Genome-wide Analysis of Androgen Receptor Function in Prostate Cancer

Biswa Jayoti Sahu

Helsinki 2012
Genome-wide Analysis of Androgen Receptor Function in Prostate Cancer

Biswa Jayoti Sahu

Institute of Biomedicine, Physiology, Faculty of Medicine
Helsinki Graduate Program in Biotechnology and Molecular Biology
University of Helsinki
Finland

ACADEMIC DISSERTATION

To be publicly discussed with the permission of the Faculty of Medicine, University of Helsinki, in Lecture Hall 3, Biomedicum Helsinki 1, on 19th October, 2012 at noon

Helsinki 2012
Supervisor

Prof. Olli A. Jänne, M.D., Ph.D.
Institute of Biomedicine, Biomedicum Helsinki
University of Helsinki
Finland

Thesis follow-up committee

Prof. Pirkko Vihko, M.D., Ph.D.
Institute of Clinical Medicine
University of Helsinki
Finland

Docent Outi Monni, Ph.D.
Institute of Biomedicine
University of Helsinki
Finland

Pre-examiners appointed by the faculty

Prof. Tomi P. Mäkela, M.D., Ph.D.
Institute of Biotechnology
University of Helsinki
Finland

Prof. Marja T. Nevalainen, M.D., Ph.D.
Kimmel Cancer Center
Thomas Jefferson University
Philadelphia, USA

Opponent appointed by the faculty

Dr. Jason S. Carroll, Ph.D.
Cancer Research UK Cambridge Research Institute
University of Cambridge
United Kingdom

© Biswajyoti Sahu 2012
ISBN 978-952-10-8182-8 (PDF)
ISSN 1457-8433
http://ethesis.helsinki.fi
Unigrafia Oy

Front Cover: Snapshot of ChIP-seq binding sites and AR, FOXA1 expression in primary adenocarcinomas (Sahu et al., 2011. EMBO J. 30, 3962-3975)

Back Cover: Co-crystal structure of the HNF-3/fork head DNA-recognition motif Courtesy: RCSB Protein Data Bank (PDB ID 1VTN)
(Clark et al., 1993. Nature 364, 412-420)
To my beautiful daughter Tanvi (Feby)
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ABSTRACT</strong></td>
<td>3</td>
</tr>
<tr>
<td><strong>YHTEENVETO</strong></td>
<td>5</td>
</tr>
<tr>
<td><strong>LIST OF ORIGINAL PUBLICATIONS</strong></td>
<td>7</td>
</tr>
<tr>
<td><strong>ABBREVIATIONS</strong></td>
<td>8</td>
</tr>
<tr>
<td><strong>INTRODUCTION</strong></td>
<td>10</td>
</tr>
<tr>
<td><strong>REVIEW OF THE LITERATURE</strong></td>
<td>11</td>
</tr>
<tr>
<td>1. Gene transcription and transcriptional regulation</td>
<td>11</td>
</tr>
<tr>
<td>1.1 Overview of gene transcription</td>
<td>11</td>
</tr>
<tr>
<td>1.2 Basal transcriptional machinery</td>
<td>12</td>
</tr>
<tr>
<td>1.3 Nucleosomes and chromatin structure</td>
<td>12</td>
</tr>
<tr>
<td>1.4 Chromatin modifications and functions</td>
<td>13</td>
</tr>
<tr>
<td>1.5 Genome-wide analysis of gene regulation</td>
<td>16</td>
</tr>
<tr>
<td>2 Steroid receptor-mediated signaling</td>
<td>19</td>
</tr>
<tr>
<td>2.1 Nuclear receptors and coregulators</td>
<td>19</td>
</tr>
<tr>
<td>2.2 Steroid hormones and steroid receptors</td>
<td>21</td>
</tr>
<tr>
<td>2.3 Glucocorticoid receptor</td>
<td>23</td>
</tr>
<tr>
<td>2.4 Pioneer factors for steroid receptors</td>
<td>24</td>
</tr>
<tr>
<td>2.5 Forkhead box protein A1 (FoxA1) as a pioneer factor</td>
<td>25</td>
</tr>
<tr>
<td>3 Androgens and androgen receptor</td>
<td>28</td>
</tr>
<tr>
<td>3.1 Physiological androgens</td>
<td>28</td>
</tr>
<tr>
<td>3.2 Androgen receptor gene and its regulation</td>
<td>29</td>
</tr>
<tr>
<td>3.3 Modifications of androgen receptor protein</td>
<td>30</td>
</tr>
<tr>
<td>3.4 Modular structure and functional domains of androgen receptor</td>
<td>32</td>
</tr>
<tr>
<td>3.5 Androgen receptor mediated transcriptional regulation</td>
<td>35</td>
</tr>
<tr>
<td>4 Androgen receptor in prostate cancer (PCa) and disease</td>
<td>38</td>
</tr>
<tr>
<td>4.1 Cancer of the prostate gland</td>
<td>38</td>
</tr>
<tr>
<td>4.2 Androgen signaling in prostate cancer</td>
<td>39</td>
</tr>
<tr>
<td>4.3 Androgen signaling in castration-resistant prostate cancer</td>
<td>40</td>
</tr>
<tr>
<td>4.4 Endocrine treatments of prostate cancer</td>
<td>41</td>
</tr>
<tr>
<td>4.5 Androgen receptor and other related disorders</td>
<td>43</td>
</tr>
<tr>
<td><strong>AIMS OF THE STUDY</strong></td>
<td>44</td>
</tr>
<tr>
<td><strong>MATERIALS AND METHODS</strong></td>
<td>45</td>
</tr>
<tr>
<td>1. Cell lines, xenografts and patient samples</td>
<td>45</td>
</tr>
<tr>
<td>1.1 Cell lines and cell culture procedure</td>
<td>45</td>
</tr>
<tr>
<td>1.2. Prostate cancer patient and xenografts material</td>
<td>46</td>
</tr>
</tbody>
</table>
Contents

1.3 Tissue microarray construction (TMA) and immunohistochemistry 46

2. Gene expression profiling and RNA interference 47
  2.1 RNA isolation, cDNA synthesis and quantitative RT-PCR 47
  2.2 Gene expression profiling and data analysis 47
  2.3 RNA interference (RNAi) 48
  2.4 Immunoblotting 48

3. Genome-wide approaches for gene regulation 49
  3.1 Chromatin immunoprecipitation (ChIP) and ChIP-qPCR 49
  3.2 ChIP-sequencing assay 50
  3.3 DNasel-hypersensitive site sequencing (DHS-seq) 50
  3.4 Formaldehyde-assisted isolation of regulatory elements (FAIRE) 51
  3.5 Computational data analysis for next-generation sequencing data 51

RESULTS AND DISCUSSION 53

1. Androgen receptor (AR) cistromes in prostate cancer cells and role of FoxA1 in prostate cancer (Studies I and II) 53
   1.1 Ligand as regulators of AR cistrome and transcriptome 53
   1.2 AR and FoxA1 cistromes in prostate cancer cells 54
   1.3 FoxA1 reprograms AR cistrome and transcriptome 54
   1.4 Correlation of AR, FoxA1 with active chromatin marks 56
   1.5 FoxA1 reprograms glucocorticoid signaling 57
   1.6 Relationship between FoxA1 expression and PCa outcome 58

2. FoxA1 is a key determinant of unique receptor binding events in prostate cancer cells (Studies I and II) 61
   2.1 AR cistromes in LNCaP-1F5 and VCaP prostate cancer cells 61
   2.2 AR and GR cistromes in LNCaP-1F5 cells 62
   2.3 Comparison of AR and GR cistromes in VCaP cells 64

3. Increased AR expression enhances receptor binding events in prostate cancer cells (Study III) 66
   3.1 ChIP-seq mapping of AR cistromes in AR overexpressing cells 66
   3.2 Identification of androgen-regulated target genes in CRPC 67

SUMMARY AND CONCLUSIONS 69

ACKNOWLEDGEMENTS 72

REFERENCES 74
The androgen receptor (AR) mediates the effects of the two physiological male sex hormones (androgens) – testosterone and 5α-dihydrotestosterone (DHT) – that are critical in the development and maintenance of the male sexual characteristics. Androgens and AR signaling play an important role in the growth of prostate gland and in the development and progression of prostate cancer (PCa). AR is a member of the steroid receptor family and is a ligand-inducible transcription factor. Upon binding to male sex steroids, AR acquires a new conformational state, translocates to the nucleus, homodimerizes and binds to regulatory regions called androgen-response elements on target cell chromatin, as well as interacts with coregulatory proteins, sequence-specific transcription factors, and the basal transcription machinery to regulate gene expression (transcription). AR regulates distinct transcriptional programs in androgen-dependent and castration-resistant (androgen-independent) prostate cancer (CRPC). Gene transcription is a well-orchestrated process of multiple events involving chromatin modifiers, coregulators and collaborating transcription factors, occurring in an ordered and sequential fashion. This strictly regulated process is tightly controlled at every stage to govern cell growth and homeostasis. Details of the ways by which AR regulates gene expression in a genome-wide fashion are still not clear. This warrants an in-depth study of the molecular basis of genome-wide AR function in response to cognate ligands under androgen-dependent and CRPC conditions, their recruitment to specific genomic sites and the role of other transcription factors, chromatin modifiers and coregulators in modulating AR function.

This study aimed at investigating genome-wide AR binding sites in prostate cancer cells by using ChIP-sequencing technique to understand the androgen-regulated gene expression. Genome-wide AR binding sites (ARBs) in the presence of physiological androgens and partial agonist/antagonists were profiled and respective downstream target genes were characterized. Particular emphasis was paid to investigate the role of the collaborating transcription factor FoxA1 in mediating recruitment of AR to the chromatin and in androgen signaling, and prognostic value of FoxA1 expression in PCa. Furthermore, we explored the role of FoxA1 in defining cell type-specific recruitment of AR in two different PCa cell lines, and its influence on defining site-specific occupancy
of two steroid receptors, namely AR and glucocorticoid receptor (GR), in PCa cells. Finally, the consequences of AR overexpression and its repercussions on genomic recruitment of AR to chromatin were investigated.

Our work showed that nuclear receptors, such as AR and GR, employ distal modes of transcriptional regulation, as majority of the binding sites are located far from the transcription start sites. The physiological androgen DHT was more potent in recruitment of AR to the chromatin and mediating subsequent gene expression than partial androgen agonist/antagonists. This demonstrates that agonist-bound AR forms a more favorable conformation compared to partial agonist/antagonists-bound AR, and the resulting binding profiles as well as the transcription programs exhibit quantitative differences. The genome-wide DHT-bound ARBs were significantly over-represented for FoxA1 cis-element, suggesting a global role of FoxA1 in AR function. The identification of genome-wide FoxA1 binding sites correlating with AR binding sites led to a paradigm shift in understanding the role of FoxA1 in androgen signaling. FoxA1 has previously been considered as a pioneering factor with the ability to bind compact chromatin and prime it for nuclear receptor binding. Our findings reveal that FoxA1 has a dual role in regulating AR functions and we defined three distinct classes of ARBs in PCa cells that were commensurate with the androgen-regulated transcription programs. Importantly, FoxA1 levels significantly correlate with prostate cancer-specific disease survival as low FoxA1 levels predict good clinical outcome in primary PCa patients. Further analysis of the ARBs in two different PCa cell lines revealed that FoxA1 defines unique receptor binding sites in each cell line. Moreover, AR and GR binding specificity to chromatin was determined by FoxA1, when analyzed in a single PCa cell line in response to their respective ligands. Furthermore, the results also highlight the role of glucocorticoid as an anti-androgen, which attenuates AR-dependent transcription programs. In AR overexpressing PCa cells, we demonstrated sensitization to low levels of androgens due to enhancement in receptor binding. The association between AR levels and chromatin occupancy was also validated in two PCa xenografts, one with high and another with low AR expression levels.
**YHTEENVETO**


---

Yhteenveto


Yhteenveto

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles that are referred to in the text by their Roman numerals.


II. **Biswajyoti Sahu**, Marko Laakso, Päivi Pihlajamaa, Kristian Ovaska, Ievgenii Sinielnikov, Sampsa Hautaniemi and Olli A. Jänne (2012) FoxA1 is a key determinant of unique androgen and glucocorticoid receptor binding events in prostate cancer cells. (Submitted)


Original publication III was also included in the thesis “Overexpression of androgen receptor in prostate cancer” by Alfonso Urbanucci, University of Tampere.

The original publications are reproduced with the permission of the copyright holders.
# Abbreviations

## ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTH</td>
<td>adrenocorticotropic hormone</td>
</tr>
<tr>
<td>AF</td>
<td>activation function</td>
</tr>
<tr>
<td>ADT</td>
<td>androgen-deprivation therapy</td>
</tr>
<tr>
<td>AIS</td>
<td>androgen insensitivity syndrome</td>
</tr>
<tr>
<td>AP</td>
<td>activator protein</td>
</tr>
<tr>
<td>AR</td>
<td>androgen receptor</td>
</tr>
<tr>
<td>ARBs</td>
<td>androgen receptor binding sites</td>
</tr>
<tr>
<td>ARE</td>
<td>androgen response element</td>
</tr>
<tr>
<td>BPH</td>
<td>benign prostate hyperplasia</td>
</tr>
<tr>
<td>CBG</td>
<td>corticosteroid binding globulin</td>
</tr>
<tr>
<td>ChIP</td>
<td>chromatin immunoprecipitation</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>chromatin immunoprecipitation coupled to deep sequencing</td>
</tr>
<tr>
<td>CPA</td>
<td>cyproterone acetate</td>
</tr>
<tr>
<td>CRPC</td>
<td>castration-resistant prostate cancer</td>
</tr>
<tr>
<td>DBD</td>
<td>DNA-binding domain</td>
</tr>
<tr>
<td>Dex</td>
<td>dexamethasone</td>
</tr>
<tr>
<td>DHS-seq</td>
<td>DNaseI-hypersensitive site sequencing</td>
</tr>
<tr>
<td>DHT</td>
<td>5α-dihydrotestosterone</td>
</tr>
<tr>
<td>ER</td>
<td>estrogen receptor</td>
</tr>
<tr>
<td>ERE</td>
<td>estrogen response element</td>
</tr>
<tr>
<td>eRNA</td>
<td>enhancer RNA</td>
</tr>
<tr>
<td>FAIRE</td>
<td>formaldehyde-assisted isolation of regulatory elements</td>
</tr>
<tr>
<td>FoxA1</td>
<td>forkhead box A1</td>
</tr>
<tr>
<td>GR</td>
<td>glucocorticoid receptor</td>
</tr>
<tr>
<td>GRBs</td>
<td>glucocorticoid receptor binding sites</td>
</tr>
<tr>
<td>GRE</td>
<td>glucocorticoid response element</td>
</tr>
<tr>
<td>GRIP</td>
<td>glucocorticoid receptor-interacting protein</td>
</tr>
<tr>
<td>GTF</td>
<td>general transcription factors</td>
</tr>
<tr>
<td>HAT</td>
<td>histone acetyl transferase</td>
</tr>
<tr>
<td>HDAC</td>
<td>histone deacetylase</td>
</tr>
<tr>
<td>HMT</td>
<td>histone methyl transferase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>HRE</td>
<td>hormone response element</td>
</tr>
<tr>
<td>Hsp</td>
<td>heat shock protein</td>
</tr>
<tr>
<td>LBD</td>
<td>ligand-binding domain</td>
</tr>
<tr>
<td>LHRH</td>
<td>luteinizing hormone-releasing hormone</td>
</tr>
<tr>
<td>IncRNA</td>
<td>long non-coding RNA</td>
</tr>
<tr>
<td>LSD</td>
<td>lysine-specific demethylase</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>NF-κB</td>
<td>nuclear factor kappa-light-chain-enhancer of activated B cells</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localization signal</td>
</tr>
<tr>
<td>NR</td>
<td>nuclear receptor</td>
</tr>
<tr>
<td>NTD</td>
<td>NH2-terminal domain</td>
</tr>
<tr>
<td>PCA</td>
<td>prostate cancer</td>
</tr>
<tr>
<td>PIAS</td>
<td>protein inhibitor of activated STAT</td>
</tr>
<tr>
<td>PIC</td>
<td>pre-initiation complex</td>
</tr>
<tr>
<td>PSA</td>
<td>prostate-specific antigen</td>
</tr>
<tr>
<td>PTM</td>
<td>post-translational modification</td>
</tr>
<tr>
<td>RAR</td>
<td>retinoic acid receptor</td>
</tr>
<tr>
<td>RB</td>
<td>retinoblastoma</td>
</tr>
<tr>
<td>RNA Pol II</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>real-time polymerase chain reaction</td>
</tr>
<tr>
<td>RU486</td>
<td>mifepristone</td>
</tr>
<tr>
<td>SBMA</td>
<td>spinal and bulbar muscular atrophy</td>
</tr>
<tr>
<td>SHBG</td>
<td>sex hormone-binding globulin</td>
</tr>
<tr>
<td>siRNA</td>
<td>small interfering RNA</td>
</tr>
<tr>
<td>SRC</td>
<td>steroid receptor coactivator</td>
</tr>
<tr>
<td>SRY</td>
<td>sex-determining region Y</td>
</tr>
<tr>
<td>SUMO</td>
<td>small ubiquitin-related modifier</td>
</tr>
<tr>
<td>TF</td>
<td>transcription factor</td>
</tr>
<tr>
<td>TFAS</td>
<td>transcription factor association strength</td>
</tr>
<tr>
<td>TMA</td>
<td>tissue microarray</td>
</tr>
<tr>
<td>TSS</td>
<td>transcription start site</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
</tbody>
</table>
INTRODUCTION

Prostate cancer (PCa) is the most common cancer in males worldwide (Jemal et al., 2011). In Finland, the age-adjusted PCa incidence was 89.3 per 100,000 males, and 4,716 cases were reported in 2010, indicating that PCa is the most common of all cancers. This accounted for the second leading cause of cancer-related deaths with 847 reported mortalities in 2010 (Finnish Cancer Registry, Cancer Statistics at www.cancerregistry.fi).

PCa is a complex disease characterized by multiple genetic alterations and aberrant AR signaling. AR signaling axis is central to the growth and progression of PCa, since a vast majority of tumors originate from androgen-dependent secretory epithelial cells. The organ-confined PCa is treated with radiation, surgery or watchful waiting, but if not treated or detected early enough, relapse occurs turning the tumor into a metastatic form. The first line of treatment for metastatic PCa is androgen-deprivation therapy (ADT), chemotherapy or inhibitors of androgen synthesis, but eventually PCa becomes hormone-refractory taking the castration-resistant form with active AR signaling. Glucocorticoids and antiandrogens have been used to treat advanced PCa with scant knowledge of the molecular basis of their interactions. Despite the high rate of PCa incidence, the molecular mechanisms of genome-wide AR function in response to cognate ligands and crosstalk with other signaling cascades are poorly understood.

