Distribution Agreement

In presenting the thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University. I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature: _____ _____

Nitya V. Sharma Date

Identification of Regulators Associated with Androgen Deprivation Therapy Response and Metastatic Progression in Prostate Cancer

By

Nitya V. Sharma
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Genetics and Molecular Biology

Carlos S. Moreno, Ph.D.
Advisor

Paula Vertino, Ph.D.
Committee Member

Paul Doetsch, Ph.D.
Committee Member

Ken Moberg, Ph.D.
Committee Member

Adam Marcus, Ph.D.
Committee Member

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Identification of Regulators Associated with Androgen Deprivation Therapy Response and Metastatic Progression in Prostate Cancer

By

Nitya V. Sharma
B.S., Emory University, 2006
M.S., Georgia Institute of Technology, 2010

Advisor: Carlos S. Moreno, PhD

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Graduate Division of Biological and Biomedical Sciences
Genetics and Molecular Biology
2018

Identification of Regulators Associated with Androgen Deprivation Therapy Response and Metastatic Progression in Prostate Cancer

By

Nitya V Sharma

While some patients will benefit from androgen deprivation therapy (ADT), there is another subset of patients for whom therapy is not sufficient. These patients will acquire castration resistant prostate cancer that is typically associated with metastasis. Changes in tumor biology reflect the increasingly aggressive stages of prostate cancer, and drive differential clinical responses. Complex variations in tumor transcriptional profiles depend on a precise interplay of multiple regulators and cofactors. However, the transcription factor dynamics that promote aggressive prostate cancer and metastasis are not fully understood. Here, we identified transcription factor coordinated groups that may reflect activation of compensatory regulatory mechanisms that mediate ADT response and metastatic progression. Furthermore, our data suggest that these transcription factor interactions are maintained after ADT and in metastatic tumors, despite predicted genomic re-localization. Transcription factors exhibit context dependent re-localizations to drive multiple phases of cancer. SOX4 is a transcription factor that can transform prostate cancer. We investigated a continued role of the oncogene SOX4 after transformation in the epithelial to mesenchymal transition (EMT), a critical step in the metastatic cascade. Our studies suggest that SOX4 functions downstream of the important EMT-promoting signaling pathways, TGF-β and EGF, and is sufficient to promote the EMT program. We also shed light on the potentially interdependent relationships between SOX4 and the histone acetyltransferase complex PCAF to transcriptionally activate pro-EMT genes. The analyses described in this dissertation demonstrate how regulatory activities that drive aggressive prostate cancer depend on precise combinatorial transcription factor relationships.

Identification of Regulators Associated with Androgen Deprivation Therapy Response and Metastatic Progression in Prostate Cancer

By

Nitya V. Sharma
B.S., Emory University, 2006
M.S., Georgia Institute of Technology, 2010

Advisor: Carlos S. Moreno, PhD

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Graduate Division of Biological and Biomedical Sciences
Genetics and Molecular Biology
2018

Table of Contents

Chapter 1. Introduction **pg. 1**

1.1 Prostate Cancer Introduction pg. 1

1.1.1 Overview pg. 1

1.1.2 Evaluation and Classification of Prostate Cancer pg. 2

1.2 Androgen receptor: A brief introduction pg. 4

1.2.1 Androgens and the androgen receptor pg. 4

1.2.2 AR signaling pg. 8

1.3 AR-signaling and prostate cancer pg. 10

1.3.1 Androgen deprivation therapy as a prostate cancer therapeutic pg. 10

1.3.2. Aberrant AR-signaling in castration resistant prostate cancer pg. 13

1.4 The transition to CRPC and metastasis pg. 15

1.5 SOX4 is a developmental transcription factor that is important for prostate cancer metastasis pg. 16

1.5.1 SOX4 background pg. 16

1.5.2 SOX4 in cancer pg. 18

1.5.3 SOX4 in the epithelial to mesenchymal transition, and metastasis pg. 20

1.6 Concerted action of multiple transcription factors drives complex changes in transcriptional profiles pg. 23

1.7 Dissertation objectives pg. 25

Chapter 2. The biology of castrate resistant prostate cancer pg. 26

2.1 Introduction pg. 28

2.2 The androgen receptor pg. 29

2.3 PI3K-AKT-mTOR pathway pg. 31

2.4 Receptor tyrosine kinase (RTK) growth factor pathways pg. 32

2.5 Analysis of CRPC genomes pg. 33

2.6 Epigenetic pathways pg. 34

2.7 EMT and SOX family genes pg. 36

2.8 WNT signaling pg. 38

2.9 Notch and Hedgehog pg. 39

2.10 lncRNA-mediated pathways pg. 40

2.11 Angiogenic pathways pg. 41

2.12 Biomarkers pg. 42

2.13 Conclusion pg. 44

Chapter 3. Identification of transcription factor relationships associated with androgen deprivation therapy response and metastatic progression in prostate cancer **pg. 47**

3.1 Introduction pg. 49

3.2 Materials and methods pg. 51

3.3 Results pg. 55

3.4 Discussion pg. 99

Chapter 4. The Role of SOX4 in Epithelial to Mesenchymal Transition in Prostate Cancer **pg. 104**

4.1 Introduction pg. 105

4.2 Materials and methods pg. 108

4.3 Results pg. 111

4.4 Discussion pg. 117

Chapter 5. TFBSET: A New Tool for the Identification of Enriched Transcription Factor Binding Sites in the Promoter Regions of Human Genes **pg. 121**

5.1 Introduction pg. 123

5.2 Materials and methods pg. 126

5.3 Usage and implementation pg. 130

5.4 Results pg. 131

5.5 Discussion pg. 142

Chapter 6. Discussion and future directions

pg. 146

6.1 Predicting transcriptional regulators of ADT response and metastasis

pg. 146

6.1.1 Early divergent transcriptional response to ADT could reflect an ADT mediated clonal selection

pg. 146

6.1.2 Distinct transcription factor associations may drive transcriptional programs that influence prostate cancer aggressiveness.

pg. 148

6.1.3 Future Directions

pg. 150

6.2 SOX4 promotes EMT in prostate cancer cells

pg. 153

Figures and Tables List

Figure 1.1. Prostate cancer incidence and mortality statistics**.**
pg. 3

Figure 1.2. Synthesis of DHT and subsequent binding to the AR.
pg. 6

Figure 1.3. AR locus, gene and protein structure.
pg. 7

Figure 1.4. AR signaling axis.
pg. 9

Table 1.1. Pharmacological agents that inhibit AR signaling in prostate cancer.
pg. 11

Figure 1.5. Timing of first and second line drugs in CRPC.
pg. 12

Figure 1.6. SOX4 domain structure.
pg. 17

Figure 1.7. Increased SOX4 expression correlates with aggressive prostate cancer.
pg. 19

Figure 1.8. Steps of the metastatic cascade.
pg. 21

Figure 1.9. EMT induces molecular changes that allow for epithelial cells to acquire mesenchymal capabilities.
pg. 22

Figure 2.1. Signaling pathways in castration resistant prostate cancer.
pg. 45

Figure 3.1. Hierarchical clustering and PCA of 190 significantly differentially expressed genes in 20 matched Pre-ADT Bxs and Post-ADT RPs.
pg. 66

Figure 3.2. Divergent expression of PCS1 genes in the high impact group and common loss of PCS2 and PCS3 after ADT.
pg. 69

Figure 3.3. Identification of TFCGs in the high impact group network.
pg. 86

Table 3.1. 33 oTFCGs between the high impact ADT group and Met.PCS1 networks.
pg. 97

Table S3.1. Patient and treatment characteristics.
pg. 56

Table S3.2. Significantly differentially expressed genes identified using edgeR analysis of Pre-ADT Bxs and Post-ADT RPs.
pg. 58

Table S3.3. Ingenuity Pathway Analysis identifies chemical agents associated with signaling pathways. Predicted upstream chemical agent regulators suggest inhibition of androgen driven genes, and increase in estrogen and PDGF-MAPK signaling.
pg. 63

Table S3.4. Key TFs in the high impact network.
pg. 71

Table S3.5. TFCGs in the high impact network.
pg. 87

Table S3.6. Key TFs in the Met.PCS1 network.
pg. 90

Table S3.7. TFCGs in the Met.PCS1 network.
pg. 96

Figure S3.1. Hierarchical clustering of PCS genes of 20 matched pre-ADT Bxs and post-ADT RPs again segregates three groups.
pg. 68

Figure 4.1. TGF-β and EGF initiate the EMT program and induce SOX4 expression in RWPE1 non-neoplastic cells.
pg. 112

Figure 4.2. Over expression of SOX4 induces EMT in prostate cancer cell lines.
pg. 113

Table 4.1. Identification of candidate SOX4 targets.
pg. 115

Figure 4.3. SOX4 interacts with PCAF and TRRAP in prostate cancer cell lines.
pg. 116

Figure 4.4. SOX4 expression returns after knockdown with CRISPR/Cas9.
pg. 118

Figure 4.5. Proposed mechanism.
pg. 120

Figure 5.1. TFBSET tool characteristics.
pg. 127

Table 5.1. TFBSET was run on multiple gene sets to identify enriched TFBSs in proximal promotor regions

pg. 135

Table 5.2. Comparison of TFBSET to other popular programs

pg. 139

Table S5.1 Optimal parameters for identification of Myc target sites in the Dang gene set

pg. 132

Table S5.2 Comparison of results from 6 different gene sets upregulated by Myc. The numbers in the table represent the FDR.

pg. 133

Table S5.3 TFBSET produces better results at ±5000 bp than ±1000 bp around the TSS.

pg. 134

Table S5.4 Comparison of TFBSET results of gene sets run excluding and including enhancer regions shows that the results are generally less accurate when the enhancer regions are included

pg. 137

Chapter 1. Introduction

1.1 Prostate cancer introduction

1.1.1 Overview

When J. Adams first discovered prostate cancer in 1853, it was considered “a very rare disease”. On the contrary, improvements in detection show that prostate cancer is the most commonly diagnosed cancer among U.S. males, and the third-leading cause of cancer mortality in American men (Figure 1.1). These alarming statistics may reflect the longer life expectancy, as the rate of increase in prostate cancer diagnosis with age is faster than in any other cancer type[1]. Additionally, prostate cancer has a high heritability, much like other cancers, such as lung cancer, breast cancer, and colon cancer[2]. Although advances in diagnosis and treatments have yielded improved outcomes for prostate cancer patients, there remains a subset of these men for whom current therapies are not sufficient.

1.1.2 Evaluation and classification of prostate cancer

The standard of care for diagnosis of prostate cancer is based on the histological evaluation of prostate tissue needle biopsy. This process is performed by obtaining ten to twelve tissue samples in a grid-like pattern via transrectal ultrasound. There are three variables that are used to determine risk stratification. The Gleason grading system will yield a score based on the histologic appearance of cells that correlate to cancer aggressiveness[3]. To account for the high tumor heterogeneity, the Gleason score is derived from the sum of grades from the primary and secondary dominant lesions. Using a scale of one to five based on the level of histological de-differentiation, the primary score represents the most predominant pattern among tissue samples, and the secondary score denotes the second most predominant pattern. The concentration of prostate specific antigen (PSA) is another variable for the diagnosis and risk stratification of prostate cancer. PSA is produced by the epithelial cells in the prostate gland, and is elevated in men with prostate cancer. Finally, the third variable used to determine risk is the clinical staging based on the TNM staging system, that evaluates the degree of tumor (T) invasion, the extent of spread into the lymph nodes (N), and whether there is evidence of metastasis (M). Specifically, the T score evaluates the primary tumor appearance, volume, and extent of invasion into surrounding tissues and structures. The N category assesses to what extent there is cancer presence in nearby lymph nodes, indicating the likelihood of disease spreading. An NX score specifies that the nearby lymph nodes cannot be evaluated. An N0 score denotes that is no identifiable cancer in the lymph nodes. N1, 2, and N3 scores correspond to the number of nodes, or the nodal groups, contain cancer. Higher scores suggest that the extent of cancer is greater. The M score informs whether the cancer has metastasized to distant sites, such as distant lymph nodes or beyond. A score of M1 suggests the cancer has spread to distant sites. These variables all inform a patient's prognosis.

Estimated New Cases

			Males
Prostate	161,360	19%	
Lung & bronchus	116,990	14%	
Colon & rectum	71,420	9%	
Urinary bladder	60,490	7%	
Melanoma of the skin	52,170	6%	
Kidney & renal pelvis	40,610	5%	
Non-Hodgkin lymphoma	40,080	5%	
Leukemia	36,290	4%	
Oral cavity & pharynx	35,720	4%	
Liver & intrahepatic bile duct	29,200	3%	
All Sites	**836,150**	**100%**	

Estimated Deaths

			Males
Lung & bronchus	84,590	27%	
Colon & rectum	27,150	9%	
Prostate	26,730	8%	
Pancreas	22,300	7%	
Liver & intrahepatic bile duct	19,610	6%	
Leukemia	14,300	4%	
Esophagus	12,720	4%	
Urinary bladder	12,240	4%	
Non-Hodgkin lymphoma	11,450	4%	
Brain & other nervous system	9,620	3%	
All Sites	**318,420**	**100%**	

Figure 1.1. Prostate cancer incidence and mortality statistics. Prostate cancer leads in the number of new diagnoses, and ranks third among cancer related mortality, for 2017.

Adapted from [4]

1.2 Androgen receptor: A brief introduction

1.2.1 Androgens and the androgen receptor

Androgens are a class of steroid hormones that are required for prostate development and normal prostate function in males[5]. These hormones promote male sexual differentiation and their continued presence is required for secondary male characteristics and spermatogenesis. The synthesis of androgens originates in the hypothalamus. From here, gonadotropin-releasing hormone (GnRH) moves to the pituitary and interacts with GnRH receptors, to stimulate the pulsatile release of luteinizing hormone (LH). This hormone binds to LH receptors in the testes, resulting in production of testosterone, a type of androgen. Testosterone then travels to prostate cells, and is converted to dihydrotestosterone (DHT) by the enzyme 5 α-reductase. DHT is the predominant androgen that binds tightly to the androgen receptor (AR), allowing for conformational changes and translocation to the nucleus, where it behaves as a transcription factor (Figure 1.2).

Before the 1960s, the mechanisms of androgen action were largely unknown. The first indication that this process initiated with androgen binding to a specific protein came from K.M. Anderson and Shutsung Liao in 1968, when they observed a strong selective retention of androgens in prostatic nuclei, but not in other cell types from tissues insensitive to androgen[6]. The *AR* gene is located on the X-chromosome (Xq12), belongs to the nuclear receptor gene superfamily, and is comprised of four domains (Figure 1.3). The conserved DNA binding domain contains two zinc finger motifs that are critical for binding at androgen responsive elements (ARE). The N- and C-terminal domains are points of interactions with transcriptional coregulators, and contain transactivation domains. The activation function-1 (AF-1) domain is in the N-terminal domain and has activator functions that are ligand-independent. The activation function-2 (AF-2) is in the C-terminal LBD, and activity is ligand-dependent. The degree of relative transactivation of AF-1 and AF-2 may be driven by expression and activity of

coregulators[7]. The AR nuclear localization signal sequence is in the hinge region, between the DBD and LBD[8].

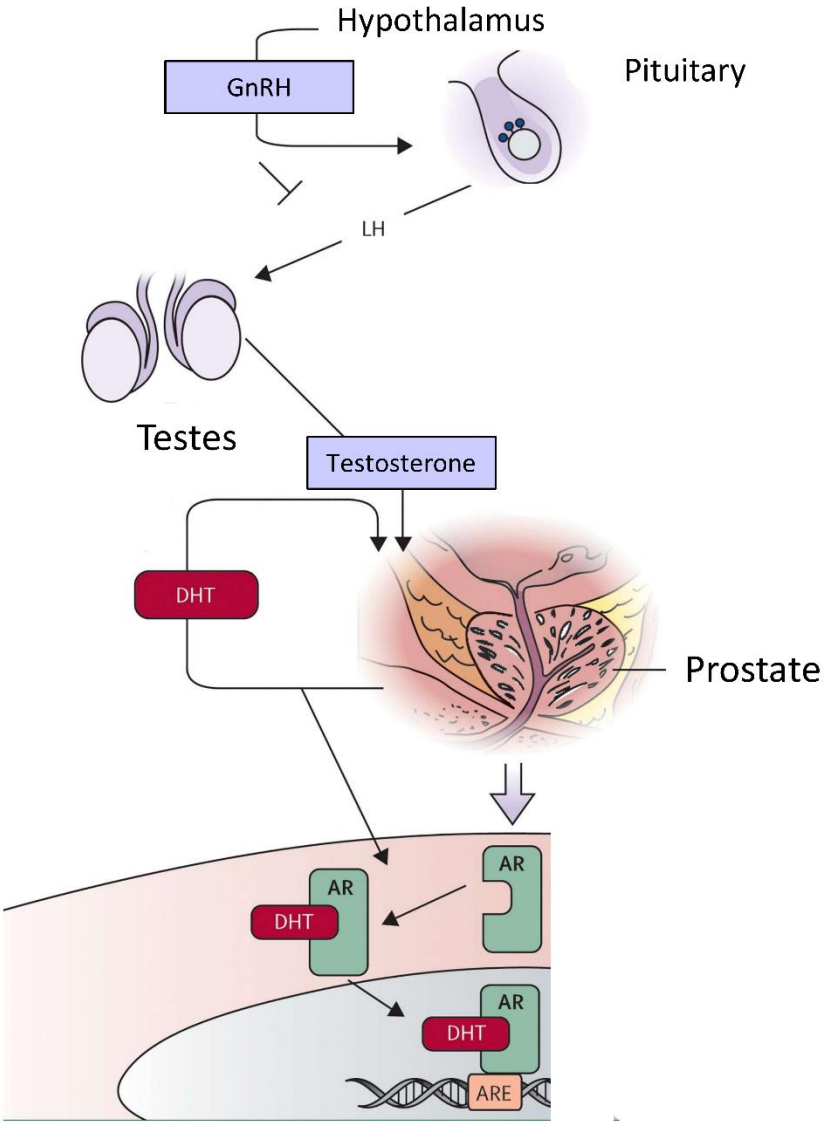


Figure 1.2. Synthesis of DHT and subsequent binding to the AR. The hypothalamus secretes GnRH to the anterior pituitary. The anterior pituitary then releases LH which travels to the testes and results in the production of testosterone, that moves to the prostate. The prostate cells can then convert testosterone to DHT. DHT can bind tightly to the AR allowing for subsequent translocation to the nucleus and binding to AREs. Abbreviations: GnRH, gonadotropin-releasing hormone, LH, luteinizing hormone, DHT, dihydrotestosterone, AR, androgen receptor, ARE, androgen response elements. *Adapted from*[9]

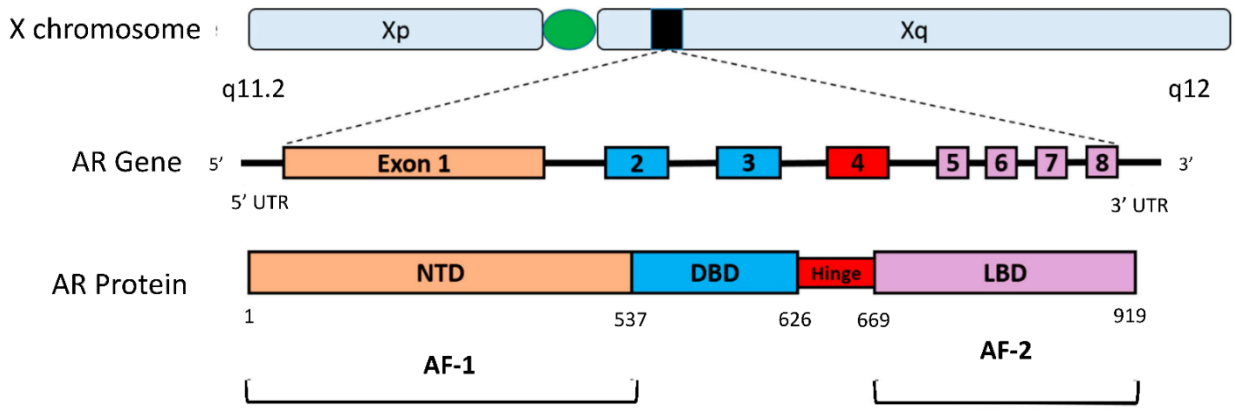


Figure 1.3. AR locus, gene and protein structure. The X chromosome (Xq11.2) contains the AR gene. Full length AR gene is made up of 8 exons. AF-1 is located in the NTD, and AF-2 is in the C-terminal LBD. The AR nuclear localization signal sequence is in the hinge region between the DBD and LBD. Abbreviations: AR, androgen receptor; AF-1, activation function 1; AF-2, activation function 2; DBD, DNA-binding domain; LBD, ligand-binding domain; NTD, N-terminal transactivation domain

Adapted from[10]

1.2.2. AR signaling

Elwood Jensen was the first to define a “two-step” mechanism for steroid hormone action, in which the steroid first binds to a cytoplasmic receptor, leading to dissociation into a smaller hormone-receptor complex, and subsequent translocation to the nucleus[11]. Upon DHT ligand activation, AR is released from sequestration by chaperones, such as heat shock protein 90 (HSP90), in the cytoplasm, translocates to the nucleus, and binds as a homodimer to androgen response elements (AREs) (Figure 1.4)[8]. AR can dimerize in a ligand-independent fashion *in vitro,* suggesting that dimerization *in vivo* is dependent on the activity of specific coregulators[12]. Ligand activation induces conformational changes in AR, and allows for associations with coregulators, such as pioneer factors and histone modifiers, to bind DNA at androgen responsive elements (AREs) and activate target genes, such as *KLK3* (PSA). Generally, the ARE palindromic repeat sequence is 5’-AGGTCA NNN TGACCT-3 , but AR can bind to motifs with minor variations[8]. Additionally, there is *in vitro* evidence of non-canonical signaling in which AR acts as a plasma membrane associated receptor to facilitate stimulation of other intracellular signaling pathways, via direct contact with signal transducers and/or activation of kinases[13, 14].

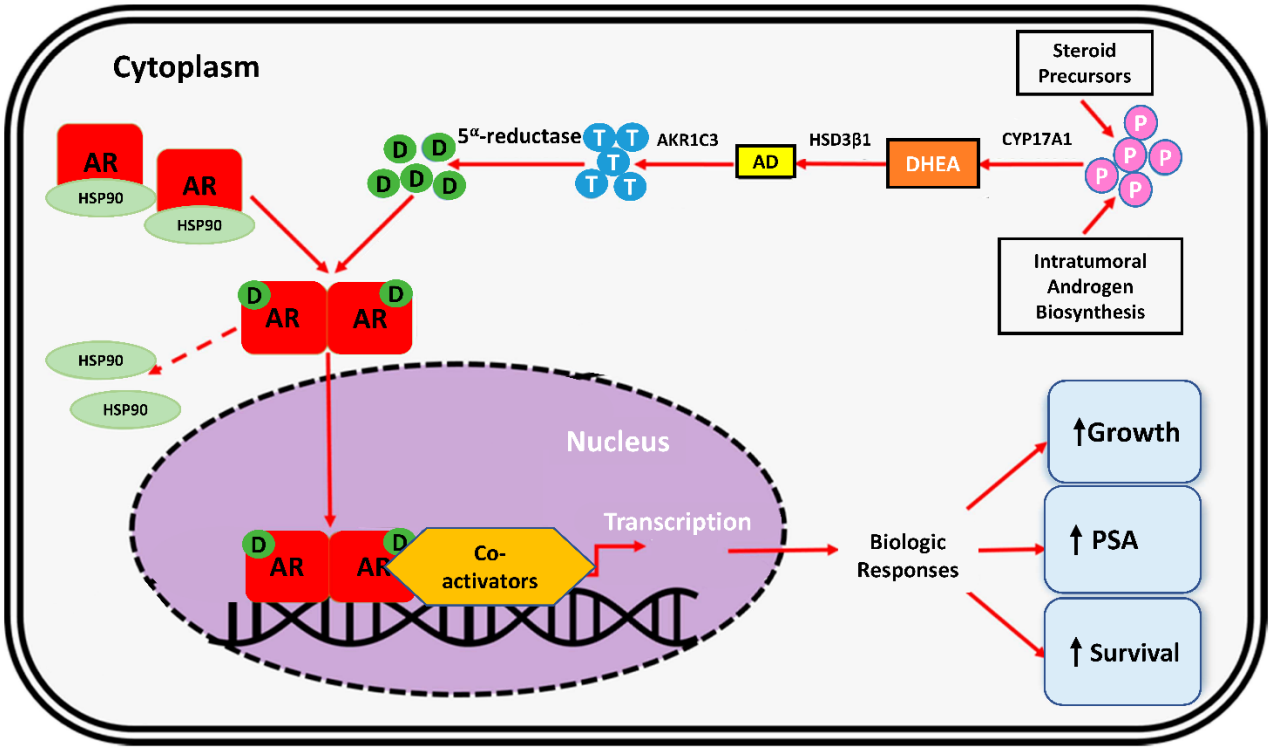


Figure 1.4. AR signaling axis. The enzyme CYP17A1 converts androgen precursors ("P" light purple circles) to DHEA, then HSD3β1 converts DHEA to AD, AKR1C3 converts AD to testosterone ("T" blue circles) then finally 5α-reductase converts testosterone to DHT ("D" green circles). DHT-ligand binding to AR causes conformational changes allowing for AR dimerization and subsequent translocation into the nucleus. Abbreviations: AD, androstenedione; AKR1C3, aldo-keto reductase family 1 member C3; AR, androgen receptor; CYP17A1, cytochrome P450 c17; DHEA, dehydroepiandrosterone; D, dihydrotestosterone; HSP, heat shock protein; HSD3β1, human 3-beta-hydroxysteroid dehydroxynase/delta5-4 isomerase type 1; P, androgen precursors; PSA,prostate-specific antigen; T, testosterone.

