REVIEW

Androgen receptor enhancer usage and the chromatin regulatory landscape in human prostate cancers

Suzan Stelloo1, Andries M Bergman2,3 and Wilbert Zwart1,4

1Division of Oncogenomics, OncoCode Institute, The Netherlands Cancer Institute, Amsterdam, The Netherlands
2Division of Oncogenomics, The Netherlands Cancer Institute, Amsterdam, The Netherlands
3Division of Medical Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
4Department of Biomedical Engineering, Laboratory of Chemical Biology and Institute for Complex Molecular Systems, Eindhoven University of Technology, Eindhoven, The Netherlands

Correspondence should be addressed to W Zwart: w.zwart@nki.nl

Abstract

The androgen receptor (AR) is commonly known as a key transcription factor in prostate cancer development, progression and therapy resistance. Genome-wide chromatin association studies revealed that transcriptional regulation by AR mainly depends on binding to distal regulatory enhancer elements that control gene expression through chromatin looping to gene promoters. Changes in the chromatin epigenetic landscape and DNA sequence can locally alter AR-DNA-binding capacity and consequently impact transcriptional output and disease outcome. The vast majority of reports describing AR chromatin interactions have been limited to cell lines, identifying numerous other factors and interacting transcription factors that impact AR chromatin interactions. Do these factors also impact AR cistromics – the genome-wide chromatin-binding landscape of AR – in vivo? Recent technological advances now enable researchers to identify AR chromatin-binding sites and their target genes in human specimens. In this review, we provide an overview of the different factors that influence AR chromatin binding in prostate cancer specimens, which is complemented with knowledge from cell line studies. Finally, we discuss novel perspectives on studying AR cistromics in clinical samples.

Transcription factor binding

By 2003, the DNA sequence of the entire human genome was annotated by The Human Genome Project (International Human Genome Sequencing Consortium 2004). However, as informative the primary DNA sequence information is on the protein-coding genome (~2% of the total human DNA), interpretation of the remaining ~98% of the human genome appears more challenging. One of the major challenges in functionally interpreting the non-protein-coding human genome is that the primary sequence itself does not explain how the DNA is packaged into chromatin and where transcription factors (TFs) bind. Some TFs bind indirectly to the DNA via protein–protein interactions by means of tethering, while other TFs recognize specific DNA sequences, often referred to as TF motifs, footprints or grammar (Fig. 1A). Many computational methods have been developed to scan for these TF motifs across the entire genome to predict the capacity of a particular TF to bind specific regions, such as

Key Words

- androgen receptor
- prostate cancer
- ChIP-seq
- cistrome
- subtypes
the SeqPos motif tool (Liu et al. 2011) and HOMER (Heinz et al. 2010). In addition, other algorithms (e.g. MEME Suite; Bailey et al. 2015) can be applied to discover DNA sequence patterns in given regions. However, the presence of a TF motif does not imply that the TF is actually capable to bind this region. For example, the AR – the main driver in prostate cancer development and progression – recognizes a palindromic dihexameric androgen-responsive element (ARE), 5′-AGAACAnnnTGTTCT-3′ (Fig. 1B), which occurs a few million times throughout the human genome. However, only hundreds to tens of thousands AREs appear to be functionally active in a given context and occupied by AR in prostate cell lines (Yu et al. 2010, Massie et al. 2011, Toropainen et al. 2016, A C

### Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Accession number</th>
<th>Factor</th>
<th>Normal cells</th>
<th>Primary tumor</th>
<th>CRPC</th>
<th>Stroma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yu et al. (2010)</td>
<td>GSE14097</td>
<td>AR</td>
<td>2</td>
<td>5</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Sharma et al. (2013)</td>
<td>GSE28219</td>
<td>AR</td>
<td>2</td>
<td>4</td>
<td>13</td>
<td>-</td>
</tr>
<tr>
<td>Chen et al. (2015)</td>
<td>GSE43791</td>
<td>AR</td>
<td>7</td>
<td>13</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Pomerantz et al. (2015)</td>
<td>GSE56288</td>
<td>AR</td>
<td>-</td>
<td>4</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Stelloo et al. (2015)</td>
<td>GSE65478</td>
<td>AR</td>
<td>-</td>
<td>3</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Chen et al. (2018)</td>
<td>GSE99378</td>
<td>VAHOBX13</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Stelloo et al. (2018b)</td>
<td>GSE120738</td>
<td>AR</td>
<td>-</td>
<td>88</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Kron et al. (2017)</td>
<td>GSE96652</td>
<td>H3K27ac</td>
<td>92</td>
<td>92</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Takeda et al. (2018)</td>
<td>GSE114385</td>
<td>H3K27ac</td>
<td>6</td>
<td>4</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Cioni et al. (2018)</td>
<td>GSE126852</td>
<td>AR</td>
<td>-</td>
<td>-</td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>
McNair et al. 2017, Stelloo et al. 2018a) and tissue samples (Sharma et al. 2013, Chen et al. 2015, Pomerantz et al. 2015, Stelloo et al. 2015, 2018b). In addition, all cells in the human body carry an identical genome; yet, TFs regulate gene expression patterns to dictate organ development and identity. Key questions to further understand AR action on a genome-wide scale are the following: Where does AR bind the genome in different contexts, and what are the biological consequences thereof? While full discussion on the biological significance of AR in prostate cancer is beyond the scope of this review, this subject has been covered elsewhere (Cilig & Santer 2014, Copeland et al. 2018, Isaacs 2018, Ken-ichi 2018). Another question is how AR chromatin binding is regulated and to what degree is this context dependent? Furthermore, with more reports describing AR cistromics in clinical samples (Table 1), what did we learn from this, and how could we use this information in the clinical setting? In particular, we focus on the technological developments (Table 2) that have made the transition from cell line models toward the study of clinical specimens. This review will address these points systematically and will highlight the potential future research directions aimed to enhance our understanding of genome regulation in prostate cancer.