The advent of next-generation sequencing-based approaches coupled with chromatin immunoprecipitation (ChIP-seq) has provided impetus to examine the role of regulatory transcription factor networks and the interaction between cis-acting elements and nuclear receptors in determining the transcriptional outcome. One such transcription factor, FoxA1, plays a pivotal role in cancer progression and nuclear receptor signaling, highlighting the significance for studying molecular mechanisms of FoxA1 in modulating AR function in PCa. The use of ChIP-seq and gene expression profiling in response to physiological androgens and partial agonists/antagonists and the role of other transcription factors such as FoxA1 and GR has provided novel insights in understanding the genome-wide mechanisms related to AR transcriptional regulation in PCa cells, which forms the core of this thesis.
REVIEW OF THE LITERATURE

1. Gene transcription and transcriptional regulation

1.1 Overview of gene transcription

Gene transcription is the first step of gene expression, an enzymatic process where the DNA strand is synthesized into complementary RNA. The stretch of DNA that is transcribed is called a transcription unit, and the enzymes facilitating transcription are RNA polymerases. The products of gene transcription, either translated proteins or RNA molecules, control and determine the ultimate fate of each cell and the organism. The process is highly cell- and tissue-specific and the events occur at a temporally and spatially regulated fashion with the regulation being achieved at multiple levels. One of the key steps is mediated by nuclear receptors (NR) that communicate the extracellular signals (hormones) to the nucleus in order to regulate target gene expression. Hormone binding translocates NRs to nucleus where they recognize cognate DNA sequences called hormone response elements (HRE). HREs are *cis*-acting DNA elements that recruit a menagerie of *trans*-acting factors which are usually proteins that bind to the *cis*-acting elements to regulate gene expression. NRs are *trans*-acting factors that in addition to their coregulatory proteins and components of the basic transcriptional apparatus including RNA polymerase II (RNA Pol II) and general transcription factors (GTFs), are required for the formation of the pre-initiation complex.

The first NR cloned was human glucocorticoid receptor (GR) by Ron Evans and his collaborators almost three decades ago (Hollenberg et al., 1985); concomitantly, human estrogen receptor (ER) cDNA was isolated and cloned by Pierre Chambon and his co-workers (Green et al., 1986). The GR and ER were the first RNA Pol II transcription factors to be cloned, and since then there has been huge accretion of information on NR-dependent transcriptional regulation (Mangelsdorf et al., 1995; McKenna and O'Malley, 2002; Metivier et al., 2006; Nagy and Schwabe, 2004). In this literature review, the emphasis is on how AR relays the signal of androgens, the male sex-steroids, resulting in AR-dependent transcription and androgen signaling in prostate cancer.
1.2 Basal transcriptional machinery

A gene is a unit of heredity that contains a single transcription unit and usually produces a single protein or RNA. This unit encompasses both the coding exons and non-coding regulatory sequences and introns. The gene also contains a regulatory region, which in a eukaryotic gene comprises a core promoter, an upstream proximal promoter region and distal enhancers with regulatory DNA sequences such as HREs. The core promoter is the minimal length of DNA, usually -35 to +35 base pairs of DNA from the transcription start site (TSS, +1), where the assembly of the basal transcriptional machinery including RNA Pol II and GTFs occurs, determining the point of transcription initiation. The proximal promoter is located about 100–2000 base pairs upstream and enhancers can be located at distant sites either upstream or downstream of the TSS of the genes. Both these regions are enriched in specific DNA motifs, where transcription factors (TFs) bind and regulate the rate of transcription. TFs coordinate with the cis-elements to form a fully competent transcription complex ensuing the target gene expression. The recruitment of transcription factors to their cognate DNA sequences and subsequent initiation of transcription is a highly complex and dynamic process. This involves remodeling of chromatin structure resulting in unpacking of the compact chromatin to make it accessible for binding of transcription factors and other associated molecules, thereby allowing the process of transcription initiation by RNA Pol II. The key player in transcription is RNA Pol II that is not only involved in transcription initiation but also is critical for transcript elongation, which occurs in the context of chromatin and nucleosomes (Selth et al., 2010). Recent genome-wide approaches using RNA Pol II have been used to define nucleosomal architecture and higher order chromosomal organization (Barski et al., 2007; Li et al., 2012a).

1.3 Nucleosomes and chromatin structure

In the eukaryotes, DNA is hierarchically organized into compact nucleo-protein complex called chromatin. The chromatin helps in packaging the DNA within the nucleus and in controlling DNA replication and gene transcription. Compaction of DNA is achieved by wrapping of ~147 base pairs of DNA around an octamer of the four core histones (H2A, H2B, H3 and H4) to form the nucleosome core particle, the basic unit of chromatin (Kornberg, 1977). The core histones are predominantly globular with unstructured N-terminal “tails”, and these tails in particular possess a large number and many types of
Review of the literature

modified residues. The nucleosome cores are efficiently condensed by subsequent association with linker histone H1 near the nucleosome dyad axis (Zhou et al., 1998), binding of corepressor complexes (Fan et al., 2004; Francis et al., 2004; Sekiya and Zaret, 2007) and folding of nucleosomes into higher order chromatin structure (Schwarz and Hansen, 1994). This arrangement of DNA into chromatin occludes the regulatory elements of DNA to any process that uses DNA as a template such as transcription, replication and repair. The nucleosome is a highly dynamic structure involving different structural rearrangements, and transitions associated with the opening/closing of chromatin structure have been described using the term “chromatin remodeling”. Thus, chromatin remodelers either densely package chromatin into nucleosomes genome-wide or dislocate and evict nucleosomes to allow the chromosomal processes (chromatin assembly, transcription, and repair) to occur in a regulated manner (Banaszynski et al., 2010; Clapier and Cairns, 2009; Kouzarides, 2007; Li et al., 2007).

1.4 Chromatin modifications and functions

The mechanisms to alter chromatin structure have greatly evolved and are necessary for meaningful interpretation of the genetic information encoded in the DNA. Three major forms of these are covalent histone modifications (Kouzarides, 2007; Li et al., 2007), chromatin remodeling by ATP-dependent remodeler complexes (Clapier and Cairns, 2009; Li et al., 2007) and the use of highly conserved histone variants (Banaszynski et al., 2010; Li et al., 2007). These heritable, non-genetic changes in chromatin packaging are referred to as “epigenetic” regulation, which is a critical determinant in defining and regulating cellular identity.

The “histone code hypothesis” proposed a decade ago by Strahl and Allis provided a model how single or multiple post-translational modifications (PTMs) on one or multiple histone tails can regulate the diverse processes associated with chromatin, such as transcription (Strahl and Allis, 2000). As per this model, there are specific writers, erasers and readers that place, remove and interpret these PTMs, respectively, and play an important role in establishment of proper spatio-temporal regulation of DNA-templated processes. To date, an extensive literature documents the vast arrays of histone PTMs, such as acetylation, phosphorylation, methylation, ubiquitination, and sumoylation, and their respective chromatin modifying enzymes (Kouzarides, 2007; Li et al., 2007).
Review of the literature

Additional layers of complexity is brought about not only by modifications that can exist in multiple states (e.g., mono-, di- or trimethylation) but also the fact that a particular amino acid can undergo more than one type of modification (e.g., lysine residues). The functional consequences of these PTMs can be either direct, causing an alteration of chromatin structure, or indirect, such as recruitment of effector proteins. In a gene-centric approach, covalent histone modifications in response to androgen and anti-androgen treatment have been studied for prostate specific antigen (KLK3) and kallikrein genes (KLK2) in LNCaP cells (Kang et al., 2004).

Histone acetylation is one of the best-studied modifications mediated by histone acetyltransferases (HATs) and histone deacetylases (HDACs) and is usually associated with accessible “euchromatin” and transcriptionally active genes. Discovery of histone methyl transferases (HMTs) led to identification of histone methylation but existence of demethylases was still elusive until recent years. The first histone demethylase reported, lysine-specific demethylase 1 (LSD1), acts to demethylate histone 3 lysine 4 (H3K4) and repress transcription (Shi et al., 2004). However, another study reported that LSD1 activates transcription when present in a complex with AR by demethylating histone 3 lysine 9 (H3K9) (Metzger et al., 2005). Interestingly, a recent elegant study has shown that at increasing androgen levels, LSD1 can repress the AR gene by getting recruited by AR to an intronic regulatory element (Cai et al., 2011). Additionally, histone methylation can be either activating or repressing depending on the precise context and modified residues (Berger, 2007). For example, high levels of methylation of histone 3 lysine 4 (H3K4me1, H3K4me2 and H3K4me3) are detected in promoter regions of active genes (Bernstein et al., 2005; Kim et al., 2005). Moreover, high-resolution ChIP-seq mapping showed the association of H3K4me2 with enhancers and transcription start sites of active genes (Barski et al., 2007; Heintzman et al., 2007). Conversely, transcriptional repression correlates well with elevated levels of methylated histone 3 lysine 27 (H3K27) (Roh et al., 2006), and large domains of dimethylated histone 3 lysine 9 (H3K9me2) organize inaccessible “heterochromatin” and are altered during the process of differentiation (Wen et al., 2009).

The advent of genome-wide approaches has led to a gradual shift from gene-centric to genome-wide view in addressing and understanding the global patterns of specific modifications and higher order chromatin structure (Fig. 1). For example, a high-
resolution genome-wide mapping of 20 histone lysine and arginine methylations, histone variant H2A.Z, RNA Pol II and insulator binding protein CTCF (Barski et al., 2007) and another independent study with histone H3, five histone modifications, RNA Pol II, TBP associated factor 1 and transcriptional coactivator p300 (Heintzman et al., 2007; Kim et al., 2005), have revealed the distinct patterns of these modifications at promoters, enhancers, insulators and transcribed regions across the human genome. These modifications ultimately determine how cell- and tissue-specificity of TFs is achieved and how gene expression is regulated in a spatio-temporal fashion. Enormous amount of information produced by genome-wide studies such as DHS-seq, FAIRE-seq for chromatin accessibility; MeDIP-seq for DNA methylation; MNase-seq for nucleosome position and turnover respectively gives a more comprehensive view of gene expression (Zhou et al., 2011).

Figure 1. Modifications associated with gene bodies, and promoter, enhancer and boundary elements. (Reprinted by permission from Macmillan Publishers Ltd: [Nature Review Genetics] (Zhou et al., 2011) copyright 2011).

The second well-studied mechanism of tinkering with chromatin is by ATP-dependent chromatin remodelers for regulated DNA accessibility during several DNA-templated processes such as replication, repair, and transcription, where nucleosomes must be correctly spaced, mobilized or ejected for the smooth running of the processes. All the four known chromatin-remodeling complexes utilize ATP hydrolysis to alter histone-DNA contacts and share a similar ATPase domain. These four families are SWI/SNF (switchingdefective/sucrose non-fermenting), ISWI (imitation switch), CHD
Review of the literature

(chromodomain, helicase, DNA binding) and INO80 (inositol requiring 80) family of remodelers, and individual families are highly conserved from yeast to human. All remodelers share five basic properties namely: a) nucleosome affinity; b) domain recognizing covalent histone modifications; c) DNA-dependent ATPase domain; d) domains/proteins that regulate ATPase domain; and e) domains and/or proteins for interaction with other chromatin remodelers or transcription factors. The mechanisms of histone eviction resulting in accessibility of nucleosomal DNA include various processes such as co-operative TF binding, chromatin remodeling, actively transcribing RNA Pol II, and presence of histone chaperons (Clapier and Cairns, 2009; Li et al., 2007).

The third mechanism that alters the epigenetic landscape of a cell is incorporation of histone variants into the nucleosome, information about which has started to accumulate over the last decade. The core histones are the canonical histones, synthesized during the S phase of the cell cycle and incorporated in a DNA replication-dependent manner. The variant histones are unique in the sense that they are expressed outside of the S phase and incorporated in a DNA replication-independent manner. In mammals, all core histones have several sequence variants with the exception of H4 (Banaszynski et al., 2010; Li et al., 2007). Recent studies suggest that the histone variants like H2A.Z and H3.3 can be deposited into nucleosomes making them intrinsically labile and thus determining the ultimate fate of nucleosome stability (Jin et al., 2009). In the context of AR binding to enhancers in PCa cells, it was shown that androgen treatment leads to eviction of the central nucleosome and the central nucleosome was also enriched for the H2A.Z variant compared to the more stable flanking nucleosomes (He et al., 2010).

1.5 Genome-wide analysis of gene regulation

The advent of new technologies such as DNA microarrays and next-generation sequencing applications over the past decade have advanced our knowledge on the interplay between the chromatin structure, genome function and gene regulation. It is clear that nucleosome and higher order chromatin organization along with the dynamic variations across cellular contexts are very tightly coupled with gene expression. With the help of these techniques, a vast amount of information has accumulated very rapidly over the last few years and has provided more comprehensive overview of genome-wide occupancy by transcription factors such as NRs, histone modifications and related
chromatin structures (Cheung and Kraus, 2010; Kim and Ren, 2006; Zhou et al., 2011).

One of the foremost methods that led to a gradual shift from gene-centric to genome-wide approach is chromatin immunoprecipitation (ChIP). Although ChIP has been used since 1988 (Solomon et al., 1988) to probe protein-DNA interactions at individual target loci in vivo, its coupling with genomic microarrays (ChIP-on-chip) and, more recently, with next-generation sequencing (ChIP-seq), provided the means to gain more precise, robust and comprehensive view of the genomic regulatory elements and the chromatin landscape. ChIP is an antibody-based technique where an antibody is used to enrich chromatin regions associated with the protein against which the antibody was raised. The chromatin with associated DNA and proteins is cross-linked with a reversible cross-linker such as formaldehyde and then sonicated to generate randomly-sized DNA fragments. The DNA thus hangs on when protein is immunoprecipitated by the specific antibody. The immunoprecipitated DNA fragments are then purified to isolate naked DNA. In ChIP-on-chip, the co-purified DNA is then hybridized to genomic microarrays known as tiling arrays, which contain probes designed to cover regions representative of the curated genome (Kim and Ren, 2006). The first global map of ER and AR binding was reported using tiling arrays (Carroll et al., 2005; Carroll et al., 2006; Wang et al., 2007; Wang et al., 2009), which highlighted two important features in the field of NR-mediated gene regulation. First, NRs, such as ER and AR, employ a distal mode of transcriptional regulation, preferentially binding to distal enhancer elements. Second, collaborating factors, such as FoxA1, mediate the recruitment of ER and AR to target genomic loci by binding to compact chromatin, enabling the estradiol- or androgen-dependent transcriptional programs.

In ChIP-seq, the immunoprecipitated DNA is deep-sequenced to produce digital maps of ChIP enrichment (Johnson et al., 2007) that allow direct, more quantitative identification of transcription factor binding sites and various epigenetic signatures with a much better resolution than ChIP-on-chip. The use of ChIP-on-chip and ChIP-seq technologies has brought novel information on steroid hormone receptor “cistromes” (genome-wide set of cis-acting DNA binding elements for transcription factors). Next-generation sequencing has allowed other assays to be also scaled genome-wide, such as RNA expression analyses (RNA-seq), DNaseI-hypersensitivity assay (DHS-seq), formaldehyde-assisted isolation of regulatory elements (FAIRE-seq), global run on assays (GRO-seq),
Review of the literature

cromatin interaction analysis by paired-end tag sequencing (ChIA-PET) and Hi-C methods for unbiased genome-wide analysis to identify long-range interactions (Kim and Yu, 2012). Thus, integrating the data from diverse genome-wide experiments has enriched partial understanding of the process of gene regulation and is essential for further comprehension of the intricate mechanisms regulating the genome structure and function (Fig. 2).

Figure 2. Overview of genome-wide approaches for understanding gene expression. In the context of AR function, differentially expressed genes are identified by microarray or RNA-seq, and AR or other transcription factor binding sites and histone modifications mapped by ChIP-seq. Chromatin structure and modifications can be elucidated by DHS-seq (DNase-seq), FAIRE-seq, MNase-seq and DNA methylation by MeDIP-seq. Moreover, ChIA-PET and Hi-C approaches can reveal intra- and inter-chromosomal interactions required for a fully competent transcription complex. (Reprinted from Kim and Yu et al., Interrogating genomic and epigenomic data to understand prostate cancer. Biochim Biophys Acta. 2012; 1825:186-96 with permission from Elsevier, copyright 2012).
2 Steroid receptor-mediated signaling

2.1 Nuclear receptors and coregulators

The NR superfamily comprises of structurally related proteins that are ligand-activated transcription factors regulating a vast repertoire of downstream target genes. The superfamily consists of 48 genes in the human genome, including those encoding receptors for steroid hormones such as ER, AR, and GR (Class I); receptors for non-steroidal ligands, such as the thyroid hormone receptor (TR) and the retinoic acid receptor (RAR) (Class II), as well as the receptors, the ligands of which are the products of lipid metabolism, such as fatty acids and prostaglandins, like peroxisome proliferator activated receptor (PPARs) and liver X receptors (LXRs). Additionally, a large number of orphan receptors (Class III) have been found on the basis of homology but with no identified ligand so far.

NRs are characterized by their modular structure (Fig. 3). The functional domains of the NRs are a central well-conserved DNA-binding domain (DBD), a variable N-terminal domain (NTD) or A/B domain, a non-conserved hinge region and a moderately conserved C-terminal ligand-binding domain (LBD).

![Figure 3](image)

**Figure 3.** Modular structure of NR depicting the major domains and associated functions. NTD denotes NH2-terminal domain with activation function-1, DBD, the DNA binding domain for sequence recognition, dimerization and nuclear transport and LBD, the ligand-binding domain for hormone binding harboring the activation function-2.

The high degree of conservation between NRs suggests that they have arisen from a common ancestry by gene duplication and genetic divergence (Aranda and Pascual, 2001; Gronemeyer et al., 2004). As small molecules activate NRs, they are promising targets for drug-development. Over the last decade, major endeavors have been made to understand NR-mediated gene regulation, to characterize receptor-selective agonists, antagonists or ligands that act in a cell type-specific manner (Selective Nuclear Receptor
Review of the literature

Modulators, SNuRMs), and to scrutinize their mechanism of action thus facilitating drug design, screening and target validation (Shang and Brown, 2002).

The functional consequence of NR transcriptional activity is achieved by interaction with components of basal transcriptional machinery. Importantly, it also requires the interaction with functionally diverse group of proteins called coregulators, either coactivators that activate transcription or corepressors that repress transcription. Coregulators ensure the transcriptional outcome by modifying chromatin structure, bridging the proximal and distal components of transcriptional apparatus and transducing cellular signals to the site of transcription, primarily achieved by PTMs. Mechanistically, the cartage of PTMs to NRs and coregulators are similar to that of histones, mostly contributed by intrinsic enzymatic activities possessed by coregulators themselves or by other enzymes involved in the signaling cascade. Conceptually though, these two processes are distinct, as histone modifications are usually directed to a specific genomic locus influencing the expression of a single gene, whereas PTMs of coregulators regulate multiple gene sets through the multitude of interactions with a variety of transcription factors. Cracking the coregulator code has been a herculean effort by the nuclear receptor researchers and to date ~300 coregulators have been reported (www.nursa.org).