Adapted from[10]

1.3 AR-signaling and prostate cancer

1.3.1. Androgen deprivation therapy as a prostate cancer therapeutic

Charles Huggins' pioneering work in 1941 discovered the benefits of androgen ablation, resulting in a Nobel Prize in Physiology and Medicine in 1966. Huggins *et al.* found that androgen deprivation via surgical castration, or estrogen therapy, in men with advanced metastatic prostate cancers yielded substantial therapeutic benefits[15]. Today, patients receive pharmacological agents that target androgen receptor signaling, directly and/or indirectly, in conjunction with or without radical prostatectomy (RP), radiation and chemotherapy (Table 1.1). Chemical castration is the chronic use of gonadotropin-releasing hormone (GnRH) agonists or antagonists, and is typically regarded as a first line therapy. This form of androgen deprivation therapy (ADT) lowers testosterone levels by suppressing testosterone secretion from the testes, via inhibition of LH production (Figure 1.1). Patients with localized disease are usually treated with either RP, or radiation with or without GnRH agonist/antagonists. Locally advanced prostate cancers (defined as no regional lymph nodes or distant metastases) are typically treated with radiation and GnRH agonists/antagonists, and patients that have hormone therapy naive metastatic prostate cancer are given GnRH antagonists/agonists plus chemotherapy or abiraterone, a selective irreversible inhibitor of the enzyme CYP17A1 that is upstream of DHT synthesis (Figure 4) (reviewed in [16]).

Failure to respond to first line ADT treatment via chemical castration is termed castration resistant prostate cancer (CRPC) that is typically lethal. Second-line therapies, such as abiraterone, are administered to patients with CRPC in conjunction with anti-androgens, such as enzalutamide, that competitively bind to the LBD, preventing ligand activation. Despite administration of second-line therapies, many CRPC prostate tumors will continue to depend on AR-signaling (Figure 1.5).

Generic or code name	Brand name or other code name	FDA approval year (or clinical development stage)	Drug class
FDA approved			
Leuprolide	Lupron Depot, Eligard	1989 (Lupron Depot), 2002 (Eligard)	GnRH agonist
Goserelin	Zoladex	1989	GnRH agonist
Triptorelin	Trelstar	2000	GnRH agonist
Histrelin	Vantas	2004	GnRH agonist
Degarelix	Firmagon	2008	GnRH antagonist
Flutamide		1989	AR antagonist
Bicalutamide	Casodex	1995	AR antagonist
Nilutamide	Nilandron	1996	AR antagonist
Enzalutamide	Xtandi	2012	AR antagonist
Abiraterone	Zytiga	2011	CYP17A1 inhibitor

Table 1.1. Pharmacological agents that inhibit AR signaling in prostate cancer.

Table from [16]

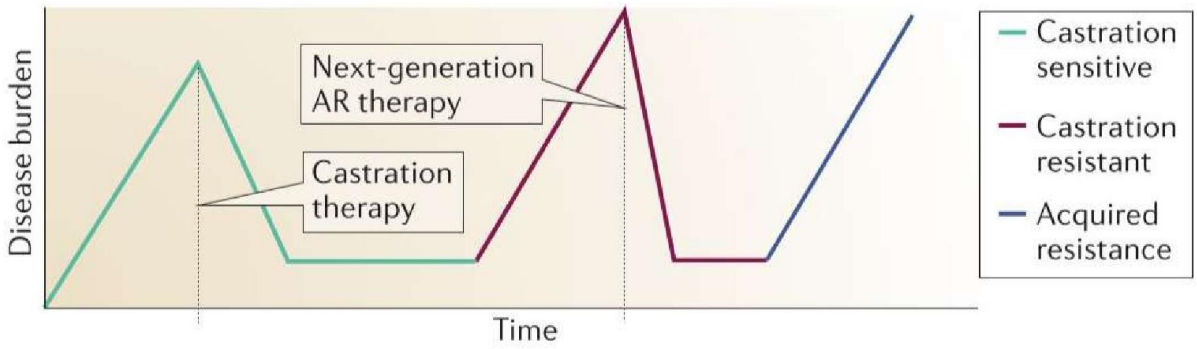


Figure 1.5. Timing of first and second line drugs in CRPC. Increased disease burden is measured by PSA levels and/or surveillance imaging. Initial increase in disease burden (green line) is treated with 1st line medical castration, GnRH agonists/antagonists. Once resistance occurs (purple line) AR inhibitors are administered, but acquired resistance to these 2nd line drugs is typically inevitable (blue line).

Figure from [16]

1.3.2. Aberrant AR-signaling in castration resistant prostate cancer

Patients with locally advanced, or hormone naïve metastatic prostate cancer, will usually respond favorably to ADT, but will frequently develop CRPC, despite therapy[17]. CRPC is typically associated with metastatic disease and poor prognosis, rendering it virtually incurable. It was thought that CRPC developed independent of AR signaling, leading to terms such as "androgen independent" or "hormone refractory". However, now it is generally accepted that there is sustained AR-signaling in CRPC tumors[18, 19].

At diagnosis, eighty to ninety percent of prostate cancers are dependent on circulating androgens for growth and proliferation[20, 21]. In fact, in early stages of prostate cancer, AR somatic mutations are rare, but this mutation rate significantly increases in advanced, castration resistant tumors[22]. AR mutations, along with AR alternative splice variants (ARVs) and AR amplifications, all sustain AR-signaling independent of levels of circulating androgens to drive CRPC[23, 24]. Constitutively active ARVs can be observed in multiple CRPC lines and patient tissues[25]. ARVs usually display structural changes in the N-terminal domain, encoded by exons 1 and 2 or exons 1 to 3, and a truncated C-terminal domain due to introns 2 or 3. ARV-7, encoded by exons 1 to 3 is the most frequently detected ARV in prostate cancer[26]. Modifications of the LBD are the major culprits in promoting ARV activity as the LBD is a common target of treatment, and can occur via mRNA alternative splicing, or chromosomal rearrangements. Li *et al.* found that *in vitro* ARVs were sufficient to confer enzalutamide (an AR-antagonist) resistance[27]. Using the CWR22 prostate cancer xenograft-derived line, Tepper and colleagues found that ARVs lacking the LBD are maintained in the nucleus, and are sufficient to restore AR-signaling, independent of ligand activation[28]. Hu *et al.* found four cryptic exons that yielded seven ARVs, all without the LBD[26].

Point mutations in the AR LBD can also lead to resistance of AR-antagonist therapies, via relaxation of ligand specificity, forcing the converting of an AR-antagonist effect to an AR-

agonist effect, yielding subsequent AR activation[22, 29, 30]. One of the most frequently observed mutations in prostate tumors occurs in the LBD, and mutation from threonine 877 to alanine (T877A). These types of point mutations are significant as they can induce hypersensitivity to other endogenous steroid hormones, and facilitate resistance to AR-antagonists[31]. Accordingly, an *in vitro* screen for mutations that confer enzalutamide resistance identified a point mutation in the LBD (F876L)[32].

DNA binding and subsequent gene expression profiles may vary between full length AR (AR-FL) and an ARV. Cao *et al.* determined that when ARV7 and AR-FL were both expressed, interactions between ARV7 and AR-FL were necessary to constitutively activate canonical AR-FL genes such as *PSA*. Notably, when this group use ChIP re-ChIP at *UBE2C*, a gene that is uniquely upregulated with over expression of ARV7, they found that ARV7 was sufficient to for transcriptional activation, without the interaction with AR-FL at the promoter[33].

AR amplifications are not typically found in hormone-sensitive cells, yet as much as 80% of CRPC cells display amplifications in the AR gene (reviewed in [10]). Prostate cancers that contain AR amplifications, thus exhibiting AR overexpression, are more responsive to decreasing levels of circulating androgens[18, 34]. Concordantly, Visakorpi *et al.* observed that the tumors of thirty percent of patients with CRPC displayed AR amplification only after ADT, as compared to matched tumors before ADT, suggesting therapy-induced selection of AR amplification as a mechanism for castration resistance[18]. Additionally, AR gene amplifications were also shown to drive an AR-antagonist to behave as an AR-agonist[19].

1.4 The transition to CRPC and metastasis

Increased PSA levels after initial treatment, known as biochemical recurrence (BCR), is the main way to determine recurrence. CRPC typically leads to bone metastases, and is associated with a poor prognosis with a mean survival time of only sixteen to eighteen months[35].

Despite the initial favorable responses to 2nd line ADT treatments that target AR, most patients with CRPC will acquire resistance, and the cancer eventually metastasizes. This inevitability is due to the CRPC tumor's continued dependence on AR-signaling. It is generally accepted that ADT forces a temporal and clonal selection for resistant clones in CRPC tumors[36]. Consequently, there is growing concern that continued long-term use of AR-antagonists might yield higher frequencies of tumors with AR deletions, but sustained AR-signaling. Thus, understanding the mechanisms that drive CRPC progression is vital.

Grasso *et al*. and Robinson *et al.* identified actionable signaling pathways, including the Wnt and PI3K-AKT- PTEN signaling pathways, as significantly altered in tumors from patients with metastatic CRPC, as compared to hormone naïve localized tumors[37, 38]. The CRPC tumor has multiple genomic alterations, mutations, chromosomal rearrangements, and changes in copy number. Thus, understanding the drivers responsible for the progression from the primary hormone naïve tumor to CRPC is essential.

1.5 SOX4 is a developmental transcription factor that is important for prostate cancer metastasis

1.5.1 SOX4 background

SOX4 is a developmental transcription factor conventionally known for its role in embryonic, and progenitor development[39, 40, 41]. SOX4 is expressed during breast and osteoblast development[42]. SOX4 is also important in early differentiation and expansion of transit amplifying progenitor cells. Knockout of SOX4 is embryonic lethal, with fatal cardiac defects at E14, and impaired lymphocyte development[43]. SOX4 is expressed in the gonads, thymus, T- and pro-B-lymphocyte lineages and is expressed at lower levels in the lungs, lymph nodes, and heart[44]. It is induced by multiple developmental pathways including the TGF-β-, Wnt- and PI3K-AKT signaling pathways[45, 46, 47] Additionally, SOX4 can also interact directly with and activate β-catenin in the Wnt pathway. The SOX4 gene contains a single exon that encodes a 47 kDa protein[48]. It has a conserved high mobility group (HMG) DNA-binding domain at the N-terminus, related to the TCF/LEF family of transcription factors, and a serine rich domain, a glycine rich-domain, and a transactivation domain (TAD), that is also a degradation domain, at the C-terminus[49] (Figure 1.6).

Figure 1.6. SOX4 domain structure. HMG domain: DNA-binding domain; GRR: glycine-rich region; SRR: serine-rich region; TAD/DD: transactivation domain/degradation domain.

Figure from [39]

1.5.2 SOX4 in cancer

In a large-scale meta-analysis of transcriptional profiles from a number of major human cancers, SOX4 was found to be one of 64 “Cancer Signature Genes”, suggesting that it has a fundamental oncogenic role[50]. SOX4 affects tumorigenesis in several ways, including inhibition of apoptosis, increased invasion and migration, and induction and maintenance of self-renewing cancer initiating cells[39, 40, 41, 48, 49, 51, 52, 53]. SOX4 is associated with cancer progression in multiple cancers. In adenoid cystic carcinoma cells, Pramoonjago *et al.* utilized siRNA to knockdown SOX4, and observed decreased proliferation and increased apoptosis[54]. In human endometrial cancer cell lines, expression of SOX4 was correlated with increased proliferative capacity, and knockdown resulted in diminished cell growth[55]. Analysis of RNA-Seq data from The Cancer Genome Atlas of 148 patients shows that SOX4 expression tends to be more highly expressed in aggressive prostate cancer (Figure 1.7A). Notably, we have previously demonstrated that in prostate cancer, SOX4 expression directly correlates with a high Gleason score, agreeing with analysis of prostate cancer RNA-Seq data from The Cancer Genome Atlas (TCGA) (Figure 1.7B)[3, 56]. Additionally, SOX4 over expression transformed immortalized non-neoplastic cell line RWPE-1, enabling anchorage-independent growth in soft agar[56, 57]. The Moreno lab generated ChIP-chip data and found 282 high-confidence direct SOX4 targets that included genes involved in microRNA processing, transcriptional regulation, developmental pathways, growth factor signaling, and tumor metastasis. Additionally, the Moreno lab found that SOX4 was necessary for expression of these genes[43].

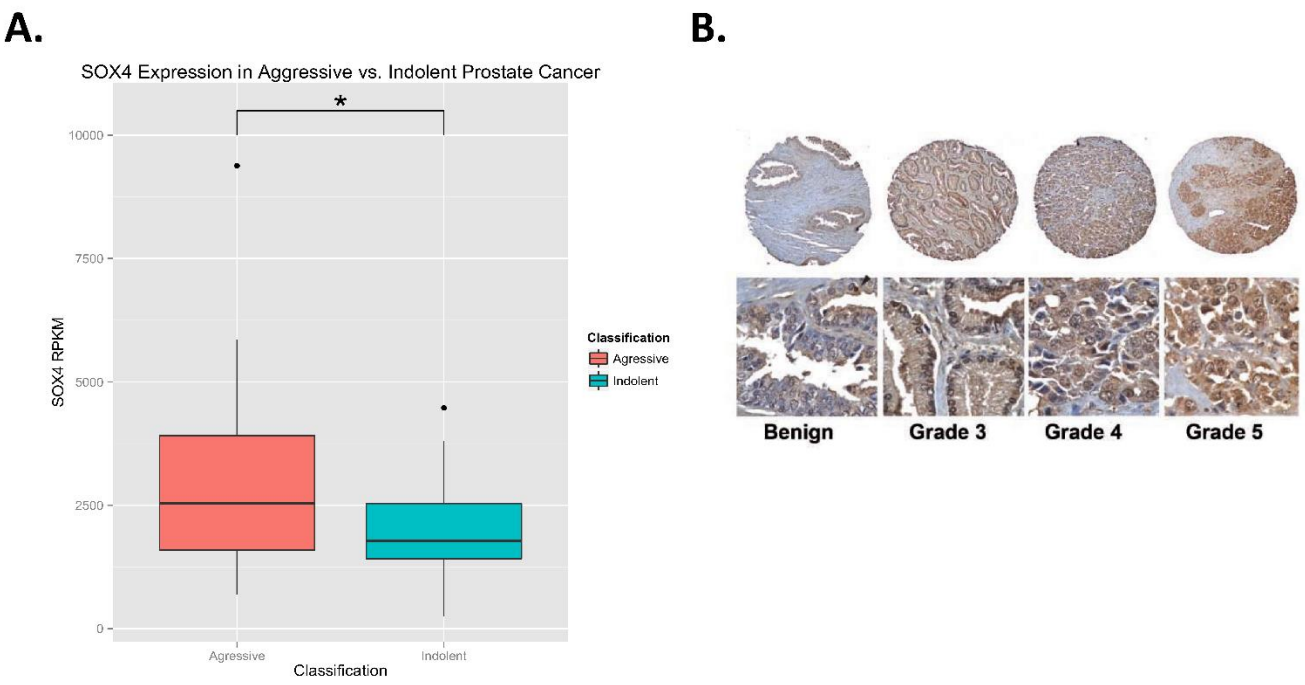


Figure 1.7. Increased SOX4 expression correlates with aggressive prostate cancer. *A)* analysis of RNA-Seq data from The Cancer Genome Atlas of 148 patients classified as having aggressive (Gleason Score > 7) or indolent prostate cancer indicates that SOX4 tends to be more highly expressed in aggressive prostate cancer. Asterisk: significance $p < 0.05$ Mann Whitney U Test. *B)* Formalin-fixed, paraffin-embedded prostate samples from Cooperative Prostate Cancer Tissue Resource (Silver Spring, MD) Gleason tissue microarray with SOX4 antisera. Dark brown: SOX4 detection; blue: nuclei counterstained with hematoxylin; Arrowhead: dark staining of nuclei from infiltrating lymphocytes (*Panel B from* [56]).

1.5.3 SOX4 in the epithelial to mesenchymal transition, and metastasis

Tavazoi *et al.* provided the first indication that SOX4 plays a role in promoting metastasis, demonstrating that knockdown of SOX4 resulted in fewer lung and bone metastases in breast cancer xenografts[58]. Additionally, we and others have found that knockdown of SOX4 in breast cancer cell lines significantly reduced migration and invasion[59, 60]. Over expression of SOX4 promotes adoption of the mesenchymal phenotype[61]. The Moreno lab found that SOX4 regulates, and is necessary for activation of, metastasis related genes, including epidermal growth factor receptor (EGFR), tenascin C (TNC), DLL1, and the metalloprotease ADAM10[43].

Metastasis is a sequential multi-step process that requires cancer cells to develop vasculature via angiogenesis, acquire migratory capabilities to detach from the primary tumor, enter blood vessels via intravasation, evade the immune system to survive in circulation, and finally extravasate and colonize a distant site (Figure 1.8). The Epithelial to mesenchymal transition (EMT) is a critical step in this process (Reviewed in [62]). EMT encompasses vast molecular changes mediated by aberrant activation of developmental signaling pathways, that drive cells to lose epithelial features and gain mesenchymal characteristics (Figure 1.9). Some of the processes that define EMT include the loss of cell-cell adhesion, loss of planar and apical-basal polarity, increased motility, and resistance to apoptosis and anoikis[62, 63]. Notably, SOX4 is associated with multiple developmental signaling pathways that are aberrantly activated during EMT, such as the TGF-β and EGF pathways[64]. These and other signaling pathways also converge on well-known transcription factors associated with EMT, such as Snail, Slug, Zeb1, Zeb2, and Twist[65, 66, 67, 68, 69, 70, 71, 72]. SOX4 may also be a critical transcription factor involved in EMT, as there is evidence shows that it is a master regulator of EMT in breast cancer via activation of the histone methyltransferase EZH2, in response to TGF-β signaling[73, 74].

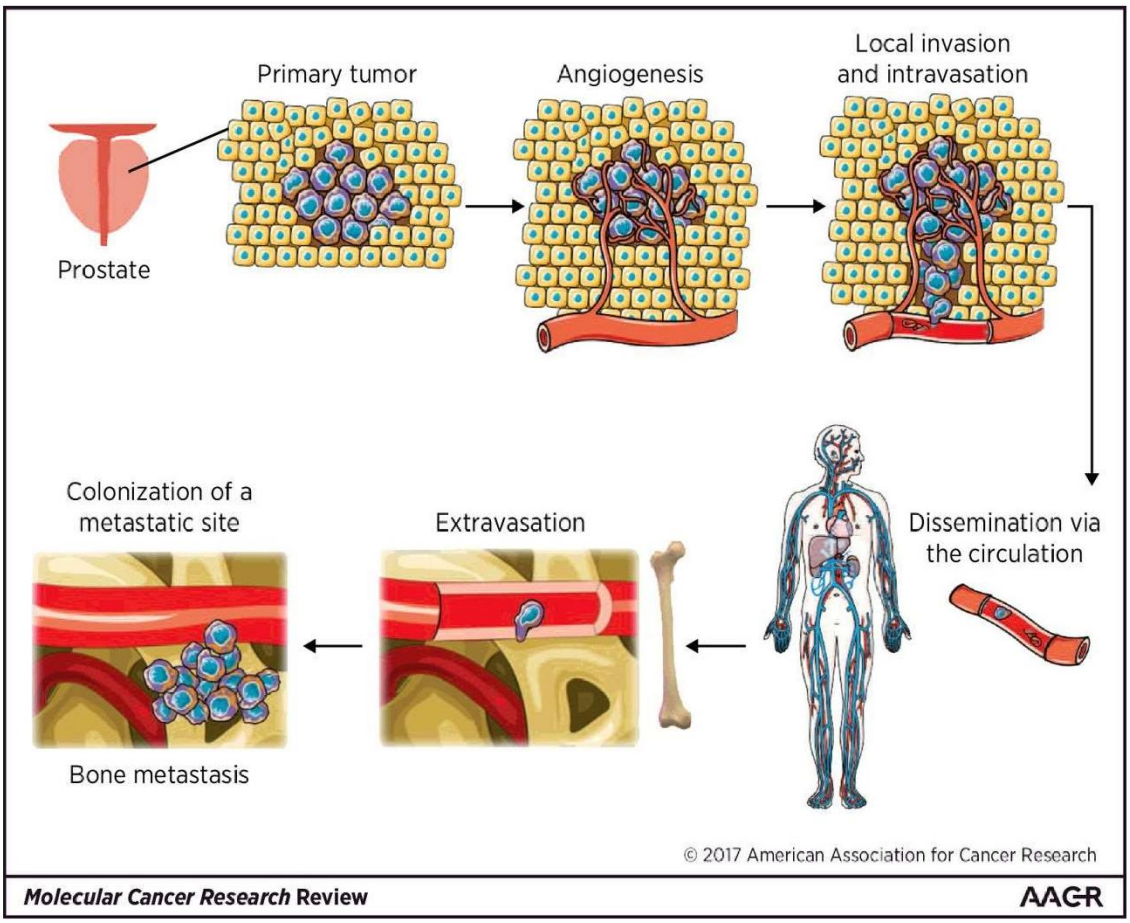


Figure 1.8. Steps of the metastatic cascade. Prostate tumor epithelial cells must acquire invasive and migratory characteristics to detach from the primary tumor, survive in the circulation, and colonize a metastatic site, typically in the bone.

Figure from [75]

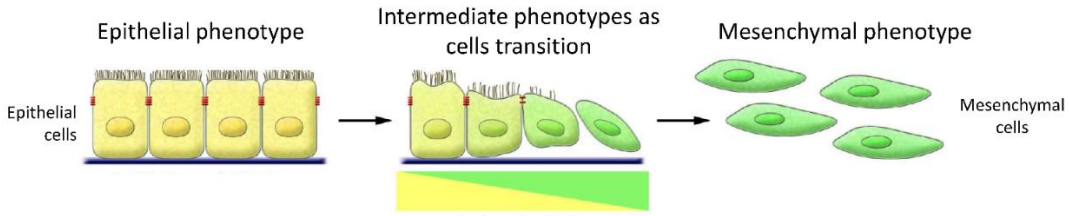


Figure 1.9. EMT induces molecular changes that allow for epithelial cells to acquire mesenchymal capabilities. Epithelial characteristics, such as cell-cell adhesions and apical-basal polarity, are lost as EMT progresses and cells develop a mesenchymal phenotype via actin reorganization and spindle-like morphology.

Adapted from [76]

1.6 Concerted action of multiple transcription factors drives complex changes in transcriptional profiles

Variations in gene signatures, such as those induced by aberrant activation of signaling pathways, are governed by combinatorial actions of a precise set of transcription factors. As noted previously, signaling pathways can cross talk and converge on a distinct group of transcription factors. A conceptual example comes from Barolo *et al.*, who observed synergistic activation of the Drosophila *eve* muscle and ear enhancer by transcription factors downstream of the developmental signaling pathways Wnt, Dpp/TGF-β, and receptor tyrosine kinase, and tissue specific mesodermal transcription factors[77]. This relationship demonstrates the "activity insufficiency" concept that describes how a signaling pathway induced transcription factor requires combinatorial interactions with other factors to activate cell type-specific genes.

AR transcriptional activity is largely dependent on the cooperation with specific co-activators, and interaction with other transcription factors can drive distinct transcriptional programs in response to ADT and transition to CRPC[78]. For example, there is evidence suggesting that alternative oncogenic signaling pathways can facilitate AR-signaling in a low androgen environment[79]. While there are many studies investigating the role of AR cofactors, how AR-independent factors that supplement, or function independently of, AR signaling to drive CRPC remains largely unknown.

Additionally, Fiore *et al.*, investigated the interdependence of interactions among pluripotency transcription factors using libraries of massively parallel reporter gene assays with multiple combinations of transcription factor synthetic *cis*-regulatory elements (CREs) in embryonic stem cells. They discovered that specific combinations of different transcription factors drove a more robust expression pattern, as compared to groups of the same transcription factor. Moreover, transcription factor combinatorial interactions that drive different phenotypic responses, such as after ADT, or during metastasis, require genomic re-localization

to new targets[80]. Taken together, these data support the idea that transcription factors do not act alone to drive transcriptional programs necessary for the transition to different molecular subtypes or biological processes.