### AR cistromics

**Where does AR bind the chromatin in cell lines?**

Prior to the era of genome-wide mapping of TF binding, AR binding was predominantly studied at the *Kallikrein-3* (*KLK3*) gene locus, encoding the commonly used serum

<table>
<thead>
<tr>
<th>Method</th>
<th>Full method name</th>
<th>Approach for</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChIP-on-chip</td>
<td>Chromatin immunoprecipitation followed by genomic tiling microarray hybridization</td>
<td>Genome-wide analysis of transcription factor binding and histone modifications</td>
<td>Outdated technology AR (Wang et al. 2007, 2009) Barski et al. (2007)</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>Chromatin immunoprecipitation followed by sequencing</td>
<td>Genome-wide analysis of transcription factor binding and histone modifications</td>
<td>AR (Wang et al. 2007, Cao et al. 2014) Fanelli et al. (2010), Cejas et al. (2016)</td>
</tr>
<tr>
<td>FFPE ChIP</td>
<td>ChiP-seq analysis on formalin-fixed paraffin-embedded tissue samples</td>
<td>Genome-wide analysis of histone modifications using FFPE material</td>
<td>Simonis et al. (2006), Zhao et al. (2006) Allahyar et al. (2018)</td>
</tr>
<tr>
<td>ChiA-PET</td>
<td>Chromatin interaction analysis by paired-end tag</td>
<td>Protein-centered view of genome architecture</td>
<td></td>
</tr>
<tr>
<td>3C</td>
<td>Chromosome conformation capture</td>
<td>Exploring the interaction between two genomic loci (one locus to one locus)</td>
<td></td>
</tr>
<tr>
<td>4C</td>
<td>Circular chromosome conformation capture on chip</td>
<td>Exploring all possible interactions with a single genomic locus (one locus to all loci)</td>
<td></td>
</tr>
<tr>
<td>MC-4C</td>
<td>Multicontact-4C</td>
<td>Studying multiway chromatin interaction between many genomic loci at single allele resolution</td>
<td></td>
</tr>
<tr>
<td>Hi-C</td>
<td>Genome-wide chromatin conformation capture</td>
<td>Exploring all chromatin interactions within the genome (all loci to all loci)</td>
<td>Lieberman-Aiden et al. (2009) Schoenfelder et al. (2018)</td>
</tr>
<tr>
<td>p-CHI-C</td>
<td>Promoter capture-Hi-C</td>
<td>Hi-C coupled with RNA bait capture probes targeting all promoters (promoters to all loci)</td>
<td></td>
</tr>
<tr>
<td>ATAC-seq</td>
<td>Assay for transposase-accessible chromatin using sequencing</td>
<td>Genome-wide analysis of accessible chromatin</td>
<td>Buenrostro et al. (2013)</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>Ribonucleic acid sequencing</td>
<td>Measuring steady-state mRNA levels</td>
<td>Mortazavi et al. (2008)</td>
</tr>
<tr>
<td>RIME</td>
<td>Rapid immunoprecipitation mass spectrometry of endogenous protein</td>
<td>Studying endogenous protein complexes</td>
<td></td>
</tr>
<tr>
<td>WGS</td>
<td>Whole genome sequencing</td>
<td>Providing the nucleotide sequence of the entire genome</td>
<td></td>
</tr>
<tr>
<td>CRISP-seq</td>
<td>Combining single-cell RNA-seq and CRISPR-based perturbations</td>
<td>Linking phenotypes at the level of the transcriptome with gene functions</td>
<td>Jaitin et al. (2016)</td>
</tr>
<tr>
<td>Perturb-seq</td>
<td>Combining single-cell RNA-seq and CRISPR-based perturbations</td>
<td>Same as CRISP-seq</td>
<td>Dixit et al. (2016)</td>
</tr>
</tbody>
</table>
biomarker prostate-specific antigen (PSA). PSA expression was shown to be upregulated upon androgen stimulation in LNCaP cells, implicating this gene as a bona fide AR target (Young et al. 1991, Henttu et al. 1992). In response to androgens, AR along with its coregulators (Shang et al. 2002) is recruited to cis elements upstream at the promoter (ARE I and ARE II) (Pang et al. 1995, Cleutjens et al. 1996) and enhancer (ARE III) (Cleutjens et al. 1997) regions of the KLK3 gene.

After extensive studies using chromatin immunoprecipitation assays (ChIP) to study AR chromatin binding at specific loci, technological advances enabled for ChIP assays followed by genomic tiling microarray hybridization (ChIP-on-chip) and later on massive parallel sequencing (ChIP-seq). These technological developments truly revolutionized the field, allowing for unbiased discovery of TF-binding sites on a genome-wide scale. The pioneering studies of AR ChIP-on-chip assays capturing 30Mb human genomic DNA (Takayama et al. 2007), two chromosomes (Wang et al. 2007) or the entire human genome (Wang et al. 2009) reported hormone-induced AR-DNA binding mostly located in non-promoter proximal regions far from transcription start sites (20–50kb). Later, AR ChIP-seq in LNCaP and VCaP cell lines confirmed that ~90% of the AR-binding sites can be found at distal cis-regulatory elements (Fig. 1C) (Yu et al. 2010, Massie et al. 2011). These AR-binding events, the AR cistrome, define the transcriptional response associated with AR activation.

How to identify AR target genes?

We and others defined direct AR target genes as those genes with an AR-binding site within 20–50kb of the genes (Wang et al. 2009, Sharma et al. 2013, Stelloo et al. 2015). This approach is a simple solution, but not well suited as AR-bound enhancer–target gene interaction can go far beyond a linear distance of 20–50kb as recently shown for the enhancer of the AR gene (650kb) (Takeda et al. 2018). In addition, promoter capture Hi-C in primary hematopoietic cell types revealed a median linear distance of 331kb between promoters and distal regulatory elements (Javierre et al. 2016). Furthermore, capture Hi-C and chromatin interaction analysis by paired-end tag sequencing (ChiA-PET) revealed that one single promoter can interact with multiple enhancers, but also revealed the regulation of multiple genes by the same enhancer (Fullwood et al. 2009, Javierre et al. 2016, Zhang et al. 2019). Recently, novel method have been developed such as multicontact 4C and single-cell Hi-C to rule out whether interactions co-occur simultaneously in individual cells or independently in different cells (Ramani et al. 2017, Allahyar et al. 2018). Moreover, to uncover combinatorial relationships between enhancers, several laboratories applied CRISPR genome editing or CRISPR interference (CRISPRi) to silence individual or multiple enhancer regions. These data on other steroid hormone receptors, such as the estrogen receptor alpha (ERα) and the glucocorticoid receptor (GR), provided evidence for cooperative interaction between several hormone receptor-bound enhancers (Carleton et al. 2017, Thormann et al. 2018). It is likely that AR functions in a comparable manner, given the close homology between AR, GR and ER, but conclusive AR-centered studies on chromatin looping are yet to be reported.