Coactivators and corepressors both possess LXXLL motifs (where L is leucine and X stands for any amino acid), also known as the nuclear receptor interaction box (NR-box). Ligand-dependent recruitment of coactivators is dependent on the AF-2 domain (cf., Fig. 3), which consists of short helical sequence within the C-terminus of LBD. Coactivators, such as steroid receptor coactivator (SRC1) or nuclear co-activator 1 (NCoA1), function by making the receptor competent for target gene activation, and the recognition is mediated by NR-boxes (Heery et al., 1997; Torchia et al., 1997). On the contrary, corepressors such as NR co-repressor (N-CoR) and silencing mediator of retinoid and thyroid receptors (SMRT), whose recognition to NRs is mediated by “CoRNR boxes” (Hu and Lazar, 1999; Nagy et al., 1999; Perissi et al., 1999), preferentially interact with unliganded nuclear receptors or receptors bound to antagonists such as tamoxifen and mifepristone for ERα and GR, respectively (McKenna and O'Malley, 2002; Perissi and Rosenfeld, 2005). Interestingly, the N-CoR and SMRT may interact with both agonist-liganded AR and in presence of antagonist and partial antagonists (Hodgson et al., 2008).
However, evidence is accumulating against typecasting the coregulators as solely positive coactivators or negative corepressors. Instead, role reversal seems to be a common mechanism for some of the reported coregulators. For instance, SRC2 or glucocorticoid receptor-interacting protein 2 (GRIP2), a coactivator, can function as a repressor for estradiol-bound ERα in TNFα–mediated transcription (Cvro et al., 2006). In another example, HDAC1, usually considered as a transcriptional repressor due to its histone deacetylase activity and its association with N-CoR and SMRT complexes, can function as a coactivator for GR-mediated transcription (Qiu et al., 2006). Importantly, this role reversal in coregulator action goes in tandem with the PTM status of the protein, and emphasizes the importance of coregulator dynamics exerted through the epigenetic switch (Lonard and O'Malley B, 2007; O'Malley et al., 2008; Perissi and Rosenfeld, 2005).

2.2 Steroid hormones and steroid receptors

Steroid hormones are small hydrophobic molecules synthesized from cholesterol in the gonads and adrenal cortex. Steroid hormones act as chemical messengers exerting wide variety of effects through slow genomic responses and rapid non-genomic responses. In spite of their diverse functions (Table 1), steroid hormones exhibit remarkable similarity in their structure (Fig. 4). Testosterone and DHT are the physiological ligands for AR, and testosterone serves as a substrate for 5α-reductase enzyme to produce the more potent DHT. Testosterone can also be aromatized by the enzyme aromatase to estradiol, the ligand for ER.

Table 1. Diverse actions of steroid hormones in humans. (Adapted from Bolander, 2004).

<table>
<thead>
<tr>
<th>Steroid Hormone</th>
<th>Main Source</th>
<th>Main Targets</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Androgens</td>
<td>testis, adrenal cortex</td>
<td>prostate, reproductive tract</td>
<td>sexual characteristics, reproduction, anabolic effects</td>
</tr>
<tr>
<td>Estrogens</td>
<td>ovary, placenta</td>
<td>uterus, breast, reproductive tract</td>
<td>reproduction, sexual characteristics</td>
</tr>
<tr>
<td>Glucocorticoids</td>
<td>adrenal cortex</td>
<td>muscle, liver</td>
<td>gluconeogenesis, energy metabolism, modulation of immune response</td>
</tr>
<tr>
<td>Progestins</td>
<td>ovary, placenta</td>
<td>reproductive tract</td>
<td>maintenance of pregnancy</td>
</tr>
</tbody>
</table>

Being hydrophobic, steroids are carried in circulation bound to serum proteins that also protect them from degradation. Androgens and estrogens are mostly bound to low affinity serum albumin, but also to specific high-affinity sex hormone-binding globulin (SHBG).
Likewise, corticosteroid-binding globulin (CBG) binds and carries other steroid hormones such as progesterone, cortisol and corticosteroids (Hammond, 2010).

![Chemical structures of steroid receptors](image)

**Figure 4.** Chemical structure of four natural and four synthetic steroid receptor ligands.

Steroid hormones act as ligands and mediate their effects through steroid receptors, which are a subclass of the NR family. For example, androgens relay their signals through AR, estrogens through ER and glucocorticoids through GR. In the absence of ligands, steroid receptors are kept in inactive forms by large heterocomplexes of heat-shock protein (Hsp) 90-based chaperone machinery. Hsp90, a molecular chaperon, is required for the proper folding, traffickging, transcriptional activation and turnover of steroid hormone receptors. Initial *in vivo* studies focused on GR and the role of Hsp90 and other molecular chaperons (e.g., Hsp70, Hsp40, Hsp organizing protein (Hop), and Hsp90 binding co-chaperone p23) on the steroid receptor function, and similar mechanism applies to other members of the family as well. Hsp90 and Hsp70, often called the master chaperones, are ATP-dependent chaperones that interact with other co-chaperones possessing a tetratricopeptide repeat (TPR) domains through a short C-terminal sequence, MEEVD and EEVD for Hsp90 and Hsp70, respectively (Echeverria and Picard, 2010; Pratt and Toft, 2003). The ligand binding induces a conformational change of the receptor that leads to dissociation of heat-shock proteins, allowing dimerization of the receptor and nuclear translocation. Inside the nucleus, GR rapidly cycles on and off on its cognate binding site, both in the presence or absence of hormone facilitated by chaperones (McNally et al., 2000). Thus, the role of Hsp90 in the case of GR is not limited to the first steps of maturation, but it is also required for assembly and disassembly on chromatin,
nuclear retention and export (Echeverria and Picard, 2010; Freeman and Yamamoto, 2002; Grad and Picard, 2007; McNally et al., 2000). In contrast, in the case of cyclical ERα recruitment on pS2 promoter, Hsp70 and not Hsp90 follow the same cyclicity as ERα on the promoter (Reid et al., 2003).

Importantly, the interaction with chaperone machinery keeps the steroid receptors in a partially folded, high-affinity ligand-binding conformation. Most of the liganded steroid receptors including ER, GR and AR, undergo nucleo-cytoplasmic shuttling with nuclear localization determined by the nuclear localization signal (NLS). The NLS is unmasked upon ligand binding allowing receptor transport to the nucleus (Echeverria and Picard, 2010; Prescott and Coetzee, 2006). Inside the nucleus, the steroid receptors bind to their cognate HREs, notably estrogen response element (ERE) for ER; glucocorticoid response element (GRE) for GR and androgen response element (ARE) for AR, and regulate the downstream target genes. The understanding of how steroid receptors selectively recognize and are recruited to their HREs is slowly evolving but far from being complete. Cistromes defined by ChIP-on-chip and ChIP-seq studies have not only emphasized the role of the receptor and chromatin structure, but they were also groundbreaking in highlighting the role of other actors such as collaborating transcription factors and coregulators involved in transcriptional regulation mediated by steroid receptors (Carroll et al., 2005; Carroll et al., 2006; John et al., 2011; Wang et al., 2007; Wang et al., 2009; Yu et al., 2010).

2.3 Glucocorticoid receptor

Glucocorticoids such as naturally occurring cortisol and the synthetic steroid dexamethasone (cf., Fig.4) mediate their effect through GR, an intracellular steroid receptor belonging to the nuclear receptor superfamily. Like other steroid receptors, liganded GR translocates to the nucleus, binds to GREs (5’-AGAACAnnnTGTTCT-3’) and mediates transactivation of target genes. GR-activated genes are involved in a wide spectrum of physiological processes with various regulatory roles in development, metabolism and inflammation to name a few (Yudt and Cidlowski, 2002; Mangelsdorf et al., 1995). Gene activation and repression mediated by GR can be achieved either by direct binding to DNA or through protein-protein interactions termed as transactivation or transrepression (Revollo and Cidlowski, 2009). The association of GR with activator
Review of the literature

protein 1 (AP1) or nuclear factor-kappa B (NF-κB) is one of the well-studied GR-mediated transrepression mechanisms (Biddie and Hager, 2009). The classical view of long-term residency of transcription factor on the DNA template to facilitate coregulator recruitment and interaction with chromatin remodelers was challenged by fluorescence imaging studies of GR in living cells, where rapid cycling of GR was demonstrated on the MMTV promoter (McNally et al., 2000). The GR has long been considered as a “pioneer factor” able to access compact DNA when other transcription factors are unable to do so. However, recent genome-wide studies using ChIP-seq have shown that GR binding is a later event and most of the chromatin-associated changes are initiated by other transcription factors. Using DNase1-hypersensitivity mapping by deep sequencing (DHS-seq), it was shown that most of the GR bound sites are constitutively open prior to GR occupancy and are pre-programmed by other transcription factors, and a very small fraction of the sites are initiated de novo by GR itself (John et al., 2008, 2011).

Studies on GR revealed another important feature in nuclear receptor action, in that it was shown that the DNA sequence directs the GR structure and function and can act as an allosteric ligand (Meijsing et al., 2009). The DNA sequence recognized by GR is also recognized by AR and partially by progesterone and mineralocorticoid receptors, and thus it remains to be established whether GR can antagonize the function of other steroid receptors by competing for the same in vivo binding sites. The characteristics of AR structure and function are described in detail in Chapter 3.

2.4 Pioneer factors for steroid receptors

Pioneer factors are a distinct class of proteins with a remarkable competence to associate with compact chromatin. The DNA is occluded by nucleosomes, higher-order chromatin structure and repressor complexes, thus inaccessible to factors involved in DNA-templated processes (Fan et al., 2004; Francis et al., 2004; Schwarz and Hansen, 1994; Sekiya and Zaret, 2007; Zhou et al., 1998). Recent genome-wide studies have revealed that most of the putative consensus DNA-binding sites are unoccupied, and transcription factors physically bind at a fraction of these known consensus sites (Joseph et al., 2010; Kaplan et al., 2011; Zaret and Carroll, 2011). The binding site analysis of histone-modifying enzymes such as HATs (Visel et al., 2009), promoter-enhancer associated active histone modifications such as H3K4 methylation (Barski et al., 2007; Heintzman et
al., 2007), and DNaseI-hypersensitive sites marking open chromatin conformation (Boyle et al., 2008; John et al., 2011) can all help in predicting and guiding the transcription factor binding to their target sequences. Nevertheless, the question remains elusive as to what are the initiating events and which factor binds first to set the chain of reactions required for initiation of DNA-associated events.

Forkhead box (Fox) family members such as FoxA1, FoxA2 (Hannenhalli and Kaestner, 2009), and members of GATA family, such as GATA-4 and GATA-6, were the first factors reported to engage the enhancers during development, and expression of these factors was indispensable for the induction of a liver differentiation program (Holtzinger and Evans, 2005; Lee et al., 2005; Watt et al., 2007; Zhao et al., 2005). In vivo footprinting assays showed that the occupancy of FoxA1 and GATA factors clearly preceded other factors recruited to enhancers in nascent liver, and the same pattern of binding was also seen in progenitor cells prior to gene activation. Thus, these initial factors termed “pioneer factors” served the function of binding to the condensed chromatin prior to gene activation and making them competent for binding of other factors (Magnani et al., 2011b; Zaret and Carroll, 2011). In the context of steroid receptors such as ER, AR and GR, forkhead proteins, especially FoxA1, have garnered significant attention over the last couple of years (Grange et al., 1991; Lupien et al., 2008). More recently, two novel putative pioneer factors for ER have been reported, namely AP-2γ and PBX1 (Magnani et al., 2011a; Tan et al., 2011). Likewise for GR, transcription factor AP1 has been proposed to act as a pioneer factor (Biddie et al., 2011). Thus, therapeutic targeting of pioneer factors may open attractive avenues for treatment of hormone-dependent cancers (Jozwik and Carroll, 2012).

2.5 Forkhead box protein A1 (FoxA1) as a pioneer factor

FoxA1/HNF-3α is a transcription factor with a winged-helix DBD closely resembling the linker histone (Clark et al., 1993; Ramakrishnan et al., 1993). The helix-turn-helix motif of DBD makes a sequence-specific DNA contact, and two flanking polypeptide loops, called wings, contact the DNA-phosphate backbone increasing the overall affinity of the protein for DNA (Cirillo et al., 2002; Clark et al., 1993). The resemblance of FoxA1 DBD to linker histone, and its ability to bind stably to a nucleosome both in vitro and in vivo resulting in hypersensitivity of the underlying nucleosome, make FoxA1 a pioneer
Review of the literature

factor (Chaya et al., 2001; Cirillo et al., 2002; Cirillo et al., 1998; Sekiya and Zaret, 2007). These properties are unique to FoxA1, distinguishing it from other pioneer factors such as GATA-4. The pioneering function of FoxA1 is attributed to high-affinity non-specific binding to nucleosome and slower nuclear mobility, allowing more probing time for chromatinized sites that many other transcription factors are not able to access (Sekiya et al., 2009). On the other hand, the role of FoxA1 is not just limited to pioneering function, as it has the ability to create heterochromatin-like condensed chromatin structure by recruiting TLE/Groucho family of proteins (Sekiya and Zaret, 2007; Watts et al., 2011), and a large proportion of the FoxA1 binding sites are reported outside the active chromatin regions (Eeckhoute et al., 2009).

FoxA1 plays a critical role in growth and differentiation of variety of organs such as liver, prostate, bladder, and breast (Friedman and Kaestner, 2006; Kaestner, 2010; Lee et al., 2005). For instance, in mouse prostate development, FoxA1 is required in both epithelial cell differentiation and ductal morphogenesis and patterning (Gao et al., 2005). In prostate cells, FoxA1 is involved in AR-mediated transcriptional regulation, such as probasin and prostate-specific antigen (PSA) genes in humans (Gao et al., 2003; Mirosevich et al., 2006). The regulatory regions of these genes are reported to have FoxA1-binding sites and AR-binding sites (ARBs) in close proximity, and FoxA1 and AR proteins have been shown to interact through their DNA-binding domains (Gao et al., 2003; Lee et al., 2008). A recent study shows that ERα-mediated resistance and AR-mediated enhancement of hepatocellular carcinoma is dependent on FoxA1 and FoxA2 and their network of downstream target genes (Li et al., 2012b). Recently, recurrent somatic mutations in the FoxA1 gene were identified in exome sequencing studies in prostate adenocarcinomas and CRPCs (Barbieri et al., 2012; Grasso et al., 2012). FoxA1 expression has been associated with clinical outcome in breast cancer, in that high FoxA1 level predicts a favorable outcome in breast cancer patients (Albergaria et al., 2009; Badve et al., 2007; Krum et al., 2008). Apparently, this opens up an interesting prospect to study whether FoxA1 plays a similar role in clinical outcome of PCa in relation to AR function.

Much insight has come from the genome-wide studies in ER-expressing (ER+) MCF-7 breast cancer cells, where FoxA1 was shown to act as a pioneer factor for ERα and also
to drive transcription in a lineage-specific manner (Carroll et al., 2005; Carroll et al., 2006; Krum et al., 2008; Lupien et al., 2008). Significant, ~50% overlap in genomic occupancy was reported for the two factors, and FoxA1-specific RNAi depletion led to decrease in ER binding (Carroll et al., 2005; Hurtado et al., 2011; Lupien et al., 2008). Additionally, reduced coregulator recruitment was observed (Eeckhoute et al., 2006), which was commensurate with decreased estrogen-dependent transcription, and loss of FoxA1 led to cell cycle arrest (Hurtado et al., 2011; Laganiere et al., 2005). More recently, FoxA1 has been reported to reprogram the ER binding events and this has been attributed to the differential ER-binding program in tumors from patients with poor clinical outcome (Ross-Innes et al., 2012). In the context of AR in PCa cells, FoxA1 has also been shown to act as a pioneer factor (Lupien et al., 2008; Wang et al., 2007). Unlike in breast cancer cells, the role of FoxA1 in modulating PCa cell proliferation is not clear. This is due to conflicting results as one study reported that FoxA1 depletion increased cell proliferation in wild-type LNCaP cells (Wang et al., 2011), whereas another recent study showed that FoxA1 depletion led to a decrease in LNCaP cell proliferation (Imamura et al., 2012). Likewise, in androgen-independent LNCaP-abl cells, FoxA1 depletion led to growth arrest (Zhang et al., 2011). This leaves intriguing questions regarding the role of FoxA1 and other pioneer factors in controlling and affecting steroid receptor-mediated gene regulation. Is there a pioneer factor code? What is the generality in the mode of action of these pioneer factors? How do pioneer factors make the decision whether to mark active or inactive state of the chromatin?
3 Androgens and androgen receptor

3.1 Physiological androgens

Androgens control the development, differentiation and maintenance of male sexual characteristics and other normal functions of male sex organs, such as prostate, epididymis and seminal vesicles. Androgens also affect other organs such as skin, skeletal muscle, bone marrow, liver and behavioral centers of the brain (Heinlein and Chang, 2002; Nef and Parada, 2000; Quigley et al., 1995). AR mediates the effects of androgens and is highly expressed in male reproductive organs. Testosterone is the physiological androgen produced by the Leydig cells of testis in males and in very small amounts in ovaries of females. Although androgens are also produced in the adrenal cortex, this tissue is not the primary source of androgens in either sex (Shen and Coetzee, 2005). Testosterone can act both in a paracrine manner in the testis or in an endocrine manner, carried by the blood to the distant tissues/organs. In some target tissues such as prostate, testosterone undergoes a reduction reaction to 5α-dihydrotestosterone (DHT). Compared to testosterone, DHT has a higher affinity for AR, as it dissociates at a slower rate, leading to the formation of a DHT-AR conformation that is more resistant to degradation (Heinlein and Chang, 2004). Testosterone can also undergo an aromatic conversion by aromatase to form estradiol, the primary female sex hormone.

Androgens and estrogens function in development of male and female phenotype, respectively, but the sexual differentiation in mammals is hormone-independent. During embryonic development, the urogenital ridge may develop into male or female sex organs. In mammals, sexual differentiation is determined by inheritance of X or Y chromosome and is directed by the expression of the sex-determining region Y (SRY) gene that initiates the development of the testis. The Leydig cells of testis then produce testosterone that leads to the downstream cascade of events culminating in development and stabilization of other male sex organs. Thus, a delicate balance of androgen level is required for proper development of male sexual characteristics, since the absence of androgens or defective AR function causes genetically male embryos to develop female phenotypic features (Nef and Parada, 2000).
3.2 Androgen receptor gene and its regulation

The gene encoding AR is approximately 90 kb in size, and it is located on the X-chromosome at Xq11-12 (Brown et al., 1989; Kuiper et al., 1989; Lubahn et al., 1988). The processed mRNA transcript is 10.6 kb long and has an open reading frame of approx. 2.8 kb, coding for the eight exons named 1–8 (Lubahn et al., 1988; Trapman et al., 1988). In addition, an alternative exon 1 has been identified, giving rise to a naturally occurring splice variant, AR45 (Ahrens-Fath et al., 2005). However, the functional significance and the relative abundance of this and other splice variants compared to the full-length AR are yet to be established. The first AR splice variant producing a novel AR protein lacking the LBD and with partially intact NTD and DBD was characterized in the CWR22Rv1 cell line (Tepper et al., 2002). In recent years, considerable attention has been paid to identifying novel splice variants, and a total of 7 different splice variants have been reported independently in the same CWR22Rv1 cell line (Dehm et al., 2008; Guo et al., 2009; Hu et al., 2009), some variants are found in VCaP cells (Dehm et al., 2008; Hu et al., 2009), and a few of them also expressed in LuCaP xenografts (Dehm et al., 2008; Sun et al., 2010). All of these variants share the same characteristic feature of being constitutively active due to lack of the LBD. Recent deep-sequencing experiments (RNA-seq) have not only confirmed the previous findings, but have identified several additional variants that, however, require the presence of full-length AR for their activity (Watson et al., 2010). The exact mechanism and the role of splice variants remains to be fully established (Dehm and Tindall, 2011).