1.7 Dissertation objectives

In this dissertation, I elucidate the dynamics of transcription factor activity, and how this contributes to dysregulation of normal transcriptional programs to influence response to ADT, and promote metastasis. The work presented here utilizes both a predictive computational, and experimental mechanistic approach to understand important regulators responsible for driving the aggressive prostate cancer tumor. The first part of this dissertation aims to identify upstream regulators responsible for changes in the tumor biology. By integrating multiple big datasets, I predict combinatorial relationships among transcription facts that mediate ADT response and metastasis. I argue that these transcription factor relationships may have the highest potential to promote metastatic CRPC. The second part of this dissertation focuses on the mechanisms critical to driving EMT, an important step in the metastatic cascade. While SOX4 is known to be a transforming oncogene in prostate epithelial cells, I contend that it continues to remain transcriptionally active in prostate cancer as a critical regulator of EMT. I provide preliminary data that demonstrate how SOX4 impacts EMT, and describe meaningful future directions to test my hypothesis. Taken together, the analyses in this dissertation examine important interactions that facilitate context-dependent transcriptional activities to drive aggressive prostate cancer.

Chapter 2. The biology of castrate resistant prostate cancer[81]

Fei Lian, Nitya Sharma, Josue D. Moran, and Carlos S. Moreno

My Contribution

As the second author of this publication, I contributed to the writing of Section 4.7: EMT and SOX family genes.

Abstract

Castrate resistant prostate cancer (CRPC) arises in recurrent prostate cancer patients who have failed androgen deprivation therapy. This review focuses on the biological pathways that are altered, disrupted, and activated in CRPC including androgen signaling, PI3K-AKT signaling, receptor tyrosine kinases, epigenetic pathways, Wnt signaling, long noncoding RNAs, and angiogenesis. These biological pathways represent molecular therapeutic targets that are in various stages of development, from basic research to clinical trials to FDA approved drugs. Prospects for novel therapies and prognostic biomarkers in CRPC are discussed.

2.1 Introduction

Prostate cancer is one of the most commonly diagnosed cancers and the second leading cause of cancer death for males in the United States[82]. Although there are recently approved therapeutic options for men with advanced and metastatic disease, the unfortunate reality is that advanced prostate cancer is inevitably fatal. Hormonal androgen deprivation therapy (ADT)[78] is the standard of care once a patient has recurrent disease following primary surgical or radiation therapy; however, the benefits from ADT are typically short-lived. Recurrent disease that follows ADT treatment is termed castrate-resistant prostate cancer (CRPC)[78], which is the most aggressive and lethal form of prostate cancer. The current treatment regime for CRPC consists primarily of chemotherapy with docetaxel, but other agents such as the immunotherapy agent sipuleucel-T, the taxane cabazitaxel, the CYP17 inhibitor abiraterone acetate, and the androgen receptor (AR) antagonist enzalutamide are also FDA approved for the treatment of CRPC[35, 83, 84, 85, 86, 87]. While there are clinical trials testing the potential therapeutic benefits of many other compounds, CRPC remains an incurable disease[35, 85, 88]. Recent advancements in our understanding of the biochemical and genetic pathways critical to CRPC progression have identified many novel potential druggable targets as well as biomarkers of disease progression[89, 90]. This review will focus on the biological aspects of aggressive prostate cancer, biomarkers, therapeutic targets, and future directions for the treatment of CRPC.

2.2 The androgen receptor

It is well-documented that prostate cancer cells depend on the androgen receptor (AR) and steroid hormones for continued oncogenic growth. Though ADT targets these pathways, this therapy can also select for tumor cells with several alternative survival mechanisms that allow AR to continue to function in the absence of circulating androgens. One such mechanism that generates androgen insensitivity seen in CRPC involves either the amplification of the AR gene or increased AR mRNA expression, along with a concomitant increase in the abundance of AR protein expression in CRPC tumor cells[91]. Another survival mechanism relies on the specific mutations in AR that enable binding and activation of this receptor by other steroidal hormones such as progesterone, hydrocortisone, estradiol[92], or even AR antagonists such as flutamide[93]. One critical pathway for CRPC is the intratumoral synthesis of steroidal hormones via upregulation of the cytochrome P450 gene CYP17A1, steroid-5-alpha-reductase, alpha polypeptide 1 (SRD5A1), aldo-keto reductase family 1, member C3 (AKR1C3), and hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2 (HSD3B2)[92]. These enzymes contribute to *de novo* synthesis of dihydrotestosterone (DHT) from endogenous cholesterol, as well as from uptake of exogenous dehydroepiandrosterone sulfate (DHEA-S), providing the rationale for treatment with the CYP17A1 inhibitor abiraterone. Furthermore, post-translational modifications such as aberrant phosphorylation of AR, by the Src tyrosine kinase on Y534 leads to increased sensitivity of the AR to low levels of circulating androgens[94]. *Wang et al.* showed through gene expression profiling that AR in CRPC selectively upregulates cell cycle M-phase genes promoting CRPC growth[78], suggesting that AR in CRPC may regulate a distinct set of genes independent of the typical AR-dependent transcriptional program.

Genome wide localization analysis via chromatin immunoprecipitation and sequencing (ChIP-seq) of AR in CRPC tissues has identified a set of genes regulated by AR that are distinct from those observed from cultured prostate cancer cell lines[95]. Analysis of DNA binding motifs

adjacent to AR binding sites determined that unlike cultured cells in which AR interacts with FOXA1 and NF-1, *in vivo* AR interacts with E2F, MYC, and STAT5[95].

2.3 PI3K-AKT-mTOR pathway

Activation of the PI3K-AKT-mTOR pathway is extremely common, if not universal, in CRPC. It is critical for the regulation of cell survival, apoptosis, cell proliferation, autophagy, metabolism, and protein synthesis, and has been extensively studied in prostate cancer[96]. PI3K is activated by a wide range of growth factor receptors and signaling pathways, including epidermal growth factor receptor (EGFR), insulin-like growth factor 1 receptor (IGF-1R), fibroblast growth factor receptor (FGFR), and platelet-derived growth factor receptor (PDGFR). Activated PI3K activates PDK1, which in turn activates AKT, while PI3K can be inactivated by the PTEN tumor suppressor[97, 98, 99]. AKT separately phosphorylates and activates mTOR, which promotes cell cycle progression and growth and provides positive feedback via phosphorylation of AKT. Aberrant constitutive AKT activation is one of the most frequent pathway alterations observed in several different cancers. Recently, *Chin et al*. demonstrated that the AKT2 isoform was specifically required for cellular maintenance in prostate cancer tumors lacking PTEN expression[100]. Genome sequencing analysis of CRPC has shown the PTEN gene is mutated in 48% of CRPC samples, and that 33% of samples have mutations in genes that interact with PTEN, resulting in 81% of samples harboring some mutation in the PTEN interaction network[38]. Several PI3K and mTOR inhibitors are currently under investigation in clinical trials for CRPC including the dual inhibitor NVP-BEZ235[101], and the mTOR inhibitor RAD001 or everolimus[102, 103]. Of these compounds, NVP-BEZ235 likely has the most promise, although everolimus may be effective in subsets of patients.

2.4 Receptor tyrosine kinase (RTK) growth factor pathways

Multiple RTK growth factor signaling pathways including EGFR[104, 105], IGF-1R[106, 107], FGFR[108], PDGFR[109, 110], and HGFR/c-MET[111, 112, 113] pathways have been investigated extensively in prostate cancer progression due to their activation of the PI3K-AKT and RAS-MAPK pathways. Combined loss of PTEN with activation of the RAS-MAPK pathway can cooperate to induce epithelial to mesenchymal transition (EMT) and metastases[114]. Several clinical trials have been conducted targeting EGFR, including two using lapatinib, one of which was negative[115], and another that was more encouraging[116] with single agent activity in a subset of patients. A third trial using cetuximab showed significant progression free survival in approximately half of the treated patients[117]. Antibodies against IGF-1R have shown some efficacy in xenografts[107] and in human trials[118].

Recent epidemiological studies have determined that the generic anti-diabetes drug metformin results in lower prostate cancer specific mortality[119], suggesting that not only insulin and IGF, but also glucose regulation may impact patient survival. RTK growth factor pathways can allow cells to meet bioenergetic demands of rapid proliferation by influencing cell signaling via phosphorylation. The Warburg effect has been recognized for nearly a century as perturbation of cancer cell metabolism, but it has only recently become the focus of intense investigation as an avenue for potential therapy[120, 121]. In the Warburg effect, cancer cells undergo aerobic glycolysis rather than mitochondrial respiration and oxidative phosphorylation to generate the necessary precursors for protein, nucleotide, and lipid synthesis that are essential to rapid cell growth. In the reverse Warburg effect, perturbations of metabolic and catabolic metabolism can lead to a shuffling of nutrients between prostate cancer epithelial cells and surrounding stromal cells, in which pyruvate and lactate are transported to prostate tumor cells by the stroma[122]. Almost all glycolysis genes are overexpressed in prostate cancer, especially in advanced disease[123]. These data have led to calls for clinical trials employing combination therapies of RTK inhibitors or PI3K-AKT inhibitors with metformin[124].

2.5 Analysis of CRPC genomes

Although primary prostate cancers are characterized by genomic translocations that place androgen-responsive promoters such as TMPRSS2 adjacent to ETS-family transcription factors such as ERG[125], ETV1[126], ETV5[127], and ETV4[128], metastatic prostate cancers can harbor rare gene rearrangements that generate fusions of ubiquitin conjugating enzymes with KRAS[129]. Integrated gene expression and copy number analysis of prostate cancers also indicates that more aggressive prostate cancers have increased copy number alterations relative to less aggressive cancers[130]. Moreover, PI3K-AKT was altered in 100% of metastases, while RAS-RAF signaling was altered in 90% of metastases in this study[130]. Exome sequencing of CRPC has identified several recurrent mutations in transcription factors and epigenetic factors that interact with AR such as FOXA1, MLL2, UTX, and ASXL1[38].

2.6 Epigenetic pathways

The hypermethylation of CpG islands in gene promoters can inhibit expression of genes such as GSTP1[131] and has been implicated in tumorigenesis and cancer progression. Recent data have suggested that methylation of microRNA genes that target AR such as miR-34a can lead to increased AR expression, and that forced expression of miR-34a can reduce AR levels[132]. Since epigenetic modification of genetic DNA is a potentially reversible process, it is an enticing target in the treatment of CRPC. One such pathway target involves the inhibition of DNA methyltransferases (DNMT) to decrease the effects of methylation-related gene silencing. 5-Aza-2'-deoxycytidine (5-Aza-CdR) has been shown to increase expression of miR-146a in LNCaP cells and delay progression of tumor xenografts[133], but more data is needed on drug efficacy in the face of docetaxel-treated CRPC.

The enhancer of zeste 2 (EZH2) protein is a histone methyltransferase and the catalytic subunit of the Polycomb repressive complex 2 (PRC2) that tri-methylates histone 3 lysine 27 (H3K27me3) to repress many developmental genes, especially in metastatic prostate cancer[134]. EZH2 is elevated in metastatic prostate cancer versus clinically localized disease or benign prostate[135, 136], and high expression of EZH2 in patients correlates with an increased probability of developing skeletal metastases, along with seminal vesicle and lymph node infiltration[137].

Recent data from exome sequencing of CRPC have shown that there are multiple mutational alterations in the chromatin and histone modifying genes. *Grasso et al.* found that in pre-treated lethal metastatic CRPC tumors, there are interactions between the MLL complex family and AR, further demonstrating that there is dysregulation in epigenetic activation, commonly seen in CRPC[38]. *Berger et al.* showed that chromatin modifying genes CHD1, CHD5, and HDAC9 were mutated in 43% of sequenced Gleason 7 or higher prostate cancer tumors[138]. Specifically, CHD1 sequencing exhibited splice site mutations as well as intragenic breakpoints—all leading to truncated protein expression[138].

Histone and chromatin-remodeling complexes are potential targets for CRPC therapy, with targeting of histone deacetylases (HDACs) that facilitate AR-mediated transcription activation and repression. However, current HDAC inhibitors, such as vorinostat[139, 140] and panobinostat[141], have shown high rates of side effects and disappointing efficacy in the treatment of docetaxel-refractory CRPC.

Nevertheless, there have been exciting developments in the development of inhibitors of bromodomain-containing proteins that recognize and bind to acetylated lysines of histone proteins. Recently, a new group of small molecules has emerged as novel inhibitors of Bromodomain Containing Protein 4 (BRD4). BRD4, along with the Mediator complex, binds at super-enhancer sites to facilitate initiation of transcription of target genes[142]. BRD4 inhibitors, such as JQ1[143], bind to bromodomains of proteins such as BRD4, and inhibit BRD4 from binding to super-enhancers of known proto-oncogenes, including MYC[144, 145]. A mouse model of aggressive prostate cancer with simultaneous loss of PTEN and p53 tumor suppressor genes, termed RapidCaP, demonstrated highly penetrant metastases and activation of MYC[146]. Moreover, these castrate resistant tumors were sensitive to BRD4 inhibition using JQ1 the inhibitor[146]. Additional studies[147] using JQ1 and the orally bioavailable BRD4 inhibitor I-BET762[148], found that BRD4 inhibitors disrupt AR signaling, and recruitment to and activation of downstream target genes such as the TMPRSS2-ERG gene fusion[147].

2.7 EMT and SOX family genes

Typically, prostate cancer mortality is often related to metastasis to the bone, adrenal gland, liver and lung[149]. The epithelial to mesenchymal transition (EMT) is a major step in the metastatic process. To metastasize cancer cells need to acquire migratory and invasive capabilities, a process that involves EMT[62]. EMT encompasses vast molecular changes including gain of mesenchymal markers such as vimentin and N-cadherin, and loss of epithelial markers such as E-cadherin, mediated by aberrant developmental signaling pathway activation that allows epithelial cells to discard differentiated characteristics and acquire migratory and invasive capabilities typical of mesenchymal cells[62]. These changes include the loss of cell-cell adhesion, planar and apical-basal polarity, increased motility, and resistance to apoptosis and anoikis (cell death due to the detachment from the extracellular matrix)[62, 63]. Among the developmental signaling pathways that are aberrantly activated during EMT is the TGF-β signaling pathway, a highly studied major inducer of EMT[64]. The canonical TGF-β pathway is stimulated via TGF-β induced receptor complex activation, leading to phosphorylation of SMAD 2/3. Subsequently, these SMADs form a trimer with SMAD4 and translocate to the nucleus and associate with other transcription factors to transcribe EMT-inducing genes[150].

Recently, it was found that SOX4 is a master regulator of TGF-ß induced EMT via induction of EZH2[73]. *Tiwari et al* demonstrated that SOX4 directly activates EZH2 expression upon TGF-β treatment and that forced expression of EZH2 can overcome SOX4 knockdown and restore TGF-β induced EMT[73]. Moreover, *Wang* et al. found that, in prostate cancer cells, SOX4 knockdown inhibited TGF-β induced EMT, while SOX4 over expression promoted adoption of the mesenchymal phenotype[151]. They also demonstrated that TMPRSS2-ERG is critical for TGF-β induction of SOX4 expression[151]. *Tiwari et al.* and *Zhang et al.* both demonstrated that ectopic expression of SOX4 could induce EMT by increasing the expression of mesenchymal markers and decreasing the expression of epithelial markers[60, 73]. In addition,

SOX4 knockdown was sufficient to cause a reversion from a mesenchymal to epithelial phenotype after a 15-day TGF-β treatment[73].

Another SOX family factor, SOX9, has also been implicated in prostate cancer progression. Deletion of SOX9 in two different mouse models (TRAM and Hi-Myc) inhibited prostate cancer initiation[152]. ERG redirects AR to a cryptic enhancer of SOX9 to activate SOX9 expression, and knockdown of SOX9 inhibits invasion and growth of VCaP cells *in vitro* and *in vivo*[153]. SOX9 cooperates with PTEN deletion to drive prostate tumorigenesis[154], and it activates expression of Wnt pathway components such as LRP6 and TCF4[155]. Like SOX9, SOX4 also plays an important role in Wnt signaling via direct interaction with β-catenin[43, 51]. SOX4 can act as an oncogene in prostate cells[56], and activates expression of additional Wnt pathway components such as FZD3, FZD5, and FZD8[43, 156].

2.8 WNT signaling

There have been many lines of evidence that suggest that Wnt signaling may be important in CRPC[157]. Wnt pathway genes are frequently mutated in metastatic CRPC[38]. WNT7B ligand is a direct transcriptional target of AR, and can induce osteoblastic bone lesions[158]. TMPRSS2-ERG directly activates LEF1 transcription and expression of ligands such as WNT1 and WNT3A[159]. Moreover, AR interacts directly with β-catenin in CRPC xenografts, whose gene expression profiles exhibit enhanced Wnt signaling[160]. Interestingly, a novel small compound (iCRT3) that inhibits β-catenin transcriptional activity blocks AR binding to target genes, inhibits growth of tumor xenografts, and interferes with self-renewal of bicalutamide resistant cells[161].

2.9 Notch and Hedgehog

The Notch and Hedgehog pathways are also developmental pathways that likely play roles in prostate cancer progression. Notch signaling is important for prostate differentiation and maintenance of prostate stem cells[162]. Hedgehog signaling has also been associated with aggressive prostate cancer, and ADT alone or combined with chemotherapy induces hedgehog pathway activation[163]. Moreover, combined inhibition of AR and hedgehog signaling synergizes to inhibit growth of CRPC xenografts[164]. Docetaxel resistant CRPC cells survive via activation of Notch and Hedgehog pathways that inhibit apoptosis, and depletion of these cells results in re-sensitization to chemotherapy[165]. Thus, combination therapies that target developmental pathways along with targeting of AR and/or conventional chemotherapy may prove effective in CRPC.

2.10 lncRNA-mediated pathways

There is much data suggesting that long non-coding RNAs (lncRNA) greater than 200 base pairs in length play a critical role in tumor biology[166, 167, 168, 169, 170]. lncRNAs are typically transcribed by RNA polymerase II and are associated with epigenetic modification of histones. These lncRNAs partner with PRC1 and PRC2, leading to downstream ubiquitination and methylation activity that inhibits gene expression[171]. Furthermore, many lncRNAs have been linked to prostate cancer[166, 167, 172, 173, 174, 175, 176, 177]. In one study, two lncRNAs, PRNCR1 and PCGEM1, were shown to bind to AR and increase ligand-dependent and ligand-independent AR-mediated gene expression, which causes downstream cellular growth signaling and cancer proliferation[177]. However, subsequent studies have contradicted these findings, and suggest that while PCGEM1 is associated with prostate cancer, neither of these lncRNAs interacts with AR signaling nor are prognostic in prostate cancer[178]. Nevertheless, other lncRNAs play important roles in prostate cancer. *Pickard et al*. demonstrated that transiently increased GAS5 lncRNA levels upregulated basal apoptosis in PC3 prostate cancer cells, and a similar effect was seen in 22Rv1 cells[172]. *Prensner et al*. found that SChLAP1 lncRNA overexpression, which is involved in SWI-SNF activity knockdown, was observed in 25% of prostate cancers and was an independent predictor of aggressive disease[173]. Other lncRNAs that have been linked to CRPC include MALAT-1[174], CTBP1-AS[175], and Linc00963[176]. Linc00963, specifically, was shown to enhance the evolution from androgen-dependence to androgen-independence through the EGFR pathway[176]. From these studies, it is clear that lncRNAs are intimately involved in prostate cancer biology and that dysregulation of lncRNA expression may lead to tumor suppressor antagonism as well as to castrate resistance. Future goals in lncRNA research include defining specific lncRNAs in the human genome, understanding which lncRNAs are altered in the course of tumorigenesis, and delineating the structure and binding mechanisms of lncRNAs to their targets[166, 168, 169, 170].

2.11 Angiogenic pathways

Angiogenesis refers to the growth of new blood vessels from existing vasculature[179]. This process is an important pathway for CRPC progression, and is correlated with both increased rate of disease progression and decreased survival[180]. The most common target of angiogenesis is vascular endothelial growth factor (VEGF) through the use of an anti-VEGF antibody, such as bevacizumab[179]. VEGF can also drive EMT[181], and this pathway is currently a target of several therapeutic clinical trials testing new drugs including itraconazole[90], bevacizumab[182], tasquinimod[183], and ramucirumab[184]. Alternative strategies include inhibiting the binding of VEGF using VEGF-like receptors, such that VEGF is unable to interact with its normal target VEGFR receptor. VEGF ligand directed drugs have seen success in improving patient survival in other cancers, such as colorectal, breast, and non-small-cell lung cancer. However, in two prospective randomized control trials, VENICE and CALGB 90401, neither docetaxel/aflibercept nor docetaxel/bevacizumab combinations showed any significant improvement in median overall survival in patients with CRPC[185, 186]. Docetaxel plus bevacizumab did show significant improvement in PSA decrease and progression-free-survival when compared to docetaxel alone, but this was not observed in the VENICE trial. A newer small molecule inhibitor of VEGFR2 and MET, cabozantinib, early on displayed a decrease in growth of breast, lung, and glioma tumor models while increasing apoptosis[187]. Later, in a prospective randomized control trial, patients with CRPC who took cabozantinib significantly increased progression-free survival when compared to placebo[188]. Furthermore, other trials demonstrated that cabozantinib decreased narcotic use, significantly improved sleep quality[189], and inhibited prostate cancer bone growth when tested on androgen-sensitive and castration-resistant cell lines[190]. Other angiogenic strategies are to target the downstream activation pathways of VEGF, or to bypass the VEGF pathway completely and focus on established CRPC vessels and nutrient delivery[179, 180].

2.12 Biomarkers

Currently, there is a critical need for biomarkers of prostate cancer that can distinguish between locally indolent and metastatically aggressive disease[191]. The only FDA approved biomarker, prostate specific antigen (PSA), has historically been used for screening and detection of prostate cancer, but not for prognosis nor detection of CRPC. A major issue with PSA is that it detects prostate cells, not prostate cancer cells, and thus can result in over diagnosis and over treatment. While results of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial showed no reduction in prostate cancer specific mortality when comparing systematic annual PSA screening to opportunistic screening[192], the European Randomized Study of Screening for Prostate Cancer (ERPSC) found a small reduction in mortality (1 death per 1000 men screened) in a PSA-screening naïve population[193]. However, several potential flaws in the ERPSC study[194] and the associated harms of over diagnosis and over treatment led the US Preventive Services Task Force (USPSTF) to recommend against PSA screening[195, 196]. Given the recent controversy in its screening effectiveness, PSA is now recommended primarily for the determination of prostate cancer progression and recurrence.

However, there are potential replacements for PSA for detection screening under development, including PCA3 and the ETS fusion gene TMPRSS2-ERG[197, 198, 199, 200, 201, 202, 203]. TMPRSS2-ERG has been shown to be a predictive indicator of prostate cancer development in patients who present with high-grade prostatic intraepithelial neoplasia[200]. FISH analyses of CRPC and metastatic prostate cancers found that TMPRSS2-ERG was detected more often in metastatic cancer, and that TMPRSS2-ERG positivity was strongly correlated to both AR and ERG expression[201]. TMPRSS2-ERG expression has also been seen in otherwise histologically benign radical prostatectomy as well as in cystoprostatectomy specimens. TMPRSS2-ERG and PCA3 can also be detected in patient urine post-digital rectal exam[204]. However, there are still conflicting data regarding the prognostic value of PCA3 or TMPRSS2-ERG. A 2011 study by *Danila et al.* found that the TMPRSS2-ERG fusion gene could be accurately assayed in

circulating prostate tumor cells present in the blood of CRPC patients, but did not show that the presence of the fusion was a significant factor in abiraterone acetate treatment response[198]. Another prospective study of 322 patients illustrated that urinary PCA3 or TMPRSS2-ERG scores were not reliable in staging advanced prostate cancer.

Recent integrated bioinformatics network analyses that have included data mining and profiling of both animal models and human specimens have generated some interesting candidate biomarker gene sets for distinction of aggressive from indolent disease[205] and drivers of aggressive prostate cancer[206]. *Irshad et al.* identified a panel of three genes (FGFR1, PMP22, and CDKN1A) that could predict indolence outcome for low Gleason score (Gleason 6) patients via meta-analysis of multiple studies followed by independent validation on a cohort of 95 mRNA samples and 44 fixed biopsy samples at the protein level[205]. *Aytes et al.* performed cross-species analysis of mouse and human prostate cancer gene expression patterns to identify FOXM1 and CENPF as synergistic regulators of aggressive disease[206]. They also found that FOXM1 and CENPF could function as prognostic biomarkers of metastasis in two independent prostate datasets[207, 208]. Another distinct 16-gene signature of AR target genes derived from ChIP-seq of AR in CRCP tissues[95] was able to accurately predict CRPC and prostate cancer recurrence in two distinct patient cohorts[130, 207].