Enhancer–gene interactions can also be predicted in clinical specimens with assays, such as Hi-C and ATAC-seq (Schmitt et al. 2016, Beagrie et al. 2017, Corces et al. 2017, Diaz et al. 2018). ATAC-seq data from prostate cancer samples has become recently available (Corces et al. 2018), while Hi-C in clinical specimens remains to be performed. ATAC-seq is a technique to assess genome-wide chromatin accessibility, and not genetic interactions per se, but enhancer–gene interactions can be studied by correlating ATAC-seq accessibility and gene expression data (Corces et al. 2018). However, Hi-C is specifically developed to measure long-range interaction on a genome-wide scale. Hi-C datasets on prostate cell lines have been generated (Rickman et al. 2012, Taberlay et al. 2016, Luo et al. 2017), albeit without hormone stimulation. Furthermore, these datasets suffer from limited resolution, which prevents functional linkage between distal regulatory elements and their target genes. Based on data from chromosome conformation capture (3C) experiments, hormone stimulation is known to enhance AR-bound enhancer looping to its target gene promoters (Wang et al. 2005, 2007, Chen et al. 2011). This is supported by genome-wide chromatin interaction studies, showing that long-range chromatin interactions exist in the absence of hormones and that the frequency of pre-existing interactions increases upon hormone (e.g. glucocorticoids) stimulation (Kuznetsova et al. 2015, D’Ippolito et al. 2018, Le Dily & Beato 2018). It would be therefore of interest to perform for example promoter capture Hi-C, both in hormone-deprived as well as hormone-stimulated prostate cancer cell lines to characterize the AR enhancer-target gene interactome along with the potential dynamics thereof. Very recently, AR-centered chromatin interactions in VCaP cells were mapped with ChIA-PET (Zhang et al. 2019).
However, AR ChIA-PET data in the absence of hormones would not be informative, as unliganded AR resides in the cytoplasm. Due to this limitation, AR ChIA-PET data cannot be used to assess whether these long-range interaction are de novo established upon hormone stimulations or pre-exist in the absence of hormones.

A further elucidation of direct AR target genes has been performed by combining AR chromatin-binding profiles with gene expression profiling. Upon androgen stimulation, hundreds to thousands of genes are being either up- or downregulated, dependent on hormone stimulation times and bioinformatics cut-offs used (DePrimo et al. 2002, Nelson et al. 2002, Hendriksen et al. 2006, Wang et al. 2007, Massie et al. 2011, Toropainen et al. 2016). Apart from protein-coding genes, non-coding transcripts are also under direct control of AR (Wang et al. 2011, Toropainen et al. 2016). Transcripts from active enhancer regions, termed enhancer RNAs (eRNAs), have been measured using Global Run-On sequencing (GRO-seq) (Fig. 1C) (Core et al. 2008). Traditional transcriptomic profiling methods (e.g. RNA-seq) cannot successfully capture these transcripts because they are very unstable. Many studies reported that eRNA levels correlated with enhancer-related chromatin features, such as H3K27ac and H3K4me1 histone modifications, and expression level of the corresponding gene (Li et al. 2016). There are enhancer-dense regions, often referred to as super-enhancers, which have exceptionally high levels of H3K27ac as well as presence of bromodomain-containing protein 4 (BRD4), the mediator complex subunit 1 (MED1) and eRNAs (Zaber et al. 2017). Of note, the number of genome-wide AR-binding sites is considerably higher than the number of differentially expressed genes upon androgen stimulation. Moreover, merely 30% of the AR-bound enhancers show eRNA production (Toropainen et al. 2016). This suggests that only a fraction of AR-bound enhancers are functional or that these enhancers can function without eRNA transcription. However, eRNA detection might be hindered due to their instability and/or short half-life and it remains to be determined what the functional and biological contributions are for the vast majority of eRNAs.

Multiple studies reported diverse molecular mechanisms by which eRNAs impact gene expression. eRNA binding in the histone acetyltransferase (HAT) domain of CBP/p300 has been reported to induce their acetyltransferase activity, increasing H3K27ac level and gene expression (Bose et al. 2017). Furthermore, others showed eRNA (e.g. KLK3 eRNA) binding to cyclin T1, a member of the positive transcription elongation factor (P-TEFb) complex, leading to increased phosphorylation of RNA polymerase II to facilitate transcription (Zhao et al. 2016). Knockdown of KLK3 eRNA resulted in reduced enhancer–promoter interactions and decreased KLK3 gene transcription (Hsieh et al. 2014, Zhao et al. 2016). Cumulatively, these studies suggest that eRNAs are not merely by-products of enhancer activity, but have a functional role in hormone-regulated gene expression. Further studies of eRNAs are necessary to uncover which individual eRNAs carry what distinct biological functions.