The 5’ untranslated region (UTR) of the AR genes has two possible transcription start sites located 1.1 kb upstream of the start codon, and their usage depends on the cellular context and milieu (Chang et al., 1995; Faber et al., 1993). In mammals, AR is expressed in most tissues and its expression is tightly regulated in a cell type-specific manner. The isolation of 2.3 kb AR promoter region (Mizokami et al., 1994) showed that it lacks the TATA-box and CCAAT-box, and binding sites for several transcription factors have been mapped (Faber et al., 1993; Mizokami et al., 1994; Tilley et al., 1990). Transcription factors upregulating AR gene transcription are CREB (Mizokami et al., 1994), SP1 (Faber et al., 1993; Yuan et al., 2005), c-Myc (Lee et al., 2009), FOXO3 (Yang et al., 2005) to name a few. In addition, there are other transcription factors that can either upregulate or downregulate AR transcription depending upon the cellular context (Shiota et al., 2011).
Review of the literature

For instance, some E2F family members, such as E2F-1, E2F-2 and E2F-3a, can activate whereas the rest of them suppress AR expression. Similarly, nuclear factor (NF)-κB that was known to suppress AR from an earlier study (Supakar et al., 1995), was later shown to activate AR (Zhang et al., 2009; Zhang et al., 2004). Recently, the AR protein was shown to downregulate AR expression in an androgen-dependent fashion by recruiting LSD1 concomitantly with H3K4me1 and H3K4me2 chromatin modifications (Cai et al., 2011). In addition, either several other transcription factors or other signaling molecules from extracellular and intracellular sources can regulate AR expression; some of these factors have been implicated in PCa progression (Shiota et al., 2011). The translated AR protein of human, mouse and rat origin are all approx. 99 kDa (unphosphorylated) and 110 kDa (phosphorylated).

3.3 Modifications of androgen receptor protein

Like several other NRs, AR is a substrate for different covalent PTMs including phosphorylation, acetylation, sumoylation, ubiquitination and methylation (Gioeli and Paschal, 2012). The PTMs have been reported at a total of 23 sites (Fig. 5) and are necessary for proper receptor function, including subcellular localization, coregulator interactions, AR turnover, and cross-talk with other signaling pathways. Out of all the PTMs, phosphorylation has been the best studied.

Figure 5. PTMs of AR. The numbering is based on NCBI Accession No. AAA51729 (Lubahn et al., 1988). (Adapted and reprinted from Gioeli and Paschal. Post-translational modification of the androgen receptor. Mol Cell Endocrinol 2012; 352:70-78 with permission from Elsevier, copyright 2012).

Phosphorylation of AR may occur both in the absence or presence of hormone. It affects AR activity by increasing or decreasing the interactions of AR with coregulators. The majority of the sites are located in the NTD, and these occur on serine (Ser), threonine (Thr) and tyrosine (Tyr) residues, with at least one site in the other major structural
domains such as Ser 578 in the DBD, Ser 650 in the hinge region and Ser 791 in the LBD (Gioeli, 2005; Gioeli and Paschal, 2012). Certain sites such as Ser 94 are constitutively phosphorylated, though most of the events take place after androgen binding, since the change in conformation makes the sites more accessible to solvents and kinases or reduces the phosphatase interactions with AR. The androgen-induced phosphorylation sites are Ser 16, Ser 81, Ser 256, Ser 308, Ser 424 and Ser 650 (Gioeli et al., 2002). Growth factor signaling can also lead to AR phosphorylation such as EGF-induced AR phosphorylation by AKT, one of the first kinases linked to AR phosphorylation (Gioeli and Paschal, 2012). EGF or IGF receptor can activate AKT, which in turn induces the phosphorylation of Ser 213 and Ser 791 (Lin et al., 2001; Wen et al., 2000). However, the functional consequences of phosphorylation at Ser 213 and Ser 791 are controversial, due to conflicting results from different labs reporting either activating or repressive effects (Lin et al., 2001; Palazzolo et al., 2007; Taneja et al., 2005; Thompson et al., 2003; Wen et al., 2000). Interestingly, PI3K/AKT-mediated phosphorylation of an AR coactivator, MED1 (TRAP220), and not AR itself was recently shown to enhance the UBE2C oncogene locus looping (Chen et al., 2011). In addition, there are several other kinases reported to phosphorylate AR such as cyclin-dependent kinases, Aurora-A, mitogen-activated protein kinases and protein kinase C (Gioeli and Paschal, 2012).

Current models suggest that AR-mediated transcription begins with androgen-induced nuclear import and culminates with ubiquitin-dependent degradation of AR by proteasome machinery (Shank and Paschal, 2005). Ubiquitination is a reversible process of tagging the substrates with an ubiquitin moiety to an acceptor lysine, subsequently forming an ubiquitin chain marking them for degradation by 26S proteasome machinery. Mass spectrometry led identification has revealed that AR is ubiquitinated at Lys 845 and Lys 847 (Xu et al., 2009), and this process is catalyzed by three ubiquitin E3 ligases namely MDM2 (mouse double minute), CHIP (carboxyl-terminus of Hsp70-binding protein) and RNF6 (ring finger protein) (He et al., 2004; Lin et al., 2002; Rees et al., 2006; Xu et al., 2009). Unlike MDM2 and CHIP that ubiquitinate AR for proteosomal degradation, RNF6 performs a non-canonical ubiquitination on AR enhancing the rate of transcription by increased recruitment of ARA54 (RNF14), an AR coactivator. Additional evidence comes from cell line experiments, where RNF6 knockdown in LNCaP cells led to attenuation of AR transcriptional activity, and upregulation of RNF6 has been reported in CRPC (Xu et al., 2009). However, AR can also be deubiquitinated by enzymes such as Usp26, which binds to AR via NR-boxes. Usp26 depletion has been shown to enhance AR-dependent transcription (Dirac and Bernards, 2010).
**Review of the literature**

Analogous to ubiquitination is another PTM termed sumoylation, which refers to the attachment of small ubiquitin-like modifiers (SUMOs) to lysine residues in a similar enzymatic cascade. AR was the first NR shown to be sumoylated at Lys 386 and Lys 520 (Poukka et al., 2000a). PIAS1 and PIASxα have SUMO E3 ligase activity towards AR (Kotaja et al., 2002; Nishida and Yasuda, 2002), and PIAS1 was shown to sumoylate both cytoplasmic and nuclear AR with a similar efficiency (Kaikkonen et al., 2009). However, unlike ubiquitination that mostly marks the proteins for degradation, sumoylation either represses or activates transcription. Sumoylation can be reversed by SUMO proteases called as SENPs. In the case of AR, SENP1 and SENP2 are shown to be the most efficient desumoylating enzymes, and both enzymes markedly increase transcriptional activity of wild-type AR compared to sumoylation-deficient AR (Kaikkonen et al., 2009).

Additionally, acetylation and methylation have been reported to covalently modify AR. AR is a direct substrate of both HATs and HDACs, and acetylation has a dual role in terms of AR transcriptional output (Gioeli and Paschal, 2012). Recently, the hinge region of AR has been shown to be methylated by SET9 methyltransferases hence acting as a positive regulator of AR function (Gaughan et al., 2011; Ko et al., 2011). All these PTMs and signaling cascades are subjected to significant cross-talk, but major knowledge gaps as to their biological significance exist at present.

### 3.4 Modular structure and functional domains of androgen receptor

The modular structure of AR comprises four major functional domains, the NTD (or A/B domain) specified by exon 1, the DBD specified by exons 2 and 3, and the hinge region and the LBD (or androgen-binding domain) specified by exon 4 through 8 (Gelmann, 2002; Quigley et al., 1995). The AR NTD regulates the formation of pre-initiation complex (PIC) by directly recruiting the basal transcription factors such as TFIIF to androgen responsive genes. This interaction makes the NTD more compact and its conformation more active, leading to further recruitment of coregulators to promote transcription (Kumar et al., 2004; Reid et al., 2002). The AR NTD contains the major transactivation function, called activation function 1 (AF-1), which is hormone-independent (Palvimo et al., 1993; McEwan, 2004; Shen and Coetzee, 2005). In the absence of LBD, AF-1 becomes constitutively active. The AR NTD has 3 conserved
motifs within the AF-1 region. The first motif is formed by amino acids 233–246, and it is very well conserved through all vertebrate species. The other two motifs in the human AR NTD, amino acids 23-FQNLF-27 and to a to a lesser degree 433-WHTLF-473, have been shown to interact and stabilize the interaction of AF-1 with the AF-2 of the LBD (Dubbink et al., 2004; He et al., 2002).

Figure 6. Schematic representation of the AR DBD composed of two zinc fingers, each with four cysteine residues in a tetrahedral arrangement with the zinc ion. The first zinc finger determines the specificity of the DNA binding, and the second is involved in receptor dimerization. (Adapted and reprinted with permission. © (2002) American Society of Clinical Oncology. All rights reserved. Gelmann, E.P.: Molecular biology of the androgen receptor. J Clin Oncol 2002; 20(13): 3001-3015).

The AR DBD is responsible for recognizing and binding to an appropriate response element, *i.e.* ARE, to mediate AR-regulated transcription. The AR DBD is composed of nine cysteine residues that form two zinc fingers, and a C-terminal extension, and it is the most conserved domain within the nuclear receptor family. The two zinc fingers are tetrahedrally coordinated to two zinc ions, and both zinc fingers form a single compact structure that binds to the major groove of DNA (Fig. 6). The first zinc finger determines the specificity of DNA binding, and the residues in the second are involved in receptor dimerization (Heemers and Tindall, 2007; Heinlein and Chang, 2002). The canonical ARE and GRE are very similar, and they represent a partial-palindrome of two inverted repeats of two hexameric core DNA sequences spaced by 3 nt. The canonical consensus sequence recognized by AR is 5’-AGAACAnnnTGTTCT-3’, and it is identical to that recognized by GR (Nelson et al., 1999). ER recognizes a slightly different 15-bp palindromic sequence element, 5’-AGGTCAnnnTGACCT-3’ (Driscoll et al., 1998). The steroid receptors can also bind to non-canonical HREs, where two half sites of an HRE
can either be partially palindromic, inverted or direct repeats or then with variable spacing between the two hexameric cores (Claessens and Gewirth, 2004). Given the high similarity of AR and GR response elements, it is of paramount importance to understand how these two steroid receptors are recruited to their specific chromatin loci \textit{in vivo} in response to their cognate ligands and regulate their respective programs.

The hinge region contains the bipartite NLS that enables the nuclear translocation of ligand-bound activated AR (Poukka et al., 2000b), and mutations in NLS results in accumulation of cytoplasmic AR. The nuclear translocation of AR is mediated through the importin-α-importin-β complex (Cutress et al., 2008). In addition, the hinge region contains the PEST sequences that might contribute to its degradation by the 26S proteasome (Tanner et al., 2004), and it also has a putative phosphorylation site (Gioeli et al., 2002). The hinge region was previously considered to be only a flexible linker between the DBD and the LBD. However, recent studies have shown that it has an additive role in DNA binding and receptor dimerization, and it was also suggested to attenuate AR transcriptional activity (Haelens et al., 2007; Wang et al., 2001). Moreover, the hinge region has a target site for acetylation, ubiquitination, and methylation, and it is important for AR binding to canonical versus selective AREs (Clinckemalie et al., 2012; Haelens et al., 2007).

The AR LBD has been well characterized by crystallographic studies that show its similarity to other NRs (Moras and Gronemeyer, 1998; Sack et al., 2001). It has 12 discrete α-helices forming a 3-layer anti-parallel “α-helical sandwich” fold. The ligand-binding pocket is formed by the helices 3, 4, 5, 7, 11 and 12, together with the β-sheet preceding helix 6 (Matias et al., 2000; Williams and Sigler, 1998). Binding of an agonist into the ligand-binding pocket has been suggested to make a conformational switch of the LBD-stabilizing helix 12. This leads to formation of a hydrophobic groove at the top of the LBD that is generally referred to as activation function 2 (AF-2) (Heemers and Tindall, 2007; Nagy and Schwabe, 2004). The AF-2 in AR is transcriptionally weak compared to other NRs (Ikonen et al., 1997; Moilanen et al., 1997), but it mediates an important ligand-dependent NTD/LBD interaction (Ikonen et al., 1997). Additionally, AF-2 is the major protein-protein interaction platform, and it modulates the recruitment of LXXLL-motif containing coactivators determining AR transcriptional activity.
Review of the literature

(Heemers and Tindall, 2007; Heery et al., 1997; Ikonen et al., 1997; Nagy and Schwabe, 2004). The hydrophobic groove in the AR-LBD interacts strongly with FXXLF motif of the NTD, whereas it interacts only relatively weakly with LXXLL motifs commonly found in the coactivators (Dubbink et al., 2004; He et al., 2000; Heery et al., 1997; Hu and Lazar, 1999). This facilitates the intramolecular and intermolecular interactions between the AR NTD and the C-terminus. In the context of PCa, most of the AR point mutations have been mapped to the LBD (Buchanan et al., 2001), but not all PCa and CRPC cases harbor these mutations (Taplin et al., 1995; Visakorpi et al., 1995).

3.5 Androgen receptor mediated transcriptional regulation

The predominant function of AR is ligand-dependent, and in response to androgens it regulates a diverse group of target genes. In the absence of androgens, AR is mainly cytoplasmic and stays in multiprotein complex with chaperones and immunophilins. The complex consists of Hsp90 as the main chaperone and the co-chaperones are typically p23 and FKBP51 or FKBP52 (Echeverria and Picard, 2010; Pratt and Toft, 1997). The chaperone complex dissociates after androgen binding to the ligand-binding pocket of AR monomer resulting in a conformational change of the receptor’s LBD. This leads to receptor dimerization and nuclear translocation. Inside the nucleus, the AR dimer recognizes and binds to its consensus ARE, either in the proximal promoter and/or distal enhancer elements of its target genes. This is followed by recruitment of p160 proteins, CBP, p300 in a sequential order, resulting in enrichment of these coregulatory proteins at distal enhancer and promoter proximal sites but not in the intermediate regions (Wang et al., 2005). Additionally, looping was validated by chromosome conformation capture assay, showing that there is direct contact between enhancer and promoter regions of the PSA gene (Wang et al., 2005). This study also detailed the spatio-temporal stoichiometry of AR and other associated molecules forming a competent AR transcriptional complex on the PSA regulatory region (Fig. 7). The looping-mediated phenomena is not unique to PSA; rather, chromatin looping has been reported for some other androgen regulated genes as well, such as TMPRSS2 (Wang et al., 2007), FKBP5 (Makkonen et al., 2009), UBE2C (Chen et al., 2011). The AR coactivators and RNA Pol II forming the AR transcription complex is recruited both to the enhancer and promoter regions of the PSA gene (Shang et al., 2002).
Figure 7. Spatio-temporal view of AR-coactivator complex assembly on PSA regulatory region. (A) Recruitment of AR and coregulators after androgen stimulation in time-dependent manner. (B) Spatial pattern of recruitment of AR and coregulators bridging the enhancer and promoter elements with tracking of RNA Pol II. (Adapted and reprinted from Wang et al., Spatial and temporal recruitment of AR and its coactivators involves chromosome looping and polymerase tracking. Mol Cell 2005; 19:631-642 with permission from Elsevier, copyright 2005).

Subsequent studies have shown that the occupancy of AR and other associated coregulators is more enriched in the enhancer as opposed to the proximal promoter region (Kang et al., 2004; Wang et al., 2005). This raises the issue of communication between enhancer and promoter proximal elements and consequently, the chromatin looping/sliding was proposed to be the most plausible mechanism. According to the looping/sliding model, the interaction between these distant elements is mediated via common coactivators referred to as mediator proteins and tracking of RNA Pol II. For example, mediator component MED1 (TRAP220) was implicated in mediating the enhancer-promoter proximity, induced by AR signaling (Wang et al., 2011). Evidence for the importance of chromosomal looping in gene regulation and gene fusions is gradually accumulating, and a variety of high-throughput and unbiased approaches have been developed to address this question (Kadauke and Blobel, 2009; Naumova and Dekker,
However, it remains unclear as to whether chromatin looping is a general event in AR-mediated gene regulation or is just operating at few specific genomic loci. To understand this, first the role of AR in one-dimension needs to be profiled, such as the whole genome binding events (cistrome) and gene expression programs (transcriptome).

Next-generation sequencing-based ChIP-seq studies from human PCa cells have allowed genome-wide mapping of AR binding sites (ARBs). The mapping has revealed that AR employs a distal mode of transcriptional regulation, as majority of the cis-elements are located beyond the boundaries of a classical promoter region. In this distal enhancer-driven mode of regulation, AR is assisted by other collaborating transcription factors such as FoxA1 (Wang et al., 2007; Wang et al., 2009). The pioneering role of FoxA1 in ERα-driven transcription in breast cancer cells (Carroll et al., 2006; Hurtado et al., 2011; Lupien et al., 2008) makes it pertinent to investigate the role of FoxA1 in AR-mediated gene regulation. FoxA1 has a clear prognostic value in breast cancer as high FoxA1 predicts good clinical outcome (Albergaria et al., 2009; Badve et al., 2007). Recently, an elegant ChIP-seq study performed on tissue material from primary breast tumors has implicated FoxA1 in mediating differential ERα cistrome with different clinical outcomes (Ross-Innes et al., 2012). Thus, delineating the role of FoxA1 in AR cistrome formation, transcription program and prognostic value in PCa could provide an opportunity to develop novel therapeutic approaches for PCa.
4 Androgen receptor in prostate cancer (PCa) and disease

4.1 Cancer of the prostate gland

The prostate gland is a walnut-sized exocrine gland located just below the bladder, surrounding the urethra. It produces important components of the seminal fluid that function in protecting the genetic material of spermatozoa. Prostatic secretion is slightly alkaline and contains simple sugars, zinc and the proteolytic enzymes such as prostatic acid phosphatase (PAP) and PSA. PSA is encoded by the KLK3 gene that belongs to the human kallikrein cluster of genes located on chromosome 19 (Lilja, 2003), and expression of PSA is tightly androgen-regulated. Androgens control the growth and development of prostate. At the histological level, the human prostate contains a pseudostratified epithelium with three differentiated cell types: luminal, basal and neuroendocrine. The luminal epithelial cells secrete PSA into the lumen and express high levels of AR, but basal and neuroendocrine cells express AR at very low or undetectable levels. In the case of PCa, high levels of PSA enter into blood circulation, and hence, it is used as a general marker in PCa screening. An increasing body of evidence points to the central role of AR in pathophysiology and progression of PCa, and PSA has been the extensively studied in the context of AR-mediated gene regulation (Abate-Shen and Shen, 2000; Balk et al., 2003).