Our research group at Emory University recently completed RNAseq analysis of 100 formalin-fixed paraffin embedded prostatectomy specimens, and identified a 24-gene biomarker panel that robustly predicts biochemical recurrence following surgery[209]. Our biomarker panel accurately predicted recurrence in an independent patient cohort[130], and outperformed previously developed RNA biomarkers developed by Myriad Genetics[210]. Whether this set of 24 biomarker genes will also detect aggressive CRPC or discriminate aggressive from indolent disease requires further research on additional patient cohorts.

2.13 Conclusion

Recent advancements in the molecular understanding of CRPC have given us a number of potential biological targets for the treatment of CRPC. Many of the pathways and targets covered in this review currently have agents that are undergoing clinical trials, and some are FDA approved for the treatment of CRPC. Unfortunately, most of these pharmaceutical agents only moderately increase survival and CRPC still remains incurable. However, recent discoveries and avenues of research may enable more effective molecularly targeted therapies as well as a better understanding of the mechanisms of CRPC.

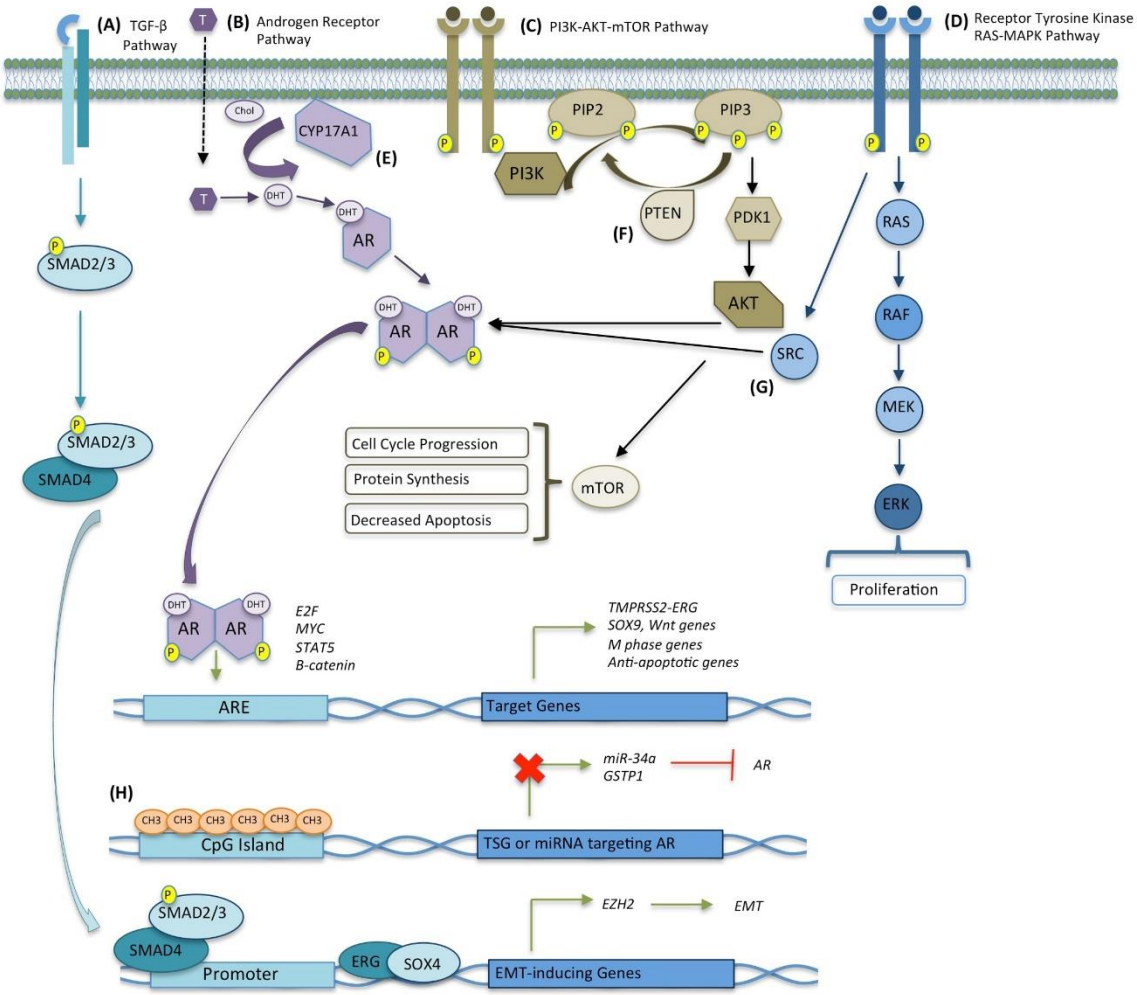


Figure 2.1. Signaling pathways in castration resistant prostate cancer.

(A) TGF-β pathway is stimulated via TGF-β induced receptor complex activation, leading to phosphorylation of SMAD2/3 and subsequently form a trimer with SMAD4 and translocate to the nucleus and associate with other transcription factors to transcribe SOX4 which activates EZH2 expression leading to EMT. (**B**) Prostate cancer cells depend on the androgen receptor (AR) pathway and steroid hormones for continued oncogenic growth. This can occur by way of AR over expression or by AR gene amplification. Mutations in AR allow binding and activation of AR by other steroidal hormones. Wnt pathway genes are frequently mutated in CRPC and are transcriptional targets of AR. (**C**) Activation of the PI3K-AKT-mTOR pathway is extremely common in

CRPC. Activated growth factor receptor tyrosine kinases (e.g. EGFR and IGF-1R) leads to activated PDK1, which in turn activates AKT. AKT separately phosphorylates and activates mTOR which promotes cell cycle progression, protein synthesis and decreased apoptosis. AKT can interact with AR in an androgen independent manner. (**D**) Activation of Receptor Tyrosine Kinase (RTK) pathway (PDGFR, HGFR/c-MET, etc.) leads to proliferation through RAS-MAPK. Combined with loss of PTEN, overactive RAS-MAPK can induce EMT. (**E**) Intratumoral synthesis of steroidal hormones from cholesterol via upregulation of the cytochrome P450 gene CYP17A1. (**F**) Loss of the PTEN Tumor Suppressor promotes aberrant PI3K-AKT-mTOR signaling. (**G**) Aberrant Y534 phosphorylation by Src increases AR sensitivity to androgens**. (H)** Epigenetic Pathways: Hypermethylation of CpG islands in gene promoters inhibits expression of tumor suppressor genes or miRNAs targeting AR.

Chapter 3. Identification of transcription factor relationships associated with androgen deprivation therapy response and metastatic progression in prostate cancer

Authors: Nitya V. Sharma[1], Kathryn L. Pellegrini, PhD[2], Veronique Ouellet, PhD[3], Felipe O. Giuste[4], Selvi Ramalingam[5], Kenneth Watanabe, PhD[5], Eloise Adam-Granger[3,] Lucresse Fossouo[3], Sungyong You, PhD[8], Michael R. Freeman, PhD[8], Paula Vertino, PhD[9,10], Karen Conneely, PhD[10,11], Adeboye O. Osunkoya, MD[3,5,10], Dominique Trudel, MD,PhD[3,6,], Anne-Marie Mes-Masson, PhD[3,7], John A. Petros, MD[2,10,12], Fred Saad, MD[3,13,14], and Carlos S. Moreno, PhD[5,10].

Author Contributions

NVS, KLP, FOG, AOO, PV, KC, and CSM conducted the experiments and contributed to data interpretation; VO, DT, AMM, EAG, LF, AOO, and FS participated in sample selection, case review, and clinical data annotation; NVS, KLP, FOG, SR, KW, SY, MRF, PV, KC, JAP, and CSM participated in data analysis, discussion and manuscript preparation; NVS, JAP, FS, and CSM designed the experiments and wrote the paper; and all were involved in manuscript editing.

This manuscript is currently under review at BMC Cancer.

Abstract

Background: Patients with locally advanced or recurrent prostate cancer typically undergo androgen deprivation therapy (ADT), but the benefits are often short-lived, and responses are variable. ADT failure results in castration-resistant prostate cancer (CRPC), that inevitably leads to metastasis. We hypothesized that differences in tumor transcriptional programs may reflect differential responses to ADT and subsequent metastasis.

Results: We performed whole transcriptome analysis of 20 patient-matched Pre-ADT biopsies and 20 Post-ADT prostatectomy specimens, and identified two subgroups of patients (high impact and low impact groups) that exhibited distinct transcriptional changes in response to ADT. We found that all patients lost AR-dependent subtype (PCS2) transcriptional signatures. The high impact group maintained the more aggressive subtype (PCS1) signal, while the low impact group more resembled an AR-suppressed (PCS3) subtype. Computational analyses identified transcription factor coordinated groups (TFCGs) enriched in the high impact group network. Leveraging a large public dataset of over 800 metastatic and primary samples, we identified 33 TFCGs in common between high impact group and metastatic lesions, including SOX4/FOXA2/GATA4, ERF/ETV5/ETV3/ELF4, and a TFCG containing JUN, JUNB, JUND, FOS, FOSB, and FOSL1. The majority of metastatic TFCGs were subsets of larger TFCGs in the high impact group network, suggesting refinement of critical TFCGs in prostate cancer progression.

Conclusions: We have identified TFCGs associated with pronounced initial transcriptional response to ADT, aggressive signatures, and metastasis. Our findings suggest multiple new hypotheses that could lead to novel combination therapies to prevent development of CRPC following ADT.

3.1 Introduction

Prostate cancer is one of the most commonly diagnosed cancers and the second leading cause of cancer death in men in the United States [4]. Currently, androgen deprivation therapy (ADT) is one component of care for patients with locally advanced prostate cancer, and advanced or metastatic prostate cancer [35, 211]. Patients with advanced and metastatic prostate cancer will usually respond favorably initially, but will frequently experience disease progression despite therapy [17]. This type of cancer is termed castration-resistant prostate cancer (CRPC), and is typically associated with metastatic disease and poor prognosis, rendering it virtually incurable [212]. However, there is a subset of patients with locally advanced prostate cancer who benefit from ADT, in conjunction with other treatments such as radiation therapy, and experience improved disease-free and overall survival [211]. In these patients, ADT forces changes in tumor biology that result in distinct molecular profiles. Currently, the upstream regulators of that characterize the differential transcriptional programs have not been comprehensively elucidated.

Androgen receptor (AR) stimulation and downstream signaling is critical for the initiation and progression of prostate cancer [213]. Upon androgen ligand activation, AR can function as a transcription factor to regulate target genes. AR-signaling is reestablished in CRPC despite initial inhibition by ADT, due to mutational adaptations of the AR gene, including gene amplification and the expression of alternative AR splice variants (AR-V) [214, 215]. Additionally, other AR-independent signal transduction pathways can be aberrantly activated facilitating crosstalk with, and/or bypass of, the AR-signaling pathway. Stimulation of AR transcriptional activity was found to be largely dependent on the cooperation with specific co-activators [78, 216, 217]. This underscores the importance of identifying key regulators that may supplement, or function independently from, AR transcriptional activity, to promote progression in an androgen deprived environment.

Here we identify novel putative transcription factor coordinated groups (TFCGs) that characterize the differential transcriptional signatures in tumors of patients who received ADT, as well as the progression from localized prostate cancer to metastatic disease. We generated whole transcriptome gene expression data from 20 patient-matched formalin-fixed prostate paraffin embedded (FFPE) needle core biopsies, taken before initiation of neoadjuvant ADT (pre-ADT Bxs), and corresponding FFPE radical prostatectomy samples, acquired after ADT (post-ADT RPs), and leveraged a large dataset (n > 800) of multiple publicly available cohorts of primary and metastatic tumors [218]. We integrated protein-protein interaction, gene expression, and DNA binding data by utilizing the PANDA (Passing Attributes between Networks for Data Assimilation) method [219, 220] to infer condition-specific relationships between transcription factors and putative gene targets. We combined these analyses to uncover groups of putative transcription factor regulators that were unique to patients demonstrating a strong transcriptional response to ADT, and to patients with metastatic prostate cancer. Our analysis leverages multiple datatypes and independent datasets to find common TFCGs that may serve as putative regulators of prostate cancer aggressiveness.

3.2 Materials and methods

Tissue specimens

Patient tissue specimens and associated clinical data were selected from the prostate cancer biobank of the Centre hospitalier de l'Université de Montréal research center (CRCHUM). All patients signed an informed consent to participate in the biobank and the Comité d'éthique à la recherche of the CRCHUM approved the study. We selected patients with matched pre- and post-androgen deprivation therapy specimens of biopsy and RP performed at the CHUM between 1993 and 2012. Following review of hematoxylin/eosin (H&E) -stained slides by a genitourinary pathologist, tumor areas were identified. For RP, the whole case was reviewed to identify the index tumor or, when it could be identified from the biopsy report, the nodule from which the biopsy was taken. The corresponding area on FFPE tumor blocks were extracted using a 0.6 mm diameter tissue arrayer needle (TMArrayer; Pathology Devices, Inc., Westminster, MD, USA) and transferred into a 1.5mL plastic tube prior to extraction.

RNA-Sequencing and differential gene expression analysis of Pre/Post ADT patient samples

Total RNA from 40 matched FFPE specimens from Pre-ADT core biopsies and Post-ADT radical prostatectomies were sequenced using Illumina's TruSeq RNA Access Library Prep kit. Sequence alignment and gene level expression quantifications were obtained using the STAR read aligner [221, 222, 223]. Differential gene expression analysis was performed with the edgeR Bioconductor package in R. The final statistical model included corrections for sequencing batch effects and sequence coverage. Datasets can be accessed in the NCBI GEO and SRA databases (accession no. GSE111177).

Constructing transcriptional networks using PANDA

To infer interactions between transcription factors and gene targets, we used three datatypes as inputs to PANDA: co-transcriptional expression, protein-protein interaction (PPI), and transcription factor binding site (TFBS) motif data, described in detail below [219]. Briefly, we used PANDA to construct a network for each condition (e.g. pre-, post-ADT) using constant PPI and protein-gene datasets, but condition-specific gene expression data. PANDA was run with default parameters (alpha = 0.2, hamming distance = 1e-05) using MATLAB. Transcription factors that were absent in either the PPI or expression data were removed.

Protein-Protein Interaction Data

Binary (direct interactions between two proteins) PPI data was obtained from the human protein reference database (HPRD) (http://www.hprd.org), and OncoPPI [224]. HPRD contains interactions that are derived from experimental evidence from the literature [225], and OncoPPI contains interactions from TR-FRET screening data.

Motif Data

The regions of H3K27Ac and DNaseI hypersensitivity within the human genome (build hg19) were obtained from ENCODE data tracks [226] in the UCSC human genome database[227]. We utilized the MATCH software [228] based on the BioBase Knowledge Library 2017.3 TRANSFAC database [229] to identify all vertebrate TFBS in these regions with a minimum core matrix score greater than 0.95. Motif data were further annotated to retrieve HGNC symbols corresponding to TRANSFAC transcription factor position weight matrix identifiers.

Expression data

Mapped read counts were normalized using DESeq Bioconductor package. Log-2 transformed normalized counts were centered by the median of all samples in the dataset. For

primary and metastatic network construction, we obtained median-centered normalized expression data of publicly available datasets containing more than 800 patient samples that were curated and subtyped by You *et al* [218].

Identification of network-specific "Key" transcription factor coordinated groups and gene target expression analysis

To identify transcriptional relationships that were significantly altered between two conditions, we first compared pairs of networks to identify interactions between transcription factors and gene targets interactions that were unique to one network versus another. Briefly, the program uses a message-passing procedure to estimate the agreement of data types by calculating a similarity score that represents the support that a gene is putatively targeted by a transcription factor. The similarity scores, or edge weights, are z-score normalized to allow for iterative updating of these edge weights across the data types [219]. The PANDA algorithms outputs z-scores that represent the support that a transcription factor targets a gene in a given network. We identified unique interactions by converting z-scores to "unique interaction probabilities" as described in Glass *et al* [220]. Briefly, we used the cumulative distribution function to generate probabilities representing whether an interaction was unique to, and strongly supported in, one network but not in another. We selected transcription factor – target gene interactions that had a probability greater than 90%. Next, we identified transcription factors that were significantly enriched for gene targets in one network versus another by employing the hypergeometric distribution and Bonferroni correction for multiple testing with a critical p-value of 0.05 (Key TFs). To uncover transcription factors that share common gene targets in a network, we performed non-reciprocal pairwise comparisons to determine the percent overlap of predicted high confidence target genes shared between two Key TFs followed by hierarchical clustering of these percentages to reveal putative coordinated groups. We defined a coordinated group as containing Key transcription factors that exclusively share at least 70% of

their targets. Finally, we performed hierarchical clustering (Euclidean dissimilarity metric, complete agglomerative clustering) of shared target gene expression.

To assess whether we would identify a given number of Key transcription factors by chance alone, we permuted the RNA-seq sample identifiers without replacement. We used the permuted RNA-seq data as one of the inputs for PANDA, and performed analyses to identify Key transcription factors. The permuted network did not contain any transcription factor – gene target interactions with a probability greater than 90%, demonstrating the significance of the identified pairs in the actual network.

3.3 Results

Differential expression analysis reveals two distinct transcriptional responses to ADT.

Total RNA was derived from matched pairs of pre-ADT and post-ADT samples from 20 patients. All patients received neoadjuvant ADT, and one patient also underwent radiation therapy. The median ADT duration time was 3 months, with a range from 1 to 8 months (Supplementary Table S3.1). Six of the patients developed metastasis. The samples included in the study were comprised of 20 matched needle core biopsies obtained before ADT, and 20 radical prostatectomies obtained after ADT. On average, we obtained 91M reads per sample, with 64x coverage of the transcriptome.

VARIABLE	VALUE	TOTAL (N = 20)
Age at diagnosis (years)	Median	63
	Mean ± SD	62.0 ± 5.6
	Range	48 – 69
PSA prior to ADT (ng/mL)	Median	11.4
	Mean ± SD	15.7 ± 12.3
	Range	5.0 – 56.7
Neoadjuvant ADT Duration (months)	Median	3
	Mean ± SD	4.6 ± 2.5
	Range	1.0 – 8.0
BCR	Yes	10 (50%)
	No	10 (50%)
Time to BCR (months)	Median	31.5
	Mean ± SD	32.0 ± 22.7
	Range	5.8 – 82.0
Mets	Yes	5 (25%)
	No	15 (75%)
Time to Mets (months)	Median	95
	Mean ± SD	108.0 ± 50.0
	Range	42.0 – 166.0
Total Follow-up (years)	Median	7.8
	Mean ± SD	9.10 ± 5.6
	Range	**1.0 – 20.0**

Supplementary Table S3.1. Patient and treatment characteristics. Clinical metadata for 24 patients analyzed by RNAseq including treatment regimens, and outcome data.

We performed differential expression analysis of RNA-Seq data and identified 190 significantly differentially expressed genes with a fold change greater than or equal to 2 (FDR < 0.05) (Supplementary Table S3.2). To gain initial insights into signaling pathways associated with all post-ADT RPs, we performed Ingenuity Pathway Analysis (IPA) on differentially expressed genes. As expected, we observed an enrichment of downregulated genes that are typically altered in response to agents that promote AR-signaling, such as dihydrotestosterone, and the AR agonist metribilone (R1881) (Supplementary Table S3.3). Additionally, we found that genes inhibited by U0126 (an inhibitor of the MAPK signaling pathway [230]), were downregulated, while genes activated by PDGF, a growth factor that stimulates MAPK signaling, were activated. Furthermore, there was enrichment of upregulated genes within the estrogen signaling pathway (Supplementary Table S3). These data support the expected repression of androgen-driven genes as well as possibly compensatory increases of estrogen and PDGF-MAPK signaling following ADT.

logFC	logCPM	F	PValue	FDR	Gene
9.80944	7.157203	60.61676	6.84E-13	1.05E-08	ENSG00000187812
5.225642	9.748443	48.53966	1.00E-10	7.70E-07	FOS
10.47814	4.835893	44.16953	5.02E-10	2.25E-06	PRAMEF17
6.424846	7.964116	44.94508	5.84E-10	2.25E-06	FOSB
10.02448	5.403963	41.21661	1.49E-09	4.58E-06	MMP8
-4.8433	9.705468	38.78748	1.07E-08	2.40E-05	KLK3
8.269041	6.361206	36.25913	1.09E-08	2.40E-05	TRIM51FP
-4.28486	9.657199	36.56898	1.50E-08	2.73E-05	KLK2
8.616832	3.829499	35.37859	1.60E-08	2.73E-05	CA5A
-4.85896	7.59083	34.16142	2.52E-08	3.87E-05	SLC45A3
-3.76951	8.951771	33.19867	3.80E-08	5.32E-05	ZNF761
7.804497	5.804847	30.99037	1.43E-07	0.000177	TNNT3
7.530255	4.514065	30.10409	1.50E-07	0.000177	RXFP2
7.73852	4.495556	27.0217	5.83E-07	0.000641	ATP13A5
8.268114	3.928617	26.68985	6.92E-07	0.000709	KRT35
8.388406	5.945793	26.95497	8.21E-07	0.000789	CFHR4
7.121862	4.383552	25.59002	1.10E-06	0.000995	BPIFB4
7.959651	4.017315	25.0722	1.43E-06	0.001224	SLC35G4
-4.16793	8.552645	25.58564	1.56E-06	0.00126	KLK4
6.827017	3.313971	24.65237	1.70E-06	0.001304	GPIHBP1
7.647576	3.804001	24.43003	1.87E-06	0.001366	SAG
-5.72891	4.330021	24.15122	2.08E-06	0.001437	TPM3P9
-2.95882	9.327148	24.07645	2.15E-06	0.001437	ABCC4
7.42675	4.241989	23.56315	2.79E-06	0.001788	POM121L3P
8.033789	4.295864	23.50988	2.92E-06	0.001797	ENSG00000213394
-2.54188	11.16429	23.3109	3.05E-06	0.001802	FASN
6.390391	2.598983	23.22295	3.26E-06	0.001855	ASTL
-5.03182	5.704731	22.62019	4.18E-06	0.002297	KCNN2
6.820939	4.509012	22.20695	5.21E-06	0.002765	SPATA31E1
-3.7631	9.063768	23.03303	5.43E-06	0.002783	ACACA
5.863747	6.163371	21.97761	6.34E-06	0.003146	XIRP2
4.706868	6.856035	21.54489	6.94E-06	0.003248	PAX6
5.858213	2.484171	21.57093	6.97E-06	0.003248	FAM71A
6.364695	6.195779	21.21752	9.78E-06	0.004404	MUC5AC
-2.73013	9.003009	20.70563	1.01E-05	0.004404	GREB1
7.223808	6.675112	22.86795	1.04E-05	0.004404	MMP20
4.6026	6.950226	21.07653	1.06E-05	0.004404	MYH2
7.80572	3.190586	20.27079	1.27E-05	0.005129	FGA
-3.13263	9.724069	20.22441	1.33E-05	0.005235	ACPP
-4.08429	7.911739	20.45775	1.37E-05	0.005258	MSMB
7.606458	4.724845	19.83012	1.58E-05	0.005919	MYT1L
6.746006	4.418654	19.65559	1.74E-05	0.006387	COL6A5
-2.61763	9.59765	19.32454	1.93E-05	0.006747	MLPH
6.597378	5.008853	19.47227	1.97E-05	0.006747	OPTC

-3.01761	8.64806	19.23548	2.02E-05	0.006747	ANO7
6.588437	4.475063	19.30993	2.02E-05	0.006747	KRT17P3
7.985559	5.231752	19.84802	2.11E-05	0.006908	ABCG8
7.329483	5.027821	19.42221	2.24E-05	0.007134	GOLGA8VP
6.776473	4.088572	19.01427	2.27E-05	0.007134	LBP
-4.68505	6.784203	18.622	2.87E-05	0.008451	TRGC1
5.580504	5.667862	18.68222	2.88E-05	0.008451	TRPA1
-4.92179	6.32399	18.48947	2.88E-05	0.008451	TRGC2
6.465318	4.695528	18.45125	2.94E-05	0.008451	C12orf40
6.438552	4.229942	18.45233	3.04E-05	0.008451	NR2E1
5.895529	3.462676	18.35283	3.08E-05	0.008451	COL20A1
7.183604	3.553876	18.40295	3.08E-05	0.008451	SHCBP1L
3.970612	9.575231	19.71377	3.17E-05	0.008561	EGR1
6.785466	6.450785	18.84798	3.65E-05	0.009404	DMRT1
5.270739	5.559578	18.02414	3.65E-05	0.009404	DNAI2
5.47394	4.72962	17.97566	3.67E-05	0.009404	OLFM4
-4.4244	6.011291	17.92096	3.76E-05	0.009467	STEAP4
-5.21462	0.725052	17.7384	4.14E-05	0.01027	ENSG00000213896
6.134218	4.327912	17.5662	4.49E-05	0.010968	CHRNA1
5.968563	4.39408	17.42944	4.76E-05	0.011432	TNR
-2.89664	8.684498	17.31126	5.02E-05	0.011879	NCAPD3
5.309175	3.888687	17.19697	5.32E-05	0.012387	ERN2
5.655449	6.831706	19.33516	5.68E-05	0.01303	ENSG00000277327
6.808717	4.151588	17.16154	5.90E-05	0.013283	TRIM51
-5.85034	1.910352	16.97905	5.96E-05	0.013283	TMSB15A
-3.10182	8.353989	17.20675	6.06E-05	0.013314	TMPRSS2
5.819206	2.119233	16.90333	6.24E-05	0.013393	HHLA1
3.052343	7.713849	16.83288	6.31E-05	0.013393	NR4A1
6.796511	3.73185	17.07251	6.51E-05	0.013393	SLC34A1
5.036915	7.209617	17.88091	6.53E-05	0.013393	CHRNA4
-2.68839	8.947872	16.76156	6.53E-05	0.013393	TRPM8
5.325871	3.680291	16.71421	6.70E-05	0.01353	CLCA2
5.093119	5.625384	16.78197	6.77E-05	0.01353	NR4A3
5.561204	3.873633	16.59487	7.14E-05	0.014085	SYCP1
-2.32818	11.72986	16.7393	7.33E-05	0.014268	SNORA73B
6.389367	5.146819	17.08134	7.55E-05	0.014521	GRM1
-2.87217	8.392086	16.40052	7.77E-05	0.014749	PDLIM5
6.466649	3.924596	16.32099	8.23E-05	0.015379	GPM6A
-5.32356	3.021481	16.26494	8.35E-05	0.015379	NBEAP1
6.754247	4.328743	16.42534	8.56E-05	0.015379	CDH8
-5.94292	1.906225	16.21313	8.60E-05	0.015379	ENSG00000213777
-3.29608	7.09355	16.17536	8.66E-05	0.015379	AZGP1
6.209054	5.299752	16.82401	8.78E-05	0.015379	DUX4L17
5.908223	4.242862	16.17548	8.80E-05	0.015379	CHRND
5.83466	3.110496	16.12765	9.05E-05	0.015635	SERPINB10