Factors affecting AR chromatin binding

AR cistromic regulation is a tightly-orchestrated process, impacting AR-responsive genes that are selectively controlled in a cell type-specific manner to fulfill distinct biological functions in various cell types and tissues. In the case of prostate cancer, the AR cistrome is altered at different stages of disease, impacting expression of a selective repertoire of responsive genes in a context-and stage-specific manner. In normal prostate epithelial cells, AR regulates expression of genes involved in differentiation and growth suppression (Isaacs 2018). In contrast, AR signaling in prostate cancer stimulates proliferation and survival of tumor cells (Isaacs 2018). Consistent with the altered AR transcriptional program in tumorigenesis, recent ChIP-seq analysis in prostate tissue has revealed a unique AR cistrome in normal prostate tissue versus primary prostate cancers (Pomerantz et al. 2015). These AR cistromes are in part reprogrammed through the action of the TFs forkhead box protein A (FOXA1) and homeobox B13 (HOXB13). Overexpression of both FOXA1 and HOXB13 in normal prostate epithelial cells with exogenously introduced AR (LHSAR) resulted in redistribution of AR to genomic regions containing forkhead and homeodomain motifs, resembling an AR chromatin-binding profile as observed in human tumor specimens (Fig. 2A and B) (Pomerantz et al. 2015, Stelloo et al. 2018a). This finding is corroborated in tissue samples, where higher expression of FOXA1 and HOXB13 was observed in tumors as compared to normal tissue (Stelloo et al. 2018a). While the biological impact of FOXA1 and HOXB13 in directing AR function is slowly becoming apparent, it remains unclear how increased expression of FOXA1 and HOXB13 in prostate cancer development is regulated.
FOXA1

FOXA1 is both sufficient and essential for a prostate cancer cell to adopt a ‘tumor-like AR cistrome’, as loss of FOXA1 expression also alters the AR cistrome repertoire in prostate cancer cell lines, as reported by three groups (Sahu et al. 2011, Wang et al. 2011, Jin et al. 2014). FOXA1 is frequently mutated in primary and metastatic castration-resistant prostate cancer (Barbieri et al. 2012, Grasso et al. 2012, Cancer Genome Atlas Research Network 2015, Robinson et al. 2015, Armenia et al. 2018), but to date, it remains elusive which effects such mutations have on the AR cistrome. The mutations in the FOXA1 gene were identified around the forkhead DNA-binding (FK) domain or in the C-terminal transactivation domain with a frequency of 3–12% in both primary prostate cancers and metastases (Barbieri et al. 2012, Grasso et al. 2012, Cancer Genome Atlas Research Network 2015, Armenia et al. 2018). The FK domain is necessary for stable FOXA1 binding to nucleosomal DNA and the C-terminal domain of the protein interacts with core histones H3 and H4 to mediate chromatin accessibility (Cirillo et al. 2002). In the TCGA cohort, primary tumors with FOXA1 mutation have been classified into one of the seven molecular subtypes, hallmarked by elevated AR target gene expression and a distinct DNA methylation profile (Cancer Genome Atlas Research Network 2015). In addition, profiling of AR, H3K27ac, H3K4me3 and H3K27me3 in primary prostate tumors identified subtype-specific chromatin landscapes, with one distinct subgroup capturing all samples bearing FOXA1 mutations (Fig. 3) (Stelloo et al. 2018b). Cumulatively, these reports demonstrated that FOXA1 mutations influence AR chromatin binding and alter its downstream transcriptional profile. Further characterization of individual FOXA1 mutations is required to elucidate whether distinct mutations would differentially impact FOXA1 function and alter the AR cistrome in different ways, and whether this would impact tumorigenesis and response to hormonal therapy. In a cohort of 28 patients receiving neoadjuvant docetaxel and androgen deprivation therapy (ADT), a pathological complete response was achieved in three patients, two of which carried a FOXA1 mutation (Beltran et al. 2017). Clinical studies with information on FOXA1 mutation status are however largely underpowered, and it remains to be determined whether FOXA1 mutations would impact hormone therapy response.
HOXB13

Homeodomain-containing proteins belong to a large class of sequence-specific TFs, of which HOXB13 is the most well studied in prostate cancer. Unlike for FOXA1, no somatic mutations in HOXB13 have been detected in primary prostate cancer or in the metastatic disease. However, a recurrent germline mutation HOXB13 (G84E) has been identified (Ewing et al. 2012), which is associated with increased prostate cancer risk (Decker & Ostrander 2014). Recently, more hereditary HOXB13 mutations conferring increased risk of prostate cancer have been reported (Maia et al. 2015). The G84E mutation occurs in the conserved MEIS protein-binding domain; however, it does not affect HOXB13-MEIS1 interaction (Johng et al. 2019). Whether the G84E mutation affects the HOXB13-AR interactions remains elusive. From computational analyses, it is predicted that the mutation results in damaging and deleterious effects on HOXB13 DNA-binding capacity and increased stability of its protein structure (Chandrasekaran et al. 2017). While hereditary HOXB13 mutations have been extensively reported in tissue, it remains to be determined whether this mutation affects HOXB13 and/or AR chromatin binding in vitro. Especially since HOXB13 functions as a key upstream regulator of AR signaling, both for the full-length (Pomerantz et al. 2015) and AR-V7 splice variant (Chen et al. 2018) (Fig. 2C), deeper mechanistic understanding of HOXB13 action in the AR signaling axis would be highly instrumental for designing novel therapeutics that would perturb this route.

ERG fusions

In 2005, Tomlins et al. identified gene fusions of transmembrane protein serine 2 (TMPRSS2) with ETS-related gene (ERG) or ETS variant 1 (ETV1) in primary prostate cancers and cell lines, using expression outlier and RNA-exon quantification analyses (Tomlins et al. 2005). Fusions of AR-regulated genes with ETS family members occur in ~50% of prostate cancers, resulting in elevated expression of these ETS TFs. ETS gene fusion-positive tumors represent a molecular subtype of prostate cancer with a distinct transcriptional profile (Tomlins et al. 2007, Setlur et al. 2008, Cancer Genome Atlas Research Network 2015). ChiP-seq experiments in cell lines revealed substantial co-occupancy of ERG and AR chromatin binding, together with transcriptional corepressors enhancer of zeste homolog 2 (EZH2) and histone deacetylases (HDACs) attenuating AR target gene expression (Kunderfranco et al. 2010, Yu et al. 2010, Chng et al. 2012). However, EZH2 has recently also been reported to activate AR signaling, which is distinct from its conventional role in epigenetic silencing (Kim et al. 2018). The non-polycomb repressive complex 2 (PRC2) function of EZH2 is consistent with previous work, which reported enriched AR motifs at EZH2-binding sites depleted of the repressive histone mark H3K27me3 (Xu et al. 2012). Besides ERG-induced transcriptional repression, ERG overexpression also redistributes HOXB13, FOXA1 and AR to novel regulatory elements marked by H3K27ac and H3K4me1 (Chen et al. 2013, Kron et al. 2017). In addition, ChiP-seq profiles for AR, H3K27ac and H3K27me3 ChiP-seq profiles stratified primary prostate cancers on ERG fusion status (Fig. 3) (Kron et al. 2017, Stelloo et al. 2018b). The profoundly different ChiP-seq and transcriptome profiles are accompanied with distinct DNA methylation patterns and changes in 3D chromatin conformation (Rickman et al. 2012, Kron et al. 2013, Geybels et al. 2015). The latter is based on Hi-C analysis in the benign prostate epithelial cell line RWPE1, but recent technological advancements (Schmitt et al. 2016, Beagrie et al. 2017, Diaz et al. 2018) now enable the study of chromatin conformation in human tissue to reflect a more clinically relevant scenario.
The widespread chromatin and transcriptional alterations as seen in ERG fusion-positive prostate cancers render ERG and its downstream targets highly interesting targets for therapy (Kron et al. 2017, Wang et al. 2017).