Most prostate malignancies originate from luminal epithelial cells and are thus classified as adenocarcinomas. The precursors of PCa, prostatic intraepithelial neoplasias, are generally characterized by hyperplasia of luminal epithelial cells and reduction in the number of basal cells, enlarged nuclei and nucleoli, and marked elevation of cell proliferation markers. A remarkable feature in PCa is the complete absence of basal cells. Human PCa displays significant heterogeneity, and it is therefore relevant to screen for new prognostic markers to have a better distinction between various PCa subtypes in conjunction with Gleason scoring for a more patient-specific treatment (Shen and Abate-Shen, 2010). Recent oncogenomic analyses are providing compelling evidence for molecularly defined subtypes in PCa having the AR in their core (Taylor et al., 2010; Tomlins et al., 2008). Several alterations have been reported in the AR gene (Linja and Visakorpi, 2004), highlighting the importance of AR signaling in PCa.
4.2 Androgen signaling in prostate cancer

Androgens and AR are indispensable for normal growth and development of the prostate gland. Androgens play a critical role in malignant growth of prostate cells, and in progression of PCa. Numerous studies using various PCa models in mice and PCa cell lines, such as LNCaP, VCaP, 22Rv1, LAPC4, MDACaP2a, MDACaP2b, and LuCaP xenografts, have led to an accretion of information about androgen signaling in PCa. Microarray-based gene expression studies estimate that the AR transcriptome is between 10,570 and 23,448 transcripts in size, but with increasing number of other transcribed products, such as microRNAs, splice variants and fusion transcripts, the full repertoire may be even larger (Dehm and Tindall, 2006, 2011). A recent microarray coupled with quantitative RT-PCR-based analysis in LNCaP cells reported 619 androgen-responsive genes, only 75% of which have a known function (Ngan et al., 2009). The androgen-responsive genes are implicated in diverse roles, such as cell proliferation, apoptosis, lipid and sterol metabolism, and regulation of cell cycle, to name a few. One such example is the androgen-regulated gene CAMKK2, a metabolic master gene involved in aerobic glycolysis and anabolism in PCa cells (Massie et al., 2011). The downstream target genes are also involved in extensive cross-talk with other signaling pathways having a profound effect on the final cellular response (Dehm and Tindall, 2006; Lamont and Tindall, 2010).

A vast amount of information about downstream target genes from multiple resources have led to the quest for understanding the mechanisms as to how AR regulates these genes and the downstream programs. The mechanistic insight in the regulation of these downstream target genes by AR has come from several genome-wide studies; one of the key approaches being ChIP-seq. The collaborating role of FoxA1 has been implicated in AR function as discussed above (Wang et al., 2007). Likewise, another cooperative transcription factor ERG, has been identified in VCaP cells, harboring androgen-driven TMPRSS2:ERG fusion proteins resulting in enhanced ERG activity (Yu et al., 2010). ERG, in turn, attenuates AR signaling by inhibiting AR expression, which is achieved by direct binding to gene-specific loci and by inducing the EZH2-mediated repressive program. EZH2 represses several genes in PCa cell lines, one of them being DAB2IP, (Ras GTPase-activating protein), recently reported to be a regulator of metastasis in PCa (Min et al., 2010).
Identification of the androgen-driven *TMPRSS2:ERG* fusion gene (Tomlins et al., 2005) and its association with more aggressive clinical outcome in case of PCa have led to a systematic screening of other fusion genes in PCa cells, the largest of which are ETS fusions (Rubin et al., 2011). Additionally, new insightful information is evolving on the androgen-dependent role of microRNAs (miRNAs) (Jalava et al., 2012; Waltering et al., 2011), long non-coding RNAs (IncRNAs) (Kino et al., 2010) and enhancer RNAs (eRNA) (Wang et al., 2011). Furthermore, AR target genes downstream can be involved in significant cross-talk with components of multiple signaling pathways, influencing tumor progression. For example, the AR target gene *NKX3.1*, a tumor suppressor, has recently been shown to cross-talk with the *MYC* oncogene, affecting prostate tumorigenesis by regulating a common set of target genes in both mouse and human, and dysregulation of which is associated with disease relapse (Anderson et al., 2012). As an interesting side note, recent evidence shows that AR binding is very similar to that of ER in ER negative (ER-) breast cancer cells (Robinson et al., 2011a). In a parallel study, it was shown that AR mediates ligand-dependent activation of WNT7B and HER3, components of Wnt and Her2 signaling pathways (Ni et al., 2011), thus expanding the outreach of AR to cancer of other tissues.

### 4.3 Androgen signaling in castration-resistant prostate cancer

The first line of treatment for metastatic PCa is androgen ablation, which results in significant inhibition of AR signaling associated with decreased serum PSA levels and concomitant tumor regression (Feldman and Feldman, 2001). However, a relapse occurs eventually resulting in cancer resistant to further hormonal manipulations termed as castration-resistant prostate cancer (CRPC), but retains active AR signaling (Chen et al., 2008; Knudsen and Kelly, 2011). Importantly, AR controls and executes a distinct transcriptional program in CRPC compared to androgen-dependent PCa cells, resulting in androgen-independent growth by upregulating M-phase and cell cycle genes (Wang et al., 2009). AR amplification and overexpression are the major features of CRPC (Chen et al., 2004; Visakorpi et al., 1995). This effect has been validated in a model LNCaP cell line expressing different levels of AR, in which AR overexpression sensitizes the cells to low levels of androgens (Waltering et al., 2009). However, AR overexpression in CRPC is only partially explained by AR gene amplification, and other mechanisms include deregulation of AR-regulating microRNAs (Ostling et al., 2011) and other transcription
factors or coregulators. For example, a recent study has shown that NF-κB mediates upregulation of AR and NF-κB inhibition led to decreased AR expression and PCa cell growth both in vitro and in vivo (Zhang et al., 2009). Stat5a/b is another transcription factor of therapeutic interest with 95% of CRPCs showing active Stat5a/b signaling. AR and Stat5a/b interact physically with each other showing transcriptional synergy where active Stat5a/b increases AR transcriptional output and vice-versa (Tan et al., 2008). In addition to the AR-mediated effects, Stat5 regulates PCa cell viability through mechanisms that are independent of the AR (Gu et al., 2010). AR overexpression is also associated with the loss of retinoblastoma protein (RB1), a tumor suppressor gene. Loss of RB1 leads to increased expression of E2F1 transcription factor, resulting in increased expression of AR. The loss of RB is highly over-represented in CRPC (Sharma et al., 2010). Other mechanisms leading to AR recurrence and overexpression are AR mutations and alternative splicing, intracrine androgen synthesis from weak adrenal androgens, PTMs and coregulator associated alterations (Knudsen and Kelly, 2011). Besides this, increasing growth factor signaling (e.g., EGFR, IGF-1R, IL-6R, HER2 receptor tyrosine kinase) can enhance AR signaling and confer castration resistant as inferred from various preclinical models (Chen et al., 2008). The AR signaling axis is crucial in both primary and castration-resistant PCa, and hence, it has been a major goal to develop effective strategies against AR signaling.

4.4 Endocrine treatments of prostate cancer

The AR axis is central in pathophysiology of PCa and disease progression, and therefore, blocking of AR signaling is a hallmark of PCa therapeutics. The classical finding of Huggins and Hodges in 1941 revealed that castration or estrogen treatment inhibits growth of PCa, whereas the growth was activated by androgen injections (Huggins and Hodges, 1941). The majority of PCas are initially androgen-dependent and organ-confined which are treated by radiation or with surgery. The first line of treatment for metastatic PCa is androgen-deprivation therapy (ADT) that involves the use of luteinizing hormone–releasing hormone (LHRH) agonist or antagonist or orchiectomy, resulting in 90–95% reduction in serum testosterone levels. Non-steroidal anti-androgens, such as bicalutamide and flutamide, are commonly used in PCa therapy (Miyamoto et al., 2004; Wilding et al., 1989), besides chemical or surgical castration. Bicalutamide and flutamide compete with testosterone and DHT for binding to AR, thereby preventing AR activation.
Bicalutamide was previously thought to be a pure anti-androgen (Furr, 1990; Verhelst et al., 1994), but in the presence of high cellular AR content it can act as a weak androgen agonist (Chen et al., 2004). One of the first anti-androgens used was a steroidal AR modulator cyproterone acetate (CPA) used for PCa treatment (McLeod, 1993) but it also has androgen agonistic actions (Wilding et al., 1989). Mifepristone (RU486), in addition to being antiprogestin and antiglucocorticoid, has androgenic/antiandrogenic properties as well (Kang et al., 2004; Kuil et al., 1995; Wiechert and Neef, 1987). All these compounds antagonize AR function by competing with T or DHT for binding to the ligand-binding domain (LBD) of AR.

The use of steroidal anti-androgens has also been found to induce somatic AR mutations, releasing the transcriptional inhibition of AR (Chen et al., 2004; Loblaw et al., 2007). The intracrine androgen synthesis and AR amplification in CRPC further complicate the treatment options, which has led to development of secondary hormonal therapies intended for complete inhibition of AR signaling. One of the promising candidates, currently in phase III clinical trials, is abiraterone acetate, which targets both adrenal and intracrine androgen synthesis. It is an irreversible CYP17 inhibitor that blocks the synthesis of androgens from C21 steroid precursors (Attard et al., 2008; Ryan et al., 2010). It targets events preceding receptor-ligand binding; concurrently, new anti-androgens are being developed to target the postreceptor-ligand binding events as well. For example, the super anti-androgen MDV3100 is currently in phase II clinical trials and has been shown to bind AR with 10-fold higher affinity than the anti-androgen bicalutamide leading to 10-fold reduction in PSA secretion. Subsequent studies have also shown its efficacy in inhibiting growth of castration-resistant xenografts (Chen et al., 2008; Tran et al., 2009).

Continuous CYP17 inhibition leads to a rise in adrenocorticotropic hormone (ACTH) levels causing accumulation of steroid precursors upstream of CYP17. This results in excess of mineralocorticoids characterized by fluid retention, hypertension and hypokalemia. This problem has been ameliorated by using either mineralocorticoid antagonists or a low dose of glucocorticoids (dexamethasone), which prevented the rise of ACTH and levels of steroids upstream of CYP17 (Attard et al., 2009). Glucocorticoids act as the ligand for GR, and GR plays an important regulatory role in inflammation and cancer progression. Glucocorticoids have been used in the treatment of CRPC (Fakih et
al., 2002; Venkitaraman et al., 2008) and GR signaling exhibits tumor suppressor activity in CRPC cells (Yano et al., 2006; Yemelyanov et al., 2007). Limited evidence suggests that promoting GR signaling can be beneficial in treatment of CRPC, since GR antagonizes AR action in CRPC. For instance, it is known that GR interferes with transcriptional activity of NF-κB and AP-1. It has been shown that NF-κB enhances AR expression and regulates PCa growth (Zhang et al., 2009). Thus, GR-mediated inhibition of NF-κB may antagonize this effect. Importantly, most of these studies have been performed in AR-negative PCa cells, such as DU145 or PC3 cells (Kassi and Moutsatsou, 2011). The information on mechanism of GR signaling in AR-positive PCa cells is still limited. Hence, it will be of great importance to characterize GR signaling vis-à-vis AR signaling in androgen-responsive PCa cells.

4.5 Androgen receptor and other related disorders

Genetic defects of AR are implicated in several X-linked pathogenic states, including androgen insensitivity syndrome (AIS), a motor neuron disease known as Kennedy’s disease or spinal and bulbar muscular atrophy (SBMA) and male infertility (Li and Al-Azzawi, 2009; Quigley et al., 1995). Individuals with AIS are characterized by 46 XY karyotype normal for males, but they develop secondary female sexual features. In the most severe cases of this syndrome, the patients produce androgens, but their AR is inactive and fails to respond, which is mostly attributed to single mis-sense mutations in the LBD leading to loss or weakening of LBD function.

The molecular basis of SBMA is the expansion of polyglutamine tracts (CAG repeats) in the transactivation domain located in the NTD of AR (La Spada et al., 1991). Normal repeat length varies from 8 to 31, with an average of 20, but individuals with SBMA can have the repeat expansion greater than 40 (Gelmann, 2002; Quigley et al., 1995) leading to a gain-of-function of AR, which results in wasting of proximal limb muscles, changes in facial muscles and motor neuron damage. In contrast, shortening of the glutamine repeat gives rise to more transcriptionally active AR and has been suggested to be linked with a predisposition to prostatic neoplasia (Giovannucci et al., 1997). Furthermore, evidence is accumulating about the roles of PTMs in various human pathophysiological conditions, such as the role of AR phosphorylation and sumoylation in PCa and in SBMA, AR acetylation in PCa and AR ubiquitination in SBMA (Anbalagan et al., 2012).
Aims of the study

AIMS OF THE STUDY

The AR pathway is central to growth and progression of PCa and CRPC. The aim of this study was to characterize genome-wide AR binding sites (cistromes) and the downstream target genes of AR in prostate cancer cells. Over-representation of forkhead motifs in AR cistromes led us to examine whether the collaborating factor FoxA1 regulates and/or modulates AR function in prostate cancer cells. AR and GR recognize almost identical DNA sequence on chromatin, and the aim was to study the role of GR binding in PCa cells and its effects on AR-mediated gene expression. In quest of this, we have utilized chromatin immunoprecipitation coupled with massively parallel sequencing (ChIP-seq), whole genome transcriptomics and RNA interference to understand the AR-mediated gene expression in prostate cancer cell lines and tissue microarrays from primary prostate adenocarcinomas for the prognostic significance of AR and FoxA1 expression.

The following specific aims were being addressed:

**Aim 1** Delineation of genome-wide AR cistromes by ChIP-seq in response to AR ligands: $5\alpha$-dihydrotestosterone (DHT), cyproterone acetate (CPA), and mifepristone (RU486).

**Aim 2** Delineation of the cistrome of the collaborating transcription factor FoxA1; its role in AR-mediated signaling and androgen-dependent gene expression, and correlation of FoxA1 expression to prostate cancer-specific survival.

**Aim 3** Delineation of GR cistrome in response to dexamethasone (Dex) and GR signaling vis-à-vis AR cistrome and signaling in androgen-responsive PCa cell lines.

**Aim 4** Delineation of AR cistrome in an AR overexpression model of CRPC and consequences of AR overexpression on genome-wide recruitment of AR to chromatin.
MATERIALS AND METHODS

1. Cell lines, xenografts and patient samples

1.1 Cell lines and cell culture procedure

LNCaP-1F5 cells genetically engineered to express rat GR (Cleutjens et al., 1997), a kind gift from Professor Jan Trapman (Erasmus Medical Centre Rotterdam, The Netherlands), were grown in RPMI-1640 supplemented with 10% fetal bovine serum (FBS), 2 mM L-glutamine, 10 mM HEPES and antibiotics (penicillin and streptomycin). VCaP cells (Korenchuk et al., 2001), obtained from American Type Culture Collection (ATCC, VA, USA), were grown in ATCC-formulated Dulbecco's Modified Eagle's Medium (Cat. # 30-2002) supplemented with 10% FBS and antibiotics. The LNCaP-1F5 cells and VCaP cells were used in studies I and II.

The AR overexpressing LNCaP cells used in study III have been described in original communication III and in Waltering et al. (2009). Briefly, the parental LNCaP cells (Horoszewicz et al., 1983), obtained from ATCC (MO, USA) were transfected either with the pcDNA3.1(+) empty vector (Invitrogen) or with the pcDNA3.1-AR (containing the AR coding region) vector using Lipofectamine Plus (Invitrogen) as per the manufacturer’s instructions. Transfected clones were selected under 400 µg/ml Geneticin (G418, Invitrogen) for two weeks. Two clones expressing moderate and high levels of AR protein (LNCaP-ARmo and LNCaP-ARhi, respectively) were selected for further studies and were cultured in ATCC-formulated RPMI-1640 medium (Cat. # 30-2001) supplemented with 10% FBS containing 200 µg/ml Geneticin (Waltering et al., 2009).

In studies I, II and III, for total RNA isolations and gene expression profiling, the cells were seeded in 10% FBS treated with dextran-coated charcoal (DC-FBS) and antibiotics. In studies I and II, for RNA silencing experiments, the cells were seeded in DC-FBS without antibiotics.
**Materials and methods**

**1.2. Prostate cancer patient and xenografts material**

The Institutional Review Boards of the University Central Hospitals of Helsinki and Tampere approved the use of the tissue specimens and patient information. Prostate cancer patient material was used in studies I and III. In study I, a cohort of 350 patients obtained from Helsinki University Hospital were used with a median follow-up time of 13.3 years (range, 11.3–25.0 years). Patient cohort details are described in S10 of original communication I. In study III, a cohort of 277 patients obtained from Tampere University Hospital was used, and details are described in original communication III. Two PCa xenografts used in study III, LuCaP69 and LuCaP73, were grown in intact immunocompromized male mice and were a kind gift from Professor Robert L. Vessella (University of Washington, Seattle, WA, USA).

**1.3 Tissue microarray construction (TMA) and immunohistochemistry**

In study I, TMAs were constructed from formalin-fixed paraffin-embedded blocks of 350 patients who underwent prostatectomy at the Helsinki University Central Hospital. To account for tumor heterogeneity, two TMA cores were from the most dominant Gleason grade area and one from the second most dominant Gleason grade area. One core of each patient contained an adjacent non-cancerous area. The antibodies used were FoxA1 (ab23738, Abcam), AR (NCL-AR-318, Immuno Diagnostics). In study III, TMAs containing 185 formalin-fixed paraffin-embedded prostatectomy and 92 CRPC (TURP) specimens obtained from Tampere University Central hospital were stained with FEN1 antibody (mAb clone 4E7, LifeSpan Biosciences Inc.). The details of TMA construction, antigen staining, and evaluation are described in original communications I and III.

