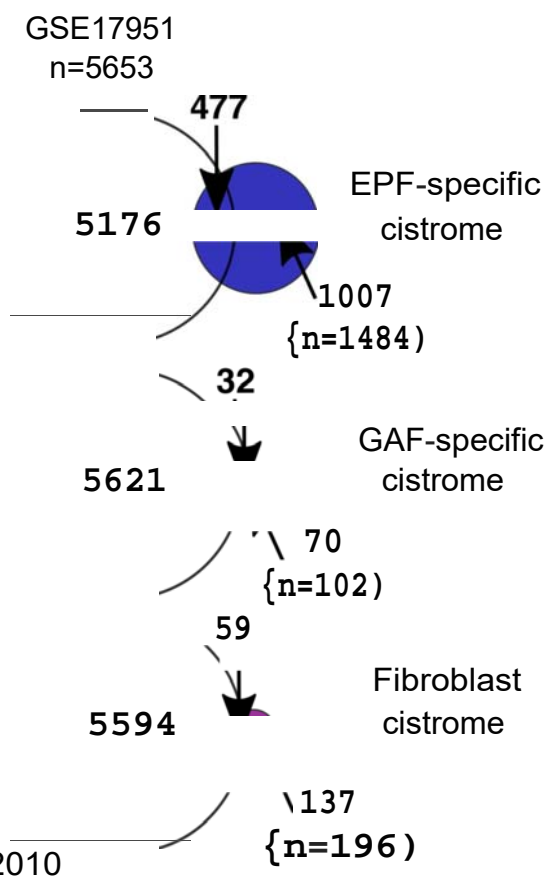
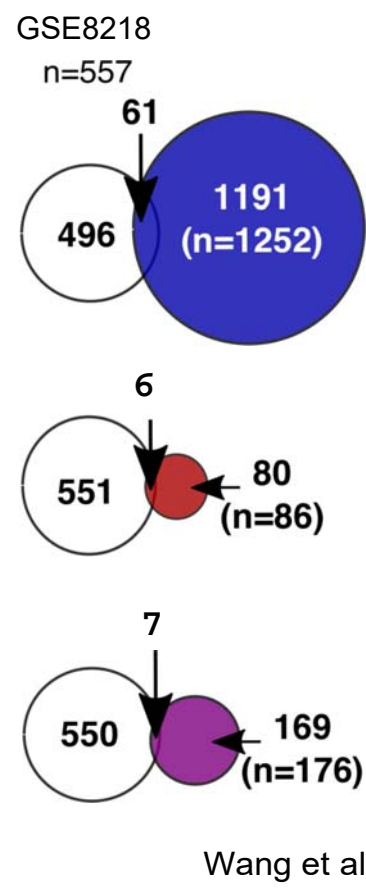
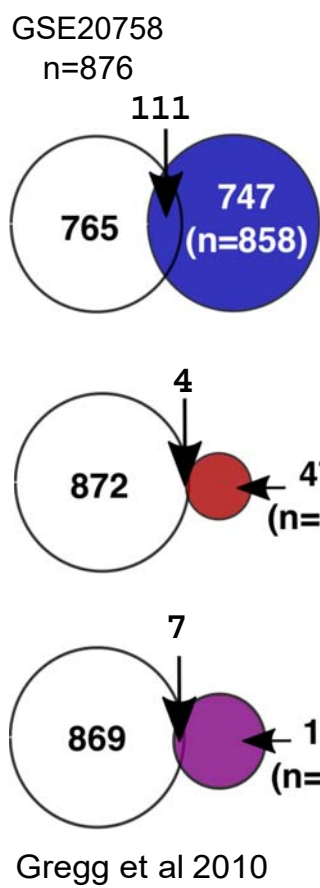


Figure 6

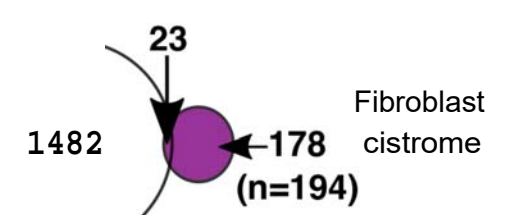
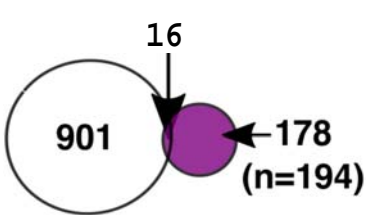
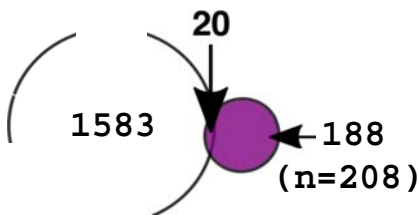
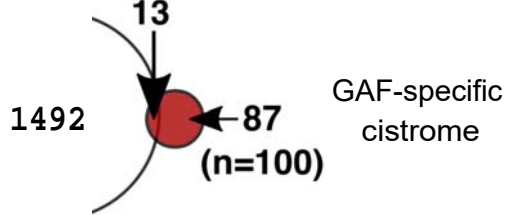
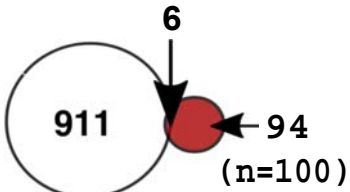
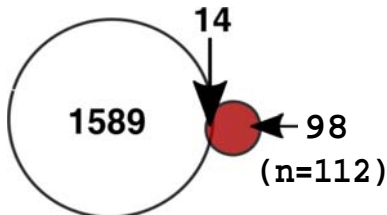
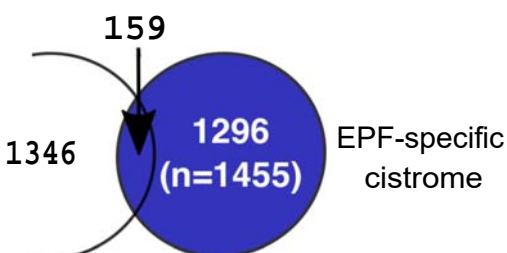
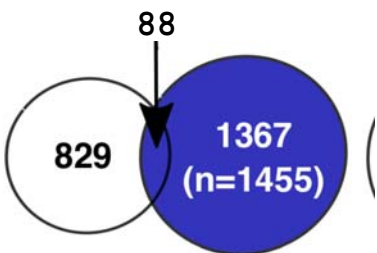
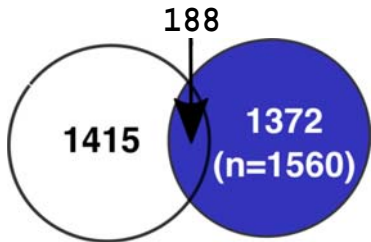
**A**