**CHD1**

One other frequent alteration reported in primary prostate cancer is loss of *chromodomain helicase DNA-binding protein 1* (*CHD1*), which is mostly observed in ERG fusion-negative tumors and not ERG fusion-positive tumors. The mutual exclusivity of *CHD1* loss with ERG rearrangements might be explained by the fact that loss of *CHD1* acts synthetically lethal with loss of *phosphatase and tensin homolog* (*PTEN*) in human prostate cancer cell lines (Zhao et al. 2017), and *PTEN* loss frequently co-occurs with *ERG* gene fusion (Taylor et al. 2010). However, this result was not observed in murine prostates, where loss of *CHD1* and *PTEN* led to the development of invasive carcinoma, while *CHD1* loss alone was insufficient to drive tumorigenesis (Augello et al. 2019). The cause of this discrepancy is unknown, although speculative it may arise from mouse/human species differences.

Knockdown of *CHD1* prevents androgen-induced chromosomal rearrangement resulting in the absence of TMPRSS2-ERG fusion transcripts (Burkhardt et al. 2013, Metzger et al. 2016). This finding might be explained by the role of CHD1 in DNA repair. Both non-homologous end joining (NHEJ) and homologs recombination (HR) DNA repair processes are involved in transcription-associated DNA damage, but which repair pathway is affected by CHD1 is still under debate (Hedayati et al. 2016, Kari et al. 2016, Shenoy et al. 2017). In addition to its role in DNA repair, CHD1 maintains chromatin in an accessible state for AR and other TF to bind. Knockdown of CHD1 impairs AR recruitment to promoters of AR-responsive genes (*KLK3, TMPRSS2, FBI2, ELK4* and *KLK2*) (Burkhardt et al. 2013). The genome-wide experiments further demonstrated overlapping CHD1- and AR chromatin-binding sites (Metzger et al. 2016, Augello et al. 2019). In addition, a large proportion of the CHD1 chromatin-associated protein complex represents members of the AR interactome, suggesting a role of CHD1 in AR signaling. Indeed, loss of CHD1 was recently reported to rewire AR signaling by redistributing AR binding to HOXB13-enriched sites (Augello et al. 2019). This was confirmed with AR ChIP-seq on tissue samples deficient for CHD1 (Fig. 3) (Augello et al. 2019).

Previous work demonstrated that a substantial number of AR-binding sites are accessible prior to hormone stimulation, while other sites become accessible after hormone stimulation (He et al. 2012, Tewari et al. 2012), providing a plausible explanation why not all AR sites are affected by CHD1 loss. However, whether tumors with loss of CHD1 have distinct nucleosome remodeling patterns needs to be determined.

**Others**

Genetic alterations in other epigenetic regulators, including *Lysine (K)-specific methyltransferase 2C* (*KMT2C*) and *zinc finger MYM-type containing 3* (*ZMYM3*), are less frequent in primary prostate cancer (Barbieri et al. 2012, Cancer Genome Atlas Research Network 2015). Nonetheless, the interplay between AR and epigenetic pathways as well as studies on other hormone receptors hint toward a role of these proteins in AR biology. For example, knockdown of KMT2C in breast cancer cells has been shown to suppress estrogen-dependent gene expression via loss of H3K4me1 and H3K27ac selectively at Eρα enhancers (Jozwik et al. 2016, Gala et al. 2018). However, KMT2C knockdown did not result in a major loss of Eρα chromatin binding but likely impairs recruitment of cofactors to the Eρα complex (Gala et al. 2018). Indeed, the recruitment of cofactors to AR-bound enhancers is crucial for AR-mediated transcriptional activation or repression. New technologies such as rapid immunoprecipitation mass spectrometry of endogenous protein (RIME) (Mohammed et al. 2013) led to increased knowledge of the AR-enhanceosome – the protein complex at AR-bound enhancers. We and others have identified novel AR protein interactions in prostate cancer cell lines (Paltoglou et al. 2017, Stelloo et al. 2018a); however, the functional impact of all newly identified AR interaction partners still needs to be unraveled. A recently developed technique, quantitative multiplexed method (qPLEX-RIME) (Papachristou et al. 2018), may shed light on the composition of these AR multiprotein complexes in human tissue samples.

**Factors affecting AR chromatin binding in castration-resistant prostate cancer**

Castration-resistant prostate cancer (CRPC) refers to the disease stage in which the tumor progresses despite reduced serum testosterone levels due to ADT. As a consequence of selection pressure from ADT or anti-androgens on AR signaling axis, AR-dependent resistance mechanisms emerge including AR amplification, overexpression, somatic point mutations and constitutively active splice
variants (Claessens et al. 2014). Moreover, genome-wide selectivity of AR-binding patterns has been reported in CRPC (Sharma et al. 2013, Stelloo et al. 2015), suggesting CRPC-specific factors to configure AR reprogramming.

**AR amplifications**

AR amplifications are significantly enriched in metastatic CRPC as compared to primary prostate cancer. These AR amplifications involve copy number gain of the AR enhancer (Fig. 2E), the AR gene or both resulting in increased AR expression (Takeda et al. 2018, Viswanathan et al. 2018). In vitro, overexpression of AR allows the cells to activate AR signaling in low androgen environment resulting in increased AR chromatin-binding sites and downstream regulated genes (Urbanucci et al. 2012). Whether this effect is also observed in clinical samples remains to be determined.