Kaplan-Meier analysis was performed in studies I and III for either disease-specific or biochemical progression-free survival for FoxA1 and FEN1 expression levels, respectively. The $p$-values for survival analysis were calculated by Mantel-Cox (logrank) test. In study I, the univariate and multivariate regression analysis of tumor characteristics were performed by Cox proportional hazard model and the pairwise associations between FoxA1 expression and tumor characteristics by Pearson’s product moment correlation in a series of 350 patients with prostate cancer as described in S11 and S12 of original communication I.
2. Gene expression profiling and RNA interference

2.1 RNA isolation, cDNA synthesis and quantitative RT-PCR

For RNA isolation, LNCaP-1F5 cells (300,000) and VCaP cells (500,000) were seeded in 2 ml of DC-FBS medium and antibiotics. After 3 days, cells were treated for 24 h with ligands [100 nM DHT, 100 nM Dex or vehicle (for both LNCaP-1F5 and VCaP) and 1 µM CPA, 1 µM RU486 (for LNCaP-1F5 only)] prior to cell lysis. Isolation of total RNA was done from these lysates using RNeasy MinElute cleanup kit (Qiagen Inc., California, USA) as per the manufacturer’s instructions. cDNA synthesis was performed from 2 µg of RNA using random hexamers supplied with Transcriptor High Fidelity cDNA synthesis kit (Roche Inc., Mannheim, Germany) as per the manufacturer’s instruction. Quantitative RT-PCR (qRT-PCR) was performed using SYBR Green mastermix (Roche Inc., Mannheim, Germany) according to the manufacturer’s instructions. The primers for qRT-PCR were designed using Primer 3 (http://frodo.wi.mit.edu/). Relative expression levels were calculated using the delta Ct method using the formula $2^{-\Delta Ct}$, where $\Delta Ct$ is Ct(Target gene)-Ct(Reference gene). The primer sequences are listed in the respective original communications.

2.2 Gene expression profiling and data analysis

Cells were cultured in DC-FBS supplemented with antibiotics for 3 days followed by ligand treatment for 24 h. In the case of RNAi experiments, 72 h after transfection, the cells (both parental and siFoxA1 cells) were exposed to 100 nM DHT or 100 nM Dex for 24 h. Total RNA was isolated using RNAeasy kit (Qiagen). RNA samples from biological replicates were hybridized to Illumina HumanHT-12 v3 Expression BeadChip Kits at Biomedicum Functional Genomics Unit (FuGU) according to the manufacturer’s instructions. The data analyses were performed using Anduril software (Ovaska et al., 2010) together with ‘R’ (http://www.r-project.org/) and Bioconductor ‘lumi’ package (http://www.bioconductor.org). Raw intensity values were normalized independently between arrays for each sample using quantile normalization. The median value of sample replicates was used to calculate differentially expressed genes. Fold changes of $\leq 1.7$ and $\geq 1.7$ were set as the cutoff values and expression heatmaps were generated by unsupervised hierarchical clustering within the Anduril framework. Pathway analyses for the differential expressed genes were carried out using the WebGestalt.
**Materials and methods**

(http://bioinfo.vanderbilt.edu/webgestalt) in studies I and II. Gene Ontology (GO) and KEGG pathway analyses to find enriched categories were done against the reference human genome using hypergeometric test with Benjamini and Hochberg (1995) multiple test adjustment. In study III, the GO analysis was performed with GeneTrail based on genes that showed AR binding in a window of 25 kb in LNCaP-pcDNA3.1 cells treated with 1 nM DHT and LNCaP-ARmo and LNCaP-ARhi cells treated with 1 nM DHT for 2 hours.

### 2.3 RNA interference (RNAi)

Cells were cultured in DC-FBS without antibiotics to 60% confluence and then transfected with siRNAs (ON-TARGETplus™ SMARTpool siRNA, Dharmacon, Thermo Scientific) with Dharmafect-3 transfection reagent according to the manufacturer's instructions. The cells were incubated for 72 h and then exposed to steroid (DHT or Dex) for 2 h (ChIP and ChIP-seq) or for 24 h (gene expression profiling). The FoxA1 siRNA SMARTpool and additional details are described in original communication I, and details of siRNAs for *FEN1, ZWINT* and *SNAI2* are described in original communication III.

### 2.4 Immunoblotting

For immunoblot analysis, cells were lysed in RIPA2 buffer (50 mM-Tris-HCl, pH 7.8, 150 mM NaCl, 15 mM MgCl₂, 5 mM EDTA, 0.5% NP-40 and 0.3% Triton X-100) and mixed with 2X SDS sample buffer (Laemmli buffer, Laemmli, 1970) in 1:1 ratio and heated at 95°C for 5 min. The samples were then resolved on 10% SDS-PAGE gel and transferred to nitrocellulose membrane (GE Healthcare). Immunoblottings were done by incubating the filters with anti-AR (Kang et al., 2004; Thompson et al., 2006), rat anti-GR (Widen et al., 2000), anti-FoxA1 (ab23738, Abcam) and anti-GAPDH (sc-47724, Santa Cruz) primary antibodies. After washing, the filters were incubated with either a secondary anti-mouse or anti-rabbit IgG antibody conjugated to horseradish phosphatase. Immunodetection was done using ECL western blot reagents (GE Healthcare) as per the manufacturer’s instruction.
3. **Genome-wide approaches for gene regulation**

3.1 **Chromatin immunoprecipitation (ChIP) and ChIP-qPCR**

Five million LNCaP-1F5 cells were cultured on 150-mm dishes in DC-FBS for 4 days and then exposed to the ligands (100 nM DHT, 1 μM CPA, 1 μM RU486 and 1 μM bicalutamide, 100 nM Dex or vehicle) for 2 h. VCaP cells were treated with 100 nM DHT, 100 nM Dex or vehicle for 2 h. Cells were fixed in 1% formaldehyde (Merck KGaA, Darmstadt, Germany) for 10 min at room temperature and washed twice with ice-cold PBS. Cells were scraped in hypotonic lysis buffer (5 mM PIPES, pH 8.0, 85 mM KCl, and 0.5% NP-40) containing Complete Protease Inhibitor Cocktail (PIC, Roche Inc., Mannheim, Germany). After centrifugation cells were sonicated in RIPA buffer (1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS in 1X PBS) containing PIC in a volume of 400 μl to shear DNA to an average fragment size of 100–500 bp using Bioruptor UCD-300-TO (Diagenode) and cleared by centrifugation. For each immunoprecipitation (IP), 100 μl of sonicated chromatin was diluted with 900 μl RIPA buffer containing PIC and 100 μl was stored as input fraction. To the remaining, 100 μl of antibody-coupled Dynal protein-G magnetic beads (Invitrogen Inc., Carlsbad, CA, USA) were added and incubated on a rotator overnight at 4°C. After incubation, beads were washed 5 times with LiCl wash buffer (100 mM Tris-HCl, pH 7.5, 500 mM LiCl, 1% NP-40, and 1% sodium deoxycholate) and twice with 10 mM Tris-HCl (pH 8.0) containing 1 mM EDTA. Co-immunoprecipitated DNA was eluted from beads by incubating for 1 h at 65°C in IP elution buffer (1% SDS, 0.1 M NaHCO3), and reverse cross-linked overnight at 65°C followed by purification using QIAquick PCR purification kit (Qiagen Inc., California, USA). The primer and antibody details are listed in the respective original communications I, II and III.

ChIP-qPCR was carried out using the co-immunoprecipitated DNA (and input DNA as a quantification control) using SYBR Green mastermix (Roche Inc., Mannheim, Germany) as per manufacturer’s instructions. The primers for ChIP-qPCR were designed using Primer 3 (http://frodo.wi.mit.edu/). The enrichment of ChIP relative to input was calculated using the formula $2^{-\Delta Ct}$, where Ct is the threshold cycle in PCR and ΔCt is Ct(IP)-Ct(Input).
3.2 ChIP-sequencing assay

For deep sequencing, immunoprecipitated DNA samples together with input samples were processed for library preparation according to Illumina’s instructions. In brief, DNA samples were blunt-ended and ligated to sequencing adapters. Adapter-ligated DNA fragments (size range, 150–300 bp) were excised from agarose gel and purified with Qiaquick gel extraction kit (Qiagen Inc., California, USA). Isolated DNA was amplified by PCR with adapter primers (15–20 cycles) and purified using Qiaquick MinElute PCR purification kit (Qiagen Inc., California, USA). The purified DNA library was quantified on Agilent High Sensitivity DNA kit (Agilent Technologies, Santa Clara, California, USA) and sequenced on Illumina Genome Analyzer II at Biomedicum Functional Genomics Unit (FuGU). ChIP-seq reads were filtered using the Illumina chastity filter during the base-calling process and ELAND (Illumina) or Bowtie were used to align 30-bp reads to human genome (hg19, Genome Reference Consortium GRCh37), allowing up to two mismatches per read. All ChIP-seq experiments were performed in biological duplicates showing >80% overlap between the replicates.

3.3 DNaseI-hypersensitive site sequencing (DHS-seq)

The DNaseI-hypersensitivity assay and deep sequencing were performed essentially as described by Song and Crawford (2010). Briefly, cells were lysed with reticulocyte standard buffer (RSB, 10 mM Tris-HCl, pH 7.4, 10 mM NaCl and 3 mM MgCl₂) containing 0.1% NP-40 to isolate intact nuclei. Intact nuclei were treated with different concentrations of DNaseI (Roche Inc., Mannheim, Germany) for 15 min at 37°C, and reactions were stopped with 0.1 M EDTA. The digested high-molecular-weight DNA was embedded in low-melt agarose, washed twice with LIDS buffer (1% lauryl sulfate, 10 mM Tris-HCl and 100 mM EDTA) at room temperature followed by overnight incubation at 37°C to remove proteins. Subsequently, the DNA plugs were washed five times with 50 mM EDTA and run through pulsed field gel electrophoresis system to identify the optimal DNaseI digest (DNA fragment size 100–1000 kb). DNaseI-digested ends were blunt-ended in gel with T4 DNA polymerase, melted at 65°C, and DNA was purified by phenol extraction and ethanol precipitation followed by library construction for deep sequencing. The library was constructed by ligating linkers with Mmel sites (TCCGAC) to the blunt-ended DNaseI-digested ends followed by digestion with Mmel. After purification of DNaseI-digested ends on streptavidin beads, a second set of
phosphorylated annealed linkers containing a two-base degenerate overhang was ligated to the dephosphorylated MmeI ends. Inserts, along with the linkers, were amplified by PCR, and the products were PAGE-purified to minimize adaptor contamination and cut and purified by phenol extraction and ethanol precipitation. The purified DNA was sequenced at FuGU with custom sequencing primers. The detailed protocol for sample preparation, primer sequences and sequencing data analysis is described in original communication I.

3.4 Formaldehyde-assisted isolation of regulatory elements (FAIRE)

FAIRE assay were performed as described (Giresi et al., 2007) with minor modifications. In short, 2 x 10^6 LNCaP-1F5 and VCaP cells were cultured in 10% DC-FBS medium for 3 days (60–70% confluence), after which they were exposed to 100 nM DHT for 2 h and cross-linked with 1% formaldehyde for 10 min. Cells were lysed in lysis buffer [1% SDS, 10 mM EDTA, 50 mM Tris-HCl (pH 8.1), and PIC (Roche Inc., Mannheim, Germany)]. Chromatin was sonicated and cleared by centrifugation. The soluble chromatin was subjected to three consecutive phenol-chloroform-isoamyl alcohol (25:24:1) extractions and the aqueous phase recovered. Samples were treated with RNase A for 1 h at 37°C, followed by overnight proteinase K digestion and reverse cross-linking at 65°C. DNA was purified using MinElute PCR purification kit (Qiagen Inc., California, USA) for qPCR. The detailed protocol and primer sequences are described in original communication II.

3.5 Computational data analysis for next-generation sequencing data

Peak calling for ChIP-seq datasets were performed using MACS (Model-based analysis for ChIP-seq; Zhang et al., 2008) and subsequently the binding sites were mapped to the nearest RefSeq gene by building a transcription factor association strength (TFAS). TFAS was built by computing the weighted sum of the corresponding ChIP-seq signal strength, where the weights reflect the proximity of the signal to the gene (Ouyang et al., 2009).

The sequencing reads for DHS-seq were filtered using the default Illumina chastity filter during the base-calling process and ELAND (Illumina) was used to align 20-bp reads to human genome (hg19), allowing up to two mismatches per read. The clustering of
Materials and methods

sequence tags and the DHS-seq peaks were analyzed using F-seq algorithm (Boyle et al., 2008). F-seq uses kernel density estimation that provides both a discrete and a continuous probability landscape across the whole genome. These kernel density estimation based probabilities are computed at each base and are directly proportional to the probability of seeing a sequence read at that location.

Pre-defined motifs of transcription factors were obtained from JASPAR (Portales-Casamar et al., 2010). Possible over-representation of motifs in ChIP-seq binding sites was analyzed by comparing the frequency to similar-sized genomic background sequences randomly collected from regions other than binding sites (Wang et al., 2007). De novo motif analysis was performed using MEME (Bailey and Elkan, 1994). The overlap analyses, CEAS analysis (Cis-regulatory Element Annotation System), genome-wide correlation, aggregation plots, motif analysis using seqpos algorithm and density maps were performed using Cistrome (Liu et al., 2011). The data visualization was done using Integrative Genomics Viewer (Robinson et al., 2011b). The detailed analysis and methods are described in original communication I, II and III.
RESULTS AND DISCUSSION

1. Androgen receptor (AR) cistromes in prostate cancer cells and role of FoxA1 in prostate cancer (Studies I and II)

AR cistromes and transcription factors that modulate AR signaling have been identified using ChIP-on-chip or ChIP-seq approaches in LNCaP cells or clonal derivatives using either DHT or synthetic androgen R1881 (Jia et al., 2008; Massie et al., 2007; Massie et al., 2011; Wang et al., 2007; Wang et al., 2009; Yu et al., 2010). However, information on AR cistromes in response to other compounds with androgen agonistic/antagonistic properties is lacking. Importantly, there is dearth of information on the mechanisms that define the differences in AR cistromes and androgen-regulated transcription programs between different cell lines. To elucidate the underlying differences, we examined AR cistromes and transcriptome in response to physiological androgen DHT, partial agonists/antagonists (CPA and RU486) and the antagonist bicalutamide.

1.1 Ligand as regulators of AR cistrome and transcriptome

To delineate the role of ligands in AR recruitment to the chromatin, we performed AR ChIP-seq in LNCaP-1F5 cells exposed to 100 nM DHT, 1 µM CPA, 1 µM RU486, 1 µM bicalutamide or vehicle for 2 h. All the ChIP-seq experiments were performed in biological duplicates and peak calling was done by MACS algorithm (Zhang et al., 2008). After DHT treatment, occupancy of AR on DNA was noted on 8,603 sites (FDR <2%) followed by CPA and RU486, whereas anti-androgen bicalutamide treatment brought about very limited AR binding to chromatin resembling to that achieved by vehicle only. Importantly, AR recruitment after exposure to different ligands showed only quantitative differences. Genome-wide distribution of AR binding events mostly occurred on distal enhancer elements (40-43%), followed by intronic regions (40-42%). Only 8–10% of binding was in the proximal promoter region (≤ 3 kb), suggesting a distal mode of transcriptional regulation.

The transcriptome profiles in LNCaP-1F5 cells treated with the three ligands for 24 h was consistent with the AR binding results. Exposure to DHT yielded the most robust transcriptional response followed by exposure to CPA and RU486. The partial
transcriptional responses brought out by CPA and RU486 is due to the fact that they possess partial agonistic/antagonistic properties for AR function. Importantly, unsupervised hierarchical clustering revealed both common and distinct signatures of genes regulated by the three ligands. Irrespective of the ligand, the genes that were up-regulated contained a higher frequency of AR-binding sites (ARB) compared to non-regulated genes. However, for down-regulated genes only DHT-downregulated genes could be mapped with statistical significance. The de novo and motif enrichment analysis for the ARBs revealed 15-bp canonical ARE-like sequences as the most over-represented (Fig. 8a). Of note, the cis-element for forkhead box proteins was significantly enriched adjacent to the ARBs (Fig. 8b).

1.2 AR and FoxA1 cistromes in prostate cancer cells

Our observation that the forkhead motif was over-represented in the ARBs (cf., Fig. 8b) together with the previously identified role of FoxA1 as a pioneer factor in a genome-wide fashion for ERα (Hurtado et al., 2011) and for a subset of AR target genes (Lupien et al., 2008), impelled us to investigate the genome-wide FoxA1 binding sites and their relation to ARBs in LNCaP-1F5 cells. Peak calling by MACS and subsequent mapping by TFAS (See Material and Methods) identified 6,215 ARBs and 18,281 FoxA1-binding sites in LNCaP-1F5 cells in response to DHT treatment. Overlap analysis of AR and FoxA1 cistromes revealed that ~71% of ARBs share at least a 1 bp overlap with FoxA1-binding sites (Fig. 8c), with the median distance of 43 nt between the center of respective binding sites. This high degree of overlap suggests a global role of FoxA1 in androgen signaling and provides additional credence to the role of FoxA1 in AR binding to chromatin (Wang et al., 2007; Lupien et al., 2008).

1.3 FoxA1 reprograms AR cistrome and transcriptome

To analyze the global role of FoxA1 in more detail, we depleted FoxA1 protein levels up to ~80% in LNCaP-1F5 cells using FoxA1-specific siRNA for a period of 72 h. The cells treated with non-targeting control or specific siRNAs were exposed to 100 nM DHT for 2 h followed by AR and FoxA1 ChIP-seq to define the AR and FoxA1 cistromes. FoxA1 depletion resulted in remarkable increase (17,022; over 2.5-fold) in the number of ARBs. Overlap analysis of AR cistromes in parental (siControl) and siFoxA1 cells revealed that 57% of the original ARBs remain unchanged in siFoxA1 cells, whereas 43% of the
Results and discussion

Parental ARBs were lost upon FoxA1 depletion. Importantly, 13,505 completely new ARBs were found in siFoxA1 cells, which is more than twice the number of parental ARBs (Fig. 8d). These new ARBs show very weak concordance (<10%) to published ARB sets in different PCa cell lines (Massie et al., 2011; Wang et al., 2009). This result shows that FoxA1 determines AR binding to the chromatin, and FoxA1-depletion leads to remarkable redistribution of ARBs. We validated this observation in another PCa cell line, VCaP cells, where FoxA1 depletion elicited a similar redistribution of ARBs.