4.897723	4.798855	15.98842	9.50E-05	0.016226	SLC5A2
4.91132	6.206395	15.86579	0.000106	0.017761	ASAH2
-2.64992	9.787211	16.02952	0.000106	0.017761	FOXA1
-2.89224	8.803956	16.06477	0.000107	0.017772	GRHL2
-4.35403	4.686995	15.59936	0.000114	0.01871	ST6GALNAC1
3.914344	6.154782	15.49124	0.000121	0.019509	SERPINE1
4.169419	6.310503	15.27837	0.000134	0.021406	ATF3
5.215406	5.377635	15.2827	0.000137	0.021776	KIAA2012
-2.93881	7.860062	15.20065	0.000139	0.021776	ELK4
5.520238	5.692675	15.22056	0.000141	0.021869	PRDM7
-2.71907	7.73823	15.13736	0.000143	0.021869	DHCR24
5.342771	5.412488	15.27894	0.000144	0.021869	OTOA
6.410916	4.381069	15.14623	0.000147	0.022102	FCRL4
5.83721	5.674748	15.27092	0.000151	0.022573	PRR29
6.505485	5.085725	15.56463	0.000156	0.022805	SLC2A2
5.607438	4.489859	15.02398	0.000156	0.022805	C8A
4.933518	6.089877	15.08008	0.00016	0.023249	ENSG00000280194
-2.33471	9.030534	14.75056	0.000173	0.024831	KIAA1244
5.845233	4.151665	14.75842	0.000181	0.025704	ACSM2A
5.22382	5.466074	14.81881	0.000188	0.026346	CCDC155
6.053148	4.307704	14.73055	0.000189	0.026346	TRIM51CP
5.243133	4.720861	14.65411	0.00019	0.026346	CYP24A1
-2.87093	8.023075	14.42155	0.000203	0.027702	SRSF6
6.169657	5.777518	14.74071	0.000204	0.027702	HEPHL1
5.064289	3.528666	14.38351	0.000208	0.027702	SPHKAP
4.767222	6.091078	14.41455	0.000211	0.027702	MYPN
6.475271	3.646669	14.55316	0.000212	0.027702	NLRP5
6.726437	5.709496	15.25378	0.000213	0.027702	PDE6C
3.472242	7.205691	14.32657	0.000213	0.027702	NR4A2
7.26853	8.906266	19.56058	0.000219	0.028283	GOLGA6L6
6.089405	3.451606	14.50252	0.000222	0.028492	TFAP2D
-2.15664	15.46967	15.70492	0.000226	0.028653	MALAT1
4.509536	6.520355	14.70609	0.000227	0.028653	DMBT1
3.164909	7.797766	14.25901	0.00023	0.0288	COL17A1
5.316798	2.964249	14.03375	0.000246	0.030334	ENSG00000270442
5.887298	4.772946	14.38569	0.000248	0.030334	GIF
4.89735	3.825236	14.02226	0.000249	0.030334	POU6F2
5.699416	3.717087	13.95483	0.000257	0.031107	PADI4
-4.19568	4.268364	13.91469	0.00026	0.031162	RAB3B
4.628314	2.465099	13.91939	0.000261	0.031162	PCP4L1
5.121115	5.02302	13.96193	0.000265	0.03137	TMPRSS15
5.095346	4.773446	13.87084	0.000267	0.03137	ALOX12B
-3.47069	6.922869	13.81175	0.000274	0.031504	SORD
4.54143	5.478813	13.82327	0.000274	0.031504	ENSG00000262533
5.678622	4.285569	13.97015	0.000275	0.031504	KCNT1

-2.6013	8.035879	13.76127	0.000281	0.031961	CANT1
-4.92731	1.118655	13.75473	0.000283	0.031991	ENSG00000251648
-4.88546	7.010979	14.51691	0.000296	0.033218	NPY
-3.15754	7.094782	13.61944	0.000301	0.033529	ARHGEF38
4.0459	7.098254	14.15855	0.000305	0.033536	RAG1
5.192566	3.283612	13.60811	0.000307	0.033536	PRKAG3
-3.50856	6.104544	13.57872	0.000307	0.033536	PPM1H
-2.77831	10.34163	14.54125	0.000312	0.033578	NEFH
-2.30594	9.900136	13.68389	0.000314	0.033578	ZBTB16
-2.48201	8.890341	13.5706	0.000314	0.033578	ERGIC1
4.534035	3.166535	13.50513	0.000319	0.033603	DRD2
3.943552	3.743223	13.50118	0.000319	0.033603	LGR6
4.698854	5.557062	13.50885	0.000322	0.033603	KRT16
-2.831	7.425252	13.47315	0.000323	0.033603	ZNF841
4.759802	3.407541	13.4772	0.000327	0.033706	ENSG00000166104
-4.94752	3.445879	13.42586	0.000332	0.03391	AZGP1P1
3.114183	7.101847	13.41405	0.000333	0.03391	MUC3A
-2.81936	7.950859	13.37183	0.00034	0.034264	PPP3CA
4.836637	2.676162	13.36958	0.000341	0.034264	MIR8078
5.767515	4.798619	13.43881	0.000343	0.034264	CPB2
-4.52466	1.699914	13.33005	0.000349	0.034603	RAB6C
6.165667	5.816233	14.05188	0.000365	0.036022	PAX2
5.151501	4.371465	13.20591	0.00037	0.036046	ZBBX
6.337041	4.397023	13.43156	0.00037	0.036046	SPDYE6
-3.47405	6.143933	13.17711	0.000374	0.036208	ZNF577
5.231231	2.982827	13.18383	0.000377	0.036271	AFP
4.92431	6.568708	13.71465	0.000386	0.036838	RP1
-3.05021	7.358717	13.06085	0.000397	0.037644	ANKRD10
-4.98812	3.009917	13.01242	0.000406	0.038285	FAR2P4
5.743347	5.966936	13.84361	0.000408	0.038285	ERICH6B
-1.86501	10.93026	12.94554	0.00042	0.038924	SRRM2
5.032557	5.643462	13.04254	0.00042	0.038924	SPATA31D5P
6.187805	2.980251	13.14429	0.000436	0.040122	PNLIPRP2
-2.82507	7.136792	12.79746	0.000452	0.041369	ACSL3
4.176364	5.328126	12.7801	0.000461	0.041924	MPO
-3.11983	6.796467	12.74638	0.000464	0.04197	BMPR1B
4.49679	3.960077	12.73781	0.000468	0.04197	PRPH
-5.14987	2.073573	12.73625	0.000469	0.04197	NPY4R
-2.16942	9.437424	12.68283	0.000478	0.042529	NDRG1
5.749017	6.905652	14.08201	0.000485	0.042841	AREG
-4.92712	1.883295	12.64973	0.000489	0.042841	BHLHA15
5.045605	3.966348	12.72336	0.00049	0.042841	GPR123
5.643236	6.576221	13.59543	0.000518	0.044993	MYH4
5.568753	7.364999	14.0172	0.000535	0.046049	PCDH15
4.50161	4.852851	12.46335	0.000536	0.046049	SCEL

5.532265	4.407161	12.5443	0.000541	0.04622	KCNIP4
-2.55216	8.970955	12.45474	0.000548	0.046523	P4HB
4.251554	3.546035	12.36393	0.000562	0.047522	TTBK1
3.432127	7.048095	12.35895	0.000569	0.047556	CPS1
3.946608	5.564466	12.33361	0.000569	0.047556	PZP
-6.20839	3.168502	12.33125	0.000572	0.047556	CYP4F30P
-2.85625	6.97284	12.31332	0.000575	0.047556	PMEPA1
5.803026	3.657818	12.29968	0.000584	0.047905	CCDC37
-2.68464	9.430791	12.44553	0.000586	0.047905	POTEE
4.765913	5.791708	12.5183	0.0006	0.048828	MUC19
4.802487	4.489191	12.19718	0.000612	0.049518	HTR3A
-2.54288	8.648817	12.51931	0.000615	0.049518	MCCC2

Supplementary Table S3.2. Significantly differentially expressed genes identified using edgeR analysis of Pre-ADT Bxs and Post-ADT RPs.

Upstream Regulator	Molecule Type	Predicted Activation State	Activation z-Score	p-value of Overlap
Metribolone (R1881)	chemical reagent	Inhibited	-4.223	4.18E-20
Dihydrotestosterone (androgen)	chemical - endogenous mammalian	Inhibited	-2.506	1.09E-09
Bicalutamide	chemical drug		1.709	8.13E-07
beta-estradiol (estrogen)	chemical - endogenous mammalian	Activated	2.493	2.69E-04
U0126	chemical - kinase inhibitor	Inhibited	-2.798	2.87E-05
PDGF-BB	complex	Activated	2.999	7.92E-06

Supplementary Table S3.3. Ingenuity Pathway Analysis identifies chemical agents associated with signaling pathways. Predicted upstream chemical agent regulators suggest inhibition of androgen driven genes, and increase in estrogen and PDGF-MAPK signaling.

Hierarchical clustering yielded three groups of samples based on the expression of two clusters of upregulated and downregulated genes that defined the pre- and post-ADT conditions (Figure 3.1A). There was a common decrease in the downregulated genes among all but one of the post-ADT RP samples, though the degree of repression was strikingly more pronounced in one group. One RP sample clustered closely with the pre-ADT samples mainly due to relatively higher expression of downregulated, discriminating it from the rest of the post-ADT RP samples. The upregulated genes further segregated the post-ADT samples based on increased expression in one subgroup (high impact group) but relatively unchanged expression after ADT in the other subgroup (low impact group).

Principal component analysis (PCA) of differentially expressed genes confirmed the observations of the hierarchical clustering. PCA revealed three distinct groups of samples, the high impact group, low impact group and pre-ADT Bx group. Three outlier RP samples did not cluster with any group, and one RP sample closely segregated with the pre-ADT Bx (Figure 3.1B).

Although there was no significant difference in ADT exposure time in the high and low impact group, we observed striking differences in transcriptional signatures after ADT. For example, *KLK3* (Prostate Specific Antigen or PSA) gene expression was more significantly decreased in the high impact group than in the low impact group after ADT ($p < 2.14 \times 10^{-3}$ Mann-Whitney test), concordant with the IPA results suggesting an decrease in androgen driven genes (Figure 3.1C). To biologically characterize the transcriptional changes specific to the low and high impact groups, we employed a subtyping method developed by You *et al* [218] that is associated with good and poor outcomes. Briefly, these subtypes utilize gene signatures to segregate patients into three groups, prostate cancer subtype 1, 2 or 3 (PCS1-3). The PCS1 subtype is enriched for genes involved in androgen receptor variant (AR-V) (ligand-independent, constitutively active) activation, and is associated with a poor prognosis. The PCS2 subtype is

enriched for genes indicative of AR activation, and has a variable prognosis. PCS3 subtype is characterized by low activation of AR or AR-V associated genes (AR-suppressed), and has a variable prognosis.

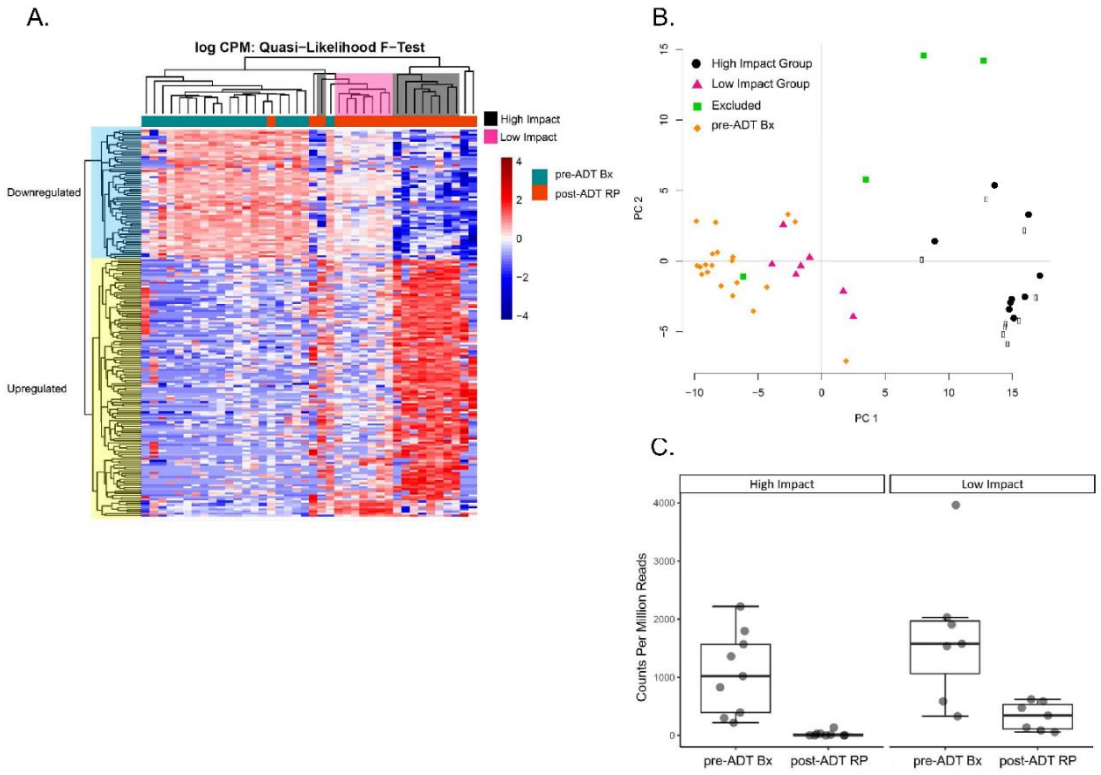
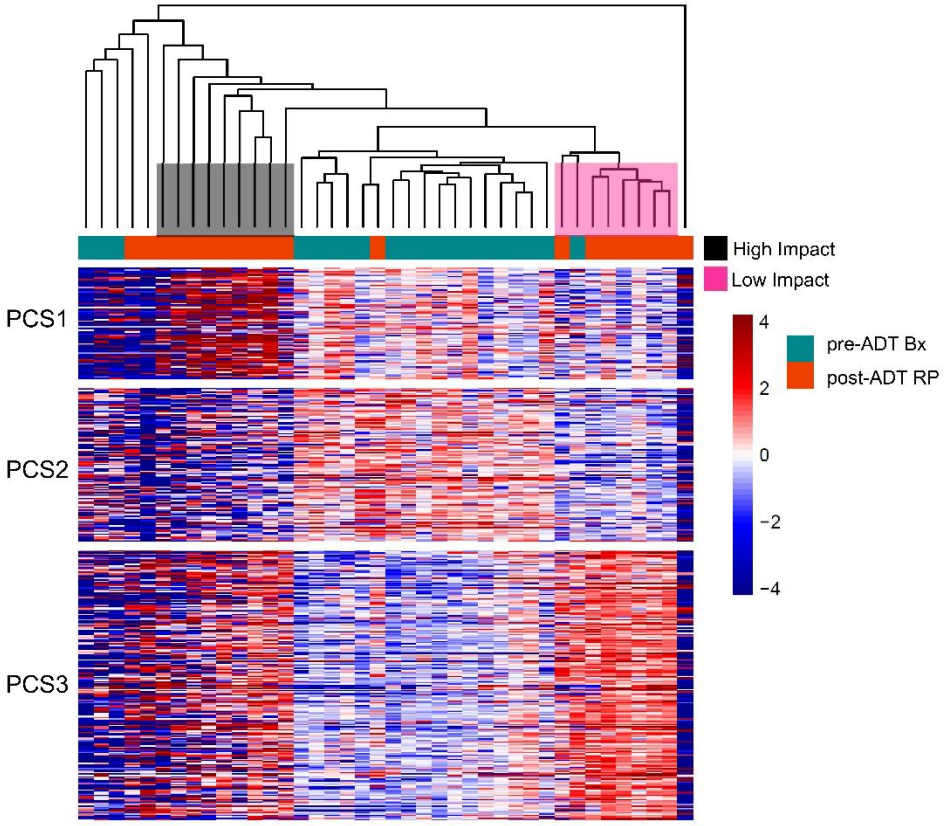


Figure 3.1. Hierarchical clustering and PCA of 190 significantly differentially expressed genes in 20 matched Pre-ADT Bxs and Post-ADT RPs. A) Clustering reveals two groups of Post-ADT RPs displaying segregated based on expression of upregulated and downregulated genes (high- or low-impact groups, respectively). B) PCA reveals 4 post-ADT RP samples as not clustering with either the high or low impact groups. C) Boxplot depicting *KLK3* expression in counts per million mapped reads demonstrates that the decrease in KLK3 expression is significantly more pronounced in the high impact group than the low impact group.

To classify the subtype makeup of the Pre-ADT Bx and Post-ADT RP samples, we performed hierarchical clustering on median-centered log-normalized counts of PCS subtype signature genes only. Interestingly, we found that clustering of the signature genes also segregated the high and low impact groups (Supplemental Figure 3.1), consistent with the clustering of differentially expressed genes. We evaluated whether the expression of subtype genes before ADT in the low impact and high impact groups would inform the differences in subtype signature observed after ADT (Figure 3.2). Specifically, we compared the fraction of subtype genes that were greater than one-fold above the median expression of all samples. We observed a common decrease in the, AR-dependent, PCS2 gene expression in both groups, and a similar increase in PCS3 gene expression, but there was a striking divergence in the expression of the aggressive PCS1 signature (Figure 3.2A). We found that after ADT treatment, the transcriptional signature indicative of the more aggressive, ADT-resistant, androgen receptor-independent subtype (PCS1) was not only retained, but also significantly increased, in the high impact group ($p < 4.66 \times 10^{-3}$ Mann-Whitney test). On the other hand, the low impact group exhibited a relative loss of the PCS1 signature, and only the proportion of genes characteristic of the AR-suppressed PCS3 drastically increased (Figure 3.2B). Taken together, these data suggest a relative shift in subtype makeup that correlates with the differential intensity in transcriptional reaction to ADT.



Supplemental Figure 3.1. Hierarchical clustering of PCS genes of 20 matched pre-ADT Bxs and post-ADT RPs again segregates three groups. Hierarchical clustering of median-centered log-normalized counts reveals two distinct groups. High impact group (black) predominantly express PCS1 and PCS3 subtype genes. Low impact group (magenta) predominantly express PCS3 genes.

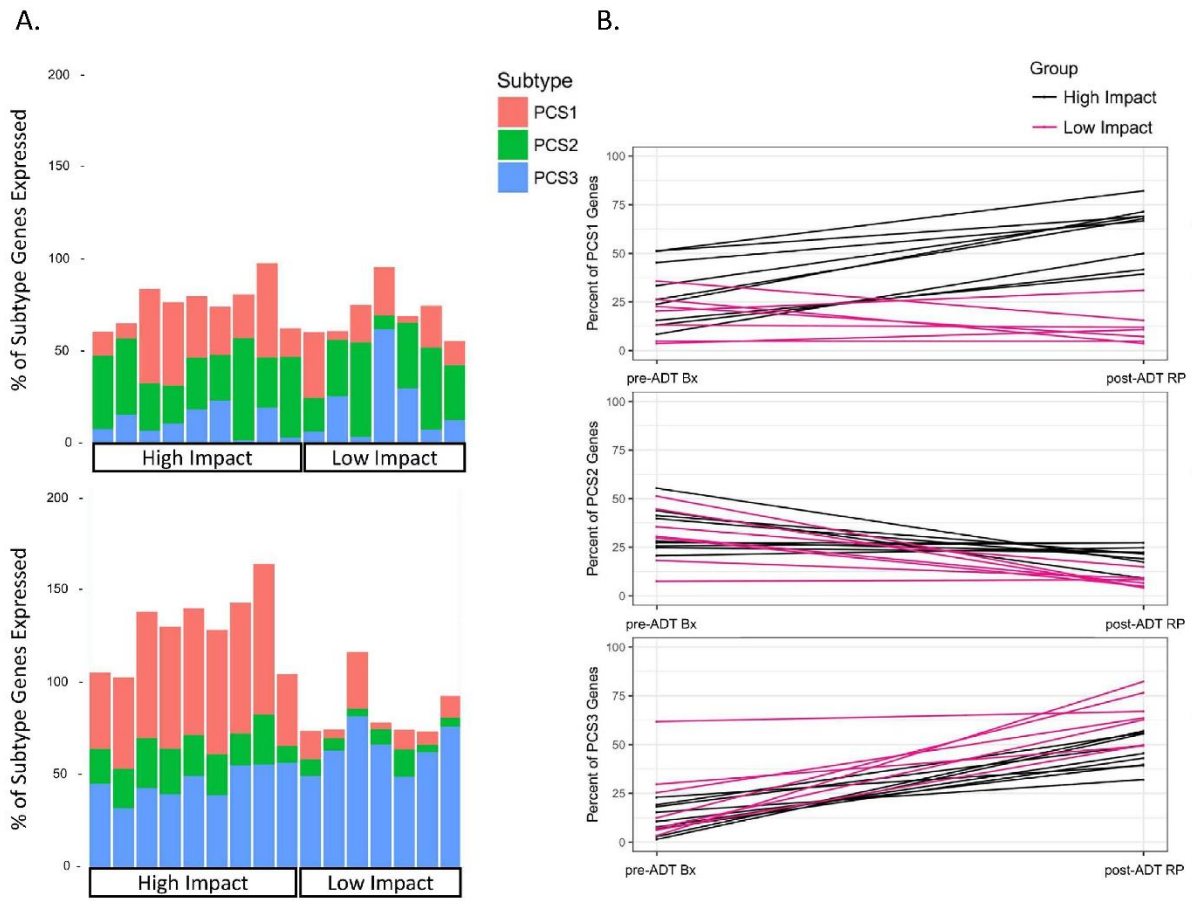


Figure 3.2. Divergent expression of PCS1 genes in the high impact group and common loss of PCS2 and PCS3 after ADT. A) Bar plots show the fraction of subtype genes expressed more than one-fold above the median across all samples. Both the high and low impact groups lose expression of PCS2 genes after ADT, but the high impact group samples display a retention and increase in PCS1 signature after ADT, while the low impact group loses PCS1 signature but displays increase in PCS2 gene expression. B) Plots depict the percent change of subtype genes that are expressed before and after ADT. The PCS1 gene signature is significantly increased in the high impact group after ADT.

Identifying transcription factors (TFs) enriched for unique targets in the high impact network.

To elucidate putative transcriptional regulators associated with the transcriptional changes unique to the high impact group, we ascertained regulatory networks. To accomplish this, we utilized the PANDA algorithm [219] that integrates RNA expression data, protein-protein interaction data and DNA binding motif data to reverse engineer transcriptional networks. We used PANDA to integrate protein-protein interaction data from the human protein reference database and experimentally validated interactions among known cancer associated drivers and tumor suppressors [224, 225], DNA binding motif data found within H3K27ac and DNase1 hypersensitivity regions [228], and RNA-Seq data from 16 Post-ADT RPs. To find TFs that putatively regulated the differential transcriptional response to ADT, as opposed to simply before and after ADT, we focused our analysis on the differences between the high impact and low impact networks. We converted PANDA z-score normalized edge weights to unique interaction probabilities that estimate how likely a transcription factor regulates a given gene, and identified transcription factors (Key TFs) that were enriched for gene targets in the high impact network as compared to the low impact network using the cumulative distribution function [219]. We identified 394 out of 725 Key TFs that were significantly enriched for unique targets in the high impact network, as compared to that of the low impact network, using the hypergeometric distribution and Bonferroni correction for multiple testing with a critical p-value of 0.05 ($p < 7.85 \times 10^{-5}$), as performed by Glass *et al* [220] (Supplementary Table S3.4). Notably, AR, and multiple transcription factors extensively reported to be involved in prostate cancer aggressiveness, such as ETV5 and ETV1 [38] were identified as Key TFs, supporting the biological relevance of our analysis.