**B**

UGS  
DHT-regulated  
n=1603

PCa ADT responsive  
GSE72920  
n=917

Benign ADT responsive  
GSE70768  
n=1505



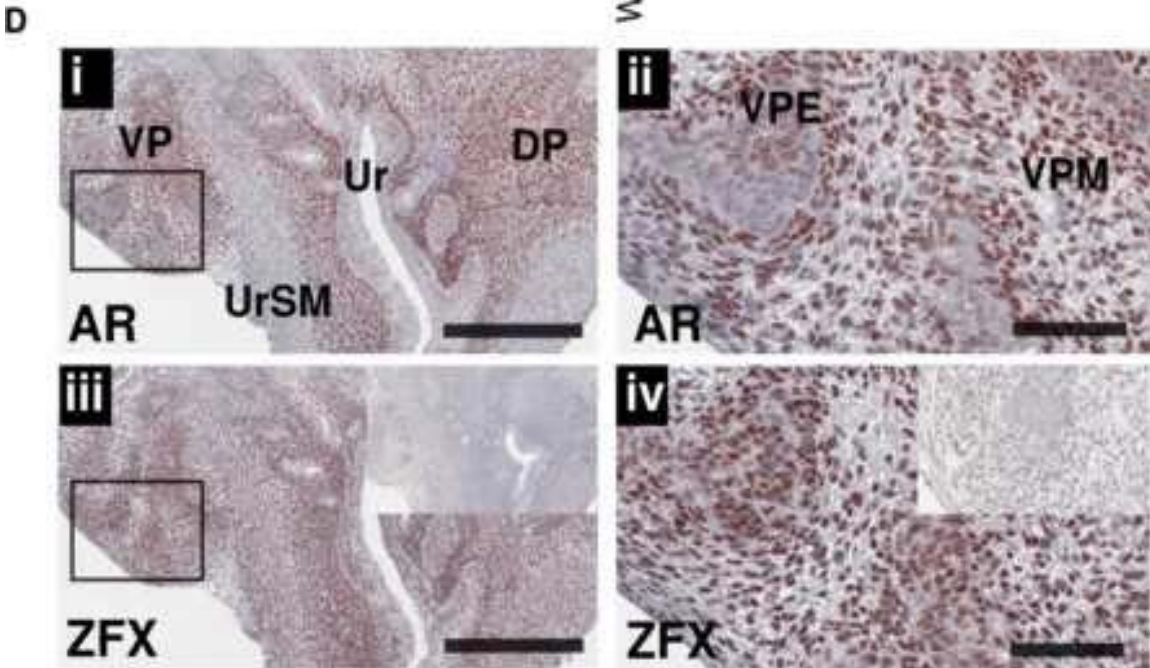
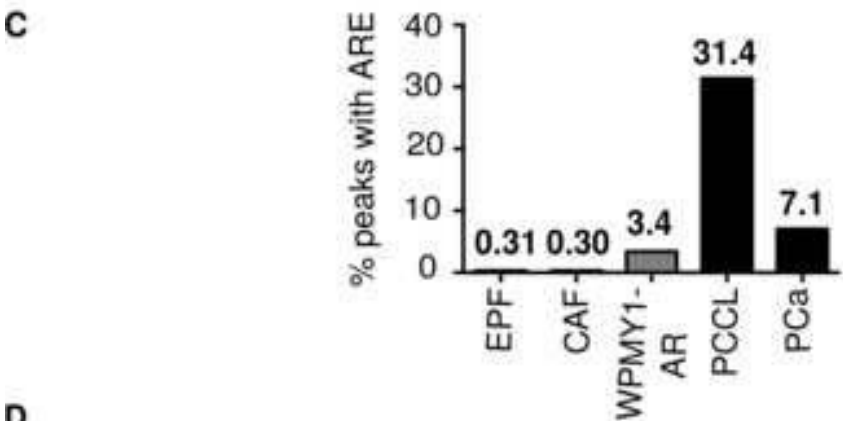
Schaetteretal2008

Figure 7  
[Click here to download high resolution image](#)



**B**

	EPF	CAF	WPMY1-AR
No. and % of peaks with motif	8904 (81.5%)	25 (1.1%)	367 (6.9%)
Best match expressed TF	ZFX	SWI5	ATF3
Transcript TPM (RNAseq)	15.5	32.9	N/A



**Supplementary excel data**

[Click here to download Supplementary Material: Supplementary-Data-File Nash.xlsx](#)

**Supplementary data and tables**

[Click here to download Supplementary Material: Supplementar Figs and tables.pdf](#)