Figure 8. AR and FoxA1 cistromes in LNCaP-1F5 cells and reprogramming of AR cistrome by FoxA1 depletion. (A) Top-scoring cis-element for ARBs in LNCaP-1F5 cells. (B) Consensus FoxA1 cis-element found over-represented within ARBs. (C) Overlap between AR- and FoxA1-binding sites in LNCaP-1F5 cells (FDR <2%). (D) Overlap of ARBs (FDR <2%) in parental and siFoxA1 LNCaP-1F5 cells. (E) Class I ARBs are independent of FoxA1, shown for TMPRSS2. (F) Class II ARBs are pioneered by FoxA1, shown for LPAR3. (G) Class III ARBs involving FoxA1 as a repressor; AR binding occurs only in siFoxA1 cells, shown from EVL. (H) Violin plots summarizing the changes in androgen-dependent transcriptome in the three classes of genes defined by FoxA1. (I) Top scoring cis-element for FoxA1 pioneered ARBs. (Adapted and reprinted from Sahu et al., 2011).
*Results and discussion*

FoxA1 defines three distinct categories of ARBs as evidenced by FoxA1 depletion results: (i) sites that are independent of FoxA1 (Fig. 8e), (ii) sites that require FoxA1 as a pioneer factor to mediate AR recruitment (Fig. 8f) and (iii) sites that are masked by FoxA1 and become accessible for AR binding upon FoxA1 depletion (Fig. 8g). This receptor binding profile dictated by presence or absence of FoxA1 was commensurate with the gene expression profiling, falling into three categories in parental cells and siFoxA1 cells after a 24 h exposure to 100 nM DHT. The differentially expressed genes unique to parental cells lost their androgen regulation in siFoxA1 cells concomitantly with the loss of ARBs, whereas the genes whose ARBs were FoxA1-independent maintained androgen responsiveness also in siFoxA1 cells. Interestingly, the new ARBs appearing upon FoxA1 depletion also defined a new signature of androgen-regulated genes that were not androgen-regulated in parental cells (Fig. 8h). Furthermore, the *de novo* motif analysis for the FoxA1-independent and the new ARBs appearing upon FoxA1 depletion revealed a canonical ARE (*cf.*, Fig. 8a), but the FoxA1-pioneered ARBs that were lost upon FoxA1 depletion showed a novel, distinct *cis*-element consisting of an ARE half-site and a partial FoxA1 motif separated by 4 nt spacer (Fig. 8i). These results suggest that the ARBs in PCa cells exhibit remarkable plasticity and their location is determined by the presence (or concentration) of another co-operating transcription factor, FoxA1 in this case. Importantly, FoxA1 depletion relieves chromatin from a repressive feature, leading to emergence of novel androgen-regulated programs in PCa cells. However, this adaptability in AR binding occurred without consistent changes in the cellular AR protein levels in LNCaP-1F5 or VCaP cells.

**1.4 Correlation of AR, FoxA1 with active chromatin marks**

Dimethylated histone 3 lysine 4 (H3K4me2) marks are enriched in enhancers and transcription start sites (TSS) of active genes (Barski et al., 2007; Heintzman et al., 2007), and this modification is instrumental in guiding FoxA1 and AR to cognate chromatin sites (Lupien et al., 2008; He et al., 2010). H3K4me2 ChIP-seq in LNCaP-1F5 cells was in agreement with a previous report, as ~70 % of the shared AR and FoxA1 sites were marked by H3K4me2 modification. However, FoxA1 depletion did not affect the H3K4me2 marks in a genome-wide fashion, suggesting that AR can recognize the H3K4me2 marks either on its own or possibly through another collaborating factor. Similarly, for FoxA1-pioneered loci, the H3K4me2 modification was intact even after the
Results and discussion

loss of FoxA1 and AR occupancy. Interestingly, in many instances FoxA1 depletion led to appearance of new H3K4me2-marked chromatin sites concomitant with the ARBs in siFoxA1 cells. Nevertheless, the incomplete concordance of H3K4me2 marks with AR and FoxA1 sites in both parental and siFoxA1 cells suggests the role of other epigenetic marks and/or other collaborating factors.

Active accessible chromatin regions can be demarcated from inactive heterochromatin by detecting their sensitivity to nucleases. These regions are highly cell-specific, and often dictated by array of chromatin remodelers and other factors (John et al., 2008; Boyle et al., 2008). DHS-seq in LNCaP-1F5 cells revealed a picture similar to that of H3K4me2 modification, in that most of the AR- and FoxA1-binding sites associate with open chromatin conformation, and majority of them (~70%) were constitutively open already prior to FoxA1 and AR binding. However, FoxA1 depletion also led to alteration in chromatin conformation forming de novo DHS sites in many instances, which were concomitant with the appearance of new ARBs linked to new androgen-regulated genes in FoxA1-depleted conditions. These new ARBs occupying de novo DHS sites comprise 16% of all ARBs mapped in siFoxA1 cells and might be formed by coordinated recruitment of chromatin modifying enzymes in an AR-dependent manner.

1.5 FoxA1 reprograms glucocorticoid signaling

To investigate whether FoxA1-mediated reprogramming of androgen signaling in PCa cells was unique to AR, we analyzed the effects of FoxA1 depletion on glucocorticoid receptor (GR) binding and transcriptional activation. GR ChIP-seq in parental and siFoxA1 cells after a 2 h Dex treatment identified 5,971 and 9,844 GRBs, respectively. Similar to AR cistromes, FoxA1 defines three distinct classes of GRBs, in that 60% of the GRBs were maintained and 40% lost upon FoxA1 depletion, and importantly, 6,287 completely new GRBs emerged in siFoxA1 cells. The de novo motif analysis for the three classes revealed identical cis-elements, unlike in the case of ARBs where FoxA1-pioneered ARBs were enriched for a composite element. This redistribution of GRBs was in parallel with changes in GR-dependent gene expression programs. Importantly, the new AR- and GR-dependent transcription programs upon FoxA1 depletion were distinct, as only a small proportion (~15%) of the androgen-regulated new transcripts in siFoxA1 cells are regulated by glucocorticoids in the same cells. Collectively, our results on
reprogramming of androgen and glucocorticoid signaling indicate that binding of their
cognate receptors to chromatin is highly plastic and significantly influenced by another
transcription factor, FoxA1. Consequently, the hormone, the receptor and the cognate
DNA sequence are necessary but in most cases not sufficient to guide the recruitment of
AR and GR to the appropriate genomic loci, in order to execute the intended hormonal
signaling.

1.6 Relationship between FoxA1 expression and PCa outcome

The ability of FoxA1 to reprogram androgen signaling and high expression levels of
FoxA1 mRNA in normal and malignant cases prompted us to investigate the FoxA1
protein expression levels in primary tumors of 350 PCa patients. Increased nuclear AR
protein expression in either diagnostic biopsy and/or radical prostatectomy samples is
associated with a reduced time to PCa-specific mortality (Donovan et al., 2008, 2010).
This earlier finding of increased AR protein expression was reproducible in the patient
cohort examined in this study.

![Figure 9](image-url). FoxA1 protein expression in prostate cancer tissue specimens and prostate cancer-
specific survival. (A) Disease-specific survival of 350 patients with prostate cancer according
to the staining intensity of FoxA1 in the primary tumor specimens. $\chi^2 = 4.36, P = 0.04$ (log-
rank test). (B) Disease-specific survival of patients with high AR staining classified
according to FoxA1 staining in the same specimen; either negative-moderate ($n = 18$) or high
($n = 222$). $\chi^2 = 2.57, P = 0.10$, log-rank test. (Reprinted from Sahu et al., 2011).

FoxA1 staining was predominantly nuclear and almost all primary PCa specimens
expressed FoxA1 protein. FoxA1 staining intensity showed positive correlation with that
of AR. On the basis of FoxA1 staining intensity, 24 samples (6.9%) were negative or
weakly staining, 71 samples (20.3%) moderate and 255 samples (72.9%) with strong
FoxA1 staining. Strong nuclear FoxA1 staining in PCa was associated with poor prognosis, that is, a reduced time to PCa-specific mortality (Fig. 9a). Importantly, low FoxA1 was associated with good prognosis, even in the presence of high AR protein that, in and of itself, spells poor disease outcome (Fig. 9b). This later finding agrees with our results from cell line experiments, giving credence to the fact that level of FoxA1 determines the reprogramming of AR signaling pathway.

Another group concomitantly reported FoxA1-mediated AR reprogramming in a related cell line (LNCaP) showing that FoxA1 can simultaneously repress or facilitate AR binding and were commensurate with gene expression changes and the production of eRNAs (Wang et al., 2011). FoxA1 depletion had little effect on DNaseI-hypersensitivity or H3K4me2 modifications; and the results were consistent with the finding that loss of FoxA1 had little effect on H3K4me1, H3K27Ac histone marks and H2A.Z ChIP-seq suggested that nucleosome remodeling was not a primary event for these new reprogrammed AR sites (Wang et al., 2011).

The results suggest that the AR and GR binding to chromatin are remarkably fluid and FoxA1 plays a dual role in regulating the accessibility to chromatin defining a tripartite distribution of AR and GR cistromes. FoxA1 functions either as a pioneer factor for a subset of binding sites or occluding AR/GR binding by making an inaccessible chromatin environment perhaps via recruitment of corepressor complexes (Sekiya and Zaret, 2007; Eeckhoute et al., 2009). Moreover, there is a subset of binding sites that are independent of FoxA1, which either are directly bound by receptor or pioneered by other transcription factors. This reprogramming of AR and GR cistrome upon FoxA1 depletion shows remarkable correlation with changes in androgen- and glucocorticoid-dependent transcription programs. FoxA1-mediated reprogramming in PCa cells is in contrast to the role of FoxA1 in ER binding and estrogen signaling in breast cancer cells. FoxA1 depletion in breast cancer cells led to a decrease of more than 90% in ER binding intensity across 50% of all the binding sites globally affecting estrogen-regulated program (Hurtado et al., 2011). Importantly, the favorable predictive outcome from low FoxA1 expression in PCa patients is of significance and is in sharp contrast to the role of FoxA1 in breast cancer, where high FoxA1 expression predicts favorable prognosis (Albergaria et al., 2009; Badve et al., 2007). The predictive power of FoxA1 in PCa was later reproduced by a separate study showing that patients with high FoxA1 levels have
Results and discussion

shorter time to biochemical recurrence (Gerhardt et al., 2012), and another study demonstrating high FoxA1 expression in 90% of prostate cancer metastases (Jain et al., 2011). These data suggest that high FoxA1 expression sustains and regulates signaling pathways important in PCa progression and in metastasis as well. This notion is consistent with changes in FoxA1-dependent transcriptome in PCa cell lines where FoxA1 depletion led to up-regulation of tumor suppressors such as FOXO1, FOXO3 and down-regulation of oncogenes such as PIM1 and c-Myc. Overall, these results are of special clinical significance in the management of PCa, given the central role of AR signaling axis in both PCa and castration-resistant prostate cancer (CRPC). Moreover, these results reveal an additional layer of complexity in the role of FoxA1 in recruitment of steroid receptors to chromatin resulting in differential gene expression program as schematically shown in the proposed model (Fig. 10).

Figure 10. Schematic model showing the dual role of FoxA1 in recruitment of AR to chromatin in prostate cancer cells either serving as a pioneer factor or as a repressor.
2. FoxA1 is a key determinant of unique receptor binding events in prostate cancer cells (Studies I and II)

Cistromes of steroid receptors such as GR, ER and AR are lineage- and cell-type specific (Krum et al., 2008; Lupien et al., 2008; So et al., 2007), but in some cases one steroid receptor can substitute for another as shown for AR in ER-negative MDA-MB-453 breast cancer cells (Ni et al., 2011; Robinson et al., 2011a). To investigate such potential overlap in PCa cells, we performed a comprehensive analysis of AR cistromes from two prostate cancer cell lines, originating from prostate epithelial cells, LNCaP-1F5 and VCaP cells. Furthermore, to understand how the receptor specificity is achieved, we compared AR and GR cistromes in these two androgen-responsive PCa cell lines.

2.1 AR cistromes in LNCaP-1F5 and VCaP prostate cancer cells

The AR gene is amplified in VCaP cells and expresses ~10 fold higher AR protein level than LNCaP cells (Makkonen et al., 2011). AR cistrome in DHT-treated VCaP cells comprised 44,879 ARBs (FDR <2%), which is approximately 5 times more than in LNCaP-1F5 cells, and most likely due to AR overexpression in VCaP cells. Meta-analysis of ARBs in LNCaP-1F5 cells revealed substantial overlap with previous ARBs reported in LNCaP-1F5 cells by ChIP-on-chip (Wang et al., 2009) and by ChIP-seq (Massie et al., 2011). Similarly, comparison of DHT-bound ARBs in VCaP cells exhibited high degree of overlap (~74%) with the R1881-occupied ARBs reported by Massie et al., (2011). Most of the ARBs (~85%) in LNCaP-1F5 cells were found in VCaP cells, but interestingly, there were also sites unique to LNCaP-1F5 cells (Fig. 11a). These unique sites were enriched for the AR-FoxA1 composite element and FoxA1 cis-element as identified by de novo motif analysis (Fig. 11b), but no canonical ARE was found. This suggests that these unique AR binding events are mediated by tethering to FoxA1, and required prior binding of FoxA1 to these sites. FoxA1 binding to the LNCaP-1F5 unique sites was almost non-existent in VCaP cells, despite the higher FoxA1 protein levels in the latter cells.

To explain the unique AR binding events in LNCaP-1F5 cells, we performed FAIRE analysis to scrutinize the arrangement of underlying regulatory elements in LNCaP-1F5 and VCaP cells. The analysis revealed that the unique ARBs in LNCaP-1F5 cells are located in an inaccessible chromatin environment in VCaP cells, with a different
Results and discussion

epigenetic make-up. The epigenetic signature associated with concomitant AR and FoxA1 binding involves eviction of central nucleosomes marked by decreased H3K4me2 levels and destabilization of histone H2A.Z variant (He et al., 2010). H3K4me2 and H2A.Z occupancy by ChIP-qPCR at the randomly selected loci, representative of unique ARBs in LNCaP-1F5 cells, showed that central nucleosome is stabilized in VCaP cells occluding the AR/FoxA1 binding site. On the contrary, the ARBs unique to VCaP cells were predominantly enriched for canonical ARE like cis-elements. Additionally, differential motif enrichment analysis revealed cis-elements for ETS family members to be highly over-represented in the ARBs unique to VCaP cells, but not in LNCaP-1F5 cell unique ARBs.

2.2 AR and GR cistromes in LNCaP-1F5 cells

GR plays an important role in inflammation and cancer progression, and GR signaling exhibits tumor suppressor activity in PCa cells (Yano et al., 2006; Yemelyanov et al., 2007). LNCaP-1F5 cells are genetically manipulated to express the rat GR and expression

Figure 11. AR and GR cistromes in LNCaP-1F5 and VCaP cells. (A) Overlap analysis of ARBs in LNCaP-1F5 and VCaP cells (FDR <2%). (B) Top scoring cis-element for LNCaP-1F5 unique ARBs. (C) Overlap between AR- and GR-binding sites in LNCaP-1F5 cells (FDR <2%). (D) Tag density maps for binding sites unique to AR or GR or shared by the two receptors. (E) ELL2 gene with shared binding sites and RNA Pol II occupancy is regulated both by androgen and glucocorticoid. (F) PER1 with GR binding events and only regulated by glucocorticoid. (G) C1orf116 with AR binding events and only regulated by androgen.
Results and discussion

level is ~4 times that of AR (Cleutjens et al., 1997). This provided us a suitable platform to study GR signaling vis-à-vis AR signaling in the context of chromatin in a single cell line. GR ChIP-seq in LNCaP-1F5 cells exposed to Dex for 2 h identified 14,103 GRBs (FDR <2%). Overall comparison of AR and GR cistromes revealed that almost half of the ARBs (46%) overlapped with the GRBs (Fig. 11c). In many instances, both AR and GR occupancy was seen to occur at the same loci and this occupancy was also reflected in the up-regulation of nearby genes such as TMPRSS2, ELL2 which are both up-regulated by DHT and Dex. Importantly, there are receptor unique binding events as seen in the tag density maps (Fig. 11d), despite the fact that both AR and GR recognize nearly identical consensus elements. This result suggests that cells have devised a mechanism to ensure receptor specificity that determines and defines the final outcome of a particular hormone response. The cistrome profiles of AR and GR with the shared and unique binding events were also reflected in gene expression profiling and their associated pathways seen by unsupervised hierarchical clustering. Of note, the LNCaP-1F5 cells were more responsive to Dex as oppose to DHT as seen by the number of differentially regulated genes.

The shared and unique ARBs or GRBs were mapped significantly more often within ±100 kb of TSS for the androgen- and glucocorticoid-regulated genes (both up- and down-regulated), respectively, compared to the stably expressed genes. In particular, the shared AR/GR binding sites were also significantly enriched for the genes regulated by both androgen and glucocorticoid over the stably expressed genes, implying that, in these instances, AR and GR occupied by their cognate ligands are capable of using the same regulatory cis-elements to regulate transcription programs. This was additionally supported by the RNA Pol II occupancy along the gene body in the presence of respective ligands, as in case of ELL2 (Fig. 11e). Interestingly, however, this was not a general rule, as in a number of instances where AR and GR were loaded into the same loci, but RNA Pol II occupancy, elongation and subsequent transcript accumulation were mainly regulated by one steroid only, as depicted by glucocorticoid for PER1 (Fig. 11f) and by androgen for C1orf116 (Fig. 11g).

DNA sequence in and of itself can act as an allosteric ligand (Meijsing et al., 2009) and the distinct profiles of AR and GR cistrome led us to examine the nature of ARBs and GRBs that were either shared or unique to either one of the two receptors. The 15-bp consensus canonical ARE/GRE was identified by de novo motif search for the shared
Results and discussion

ARBs/GRBs and for the unique GRBs. Interestingly, the unique ARBs in LNCaP-1F5 cells revealed the AR-FoxA1 composite *cis*-element (Fig.1b), a 20-bp *cis*-element identified previously for the FoxA1-pioneered ARBs and also defining the AR binding events unique and specific to LNCaP-1F5 cells as opposed to VCaP cells. This composite *cis*-element was not present in the shared binding sites or in those unique to GR. This finding adds another function to the diverse roles of FoxA1 in defining AR specificity in LNCaP-1F5 cells when compared to GR, both of which recognize the same consensus element in response to their cognate ligands.

2.3 Comparison of AR and GR cistromes in VCaP cells

The fact that LNCaP-1F5 cells do not express human GR, and the results of GR signaling described above, led us to examine these findings in another PCa cell line, the VCaP cells with endogenous GR expression. GR ChIP-seq in VCaP cells treated with Dex for 2 h identified 7,101 GRBs (FDR <2%). Similar to the distribution in LNCaP-1F5 cells, there were sites unique to AR or GR and sites shared by both the receptors in VCaP cells. Unexpectedly, the unique GRBs in VCaP cells lacked the consensus canonical GRE/ARE, rather, *de novo* motif analysis revealed over-representation of FoxA1-like *cis*-element. The shared binding sites for AR and GR possessed both FoxA1 and GRE/ARE motifs, and the unique ARBs in VCaP cells were over-represented for a canonical ARE.

The FoxA1-dependent recruitment of GR in VCaP cells was further substantiated by the comparison with GR cistrome in LNCaP-1F5 cells, as only a quarter of the GR binding events overlapped in the two cell lines. Furthermore, motif enrichment analysis comparison for GRBs in LNCaP-1F5 and VCaP cells revealed a pattern of enrichment that was a mirror image of the ARBs in these two cell lines. Previous reports have shown that AP1 co-occupancy occurs in 50% of the GRBs in 3134, a murine mammary epithelial cell line, suggesting AP1 as a pioneer factor for GR (Biddie et al., 2011; John et al., 2011) and in human HeLa cells (Rao et al., 2011). However, our analysis in LNCaP-1F5 and VCaP cells did not show enrichment of AP1 motifs either by *de novo* or by motif enrichment analysis, suggesting that these collaborating factors operate in a cell type and context-specific manner. Importantly, this indicates that FoxA1 and probably other forkhead family members are more important in GR signaling in PCa cells than AP1. This finding is consistent with previous studies in the pituitary cell line AtT20, where enrichment of FoxA1-like elements and not AP1 were prevalent among the GRBs (John
Results and discussion

et al., 2011). The role of FoxA1 as a determinant in ensuring receptor-specific chromatin binding of AR and GR in two PCa cell lines, i.e., LNCaP-1F5 and VCaP, respectively, is represented in the schematic model (Fig. 12).