	Low Impact	High Impact	TFEdgeTotal	p-val
AHR	30	848	878	3.50E-138
AP3D1	120	1362	1482	5.95E-161
AR	37	1132	1169	3.32E-187
ARHGEF7	40	824	864	1.12E-123
ARNT	36	931	967	2.12E-148
ARNT2	20	579	599	2.47E-95
ARNTL	2	186	188	6.16E-37
ASCL2	2	53	55	1.20E-09
ATF1	18	666	684	5.72E-115
ATF2	42	1104	1146	2.16E-176
ATF3	24	761	785	2.71E-127
ATF4	34	921	955	1.80E-148
ATF6	14	312	326	7.87E-49
ATM	33	1087	1120	1.33E-182
ATOH1	6	89	95	3.56E-13
BANP	31	676	707	1.97E-103
BATF	17	384	401	5.54E-60
BCL6	74	1076	1150	1.27E-142
BCL6B	55	1051	1106	1.03E-153
BDP1	6	406	412	6.58E-77
BHLHE40	27	690	717	4.65E-110
BRCA1	65	1189	1254	3.06E-171
CABLES2	28	920	948	1.07E-154
CBFB	14	448	462	1.11E-75
CD40	40	824	864	1.12E-123
CEBPA	70	1146	1216	7.27E-159
CEBPB	72	1199	1271	3.64E-167
CEBPD	67	1107	1174	5.03E-154

CEBPE	66	1074	1140	1.17E-148
CEBPG	66	1053	1119	1.01E-144
CEBPZ	43	1139	1182	2.82E-182
CHURC1	54	971	1025	2.44E-139
CLOCK	6	198	204	2.10E-34
CNOT3	15	467	482	2.19E-78
CREB1	19	659	678	2.05E-112
CREM	20	619	639	2.69E-103
CRTC2	1	42	43	1.94E-08
CTBP1	2	101	103	2.44E-19
CTCF	11	274	285	2.65E-44
CUX1	45	1198	1243	7.14E-192
DBP	70	1178	1248	6.97E-165
DDIT3	28	739	767	1.07E-118
DEAF1	20	430	450	4.39E-66
E2F1	79	1141	1220	9.97E-151
E2F2	60	1087	1147	3.48E-156
E2F3	76	1054	1130	4.87E-137
E2F4	76	1038	1114	4.12E-134
E2F5	61	1086	1147	3.94E-155
E2F6	59	1231	1290	1.10E-184
E2F7	61	1074	1135	7.42E-153
EBF1	3	60	63	4.10E-10
EFNA2	88	736	824	1.72E-73
EGR1	41	964	1005	6.61E-150
EGR2	51	928	979	9.11E-134
EGR3	42	891	933	9.19E-135
EGR4	7	294	301	1.24E-52
ELF1	62	1030	1092	1.05E-143

ELF2	16	719	735	4.11E-128
ELF4	121	1085	1206	1.59E-112
ELF5	58	1031	1089	2.80E-147
ELK1	78	1013	1091	4.48E-128
ELK3	44	1018	1062	1.56E-157
ELK4	42	1009	1051	1.00E-157
EP300	91	1381	1472	1.01E-185
EPAS1	11	579	590	7.87E-106
ERF	102	1090	1192	3.67E-125
ERG	58	1072	1130	4.50E-155
ESR1	25	987	1012	1.20E-171
ESR2	14	444	458	7.02E-75
ESRRA	3	89	92	8.35E-16
ESRRB	2	60	62	4.90E-11
ETS1	76	1127	1203	1.75E-150
ETS2	60	1101	1161	7.55E-159
ETV1	36	1012	1048	2.09E-164
ETV3	198	1096	1294	9.99E-77
ETV4	68	978	1046	3.30E-129
ETV5	67	1012	1079	2.80E-136
ETV6	42	1046	1088	5.59E-165
ETV7	36	1038	1074	1.45E-169
FERD3L	22	662	684	1.04E-109
FEV	25	909	934	6.96E-156
FGF9	10	279	289	2.68E-46
FLI1	59	1072	1131	3.35E-154
FOS	67	1170	1237	6.58E-166
FOSB	58	1087	1145	6.20E-

				158
FOSL1	58	1096	1154	1.18E-159
FOSL2	60	1096	1156	6.75E-158
FOXA1	65	1301	1366	1.13E-192
FOXA2	250	1441	1691	1.43E-104
FOXD3	131	1390	1521	1.53E-158
FOXG1	409	1557	1966	9.80E-66
FOXJ3	459	1515	1974	5.69E-49
FOXL1	620	1577	2197	1.79E-25
FOXM1	66	1225	1291	3.18E-177
FOXN1	5	177	182	2.74E-31
FOXO1	485	1582	2067	8.50E-50
FOXO4	491	1537	2028	3.13E-44
FOXP3	69	1321	1390	6.53E-193
GABPA	58	1034	1092	7.56E-148
GATA1	78	989	1067	9.82E-124
GATA2	62	1121	1183	6.50E-161
GATA3	65	1088	1153	4.07E-152
GATA4	165	1437	1602	1.02E-145
GATA5	74	1281	1355	5.97E-181
GATA6	124	1365	1489	8.71E-159
GC	32	784	816	1.59E-123
GCM1	6	106	112	2.29E-16
GCM2	5	144	149	1.09E-24
GDNF	26	386	412	1.29E-52
GEN1	79	992	1071	1.49E-123
GLI1	23	508	531	2.18E-78
GLI2	17	351	368	1.36E-53
GLI3	17	331	348	9.50E-50
GLIS2	55	872	927	7.61E-120
GMEB1	74	965	1039	2.62E-122

GSC	64	759	823	2.07E-92
GTF2A1	18	521	539	6.28E-86
GTF2A2	18	519	537	1.57E-85
GTF2I	60	1050	1110	3.64E-149
GTF2IRD1	42	1049	1091	1.44E-165
GTF3C2	28	586	614	8.13E-89
HAND1	33	921	954	1.65E-149
HBP1	507	1419	1926	6.22E-31
HDAC1	37	667	704	2.70E-96
HES1	38	977	1015	1.97E-155
HIC1	61	1049	1110	3.98E-148
HIC2	58	993	1051	4.35E-140
HIF1A	42	1057	1099	3.86E-167
HINFP	3	267	270	2.39E-52
HIVEP2	5	117	122	2.27E-19
HMGA1	54	1271	1325	3.27E-197
HNF1A	69	1268	1337	8.61E-183
HNF4A	28	834	862	1.61E-137
HNF4G	17	381	398	2.12E-59
HOXA10	118	1390	1508	2.69E-167
HOXA3	47	1086	1133	6.80E-168
HOXB2	898	1670	2568	3.57E-05
HSF1	69	1281	1350	2.87E-185
HSF2	103	1327	1430	1.16E-166
HSF4	65	1036	1101	2.33E-142
ID4	2	173	175	3.15E-34
IKZF1	44	958	1002	7.29E-146
IL10	103	270	373	3.94E-06
ING4	10	205	215	8.00E-32
IRF3	73	1042	1115	3.94E-137
IRF4	52	1215	1267	3.21E-188

IRF5	484	1393	1877	8.83E-33
IRF8	770	1627	2397	7.13E-12
JUN	63	1300	1363	2.60E-194
JUNB	58	1154	1212	9.11E-171
JUND	62	1149	1211	3.04E-166
KCNH8	9	401	410	5.71E-72
KLF1	1	84	85	4.22E-17
KLF12	53	1031	1084	1.14E-151
KLF13	46	965	1011	2.38E-145
KLF15	4	101	105	3.45E-17
KLF16	5	155	160	7.03E-27
KLF2	13	387	400	1.36E-64
KLF3	5	198	203	1.58E-35
KLF4	78	1005	1083	1.26E-126
KLF5	9	287	296	6.01E-49
KLF6	85	995	1080	7.36E-120
KLF7	11	420	431	1.95E-73
KLF8	1	52	53	1.76E-10
LEF1	74	1272	1346	3.00E-179
LTF	34	849	883	2.75E-134
MAF	19	593	612	3.41E-99
MAFA	64	1009	1073	3.97E-138
MAFB	68	1031	1099	5.18E-139
MAFF	22	617	639	9.01E-101
MAFG	23	714	737	5.36E-119
MAFK	42	1026	1068	4.69E-161
MAX	30	885	915	1.57E-145
MAZ	45	813	858	4.12E-117
MECOM	137	684	821	2.24E-43
MEIS1	51	1140	1191	1.44E-174
MEIS2	50	1080	1130	7.23E-164

MEIS3	50	1099	1149	1.49E-167
MITF	24	705	729	3.96E-116
MNT	11	704	715	1.44E-131
MTF1	49	939	988	1.26E-137
MXD1	18	533	551	2.59E-88
MXI1	18	511	529	6.05E-84
MYB	67	1065	1132	3.75E-146
MYBL1	239	1073	1312	1.12E-58
MYBL2	65	1045	1110	4.82E-144
MYC	22	735	757	2.67E-124
MYCN	18	513	531	2.43E-84
MYF5	42	908	950	4.71E-138
MYF6	73	977	1050	3.15E-125
MYOD1	44	988	1032	1.09E-151
MYOG	44	932	976	7.89E-141
MZF1	60	996	1056	5.62E-139
NEUROD1	28	422	450	1.03E-57
NF1	27	917	944	3.25E-155
NF1A	75	1009	1084	1.44E-129
NF1B	14	282	296	4.95E-43
NFAT5	47	1133	1180	4.44E-177
NFATC1	49	1138	1187	4.21E-176
NFATC2	48	1157	1205	8.57E-181
NFATC3	46	1109	1155	2.30E-173
NFATC4	47	1085	1132	1.07E-167
NFE2	28	837	865	4.06E-138
NFE2L1	53	1118	1171	2.03E-168
NFE2L2	29	935	964	1.36E-156

NFIX	75	994	1069	7.76E-127
NFKB1	42	945	987	3.08E-145
NFYA	43	1204	1247	3.92E-195
NFYB	42	1207	1249	8.81E-197
NFYC	41	1161	1202	1.05E-188
NHLH1	18	547	565	4.23E-91
NKX2-1	78	1037	1115	1.96E-132
NKX2-5	42	1188	1230	5.21E-193
NKX3-2	52	1133	1185	2.93E-172
NR0B1	5	262	267	1.42E-48
NR1H2	17	583	600	2.11E-99
NR1H3	17	564	581	1.37E-95
NR1H4	24	811	835	2.62E-137
NR1I2	17	575	592	8.51E-98
NR1I3	17	586	603	5.26E-100
NR2C1	10	206	216	5.13E-32
NR2C2	57	958	1015	2.50E-134
NR2F1	33	971	1004	1.95E-159
NR2F2	43	1007	1050	2.34E-156
NR2F6	16	419	435	8.10E-68
NR3C1	77	1124	1201	3.80E-149
NR4A1	9	177	186	2.47E-27
NR4A2	28	907	935	4.29E-152
NR5A1	41	801	842	2.40E-118
NR5A2	32	540	572	1.60E-76
NR6A1	8	81	89	2.33E-10
NRF1	21	295	316	1.07E-39
NRL	40	922	962	1.11E-142
OSR1	35	1019	1054	7.52E-167
PARP1	893	1653	2546	6.08E-05
PAX2	52	1074	1126	7.59E-

				161
PAX3	18	522	540	3.97E-86
PAX4	76	1010	1086	5.26E-129
PAX5	4	50	54	1.80E-07
PAX6	4	126	130	3.80E-22
PAX8	40	1071	1111	6.28E-172
PBX1	756	1586	2342	3.59E-11
PDX1	647	1629	2276	2.36E-25
PGR	41	1001	1042	3.81E-157
PITX2	186	1251	1437	6.83E-105
PLAG1	3	133	136	1.12E-24
PLAGL1	26	618	644	6.22E-97
PLAGL2	4	365	369	2.78E-71
PLAU	26	386	412	1.29E-52
POU1F1	140	1501	1641	4.99E-172
POU2F1	59	1272	1331	1.28E-192
POU3F3	74	1362	1436	2.51E-196
POU5F1	51	1300	1351	7.10E-206
PRDM16	28	413	441	4.80E-56
PRRX2	623	1585	2208	1.29E-25
PURA	81	1003	1084	4.23E-124
RARA	18	678	696	2.16E-117
RARB	17	569	586	1.36E-96
RARG	17	577	594	3.38E-98
RBPJ	18	320	338	8.83E-47
REL	7	251	258	6.13E-44
RELA	13	518	531	6.56E-91
REST	2	47	49	1.82E-08
REXO1	5	389	394	8.39E-75
RFX1	65	1069	1134	1.52E-148
RFX2	50	1085	1135	7.77E-165
RFX3	50	1113	1163	2.85E-170
RFX5	46	1089	1135	1.88E-169
RORB	5	250	255	4.08E-46

RREB1	12	410	422	2.80E-70
RUNX1	8	294	302	1.82E-51
RUNX2	37	1119	1156	1.30E-184
RUNX3	36	1000	1036	5.00E-162
RXRA	63	1181	1244	1.81E-171
RXRB	17	623	640	1.88E-107
SALL1	42	1041	1083	5.36E-164
SALL2	0	58	58	4.37E-13
SIN3A	16	453	469	1.54E-74
SLC6A2	9	401	410	5.71E-72
SMAD1	88	1070	1158	3.77E-131
SMAD2	68	1161	1229	2.33E-163
SMAD3	59	1224	1283	2.46E-183
SMAD4	69	1152	1221	8.00E-161
SMAD5	70	1008	1078	3.46E-133
SMAD6	57	1012	1069	1.56E-144
SMAD7	57	1028	1085	1.43E-147
SMAD9	55	952	1007	6.98E-135
SNAI1	0	64	64	2.30E-14
SNAI2	19	491	510	5.76E-79
SOX10	70	1184	1254	5.16E-166
SOX17	52	1182	1234	8.76E-182
SOX18	71	1143	1214	1.77E-157
SOX2	427	1483	1910	3.88E-53
SOX3	245	1029	1274	2.44E-51
SOX4	232	1403	1635	9.64E-107
SOX5	60	1185	1245	6.52E-175
SOX6	376	1459	1835	1.06E-63
SOX8	62	1314	1376	6.06E-198
SOX9	72	1196	1268	1.34E-

				166
SP1	72	1282	1354	7.52E-183
SP100	425	1162	1587	5.06E-24
SP2	66	1035	1101	2.32E-141
SP3	68	1077	1145	1.40E-147
SP4	67	1060	1127	3.23E-145
SPDEF	50	1115	1165	1.16E-170
SPI1	73	1166	1239	3.79E-160
SPIB	79	1022	1101	5.66E-129
SREBF1	71	1074	1145	1.30E-144
SREBF2	81	997	1078	5.06E-123
SRF	4	354	358	5.46E-69
SRY	106	1347	1453	4.27E-168
STAT1	54	1222	1276	1.18E-187
STAT2	8	244	252	1.93E-41
STAT3	36	1165	1201	6.76E-195
STAT4	87	1245	1332	1.06E-163
STAT5A	14	467	481	1.75E-79
STAT5B	10	279	289	2.68E-46
STAT6	22	925	947	1.13E-162
TAF1	13	685	698	5.69E-125
TAL1	40	1016	1056	4.35E-161
TBP	300	1587	1887	4.87E-105
TBX2	2	127	129	1.10E-24
TBX4	4	325	329	5.93E-63
TBX5	27	687	714	1.81E-109
TCF12	46	941	987	1.02E-140
TCF3	44	998	1042	1.23E-153
TCF4	41	954	995	5.93E-148

TCF7	73	1201	1274	1.04E-166
TCF7L1	62	1085	1147	4.39E-154
TCF7L2	64	1023	1087	9.60E-141
TEAD1	67	1167	1234	2.43E-165
TEAD2	46	969	1015	4.02E-146
TEAD3	6	433	439	1.69E-82
TEAD4	16	704	720	4.66E-125
TELO2	5	377	382	2.60E-72
TFAP2A	55	909	964	8.87E-127
TFAP2B	57	906	963	1.43E-124
TFAP2C	71	956	1027	6.80E-123
TFAP4	38	860	898	1.58E-132
TFCP2	69	1009	1078	3.78E-134
TFCP2L1	14	692	706	4.09E-125
TFDP1	77	1085	1162	5.80E-142
TFE3	20	621	641	1.07E-103
TFEB	18	507	525	3.76E-83
TFEC	4	777	781	7.58E-158
THAP1	63	1033	1096	1.93E-143
TNFRSF14	10	193	203	1.62E-29
TP53	65	1183	1248	4.25E-170
TP63	61	1037	1098	7.33E-146
TP73	62	1062	1124	9.79E-150
TRIM28	9	366	375	7.38E-65
TRPS1	235	512	747	1.88E-05
TSPAN1	9	401	410	5.71E-72
TSPAN15	9	401	410	5.71E-72
TWIST1	22	697	719	1.08E-116
UBE2N	44	935	979	2.08E-141

UBP1	26	656	682	2.22E-104
USF1	28	815	843	9.64E-134
USF2	28	812	840	3.80E-133
VDR	8	145	153	4.34E-22
WT1	70	1033	1103	8.38E-138
XBP1	35	339	374	5.62E-38
YBX1	43	1066	1109	6.59E-168
YY1	63	1116	1179	4.22E-159
ZBTB14	81	1005	1086	1.85E-124
ZBTB2	39	806	845	4.00E-121
ZBTB24	7	102	109	8.05E-15
ZBTB3	1	286	287	1.24E-59
ZBTB33	27	615	642	2.22E-95
ZBTB44	23	628	651	6.55E-102
ZBTB5	0	309	309	1.38E-66
ZBTB7A	65	998	1063	2.85E-135
ZBTB7B	22	434	456	5.03E-65
ZEB1	53	1041	1094	1.38E-153
ZFP161	14	358	372	8.00E-58
ZFP64	7	274	281	1.40E-48
ZIC1	62	1041	1103	8.92E-146
ZIC2	34	825	859	1.40E-129
ZIC3	50	950	1000	7.71E-139
ZNF101	39	992	1031	2.25E-157
ZNF143	0	74	74	1.70E-16
ZNF208	4	695	699	1.51E-140
ZNF263	15	287	302	4.24E-43
ZNF426	13	139	152	2.74E-17
ZNF557	16	117	133	6.91E-12
ZNF580	9	675	684	2.16E-128
ZNF675	8	936	944	2.04E-184

ZNF691	15	473	488	1.39E-79
ZNF709	10	460	470	9.28E-83
ZNF75D	9	265	274	1.46E-44
ZNF826P	66	773	839	1.63E-93
ZNF91	5	514	519	7.39E-101
ZSCAN21	4	174	178	7.85E-32
ZSCAN4	6	192	198	3.34E-33

Supplementary Table S3.4. Key TFs in the high impact network. 394 Key TFs with significant enrichment for predicted genes in high impact group as compared to low impact group.

Finding transcription factor coordinated groups (TCFGs) in the high impact network.

Next, we identified transcription factors that might function in a coordinated fashion by determining groups of Key TFs that were predicted to regulate the same target genes. We hypothesized that these transcription factor groups that are uniquely enriched in the high impact network, and that were predicted to collaboratively regulate the same genes, might provide insights into the biology of this set of patients with pronounced transcriptional responses to ADT. For each Key TF, we calculated the percent of target genes that overlapped with the target genes of every other Key TF. We then performed hierarchical clustering of the pairwise percent overlaps to find TFCGs in which all the Key TF members mutually shared at least 70% of their target genes (Figure 3.3). We found 34 TFCGs in this analysis, some of which contained both known oncogenic factors with other factors not previously associated with prostate cancer (Supplementary Table S3.5).

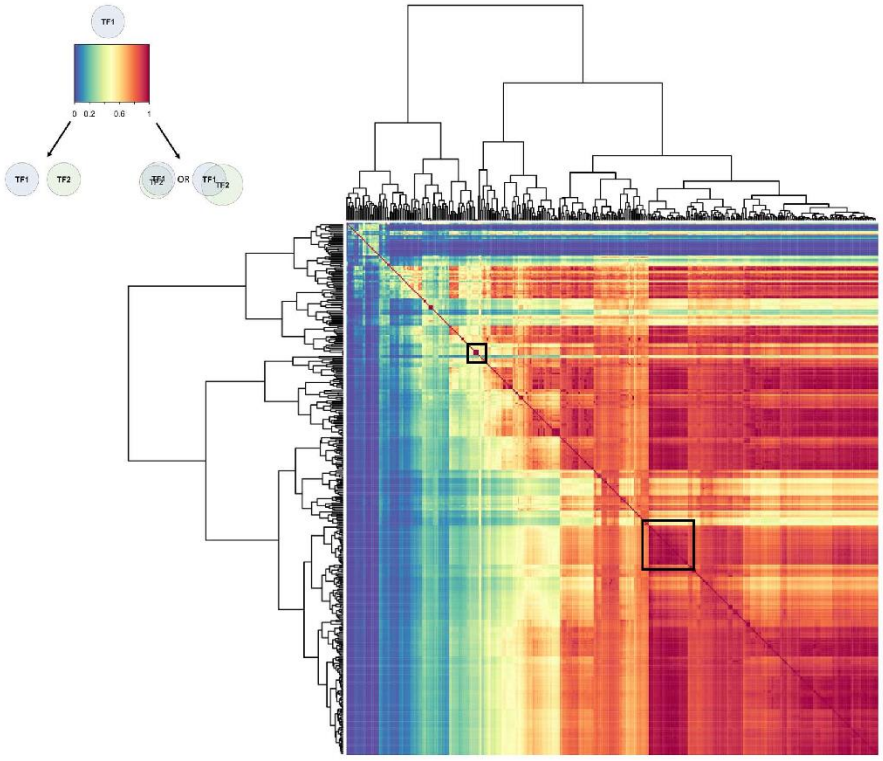


Figure 3.3. Identification of TFCGs in the high impact group network. The heatmap displays hierarchical clustering of putative gene target percent overlap of one Key TF as compared to all others. The dark blue to dark red color gradient denotes the degree of shared target overlap. Because the degree of target overlap between a pair of Key TFs may be non-reciprocal, dendrograms are ordered based on mutual relationships, and are oriented identically on the x- and y-axis. The diagonal represents a Key TF compared to itself. Only reciprocal relationships between groups of Key TFs were considered TFCGs (white boxes demarcate two representative TFCGs as symmetrical squares on the diagonal).