**AR splice variants**

Besides AR amplification, treatment selection pressure by ADT induces aberrant AR mRNA splicing. Over 20 AR variants are described to date, including the most intensively studied variant AR-V7. AR-V7 is hardly expressed in primary prostate cancer samples and associated with increased total AR expression (Sharp et al. 2018). As overall AR is often overexpressed in CRPC, AR-V7 is consequently more frequently observed in CRPC samples as well. Most models studied thus far either express full-length AR or a variant, which does not represent the clinical situation where different types of AR-Vs are often simultaneously coexpressed with full-length AR. With the development of AR-V specific antibodies, it is now possible to perform ChIP-seq with AR-V-specific antibodies and C-terminal antibodies to distinguish the variant from the full length in more physiologically relevant model systems as well as in human specimens (Chen et al. 2018).

As AR-V7 lacks the ligand-binding domain, the receptor is constitutively active, with nuclear localization and chromatin binding in the absence of ligand. AR-V7 mostly binds regions with canonical AREs, as a homodimer (Fig. 2C) (Chan et al. 2015, He et al. 2018). In addition, AR-V7 and full-length AR have been reported to heterodimerize, as determined with ChIP-re-ChIP experiments (Cao et al. 2014) and co-immunoprecipitations (He et al. 2018). AR-V7 chromatin binding appeared to be less dependent on FOXA1, but rather selectively dependent on HOXB13 for its chromatin interactions instead (He et al. 2018).

In line with this, silencing of HOXB13 significantly decreases AR-V7 chromatin binding (Chen et al. 2018). Whether this is unique to AR-V7 could not be addressed, as full-length AR protein levels were significantly decreased in HOXB13-silenced LNCaP cells causing examination of full-length AR cistrome non-informative (Pomerantz et al. 2015).

Multiple reports have shown AR-V7 preferential chromatin-binding sites and differential gene expression as compared to full-length AR (Hu et al. 2012, Cao et al. 2014, Krause et al. 2014, Lu et al. 2015, He et al. 2018), but this was contradicted by one other study (Chan et al. 2015). These AR-V7 gene signatures as well as the expression of AR-V7 itself are associated with disease progression (Lu et al. 2015, Sharp et al. 2018). However, it remains unclear whether AR-V7 contributes to resistance. Although AR-V7 is a predictive biomarker for conventional AR-targeted therapies, most likely because cells expressing AR full length are sensitive to castration and antiandrogen treatment, whereas AR-V7-expressing cells are indifferent to therapy (Guo et al. 2009, Chan et al. 2015). Therefore, the next challenge is discovering and testing AR-V7-targeted therapies, such as alternative druggable regions on AR or by targeting its cofactors. Dalal et al. recently published compounds that block the AR-DNA-binding domain (DBD) dimerization interface, inhibiting both AR and AR variant transcriptional activity (Dalal et al. 2018). Targeting other functional domains such as the transactivation domain are promising as well, for example, the EPI compounds of which EPI-506 is currently evaluated in a phase I dose escalation clinical trial (NCT02606123) (Andersen et al. 2010). Besides small molecules, an alternative approach can be the expression of decoy peptides representing the AR N-terminal domain. These decoys have been shown to inhibit both androgen-dependent prostate cancer and CRPC growth, most likely by competitive binding to AR-interacting proteins required for AR transcriptional activity (Quayle et al. 2007, Myung et al. 2017). In fact, many more inhibitors targeting the AR have been reported and are reviewed by others (Centenera et al. 2018).

**Single nucleotide variants in enhancers**

Another level of AR cistrome deregulation is reflected in single nucleotide variants (SNVs), including SNPs and somatic point mutations at cis-regulatory elements (Fig. 2E). With the growing number of genome-wide association studies (GWAS) and whole genome
sequencing (WGS) efforts, millions of SNVs have been identified, most with unknown significance and >90% found in non-protein regions (1000 Genomes Project Consortium et al. 2012, Baca et al. 2013, Espiritu et al. 2018, Quigley et al. 2018, Schumacher et al. 2018). Interestingly, prostate cancer risk SNPs are found enriched at enhancer elements including those bound by AR (Lu et al. 2012, Bu et al. 2016, Whittington et al. 2016, Dadaev et al. 2018). Causal regulatory SNPs altering AR binding has been demonstrated in the enhancer of sex determining region Y box 9 (SOX9) gene (Zhang et al. 2012), an intron of the melanophilin (MLPH) gene (Bu et al. 2016) and the promoter of KLK3 (Lai et al. 2007). Moreover, an increased mutational load proximal to AR-binding sites and other transcriptional regulators such as HOXB13 and FOXA1 has been reported (Mazrooei et al. 2018, Morova et al. 2017). However, only ~20% of the somatic SNVs which lie within regulatory elements marked with H3K27ac and binding of AR, FOXA1, HOXB13 are predicted to impact binding of these TFs (Mazrooei et al. 2018). Knowing whether a regulatory SNV potentially disrupts or creates new binding sites would aid to elucidate their target genes and their potential functional impact on prostate cancer. In addition, in vitro validation of these variants would be crucial to distinguish driver events from non-pathogenic passenger variants. As it is difficult to study the consequence of variants on TF binding in clinical samples, a functional impact of variants is generally studied by computational approaches and/or in vitro experimental methodologies. Given the growing number of genomic datasets (e.g. ATAC-seq, WGS, ChIP-seq) on clinical samples, it may become possible in the future to functionally connect variants with TF binding in human tumor samples, although larger sample sizes with matched genomic datasets would be required. Using computational approaches, TF-DNA-binding specificity can be estimated based on deviations of the position weight matrices (PWM) (Fig. 1B), which is a matrix of scores which correspond to the frequencies of the nucleotides at each position in the TF motif (Deplancke et al. 2016) Other methods rely on integration of multidimensional data including accessible chromatin regions, chromatin marks and TF-binding profiles (Deplancke et al. 2016). Furthermore, to functionally validate the significance of SNVs on TF binding in vitro, typically assays such as electrophoretic mobility shift assays (EMSA) are used, aimed to study DNA–protein interactions as well as luciferase reporter assays. Recently, the development of a massively parallel reporter assay made it possible to test different SNVs simultaneously, as shown for red blood cell trait-associated variants (Ulirsch et al. 2016). Additionally, the advent of CRISPR technology enables the study of variants in a native genomic context either by editing one single locus or by performing enhancer screens. Using the latter strategy, Korkmaz et al. identified ERα-binding sites that are required for breast cancer proliferation (Korkmaz et al. 2016). Given the work on ERα, the very same approach can be applied to analyze AR-binding sites and thereby infer causal SNVs. Alternatively, one can envision a genetic screen targeting AR-bound enhancers with transcriptional profiles as readout to infer enhancer-target gene pairs with methods such as Perturb-seq (Dixit et al. 2016) or CRISP-seq (Jaitin et al. 2016) which combine CRISPR-based perturbations with single-cell RNA-seq. This framework has recently been shown to be successful in detecting enhancer-gene pairs (Xie et al. 2017, Gasperini et al. 2019). Such efforts for AR-binding sites will characterize downstream biological consequences of genetic variants.