**Figure 12.** Schematic model showing FoxA1 as a key determinant of unique AR and GR binding events in two cancer cell lines originating from prostate epithelial cells.

AR and GR binding events and transcription programs exhibited significant overlap and GR regulated a large number of genes typically considered specific to the AR pathway. AR and GR recognize highly similar consensus elements, and this raises the possibility of co-operation or competition between nuclear receptors for the same binding site on chromatin. Importantly, in the presence of androgen, expression of transcripts linked to binding sites shared by AR and GR is often – but not always – inhibited by the concomitant exposure to both the ligands, indicating that, under these conditions, Dex-occupied GR functions to attenuate AR signaling. Thus, GR regulation of the AR pathway was dependent on the cellular androgen level, in that it opposed AR signaling in the presence of androgen but activated a number of AR-regulated genes in the absence of androgen. These findings bring up the question as to the role of GR in maintaining the AR pathway under androgen-deprived conditions in castration-resistant patients.
Results and discussion

3. Increased AR expression enhances receptor binding events in prostate cancer cells (Study III)

The AR gene amplification and overexpression is a hallmark in the majority of CRPCs (Visakorpi et al., 1995), and cancer cells typically maintain an active AR signaling axis (Chen et al., 2008; Knudsen and Kelly, 2011). An LNCaP-based model system expressing AR protein in moderate (LNCaP-ARmo) or high levels (LNCaP-ARhi) compared to control cells (LNCaP-pcDNA3.1) showed that AR overexpression sensitizes the cells to low levels of androgens (Waltering et al., 2009). To examine the effect of increased AR expression and its consequence on recruitment of AR to chromatin, we employed ChIP in this AR overexpression model system.

3.1 ChIP-seq mapping of AR cistromes in AR overexpressing cells

AR ChIP-seq profiling was performed in LNCaP-pcDNA3.1, LNCaP-ARmo and LNCaP-ARhi cells treated with 1 nM, 100 nM DHT or vehicle for 2 hours. Comparison of the AR cistromes revealed that the number of ARBs was significantly higher in LNCaP-ARhi cells with relatively low concentration of androgens (1 nM DHT) as compared to parental LNCaP-pcDNA3.1 cells with higher concentration of androgens (100 nM DHT). This correlation between AR levels and the AR binding events was further validated using an independent AR overexpression model, PCa xenografts LuCaP69 and LuCaP73. LuCaP69 is known to harbor AR gene amplification with ~10-fold higher AR protein levels than in LuCaP73 (Linja et al., 2001). ChIP-seq in the xenografts identified 19,000 ARBs in LuCaP69 and 7,000 in LuCaP73, giving further credence to the correlation between AR levels and ARBs in vivo.

Direct ChIP-qPCR validation of ChIP-seq data on PSA regulatory regions (enhancer and promoter) with similar treatment in the LNCaP-overexpression model cells showed that increased AR expression sensitized the AR binding by up to 100-fold. Similarly, AR enrichment was significantly stronger on the PSA regulatory region in LNCaP69 cells when compared to LuCaP73 cells. This result is concordant with our findings on LNCaP-1F5 and VCaP cells, where AR-overexpressing VCaP cells were found to have a five times higher number of ARBs than in LNCaP-1F5 cells. AR binding profiles from different cell lines upon stimulation with different DHT concentrations were analyzed, and the most frequently overlapping ARBs were used to construct a high-confidence
Results and discussion

The ARB map comprising 1,833 ARBs. The genome-wide distribution was similar to that of AR cistromes in studies I and II, and that reported for other steroid receptor cistromes such as ER, GR, employing a distal model of transcriptional regulation (Cheung and Kraus, 2010). Additionally, DNA sequence analysis showed presence of cis-elements for FoxA1 and ETS family members among the high confidence ARBs, suggesting their collaborating or auxiliary function in mediating transcriptional regulation. The role of FoxA1 in androgen signaling and in defining unique receptor binding events has been a key finding of our previous studies, and it is also in agreement with previous published reports (Massie et al., 2007; Massie et al., 2011; Lupien et al., 2008; Wang et al., 2009; Yu et al., 2010).

3.2 Identification of androgen-regulated target genes in CRPC

The LNCaP AR-overexpression model system has been profiled for differentially expressed genes by microarray (Waltering et al., 2009). Gene expression data was curated and combined with the high-confidence ARB map to identify novel AR target genes that might be important in the progression of PCa to CRPC. The microarray data included expression profiles of LNCaP-pcDNA3.1, LNCaP-ARmo and LNCaP-ARhi cells stimulated for 4 h and 24 h with increasing DHT concentrations (0, 1 and 100 nM). The enrichment p-values were computed by employing hypergeometric distribution for differentially expressed genes and AR-bound genes. LNCaP-ARhi cells are known to have a growth advantage at low androgen concentration (1 nM DHT) compared to control cells (Waltering et al., 2009), thus the focus was on genes showing differential expression and AR binding at that concentration. Using a fold-change cutoff of 1.5, we identified 346 genes that had ARBs in a window of 250 kb and were androgen up-regulated only in LNCaP-ARhi and LNCaP-ARmo cells but not in control cells, at 4 and/or 24 h time points. This expression signature list was curated against clinical PCa specimens from 14 independent studies with array-based expression profiling. This analysis identified 38 genes out of 346, which were overexpressed in PCa compared to benign prostate hyperplasia (BPH) or normal adjacent material in at least one of the studies.

Among the 38 identified genes, 5 putative target genes (FEN1, ZWINT SKP2, SNAI2, and AZGP1) were selected for further investigation on the basis of previous literature, and their androgen regulation was validated in our model cell lines. Quantification of gene
Results and discussion

expression by qRT-PCR in BPH, untreated PCa, and CRPC specimens revealed significant overexpression of SKP2, ZWINT and FEN1 transcripts in CRPC as compared to untreated PC and/or BPH. However, SNAI2 and AZGP2 levels were lower in CRPC than in untreated PCa. We validated the closest ARBs mapped to SKP2, ZWINT and FEN1 by direct ChIP assays and found higher enrichment of AR at these sites in LNCaP-ARhi cells as opposed to control cells.

To further elucidate the role of FEN1, ZWINT and SNAI2 in PCa, we depleted the pcDNA3.1 and LNCaP-ARhi cells of these targets by using respective mRNA-specific siRNAs. FEN1 depletion led to decreased cell proliferation in both cell lines, whereas ZWINT and SNAI2 depletion conferred growth advantage specifically onto control cells but not onto LNCaP-ARhi cells. Interestingly, increased expression of FEN1 protein was associated with LNCaP-ARhi and LNCaP-ARmo cells compared to the control cells, which led us to examine FEN1 protein expression in clinical samples. FEN1 antigen staining was performed on 185 untreated prostatectomy specimens and on 92 CRPC samples. Although the cytoplasmic staining was equal in PCa and CRPC, the CRPC samples showed a significantly stronger nuclear staining (p < 0.0001) than untreated PCa samples. Only 3% of prostatectomy samples showed nuclear staining, and this was associated with a shorter time for biochemical recurrence. These results suggest the importance of FEN1 as an AR downstream target gene that is overexpressed in CRPC, and its association with aggressive disease phenotype remains to be established.

Overall, these results on the importance of the AR level in androgen sensitivity are in agreement with the fact that the nuclear receptor-hormone interaction follows the law of mass action (Jänne and Bardin, 1984), and that the final outcome of receptor binding is determined both by the ligand concentration and the amount of the receptor. Thus, overexpression of AR in CRPC cells allows these cells to activate AR signaling at a low androgen concentration, a potential mechanism for retaining an active AR signaling axis in CRPC.
SUMMARY AND CONCLUSIONS

Genome-wide approaches have allowed mapping and characterization of nuclear receptor (NR) cistromes and given insights into the sets of NR-regulated genes and their distant regulatory cis-elements. High-throughput approaches have generated massive datasets, which requires advanced computational processing and insightful downstream analysis to interpret patterns and mechanisms that are biologically relevant. The caveat still lies in determining which binding sites are mandatory for the regulation of a particular gene. New technologies, such as ChIA-PET and Hi-C approaches are constantly evolving and developing to address these interactome networks (Kim and Yu, 2012). Despite these limitations, genome-wide methods have significantly shaped the understanding of complex chromatin-receptor interactions, and one such mechanism observed very early was the distal mode of regulation employed by NRs highlighting the importance of auxiliary transcription factor networks (Cheung and Kraus, 2010).

The classical view of gene regulation by NRs involves ligand-specific activation of the receptor, nuclear translocation and binding to the regulatory regions of target genes in collaboration with other transcription factors, which ultimately determine the specific transcriptional outcome. The results presented in this thesis work examined several aspects of this process by using AR as a model transcription factor. Events other than the nature of the ligand (agonist or partial agonist/antagonist), such as stability of ligand-AR complex and nuclear translocation of the receptor might be more important in determining the abundance of AR loading onto regulatory loci on chromatin. A remarkable feature in AR-chromatin interaction is that AR binding to chromatin is highly dynamic and variable, and dependent on the presence or level of FoxA1. FoxA1 depletion led to redistribution of AR and GR binding events in a way that is more complex than that observed for ER-FoxA1 interaction in breast cancer cells (Hurtado et al., 2011). FoxA1 defined three distinct classes of ARBs, and the pioneered sites contained a composite AR-FoxA1 cis-element, a compilation of AR half-site and the FoxA1 cis-element. FoxA1-mediated reprogramming was observed for two NRs, AR and GR, in two PCa cell lines and was commensurate with androgen- and glucocorticoid-mediated transcription programs. The majority of the receptor-binding sites were located in an open chromatin environment, as judged by genome-wide analysis of H3K4me2 modifications.
Summary and conclusions

and DNaseI-hypersensitive sites in parental and FoxA1-depletion conditions. Thus, the hormone, the receptor and the cognate DNA sequence are necessary but in most cases not sufficient to guide binding of AR and GR to the appropriate genomic loci, in order to execute the intended hormonal signaling.

FoxA1 was shown to be a critical collaborating factor for unique AR and GR binding events in LNCaP-1F5 and VCaP cells. A composite AR-FoxA1 cis-element previously identified for FoxA1-pioneered ARBs was required for the unique AR binding events in LNCaP-1F5 cells as opposed to VCaP cells. The same composite element specified unique ARBs relative to GRBs in LNCaP-1F5 cells. The ARBs and GRBs in LNCaP-1F5 and VCaP cells showed a differential enrichment for cis-elements for forkhead and ETS family members, with the patterns being like mirror image of each other. Despite their discrete receptor binding events in PCa cells, the cistromes and transcription programs regulated by androgens and glucocorticoids exhibited significant overlap, indicating that glucocorticoid-occupied GR can activate a subset of androgen-regulated genes, raising the question as to whether GR can maintain the AR pathway under castration-resistant and androgen-deprived conditions. In the presence of androgens, however, ligand-occupied GR attenuated AR-dependent transcriptional program.

At present, biochemical analysis of the PSA levels and histopathological grading by Gleason score are the mostly used decisive methods available for diagnosis of PCa and for monitoring its progression. AR signaling is crucial for PCa cells. This work demonstrated that FoxA1 expression level in the primary tumor is associated with disease progression. Patients with high FoxA1 protein-expressing tumors had significantly higher PCa-specific mortality than those with moderate or low FoxA1 levels. In PCa cell lines, high FoxA1 protein expression, via active AR signaling, sustained a transcription program with oncogenic potential by activation of known oncogenes. Identification of recurrent somatic mutations in FoxA1 gene in prostate adenocarcinomas and CRPCs further speaks for a critical role of FoxA1 in PCa (Barbieri et al., 2012; Grasso et al., 2012). AR gene amplification and over-expression of the AR protein are distinct features of CRPCs, and high receptor level was shown in this work to render the cells more sensitive to low levels of the ligand.
Summary and conclusions

Overall, the results in this work provided novel insights into the genome-wide AR function. The main conclusions highlighting these features are:

1. Selection of AR binding sites on chromatin is not markedly influenced by the nature of the ligand occupying the receptor.

2. AR binding to chromatin is remarkably fluid, and FoxA1 plays a dual role in regulating the accessibility of AR to chromatin and in reprogramming the AR pathway.

3. High FoxA1 protein level in primary PCa is of prognostic significance and predicts a non-favorable disease outcome.

4. FoxA1 acts as a key determinant for the unique AR and GR binding events in PCa cells.

5. FoxA1 depletion reprograms GR binding to chromatin and glucocorticoid-dependent transcription. AR and GR may utilize the same chromatin sites in regulating overlapping androgen- and glucocorticoid-dependent programs.

6. Glucocorticoids can act as partial anti-androgens and modulate AR transcriptional program in PCa cells.

7. AR overexpression sensitizes the receptor to low levels of androgens, resulting in enhanced AR recruitment to chromatin and activated androgen signaling at low hormone concentrations.
ACKNOWLEDGEMENTS

This work was carried out in the Androgen Receptor Laboratory in the Institute of Biomedicine, Physiology, at the University of Helsinki. I would like to extend my sincere thanks to my supervisor, Prof. Olli A. Jänne M.D. Ph.D., for the guidance and providing the excellent research facilities for the study. The staff members of the institute and the FuGU are warmly acknowledged for their support and help, FIMM and CSC for the infrastructure for high-throughput computational analysis.

This study was supported by grants from Academy of Finland, Sigrid Jusélius Foundation, Finnish Cancer Foundations, Biocentrum Helsinki, Helsinki University Central Hospital, CRESCENDO project supported by the European Union (contract No. LSHM-CT-2005-018652), ERANET SysBio+ (Synergy), Finnish Cultural Foundation and the Graduate Program in Biotechnology and Molecular biology.

The pre-examiners of my thesis, Prof. Tomi Mäkelä, M.D., Ph.D. and Prof. Marja Nevalainen, M.D., Ph.D., are sincerely appreciated for their constructive comments and for the meticulous review of the manuscript. The thesis committee members, Prof. Pirkko Vihko, M.D., Ph.D. and Docent Outi Monni, Ph.D., are warmly thanked for their inputs and suggestions during the course of studies.

I would like to express my grateful thanks to the collaborators, Docent Sampsa Hautaniemi, D.Tech., Prof. Tapio Visakorpi, M.D., Ph.D., Prof. Olli Kallioniemi, M.D., Ph.D. and Prof. Geoffrey Hammond, Ph.D., for their productive collaborations. It has always been a great pleasure to visit Tapio’s lab in Tampere, and it has been a joy to interact with him, Kati, Alfonso and other lab members. I thank all my co-authors, Marko Laakso, M.Sc., Kristian Ovaska, M.Sc., Päivi Pihlajamaa, M.Sc., Tuomas Mirtti, M.D., Ph.D., Johan Lundin M.D., Ph.D., Antti Rannikko, M.D., Ph.D., Anna Sankila, M.D., Ph.D., Juha-Pekka Turunen, M.D., Mikael Lundin, M.D., Juho Konsti, Tiina Vesterinen, M.Sc., Stig Nordling, M.D., Ph.D., Alfonso Urbanucci, Ph.D., Kati Waltering, Ph.D., Prof. Teuvo Tammela, M.D., Ph.D., Prof. Robert Vessella, M.D., Harri Lähdesmäki, D.Tech., Lenna Latonen, Ph.D., Janne Seppälä, M.Sc., Antti Larjo., M.Sc. and Ievgenii Sinielnikov, M.Sc., for their invaluable contribution to the work. Prof. Jorma Palvimo, Ph.D. and Marjo Malinen, Ph.D., are thanked for their collaboration and discussions.

The journey during my Ph.D. studies has led to an association with so many special people that a few lines cannot suffice their importance. First and foremost, I would like to acknowledge my gratitude to Olli for taking me on-board, introducing me to the challenging world of functional genomics, and guiding me on the long road to doctoral dissertation which was educational and rewarding. I am grateful for reposing his faith in me, giving me the responsibilities, and providing me with the scientific opportunities
required during the completion of the thesis. It has been a pleasure to work with you and to discuss the topics beyond the realm of science.

I thank Saija Kotola for her excellent technical help. A special thanks to Päivi for the collaboration, skilled help in making the manuscript figures, discussion on ChIP-seq and your friendship. I warmly appreciate Marko’s patience and thank his indispensable inputs to the projects. I have had a good time working in the lab with Laura, Johanna H, Tom, Ievhen, Annariikka, other past members of the ARlab, and Piia’s, Kalle’s and Taneli’s group. I wish all the best to my former students Matteo, Michaela and Heidi for their success in science and life. It has been a joy to teach and learn with you. I thank Mikko and GongHo for the discussions and sharing the resources at times. A special thanks to James for all the pints and good laughs we shared. I would also like to warmly applaud the friends outside the lab from Biomedicum and graduate schools for their good company and great parties. Anne and Johanna N’s help in administrative matters is well appreciated.

My humble thanks to all my teachers who have been the chaperones of my life and every step taken forward is the reflection of their efforts and dedication. I thank the few special ones, Gaga, Partha and Sinu bhai, whose friendship I will treasure over the years.

Big heartfelt thanks to all the friends from India in Finland, Dhakas for being an extended family. I thank Gopal for his friendship and all his help with my Linux problems. I warmly express my thanks to many other friends, Sharma, Tripathi, Ankur, Nikhil, Rohit, Ram, Laxman, Sarish, Vimal, Sanjay, Pali, Sidhesh, and Mittals, all of whom have immensely contributed to my social life and sometimes made the gloomy winter bright and memorable in Finland.

Finally I reserve my greatest accolades and sincere gratitude to my family. Baba-Bou for your love, unconditional support, encouragement, and sharing my happiness in all the endeavors. You are the true heroes of my life and I will be eternally indebted. Li’l sis Julie, Abhi bhai and other relatives are appreciated for their love and affection. My love and deepest gratitude to Mamuni, for your love, unwavering support and motivation during the course of the studies and in life. Our li’l angel Tanvi (Feby) for bestowing upon us all the happiness, and enriching our lives with her adorable charm and energy.

It’s been a long journey and I couldn’t have done it without the support from all of you. Thank you!

Helsinki, 2012
References


References


Fan, J.Y., Rangasamy, D., Luger, K., and Tremethick, D.J. (2004). H2A.Z alters the nucleosome surface to

References
promote HP1alpha-mediated chromatin fiber folding. Mol Cell 16, 655-661.


regions of the rat androgen receptor modulates
Interaction between the amino
References


Molteni, A., Rouleau, N., Ikonen, T., Palvimmo, J.J., and Jänne, O.A. (1997). The presence of a transcription activation function in the hormone-binding domain of
androgen receptor is revealed by studies in yeast cells. FEBS Lett 472, 355-358.


pioneer transcription factor primary determinant of the nuclear mobility of the 16284
stability of higher order chromatin structures. USA natural agonist dihydrotestosterone. Proc Natl Acad Sci receptor and its T877A mutant complexed with the 98 16289.
References