MNT | TFEC
TNFRSF14 | NR2C1
GCM2 | VDR
CLOCK | ING4
GTF2A1 | GTF2A2
BATF | GDNF | PLAU
GLI3 | GLI2
EGR4 | ZNF263
NF1B | CTCF | KLF5
STAT2 | STAT5B
TSPAN15 | TSPAN1 | KCNH8 | SLC6A2
ZFP161 | TRIM28 | KLF7 | ZFP161 | TRIM28 | KLF7
KLF2 | NEUROD1 | RREB1 | ZBTB7B
ESR2 | ZNF691 | CBFB | SIN3A
NR5A2 | MXD1 | NHLH1 | GLI1 | SNAI2 | TFEB | MXI1 | MYCN
EPAS1 | ARNT2
MAF | MAFF
CREM | CREB1 | ATF1
MAFG | ATF3 | ELF2
TBX5 | MYC | TWIST1 | BHLHE40 | MITF | RARA | FERD3L | HDAC1 | TFE3 | RXRB | ZBTB44 | UBP1
ZBTB33 | GTF3C2 | PLAGL1 | NR1H3 | RARB | NR1I2 | RARG | NR1H2 | NR1I3
HNF4A | AHR
ZBTB14 | ARNT | UBE2N | NFE2L2 | NR4A2 | NF1 | NRL | TFAP4 | MAX | EGR3 | TFAP2A | TFAP2B | NFE2 | NR5A1 | ZIC2 | LTF | USF2 | USF1
NR1H4 | ZBTB2 | MAZ | ARHGEF7 | CD40
TEAD1 | POU5F1 | ATF2 | RFX3
RFX2 | RFX5
FOXO4 | FOXL1 | FOXO1 | FOXG1 | AP3D1 | SOX4 | FOXA2 | GATA4 | POU1F1 | TBP | FOXJ3 | SOX2 | SOX6
SALL1 | ERF | ETV5 | ETV3 | ELF4 | GMEB1 | SP100
FEV | ATF4
PITX2 | IRF5 | STAT4 | STAT1
PBX1 | PARP1 | IRF8 | HOXB2 | PRRX2 | PDX1
EP300 | JUN | SP1 | SOX5 | SOX10 | SOX18 | JUND | JUNB | SMAD3 | TP53 | FOS | RXRA | BRCA1 | CEBPB | SPI1 | SMAD4 | SMAD2 | ETS1 | E2F1 | NR3C1 | CEBPA | CEBPD | ETS2 | GATA2 | YY1 | TCF7L1 | FOSL2 | FOSB | FOSL1 | E2F6 | TCF7 | LEF1 | MEIS1 | MEIS2 | MEIS3 | RUNX2 | SPDEF | CEBPZ
HSF2 | HOXA10 | GATA6 | GATA5 | FOXA1 | FOXP3 | SRY | POU3F3 | FOXD3 | SOX8 | HMGA1 | HSF1 | HNF1A | POU2F1 | SOX17 | CUX1 | DBP | SOX9 | FOXM1 | IRF4 | NKX2-5 | NFYC | NFYA | NFYB | NFATC3 | NFATC2 | NFAT5 | NFATC1 | STAT3 | NFE2L1 | AR | NKX3-2
ETV1 | HIF1A | E2F7 | E2F5 | E2F2 | CHURC1 | EGR1 | KLF13 | TEAD2 | NR2F1 | ESR1 | HES1 | NR2F2 | CABLES2 | NFKB1 | MYF5 | HAND1 | EGR2 | TCF4 | MYOG | TCF12 | NR2C2 | SMAD9 | MTF1 | IKZF1 | ZIC3 | TFAP2C | PAX8 | ATM | PAX2 | NFATC4 | HOXA3 | MAFK | RUNX3 | PGR | TAL1 | GTF2IRD1 | ELK4 | ELK3 | KLF12 | ETV6 | ETV7 | SMAD6

| SMAD7 | MAFA | NF1A | TFCP2 | TCF7L2 | SMAD5 | PAX4 | ELK1 | SPIB | MYOD1 | TCF3 | HIC2 | ZBTB7A | MZF1 | ETV4 | MYF6 | GATA1 | SREBF2 | NFIX | KLF4 | PURA | KLF6 | GEN1 | GATA3 | MYBL2 | MYB | YBX1 | ERG | FLI1 | CEBPG | E2F3 | RFX1 | CEBPE | SREBF1 | TFDP1 | BCL6 | SMAD1 | BCL6B | SP4 | SP3 | HSF4 | THAP1 | ZIC1 | SP2 | GTF2I | TP73 | ZEB1 | HIC1 | WT1 | TP63 | GABPA | ELF1 | ELF5 | MAFB | NKX2-1 | E2F4

Supplementary Table S3.5. TFCGs in the high impact network. 34 TFCGs were identified in the high impact network as defined as sharing >70% of predicted target genes.

Comparison of the metastatic PCS1 network and high impact group network reveals common TFCGs

We next investigated whether there were common TFCGs associated with both the pronounced transcriptional response to ADT and metastatic progression. To do this, we leveraged a large dataset of over 800 patients compiled from multiple publicly available cohorts that were normalized and subtyped by You *et al* [218] to use as the RNA expression data input to PANDA. Because our observations suggested an ADT mediated selection against PCS2, and a relative enrichment of the aggressive PCS1 signature in primary tumors, we hypothesized that there were overlapping TFCGs (oTFCG) associated with analogous changes in subtype, that were also unique to metastasis. Consequently, we elected to find oTFCGs enriched in PCS1 metastatic tumors (Met.PCS1) as compared to PCS2 primary tumors.

We first determined the Key TFs that had a significant enrichment of unique targets in the Met.PCS1 network ($p < 2.19 \times 10^{-4}$), and identified TFCGs (Supplementary Table S3.6 & Supplementary Table S3.7). We found that more than 80% of the Met.PCS1 Key TFCGs were also exclusively associated in the high impact group network. We identified 33 TFCGs enriched in both the networks (Table 3.1), and defined an “overlapping TFCG” (oTFCG) as a TFCG in the Met.PCS1 network that shared at least two Key TFs with a TFCG in the high impact group network. We found groups that contained within them known associations, such as JUN and FOS (oTFCG3, Table 3.1).

	Prim.PCS2	Met.PCS1	TFEdgeTotal	p-value
AHR	15	98	113	8.94E-17
AR	19	56	75	7.78E-06
ARHGEF7	18	68	86	1.62E-08
ARNT	14	102	116	2.33E-18
ARNT2	9	35	44	4.11E-05
ATF2	25	67	92	4.55E-06
ATM	20	103	123	2.55E-15
BCL6B	17	64	81	4.49E-08
BRCA1	15	65	80	4.32E-09
CD40	18	68	86	1.62E-08
CDX1	123	210	333	4.36E-07
CDX2	72	149	221	5.54E-08
CEBPB	22	58	80	2.45E-05
CEBPD	24	59	83	5.45E-05
CEBPE	23	58	81	4.47E-05
CEBPG	23	57	80	6.53E-05
CEBPZ	20	103	123	2.55E-15
CHURC1	20	110	130	6.05E-17
CLOCK	8	40	48	1.21E-06
CNOT3	3	24	27	2.01E-05
CRTC2	1	22	23	2.34E-06
CTCF	5	66	71	3.27E-15
CUX1	33	81	114	2.48E-06
E2F1	13	52	65	3.99E-07
E2F2	20	163	183	5.21E-30
E2F3	13	52	65	3.99E-07
E2F4	13	52	65	3.99E-07
E2F5	20	168	188	2.70E-31
E2F7	20	168	188	2.70E-31
EFNA2	7	35	42	5.75E-06
EGR1	13	83	96	3.01E-14
EGR2	17	75	92	2.06E-10
EGR3	14	76	90	4.82E-12
EGR4	9	124	133	8.34E-28
ELF1	17	58	75	7.29E-07
ELF2	13	58	71	2.02E-08
ELF4	19	132	151	8.18E-23
ELF5	17	64	81	4.49E-08
ELK1	17	51	68	1.59E-05
ELK3	25	67	92	4.55E-06
ELK4	25	66	91	6.75E-06
EP300	16	46	62	6.55E-05
ERF	22	146	168	1.68E-24
ERG	17	61	78	1.84E-07

ETS2	20	61	81	1.89E-06
ETV1	20	145	165	1.90E-25
ETV3	19	131	150	1.45E-22
ETV4	16	57	73	5.10E-07
ETV5	17	111	128	8.25E-19
ETV6	25	68	93	3.06E-06
ETV7	20	76	96	2.10E-09
EVX1	108	172	280	4.03E-05
FEV	11	58	69	2.39E-09
FLI1	16	64	80	1.83E-08
FOS	19	62	81	5.80E-07
FOSB	22	61	83	7.41E-06
FOSL1	22	65	87	1.42E-06
FOSL2	21	67	88	2.92E-07
FOXA1	41	98	139	4.15E-07
FOXA2	55	105	160	2.86E-05
FOXA3	80	136	216	4.74E-05
FOXD3	41	102	143	9.55E-08
FOXL1	67	127	194	5.35E-06
FOXM1	35	81	116	7.36E-06
GABPA	16	60	76	1.25E-07
GATA1	14	51	65	1.56E-06
GATA2	25	65	90	9.97E-06
GATA3	22	61	83	7.41E-06
GATA4	47	104	151	1.14E-06
GATA5	40	98	138	2.39E-07
GATA6	45	102	147	8.28E-07
GC	22	204	226	2.84E-39
GEN1	13	46	59	6.94E-06
GLI1	10	57	67	1.28E-09
GLI2	7	32	39	2.76E-05
GLI3	6	32	38	9.46E-06
GLIS2	30	70	100	2.63E-05
GMEB1	9	42	51	1.23E-06
GTF2I	19	53	72	2.70E-05
GTF2IRD1	21	94	115	7.83E-13
GTF3C2	11	82	93	3.51E-15
HDAC1	12	65	77	1.76E-10
HES1	23	67	90	1.23E-06
HIC1	16	53	69	3.16E-06
HIC2	18	78	96	1.23E-10
HIF1A	18	113	131	9.91E-19
HIVEP2	5	27	32	4.57E-05
HNF1A	32	74	106	1.82E-05
HOXA10	53	103	156	2.31E-05

HOXA3	36	100	136	9.81E-09
HOXB2	79	146	225	2.36E-06
HOXB5	101	179	280	8.17E-07
HOXD11	119	208	327	1.95E-07
HOXD9	110	182	292	7.07E-06
HSF1	29	70	99	1.53E-05
HSF4	16	57	73	5.10E-07
IKZF1	18	68	86	1.62E-08
IRF3	17	53	70	6.73E-06
IRF4	27	71	98	3.26E-06
JUN	20	66	86	2.12E-07
JUNB	22	61	83	7.41E-06
JUND	20	64	84	5.13E-07
KLF1	2	61	63	1.23E-16
KLF12	22	66	88	9.32E-07
KLF13	16	85	101	3.90E-13
KLF15	1	28	29	4.32E-08
KLF16	2	47	49	1.41E-12
KLF2	6	64	70	6.92E-14
KLF3	7	52	59	4.39E-10
KLF4	10	42	52	3.32E-06
KLF5	10	51	61	3.23E-08
KLF6	11	41	52	1.34E-05
KLF7	11	97	108	5.06E-19
LTF	8	35	43	1.61E-05
MAFA	17	63	80	7.21E-08
MAFK	19	60	79	1.40E-06
MAX	14	61	75	1.20E-08
MAZ	14	66	80	9.41E-10
MTF1	16	66	82	6.89E-09
MYB	24	59	83	5.45E-05
MYBL2	24	64	88	7.98E-06
MYC	10	45	55	7.31E-07
MYF5	19	52	71	4.05E-05
MYOD1	20	57	77	1.01E-05
MYOG	20	57	77	1.01E-05
MZF1	17	62	79	1.15E-07
NF1A	11	42	53	8.30E-06
NF1B	6	36	42	1.06E-06
NFIX	15	49	64	8.72E-06
NFKB1	18	66	84	4.15E-08
NFYA	21	95	116	4.72E-13
NFYB	20	94	114	2.79E-13
NFYC	20	95	115	1.67E-13
NKX2-5	32	77	109	6.18E-06

NKX3-2	26	83	109	1.14E-08
NKX6-2	81	164	245	2.64E-08
NR1H2	12	42	54	1.94E-05
NR1H3	12	42	54	1.94E-05
NR1I2	12	45	57	4.75E-06
NR1I3	12	45	57	4.75E-06
NR2C2	16	54	70	2.01E-06
NR2F2	20	61	81	1.89E-06
NR3C1	17	57	74	1.15E-06
NR4A2	16	50	66	1.19E-05
PAX2	20	66	86	2.12E-07
PAX4	14	54	68	3.75E-07
PAX8	19	79	98	1.97E-10
PDX1	76	138	214	7.16E-06
PLAG1	10	172	182	2.69E-40
PLAGL1	11	128	139	2.97E-27
POU2F1	37	83	120	1.02E-05
POU3F3	47	106	153	5.74E-07
PRRX2	77	140	217	5.95E-06
PURA	13	43	56	2.73E-05
RARA	14	50	64	2.48E-06
RARB	12	41	53	3.07E-05
RARG	12	41	53	3.07E-05
RFX1	20	67	87	1.35E-07
RREB1	8	59	67	3.09E-11
RUNX3	18	74	92	8.92E-10
RXRA	18	67	85	2.60E-08
RXRB	12	45	57	4.75E-06
SALL2	1	30	31	1.13E-08
SATB1	75	154	229	4.34E-08
SMAD2	19	58	77	3.33E-06
SMAD3	19	60	79	1.40E-06
SMAD4	14	53	67	6.05E-07
SMAD5	15	48	63	1.36E-05
SMAD6	19	60	79	1.40E-06
SMAD7	19	60	79	1.40E-06
SMAD9	19	60	79	1.40E-06
SNAI1	5	72	77	7.25E-17
SNAI2	11	42	53	8.30E-06
SOX10	18	66	84	4.15E-08
SOX17	34	77	111	1.79E-05
SOX18	22	67	89	6.09E-07
SOX4	44	90	134	2.75E-05
SOX5	26	71	97	1.75E-06
SOX6	54	102	156	4.61E-05

SOX8	34	77	111	1.79E-05
SP1	15	61	76	3.13E-08
SP100	16	116	132	1.22E-20
SP2	15	61	76	3.13E-08
SP3	16	62	78	4.81E-08
SP4	15	58	73	1.34E-07
SPDEF	17	102	119	1.25E-16
SPI1	20	55	75	2.30E-05
SPIB	16	47	63	4.31E-05
SREBF1	18	65	83	6.63E-08
SREBF2	14	55	69	2.32E-07
SRY	43	89	132	2.42E-05
TAF1	12	64	76	3.01E-10
TAL1	15	56	71	3.48E-07
TBP	59	133	192	2.28E-08
TCF12	19	51	70	6.04E-05
TCF3	20	57	77	1.01E-05
TCF4	18	53	71	1.37E-05
TCF7L1	21	69	90	1.21E-07
TCF7L2	14	57	71	8.75E-08
TEAD2	13	96	109	1.81E-17
TFAP2A	13	103	116	3.07E-19
TFAP2B	15	96	111	2.75E-16
TFAP2C	14	73	87	2.40E-11
TFAP4	11	46	57	1.18E-06
TFCP2	16	65	81	1.12E-08
TFDP1	13	52	65	3.99E-07
THAP1	20	70	90	3.47E-08
TP53	17	60	77	2.92E-07
TP63	17	62	79	1.15E-07
TP73	17	58	75	7.29E-07
TRIM28	3	63	66	3.62E-16
UBE2N	17	142	159	1.32E-26
UBP1	9	66	75	2.19E-12
USF1	14	47	61	9.83E-06
USF2	14	50	64	2.48E-06
VDR	4	64	68	1.63E-15
WT1	14	45	59	2.41E-05
YBX1	23	72	95	1.46E-07
YY1	21	59	80	8.68E-06
ZBTB14	25	203	228	5.79E-37
ZBTB2	13	44	57	1.74E-05
ZBTB33	17	114	131	1.51E-19
ZBTB7A	12	60	72	2.52E-09
ZBTB7B	6	31	37	1.62E-05

ZEB1	15	65	80	4.32E-09
ZIC1	15	59	74	8.29E-08
ZIC2	12	60	72	2.52E-09
ZIC3	12	62	74	8.75E-10

Supplementary Table S3.6. Key TFs in the Met.PCS1 network. 228 Key TFs with significant enrichment for predicted genes in Met.PCS1 specimens as compared to Prim.PCS2 specimens.

TFCG1	GLI3 | GLI2
TFCG2	KLF2 | RREB1 | GLI1
TFCG3	HOXB5 | NKX6-2
TFCG4	HOXD11 | HOXD9 | EVX1 | CDX1
TFCG5	SATB1 | PRRX2 | CDX2 | HOXB2 | PDX1
TFCG6	RARA | USF1 | USF2
TFCG7	MYC | RXRB | NR1I2 | NR1I3 | RARB | RARG | NR1H2 | NR1H3
TFCG8	FOXL1 | TBP
TFCG9	SOX4 | GATA4 | FOXA2 | HOXA10
TFCG10	SRY | SOX8 | CUX1 | POU2F1 | GATA5
TFCG11	POU3F3 | GATA6 | FOXD3 | FOXA1
TFCG12	UBP1 | HDAC1
TFCG13	ZBTB33 | PLAGL1
TFCG14	ZBTB14 | E2F2 | E2F5 | E2F7
TFCG15	KLF13 | EGR2 | EGR1 | HIC2
TFCG16	TFAP2A | TFAP2B | AHR | ARNT | TEAD2 | CHURC1
TFCG17	HIF1A | ERF | ETV1 | ETV5 | ETV3 | ELF4
TFCG18	GTF2IRD1 | ATM | CEBPZ | NFYA | NFYB | NFYC
TFCG19	MAZ | ZIC2 | MAX | EGR3 | NFKB1 | ARHGEF7 | CD40 | BCL6B | MTF1 | MZF1 | TFCP2 | SP4 | TP73 | ZIC1 | ZIC3 | TP53 | TP63 | ZBTB7A | SP3 | SP1 | SP2
TFCG20	FEV | ELF2
TFCG21	NR4A2 | GEN1 | SPIB | EP300 | ZBTB2 | TFAP4 | WT1 | KLF6 | NF1A | KLF4 | PURA | HIC1 | SMAD4 | TCF12 | SMAD5 | NFIX | ELK1 | MYF5 | TCF4 | TCF3 | MYOD1 | MYOG | GATA1 | TFDP1 | E2F4 | E2F1 | E2F3 | IRF3 | SPI1 | NR2C2 | GTF2I | PAX4
TFCG22	SP100 | SPDEF
TFCG23	NKX3-2 | HOXA3
TFCG24	PAX8 | ETV7 | TFAP2C | RUNX3 | IKZF1 | HES1 | THAP1
TFCG25	AR | MAFK | CEBPG | CEBPE | CEBPB | CEBPD
TFCG26	FOXM1 | SOX17 | NKX2-5 | HSF1 | SOX5 | HNF1A | IRF4
TFCG27	MYBL2 | ZEB1 | RFX1 | PAX2 | MAFA | KLF12 | ELF5 | TCF7L1 | YBX1 | FOSL2 | JUN | RXRA | BRCA1 | SREBF1 | GABPA | FLI1 | ETS2 | YY1 | ERG | FOS | ETV4 | ELF1 | NR3C1 | TCF7L2 | HSF4 | SREBF2 | TAL1 | ATF2 | SOX10 | ETV6 | ELK4 | ELK3 | FOSL1 | JUND | GATA2 | SOX18 | NR2F2 | GATA3 | FOSB | JUNB | SMAD2 | MYB

Supplementary Table S3.7. TFCGs in the Met.PCS1 network. 27 TFCGs were identified in the Met.PCS1 network as defined as sharing >70% of predicted target genes.

oTFCG1	NR2F2-SMAD9-PAX2-TAL1-ELK4-ELK3-KLF12-ETV6-SMAD7-MAFA-TCF7L2-ETV4-SREBF2-GATA3-MYBL2-MYB-YBX1-ERG-FLI1-RFX1-	

	SREBF1-HSF4-ZEB1-GABPA-ELF1-ELF5
oTFCG2	MYF5-TCF4-MYOG-TCF12-NR2C2-NF1A-SMAD5-PAX4-ELK1-SPIB-MYOD1-TCF3-GATA1-NFIX-KLF4-PURA-KLF6-GEN1-E2F3-TFDP1-GTF2I-HIC1-WT1-E2F4
oTFCG3	JUN-SOX10-SOX18-JUND-JUNB-SMAD3-FOS-RXRA-BRCA1-SMAD2-NR3C1-ETS2-GATA2-YY1-TCF7L1-FOSL2-FOSB-FOSL1
oTFCG4	NFKB1-MTF1-ZIC3-TFCP2-ZBTB7A-MZF1-BCL6B-SP4-SP3-ZIC1-SP2-TP73-TP63
oTFCG5	HES1-IKZF1-TFAP2C-PAX8-RUNX3-ETV7-THAP1
oTFCG6	HSF1-HNF1A-SOX17-FOXM1-IRF4-NKX2-5
oTFCG7	NR1H3-RARB-NR1I2-RARG-NR1H2-NR1I3
oTFCG8	GATA5-SRY-SOX8-POU2F1-CUX1
oTFCG9	GATA6-FOXA1-POU3F3-FOXD3
oTFCG10	EGR1-KLF13-EGR2-HIC2
oTFCG11	ERF-ETV5-ETV3-ELF4
oTFCG12	EP300-SPI1-SMAD4-E2F1
oTFCG13	SOX4-FOXA2-GATA4
oTFCG14	E2F7-E2F5-E2F2
oTFCG15	HOXB2-PRRX2-PDX1
oTFCG16	ARNT-TFAP2A-TFAP2B
oTFCG17	NFYC-NFYA-NFYB
oTFCG18	MAFK-CEBPG-CEBPE
oTFCG19	MAZ-ARHGEF7-CD40
oTFCG20	MAX-EGR3-ZIC2
oTFCG21	GLI3-GLI2
oTFCG22	ZBTB33-PLAGL1
oTFCG23	HDAC1-UBP1
oTFCG24	FOXL1-TBP
oTFCG25	KLF2-RREB1
oTFCG26	USF2-USF1
oTFCG27	CEBPB-CEBPD
oTFCG28	CHURC1-TEAD2
oTFCG29	ETV1-HIF1A
oTFCG30	ATM-GTF2IRD1
oTFCG31	MYC-RXRB
oTFCG32	SP1-TP53
oTFCG33	NR4A2-TFAP4

Table 3.1. 33 oTFCGs between the high impact ADT group and Met.PCS1 networks.

Interestingly, we observed that the transcription factors within oTFCGs maintained the exclusive associations with each other, despite regulating distinct sets of target genes in the

high impact group and Met.PCS1 networks. In some cases, oTFCGs collectively gained or lost other transcription factors between networks. For example, one TFCG comprised of ERF-ETV5-ETV3-ELF4, gained HIF1A, a well-characterized transcription factor involved in metastasis, in the Met.PCS1 network. The addition or loss of transcription factors from an oTFCG may inform the observed changes in predicted gene targets. Thus, we identified TFCGs that appear to be functioning in a coordinated fashion to achieve changes in gene expression in two distinct phases of prostate cancer progression. Taken together, these data suggest a concerted condition-dependent re-localization that maintained interactions of these transcription factors during metastasis.

3.4 Discussion

ADT is used as one component of treatment for intermediate and advanced prostate cancer. At present, it is difficult to assess transcriptional changes that are direct consequences of ADT, as current guidelines discourage the use of neoadjuvant ADT with radical prostatectomy [231]. Our cohort is novel, unique, and unusual in that all but one of the patients received neoadjuvant ADT alone, and subsequent radical prostatectomy. This allowed us to observe the direct effects of ADT using patient matched tissues.

A complete understanding of the transcriptional pathways that characterize ADT response and subsequent metastatic progression are still unclear. Previous studies have sought to find molecular mechanisms associated with prostate cancer progression, but these analyses typically rely on expression profiling alone [232] to predict putative upstream regulators. We applied the PANDA [219] algorithm to curate and integrate expression, protein-protein interaction, and DNA binding data in order to predict transcription factor-gene target interactions. We utilized these networks to identify putative cooperative and collaborative transcription factor groups that are most associated with pronounced transcriptional response to ADT, retention of aggressive subtype signatures, and development of metastatic disease. By employing a top-down approach integrating large, independent datasets, we have predicted and prioritized transcription factor groups that could serve as critical upstream regulators of ADT response and metastasis.

Differential expression analysis comparing Pre-ADT Bxs and Post-ADT RPs yielded 190 significantly differentially expressed genes. IPA analysis identified multiple therapeutic agents as putative upstream regulators of these genes. These were largely associated with androgen receptor (AR) signaling (Supplementary Table S2). Specifically, genes regulated by dihydrotestosterone (androgen), and metribilone (R1881), a widely used AR agonist that has been shown to increase expression of AR target genes [233], were both inhibited in our dataset, as

would be expected from ADT. Accordingly, KLK3 (PSA) expression was significantly lower in the high impact group than in the low impact group, despite similar ADT exposure times, underscoring the biological relevance of the differential transcriptional responses between the two groups. This may also indicate the initial AR signaling inhibition in response to ADT before eventual reactivation of this signaling pathway [214, 234]. IPA analysis also showed that beta-estradiol (estrogen) targets were activated and targets of U0126, a MAPK kinase pathway inhibitor, were inhibited in response to ADT, suggesting bypass or compensatory PDGF-MAPK and estrogen signaling. While IPA analysis provided initial insights into the initial transcriptional responses to ADT, among well-established signaling networks, we were interested in identifying novel transcriptional relationships in this setting.

We characterized the transcriptional responses in our cohort using a previously developed subtyping scheme developed by You *et al* [218]. These PCS subtypes were developed by integrating *a priori* defined prostate cancer-relevant signaling pathways, genetic and genomic alterations, and other biological characteristics of aggressive prostate cancer such as stemness and cell proliferation [218]. The PCS1 subtype is the most aggressive of the three subtypes, with poorer prognosis, shorter metastasis-free survival, and metastatic CRPC[218]. This aggressive subtype is enriched for AR-variant pathway genes, and is also associated with enzalutamide-resistance[218, 235] and metastatic-CRPC [37]. The PCS2 subtype is enriched for AR-signaling genes, and was found to be sensitive to enzalutamide [218]. The PCS3 subtype exhibits low expression of AR-signaling genes [218, 236], and is associated with gene signatures enriched in basal cells [236].

In contrast to the similar transcriptional changes in PCS2 and PCS3 signatures, we observed a striking difference in the percent of PCS1 genes expressed in high and low impact groups following ADT. The high impact group exhibited not only a retention of this signal, but in many cases an increase in expression of PCS1 genes, while the low impact group tended to

lose expression of these genes (Figure 2). These data suggest that, despite similar ADT exposure times, only the high impact group associated with a more aggressive subtype. We speculate that the high impact group tumors could resist an ADT mediated selection against aggressive components, or clonal populations by activating bypass pathways, while the low impact group were more sensitive to the effects of ADT treatment [218].