Larger cistromic datasets

In this review, several factors have been described that influence AR-binding patterns in prostate cancer. While original reports were largely focused on cell lines, gradually an increasing body of evidence is being generated that study AR chromatin binding in clinical specimens. This transition to tumor material is of added value for multiple reasons. First, cell lines are (albeit powerful and extremely useful) just a model that would always require further validation in a physiologically more-relevant setting. Secondly, in order to draw conclusions in cell lines for eventual translation of such concepts to clinical implementation, validation of such basic biology observations in tissue specimens is a crucial intermediate step. Lastly, as prostate cancer is considered a heterogeneous disease with multiple genomic (Cancer Genome Atlas Research Network 2015) and epigenetic (Stelloo et al. 2018b) subtypes, and the biological complexity of the disease is not fully covered by the limited set of available cell line models.

Large-scale ChIP-seq datasets on patient cohorts can help to better understand prostate cancer subtype-specific chromatin landscapes in a tissue context. Many AR-binding sites are patient specific (inter-individual heterogeneity), but defining common features of chromatin landscapes shared among tumors is helpful for defining subtypes. Determining subtype-specific chromatin landscapes requires differential binding algorithms such as DiffBind
(Ross-Innes et al. 2012), a ranking normalization strategy (Patten et al. 2018), ChromHMM followed by MCA (multiple correspondence analysis) factorial analysis (Lomberk et al. 2018) or matrix factorization algorithms (Stelloo et al. 2018b). New bioinformatic tools and pipelines continue to emerge, aimed at ingrating both epigenetic and genomics data to better understand specific phenotypes. Analysis of 100 primary prostate cancers revealed a relationship between chromatin landscapes and their regulation of gene expression associated with three subtypes (Fig. 3) (Stelloo et al. 2018b). In addition, two main prostate cancer subtypes related to ERG gene fusions were identified. These results are consistent with previous prostate cancer profiling studies (see above; ERG), but the results also provided unique insights, such as the discovery of a third subtype. Tumors classified in this subtype have a low AR pathway activity as reflected by the low number of AR-binding sites and low AR activity score. This opens new avenues for therapeutic decisions as tumors with low AR activity score are more prone to develop resistance to hormone therapy (Spratt et al. 2018).

Whereas the studies so far focused on inter-individual chromatin landscape variability, studying one tumor/single foci of one patient, the intra-tumor heterogeneity remains to be explored. This is of special interest as the majority of primary prostate cancers harbor multiple tumor foci (Andreoiu & Cheng 2010). These foci revealed differences on the level of gene expression, DNA methylation, copy number and mutation status (Boutros et al. 2015, Mundbjerg et al. 2017, Espiritu et al. 2018), suggesting that the observed variation might also be reflected in the chromatin landscape. Improvement of ChIP efficiency on limited quantities of tissues (Dahl & Gilfillan 2018, Singh et al. 2019) holds great promise to study multiple clones and also infer relationship between primary tumor foci and metastatic lesions with respect to the chromatin landscape. Another undoubted advancement is the usage of formalin-fixed paraffin-embedded (FFPE) specimens for ChIP (Fanelli et al. 2010, Cejas et al. 2016), which has been successfully applied to study histone marks but not yet TFs.

### AR cistrome-based classifiers for patient stratification

Even though AR chromatin binding in primary prostate cancer specimens is a determinant of the genomic landscape of prostate cancer subtypes, these binding sites did not appear to be associated with patient outcome (Stelloo et al. 2018b). However, a subset of AR-binding sites, identified through differential binding analysis in benign versus primary prostate cancer or primary prostate cancer versus progressive disease bear prognostic potential (Sharma et al. 2013, Chen et al. 2015, Pomerantz et al. 2015, Stelloo et al. 2015). To be more specific, actually the genes proximal to these AR-binding sites were tested and validated for their prognostic potential. Even though promising, as for example, our 9-gene expression signature was able to identify patients with an ultra-high risk to develop biochemical recurrence after radical prostatectomy (Stelloo et al. 2015); none of these classifiers are used in daily clinical practice yet. Application of classifiers based on differentially bound regions rather than genes would be costly and laborious, limiting the application of ChIP-based classifiers in clinical practice. The establishment of an automated workflow for both ChIP and library preparation would render the approach more cost-effective and robust. For example, a method such as ChIP-string, which is a modified NanoString’s nCounter Analysis System platform to measure enriched genomic DNA of loci of interest from ChIP experiments (Ram et al. 2011), would be more suitable for clinical application.

Clinically useful interpretation of cistrome-based classifiers requires appropriate pathology review, as bulk tumor and normal tissue samples will have varying degrees of epithelial content and stromal contamination. Fibroblasts, the most abundant cell type in stroma, express AR albeit at lower levels as compared to prostate epithelial cells (Leach et al. 2017, Cioni et al. 2018, Nash et al. 2018). In fibroblasts, it was shown that AR chromatin binding is disparate from epithelial AR cistrome, with preferential AR binding at sites enriched for motifs of FOS and JUN family members (Leach et al. 2017, Cioni et al. 2018). Knockdown of Jun reduced AR binding at three loci (Leach et al. 2017), but whether this affects global AR binding remains to be addressed. However, the data suggest that AR binding in prostate fibroblasts is dependent on the API1 complex, a complex consisting of homo/heterodimer combinations of FOS and JUN members, rather than classical pioneer factors such as FOXA1 (Fig. 2D). This phenomenon of enrichment of other motifs than AREs, forkhead and homeodomain motifs, is also seen in AR ChIP-seq on human caput epididymis (Yang et al. 2018) and mouse kidney and epididymis (Pihlajamaa et al. 2014). These observed tissue-specific AR cistromes achieved by collaborating factors further supports a rationale to study AR cistromes in other cells of the tumor microenvironment.
Future perspectives

Although ChIP-seq studies using human specimens have provided a more detailed understanding of AR chromatin binding in prostate cancer, a lot remains to be explored (Fig. 4). Comparative studies of AR cistromes can be expanded to metastatic lesions and larger cohorts to provide clues on AR functioning and its target genes during prostate cancer progression. Collecting the primary tumor(s) as well as the metastases during disease progression and at time of death is of great interest, enabling the identification of AR cistrome characteristics of the primary tumor lesion that ultimately gave rise to the lethal metastatic lesion. Despite limited to one case, Haffner et al. defined the origin of metastatic lesions by performing WGS on primary prostate foci and metastasis (Haffner et al. 2013). In addition, evaluation of AR cistromes from samples collected before, on and after treatment (NCT03297385), as well after acquisition of treatment resistance will provide valuable information to better understand how AR signaling is associated with disease progression. We can learn from research on ERα chromatin binding in breast specimens before and after neoadjuvant tamoxifen treatment (Severson et al. 2016), where drug-induced changes in ERα chromatin binding were observed.

Another interesting approach, applicable for patients with large burden of ctDNA, could be identification of nucleosome footprints in ctDNA. Analysis of two plasma samples taken 12 months apart, during which the prostate adenocarcinoma transdifferentiated to a treatment-emergent small-cell neuroendocrine prostate carcinoma (t-SCNC, AR-independent disease state), showed no longer accessibility of AR-binding sites during the disease course (Ulz et al. 2018). Given the paucity of serial sample collections at different points during the disease course, alternative models such as patient-derived xenografts (PDX), ex vivo cultures and/or cell lines are required to monitor therapy effect on transcription factor dynamics.

The analysis of 100 primary prostate cancers is the largest AR cistromic dataset on prostate cancer (Stelloo et al. 2018b) and no such large cistromic data sets of ERα in breast cancer exist yet. Future research of ideally larger sample sizes for breast cancer may be considered to study cistrome differences underlying clinical and genetic features of breast cancer subtypes.

In the era of single-cell genomics, advanced approaches enable complete characterization of the genome, transcriptome and epigenome at the single-cell level. This will enable us to assess cellular heterogeneity, cells in the tumor microenvironment and clinical specimens with too low tumor cell percentage for bulk sequencing. Also, a single-cell ChIP-seq (scChIP-seq) protocol has been published (Rotem et al. 2015), but is yet to be validated. With improvement in throughput, quantitation and sensitivity, scChIP-seq might become more valuable to analyze cell-to-cell variability in chromatin landscapes across a population. Ultimately, combining different single-cell methods to investigate for example epigenetic and transcriptome correlation...
and how these are associated with cellular phenotypes will lead towards a comprehensive understanding of the behavior of individual cells in prostate cancer.

Conclusions

To date, the vast majority of functional genomic studies on AR have used a limited panel of cell line models. Due to improved technological advances, a comprehensive understanding of AR together with a detailed picture of the epigenomic landscapes in prostate cancer specimens is beginning to emerge. Research has highlighted the involvement of numerous features beyond sequence motifs alone to be involved in AR cistrome reprogramming, such as ERG fusions and TF overexpression (e.g. FOXA1 and HOXB13). Further characterization of mechanisms underlying AR reprogramming during prostate tumorigenesis and progression – from normal to prostatic intraepithelial neoplasia, metastatic and therapy-resistant disease – will enable identification of affected target genes. Identification of those genes may enhance the translational potential of ChIP-seq data. This will stimulate the discovery of promising targets for therapeutic intervention and new biomarkers of early detection and resistance.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding

This work was supported by funding from Movember (NI1001), KWF Dutch Cancer Society (10084 ALPE), KWF Dutch Cancer Society/Alpe d’HuZes Bas Mulder Award (NIK 2014-6711) and a VIDI grant (016.156.401) from The Netherlands Organisation for Scientific Research (NWO).

Acknowledgements

The authors would like to thank Simon Linder for critical review of the manuscript.

References


Chandrasekaran G, Hwang EC, Kang TW, Kwon DD, Park K, Lee JJ & Lakshmanan VK 2017 Computational modeling of complete HOXB13 protein for predicting the functional effect of SNPs and the associated role in hereditary prostate cancer. Scientific Reports 7 43830. (https://doi.org/10.1038/srep43830)


Henttu P, Liao SS & Vlhko P 1992 Androgens up-regulate the human prostate-specific antigen messenger ribonucleic acid (mRNA), but down-regulate the prostatic acid phosphatase mRNA in the LNCaP cell line. *Endocrinology* **130** 766–772. (https://doi.org/10.1210/endo.130.2.1370795)


Ken-ichi T 2018 The biological role of androgen receptor in prostate cancer progression: advances in testosterone action. *InTechOpen* [epub]. (https://doi.org/10.5772/intechopen.76360)


subclass. *Journal of Clinical Oncology* **36** (Suppl 2) 2. (https://doi.org/10.1200/JCO.2018.36.6_suppl.2)


Zhang X, Cowper-Sal Iar R, Bailey SD, Moore JH & Lupien M 2012 Integrative functional genomics identifies an enhancer looping to


Received in final form 25 February 2019
Accepted 13 March 2019
Accepted Preprint published online 13 March 2019