We sought to find putative combinatorial relationships among Key TFs that could influence the pronounced transcriptional response to ADT. A Key TF might gain putative target genes in the high impact group network because it is either upregulated, or has increased accessibility for target genes due to changes in the local epigenetic landscape [237]. Therefore, we did not require that a Key TF exhibit a significant increase in expression. The strength of our analysis relies on the integration of context dependent transcriptional data and validated protein-protein interactions with DNA motif data to infer condition-specific TFCGs.

Next, we integrated putative regulators of transcriptional changes in the high impact group after ADT with those of metastatic samples. We integrated gene expression data from a large cohort of unmatched primary and metastatic samples (n > 800) subtyped by You *et al* [218] with our own analysis, and interrogated the transcriptional differences between primary PCS2 tumors and PCS1 metastases. TFCGs in common between the high impact group and metastatic lesions of PCS1 could shed light on potential drug targets and pathways applicable to treatment of the most aggressive forms of prostate cancer. Intriguingly, using independent gene expression datasets, we found that more than 80% of TFCGs in the Met.PCS1 network were also present in the high impact network. These observations could reflect a possible ADT-mediated clonal selection of aggressive cancer cells (reviewed in [238]).

Considerable research has been devoted to understanding how collaborative transcription factor relationships and interdependencies confer a precise temporal control of condition-specific changes in transcription (reviewed in [239]). We speculate that prediction of

associations with both the high impact group and Met.PCS1 network may reveal new regulatory relationships that coordinately mediate the shift from ADT-response to metastatic progression. Despite dynamic transcriptional changes that occur in a tumor in response to ADT and during metastatic progression, we identified common TFCGs that remain associated with each other, even though their predicted target genes are different. This possibly reflects a refinement of more robust TFCG associations that are more critical to influencing both a pronounced transcriptional response to ADT, retention of aggressive subtype pathways, and metastatic progression.

ADT resistance is associated with AR splice variants (e.g. AR-V7) [27], genomic amplification of AR [240], and mutations to the AR ligand binding domain [241] that maintain transcriptional activity in a low androgen environment. Interestingly, AR separates from a large TFCG in the high impact group network, and associates with C/EBPβ, C/EBPδ, C/EBPγ, and C/EBPε in the Met.PCS1 network. The C/EBP family of transcription factors is associated with mesenchymal gene signatures and aggressive disease in a variety of tumors including glioblastomas [242, 243, 244], esophageal squamous cell carcinoma [245], urothelial carcinoma [246], and hepatocellular carcinoma [247]. Additionally, C/EBPβ promotes oncogene-induced senescence, facilitating tumor progression and chemotherapy resistance after androgen deprivation [248]. This may reflect differential transcriptional activities of AR to both evade effects of ADT and promote metastases, and sheds light on the dynamic interactions of AR to maintain signaling throughout prostate cancer progression, despite changes in androgen availability.

The SOX4-FOXA2-GATA4 oTFCG is also of interest for multiple reasons. We have shown that deletion of SOX4 *in vivo* can inhibit prostate cancer progression [249] and that knockdown of SOX4 [56] can induce apoptosis in prostate cancer cells. Moreover, in pancreatic cancer, SOX, FOX, and GATA factors may cooperate to drive metastases [250].

Finally, FOS, FOSB, FOSL1, JUN, JUND, and JUNB, all members of the MAPK signaling pathway, were associated with each other in both high impact and Met.PCS1 networks, and FOS and FOSB were also significantly overexpressed after ADT. Notably, IPA analysis informed the biological relevance of this oTFCG as it showed that targets of U0126, a MAPK kinase pathway inhibitor, were inhibited in response to ADT. Combined with our observations of decreased *KLK3* expression, and IPA analysis indicating an inhibition of AR-signaling, these data suggest bypass or compensatory PDGF-MAPK signaling after ADT. This is consistent with previous reports of increased phospho-MAPK levels enriched in tumor tissues of patients who have undergone ADT [114]. Moreover, it has recently been shown that AR-null prostate cancers that do not undergo neuroendocrine differentiation, or 'double negative' metastatic prostate cancers have sustained FGF-MAPK signaling and that these cancers are sensitive to MEK and ERK1/2 inhibition *in vitro* and *in vivo* [251]. Thus, both the IPA analysis and our finding that these factors maintain associations in the high impact group, and metastatic tumors suggest that combination therapies that include ER, MEK, JNK, and/or ERK inhibitors may provide some benefit to patients undergoing ADT.

In conclusion, our analysis utilizes a novel, unique dataset from matched patient samples that interrogates direct effects of ADT, and integrates these data with a large cohort of primary and metastatic tumors to predict common mechanisms important to resistance to therapy and progression to metastatic disease. We have elucidated transcription factor relationships that consistently associate with aggressive ADT transcriptional responses, and metastasis, two distinct, and clinically significant phases of prostate cancer progression. This hypothesis-driving study expands what is known about important coordinated transcription factor activities that regulate prostate cancer aggressiveness, metastasis, and development of androgen-resistance.

Chapter 4. The Role of SOX4 in Epithelial to Mesenchymal Transition in Prostate Cancer

Abstract

Prostate cancer remains the most commonly diagnosed cancer in U.S. males, and ranks second in mortality with over 28,000 deaths per year. Prostate cancer mortality is due to metastases, necessitating the understanding of how metastatic disease develops. One major step for cancer cells to metastasize and acquire migratory and invasive capabilities is the epithelial to mesenchymal transition (EMT). EMT encompasses vast molecular changes in gene expression, mediated by aberrant developmental signaling pathway activation, including the loss of cell-cell, and planar and apical-basal polarity, increased motility, and resistance to apoptosis and anoikis (cell death due to the detachment from the extracellular matrix). The sex determining region Y-box 4 (SOX4) gene is a developmental transcription factor that is over expressed in prostate cancer and plays a critical role in many developmental pathways inappropriately activated during EMT, such as the transforming growth factor β (TGF-β) and epidermal growth factor receptor (EGFR) signaling pathways[2,3]. Our data suggest that concomitant treatment of TGF-β and EGF induces SOX4, and initiates the EMT program in prostate epithelial cell lines. We also demonstrate that SOX4 over expression is sufficient to induce EMT in prostate epithelial cell lines. Currently, the molecular mechanisms that govern how SOX4 initiates EMT via transcriptional activation of target genes are poorly understood. Interestingly, out preliminary data suggest that SOX4 interacts with PCAF histone acetyltransferase, and TRRAP, another subunit of the PCAF HAT complex. These data suggest that SOX4 may induce and recruit the PCAF/TRRAP HAT complex to transcriptionally activate downstream targets to initiate the EMT program.

4.1 Introduction

Prostate cancer remains the most commonly diagnosed cancer for U.S. males and ranks third among tumor site-specific mortality with other 28,000 deaths per year[252].Prostate cancer mortality is due to metastases, typically metastasizes in the bone, and relatively infrequently to the adrenal gland, liver, and lung[149]. The epithelial to mesenchymal transition (EMT) is a major step in the metastatic process. To metastasize, cancer cells need to acquire migratory and invasive capabilities, a process that involves EMT[62]. Aberrant activation of developmental signaling pathways can promote EMT to convert differentiated epithelial cells to migratory and invasive mesenchymal cells[62]. EMT encompasses changes that lead to loss of cell-cell adhesion, planar and apical-basal polarity, increased motility, and resistance to apoptosis and anoikis[63]. Among the developmental signaling pathways that are aberrantly activated during EMT is the TFG-β signaling pathway, a highly studied major inducer of EMT[64].

The sex determining region Y-box 4 (SOX4) gene is a developmental transcription factor that is one of 64 “cancer signature genes”, and plays a critical role in many developmental pathways inappropriately activated during EMT, such as the transforming growth factor β (TGF-β) and epidermal growth factor receptor (EGFR)[39, 50, 61, 253]. The SOX4 gene contains a single exon that encodes a 47 kDa protein[48]. It has a conserved high mobility group (HMG) DNA-binding domain at the N-terminus, related to the TCF/LEF family of transcription factors, and a proline, serine, and acidic residues-rich transactivation domain (TAD) at the C-terminus.

Extensive research demonstrates that there is a correlation between SOX4 expression levels and the progression of multiple cancer types[50]. In these cancers, SOX4 demonstrated a pro-proliferative, pro-metastatic, and/or anti-apoptotic role, as well as the ability to maintain cancer-initiating cells. The work of the Moreno lab and others has demonstrated that SOX4 is consistently over expressed in prostate cancer tissues as compared to benign tissues via expression array, immunohistochemistry, and immunoblotting and is correlated with tumor

grade in prostate cancer samples[56, 254]. Additionally, the Moreno lab also identified SOX4 as a transforming oncogene in non-transformed prostate epithelial cells[56]. These data suggest that over expression of SOX4 is associated with tumor aggressiveness in prostate cancer cells, and possibly increases metastatic potential.

The Moreno lab and Zhang *et al.* have observed that in a breast cancer cell line, knockdown of SOX4 significantly reduced migration and invasion[59, 60]. Furthermore, there is also evidence suggesting that siRNA mediated knockdown of SOX4 led to increased expression of the epithelial marker, E-cadherin, and decreased expression of the mesenchymal markers Vimentin, and N-cadherin[254]. Interestingly, there are also many analyses linking SOX4 and TGF-β induced EMT. In breast cancer cell lines, it has been shown that SOX4 was necessary for EMT, and that knockdown was sufficient to induce a reversion from a mesenchymal to epithelial phenotype after TGF-β treatment[73]. Additionally, SOX4 was found to be a target of two microRNAs (miRs), miR-132 and miR-212, that when expressed, could also inhibit TGF-β induced EMT, and that over expression of SOX4 was sufficient to reverse the tumor suppressive effects[255]. Tavazoi *et al.* found that restored expression of miR-335 repressed lung and bone metastasis via interference with SOX4 expression[58]. Intriguingly, Zhang *et al.* found that Metformin, a commonly used diabetes drug, upregulated miR-30a that targeted SOX4 expression ultimately inhibiting TGF-β mediated EMT[256]. Clinical evidence also suggests that SOX4 can promote the metastatic process. SOX4 expression is significantly associated with the presence of distant metastases and correlates with cancer-specific mortality in prostate cancer patients[61].

Notably, TGF-β induced EMT was found to be dependent on the synergistic activity of SOX4 in complex with ERG, the product of the highly associated prostate cancer gene fusion, TMPRSS2-ERG[151]. Although there is a great deal of evidence that suggests SOX4 is an

important mediator of EMT, less is known about how SOX4 can transcriptionally activate genes critical for EMT.

Epigenetic activities are critical to dynamic changes in global gene expression that promote cancer. PCAF (p300/CBP associated factor) is a member of the GCN5-related *N*-acetyl transferase (GNAT) family of HATs. It is highly enriched in promoters of active genes, and acetylates histone 3 on lysine 14, and less frequently, histone H4 on lysine 8[257]. Transference of the acetyl group to DNA allows increased accessibility for transcription factors and transcriptional machinery. PCAF activity and specificity is typically dictated by association with distinct protein complexes, and thus it has been shown to have opposing roles in cancer[258, 259].

In this study, our preliminary data suggest that concomitant treatment of TGF-β and EGF induces SOX4 and initiates the EMT program, and that SOX4 is sufficient to induce EMT, in prostate epithelial cells. We also present data that propose a mechanism whereby SOX4 recruits and interacts with the PCAF histone acetyl transferase complex (PCAF) to promoters to activate gene expression. Taken together, these data suggest that SOX4 is critical for EMT, and requires the PCAF complex to promote the EMT transcriptional program.

4.2 Materials and methods

Cell culture and reagents

RWPE1, non-neoplastic prostate epithelium cells, LNCaP, and ARCaPE/M lines were obtained from American Type Culture Collection (ATCC), and from Novicure, LLC, respectively. Both cell lines were cultured in Gibco Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 and MCaP-Medium, respectively. All media were supplemented with 10% fetal bovine serum (FBS) (Atlanta Biologicals), and 2mM L-Glutamine (Gibco), and 200mM Penicillin and 200mM Streptomycin antibiotics (Gibco). L-Glutamine, and 1% penicillin-streptomycin. Cells were incubated at 37^{o}C in an incubator with humidified atmosphere of 5% CO_2. Opti-MEM (ThermoFisher Scientific) was also used for transient transfection experiments.

The primary antibodies used in this study are the following: rabbit anti-SOX4 (Santa Cruz SC-20090), rabbit anti-PCAF (abcam ab12188), rabbit anti-TRRAP (abcam ab72509), mouse anti E-cadherin (Cell Signaling Technology #24E10), rabbit anti-vimentin (Cell Signaling Technology #5741).

Recombinant human TGF-β1 (R&D Systems), and recombinant human EGF (Gibco Life Sciences) were reconstituted in 4mM HCl (TGF-β) and phosphate-buffered saline with 0.1% Bovine Serum Albumin (EGF), respectively.

Western blot analysis

Whole cell lysates were prepared with lysis buffer containing 137mM NaCl, 20mM Tris-HCL (pH 8.0), 10% glycerol, 1% NP-40, and a protease inhibitor cocktail (Promega). Supernatant protein concentrations were determined using the Micro BCA Protein Assay Kit (Pierce Biotechnology). 25ug total protein was separated on a 10% SDS-polyacrylamide gel, and transferred to a nitrocellulose membrane. After transferring, the membrane was blocked in 1X PBS buffer with 0.1% Tween-20 and 5% non-fat dry milk for 1 hour at room temperature. Membranes were then

incubated with primary antibody at 4ºC. For all antibodies, except mouse anti-β-actin, the membranes were incubated with HRP conjugated anti-mouse (Cell Signaling 7076S) or HRP conjugated anti-rabbit (Abcam ab6721). for 1 hour at room temperature. Signals were visualized utilizing enhanced chemiluminescence (Pierce). After incubation with mouse anti-β-actin, the membrane was then incubated with fluorescence-conjugated IRDye 800CW goat anti-mouse at 1:5000 for 1 hour at room temperature, and signals were visualized using the Odyssey infrared imaging system (LI-COR Biosciences). For all other antibodies, signals were visualized using an enhanced chemiluminescent substrate.

Co-Immunoprecipitation analysis

Cells were washed twice with ice-cold 1X PBS and harvested in 1%NP40 lysis buffer supplemented with protease inhibitors, as described for the Western blot analysis. Whole cell lysates were pre-cleared and incubated with primary antibody overnight at 4°C. Immuno-complexes were precipitated with Protein A agarose beads (Millipore) by gentle shaking at 4°C for 2 hours. The agarose beads were washed twice with lysis buffer and three times with cold 1X PBS. Proteins were eluted by boiling in Laemmli sample buffer with 2-mercaptoethanol as a reducing agent, followed by western blot (described above).

Transfection Assays

Transfection experiments were performed as outlined in the FuGene HD Protocol Database. We used the pcDNA3.1 plasmid containing HA-tagged SOX4. The FuGene HD-to-DNA ratio was 1.5:1, with 7ug of plasmid DNA.

CRISPR/Cas9 Gene Editing of SOX4

A double nickase CRISPR/Cas9 gene editing system was used with a gRNA donor template (DNA2.0) targeting 43 base pairs of exonic sequence in *SOX4*. The Cas9 nickase mutant

system was used to introduce single-stranded breaks to promote homology-directed repair with a gRNA donor template with a 43-base pair deletion of *SOX4* exonic sequence.

4.3 Results

SOX4 is induced by canonical TGF-β signaling, a common pathway used to induce EMT *in vitro*. We found that in RWPE1 cells, TGF-β alone was not sufficient to promote EMT, but, as others have observed, required the co-treatment with EGF[260].Concomitant treatment of TGF-β and EGF in RWPE1 initiated the EMT program, by inducing expression of mesenchymal marker vimentin and inhibiting the epithelial marker E-cadherin, as well as upregulating SOX4 (Figure 4.1)[70].

Next, we determined whether SOX4 alone was sufficient to drive EMT. We tested this hypothesis using the epithelial and mesenchymal clones of androgen repressed metastatic human prostate cancer cell line (ARCaPE and ARCaPM, respectively), and RWPE1 cell lines. Ectopic expression of SOX4 in ARCaPE cells induced vimentin, and inhibited E-cadherin. ARCaP cells were isolated from a patient with bone metastasis, and display a propensity for bone metastasis[261]. The epithelium-like ARCaPE cells have a relatively decreased propensity for metastasis, but rapidly undergo EMT, upon exposure to the bone microenvironment, or growth factors, yielding the ARCaPM clone[261]. We observed that over expression of SOX4, both stable and ectopic, could induce the EMT program, reflecting what was displayed in the ARCaPM cells (Figure 4.2). As expected, E-cadherin was not expressed in ARCaPM cells, but the transiently transfection of SOX4 seem to potentiate the vimentin expression.

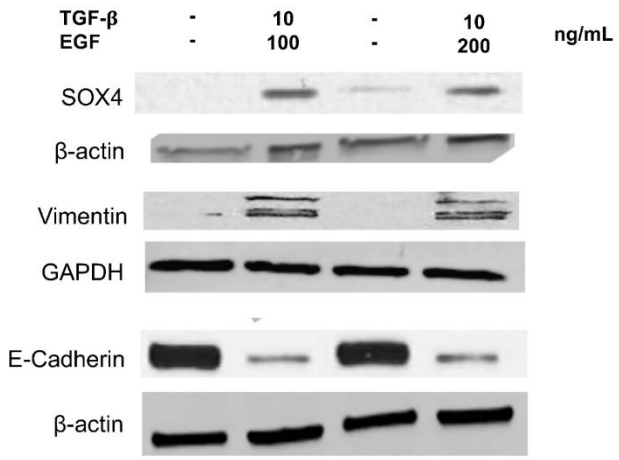


Figure 4.1. TGF-β and EGF initiate the EMT program and induce SOX4 expression in RWPE1 non-neoplastic cells. Concomitant treatment of TGF-β and EGF in RWPE-1 cells for seven days upregulates SOX4 as well as mesenchymal markers vimentin and downregulates epithelial marker E-cadherin (A). This treatment induces expression of SOX4 and vimentin transcript levels (B).

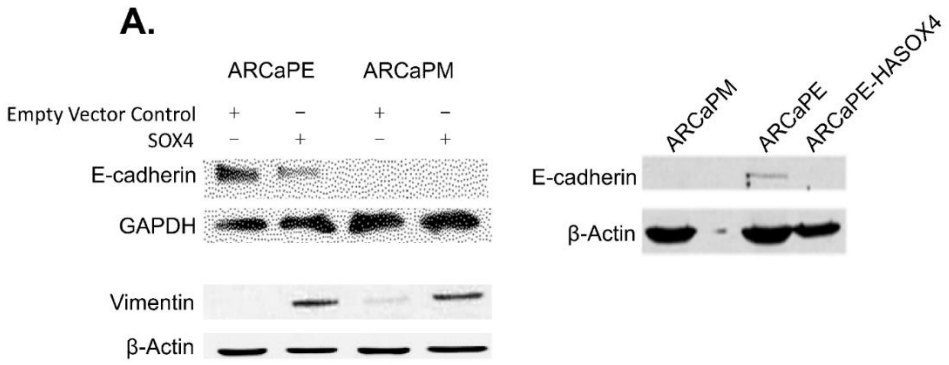


Figure 4.2. Over expression of SOX4 induces EMT in prostate cancer cell lines. A) Stable and ectopic over expression of SOX4 in ARCaPE cells leads to decreased E-cadherin, and increased vimentin. E-cadherin expression in ARCaPE with stable expression of SOX4 mimics ARCaPM. B) Stable over expression of SOX4 in non-transformed RWPE-1 prostate epithelial cells correlates with decreased E-cadherin, Note TGF-β treatment further inhibits E-cadherin expression in stable SOX4 prostate cells.

To gain insights into whether SOX4 had a ubiquitous EMT-promoting role, independent of cancer type, we analyzed RNA-seq data across nine-thousand seventy-five TCGA Pan Cancer samples. We performed ARACNE analyses as previously described to identify candidate downstream SOX4 targets[262, 263]. The ARACNE approach utilizes mutual information to predict downstream targets of putative master regulator genes based on gene expression patterns[205, 242, 262, 263, 264]. Interestingly, we found that genes involved in chromatin modification were of the most significant candidate SOX4 targets (Table 4.1). One candidate gene, KAT2B (PCAF), was of particular interest to us, as preliminary mass spectrometry data suggested that SOX4 interacted with TRRAP, a critical subunit of the PCAF histone acetyltransferase (HAT) complex, that plays an important role in the maintenance of tumorigenic cancer initiating cells, and it is mutated in several cancers, including prostate cancer[258, 265, 266]. We performed co-immunoprecipitation experiments and validated that SOX4 not only interacted with TRRAP, but also interacted with the PCAF chromatin remodeling complex. (Figure 4.3). Because TRRAP is a critical subunit of the PCAF, 2 MDa multi-subunit chromatin remodeling complex, we rationalized that SOX4 would also interact with this complex, and confirmed this via co-immunoprecipitation (Figure 4.3). Based on these data, we speculate that SOX4 transcriptional activity is mediated by interactions with the PCAF complex, and that this complex may be necessary for SOX4 to induce EMT promoting genes.

Category	Genes
BAF-SWI/SNF	ACTL6A, ARID1A, ARID2, SMARCA4, SMARCAD1, SMARCC1, SMARCD1, SMARCE1, SRCAP
Chromatin assembly	TLK2
Chromatin Remodeler	CHD4, CHD6, CHD7, CHD8
CTCF	CTCF
Deubiquitination	USP21
DNA Methyltransferase	DNMT1, DNMT3A, DNMT3B
DNA Methylhydroxylase	TET1, TET3
Histone	H2AFY2
Histone Acetyltransferase	**KAT2B (PCAF)**, TADA1, VPS72, MSL1
Histone Deacetylase	HDAC2, RBBP4, SIRT2, SUDS3, TBL1XR1
Lysine Demethylase	KDM1A (LSD1), KDM2B, KDM4A, KDM5B (JARID1B)
Lysine Methyltransferase	EHMT2, SETDB1, SETMAR, WHSC1 (MMSET)
Nucleosome Remodeling	BPTF
Polycomb	EZH2, PCGF2, RNF2
Reader	CBX2, CBX3, CBX7, JADE3 (PHF16)
Tx Termination	TTF1

Table 4.1. Identification of candidate SOX4 targets. Using ARACNE analysis based on the PanCancer dataset from TCGA. Of the most significant candidate SOX4 targets are genes involved in chromatin modification. Notably, PCAF (red), one putative target, is a member of a large histone acetyltransferase complex that contains a subunit, TRRAP, which preliminary mass spectrometry data indicated interacted with SOX4.

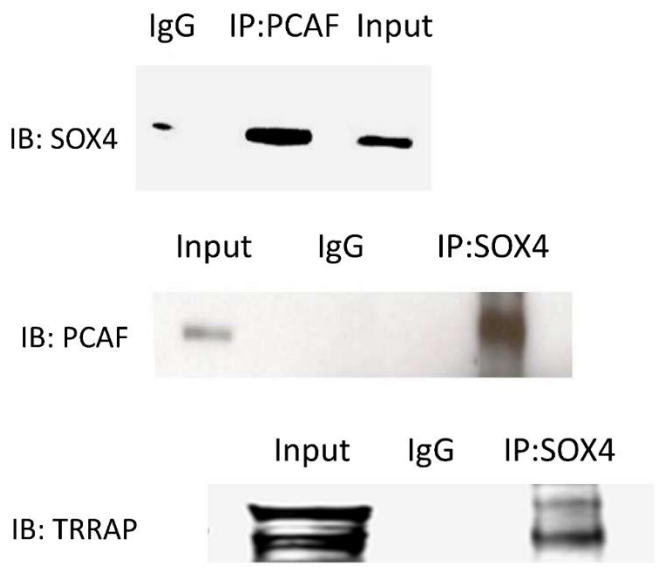


Figure 4.3. SOX4 interacts with PCAF and TRRAP in prostate cancer cell lines. *A.)* Co-immunoprecipitation (Co-IP) of SOX4 and PCAF in ARCaPE cells with stable over expression of HA-SOX4. *B.)* Reciprocal Co-IP of SOX4 and PCAF in LNCaP cells with stable over expression of SOX4. *C.* CO-IP of SOX4 and TRRAP in LNCaP cells with stable over expression of SOX4.

4.4 Discussion

In this study, we provide the foundation that supports the idea that SOX4 is an important transcription factor to drive EMT. We demonstrate that SOX4 is downstream of TGF-β signaling, as it is sufficient to initiate the EMT program. We also found that SOX4 can interact with both PCAF and TRRAP, but whether these interactions are necessary for EMT remains unknown. Interestingly, PCAF that has been shown to interact with other EMT promoting transcription factors, supporting the idea that this complex remains relevant to EMT by interacting with SOX4 [259].

Currently, how SOX4 impacts the transcriptome to promote EMT remains unclear. To elucidate the EMT-related targets that are necessary for EMT, it is necessary to inhibit the expression of SOX4. One potential hurdle is that prostate cancer cells may require SOX4 to survive (Figure 4), therefore utilizing conditional knockouts via inducible SOX4-CRISPR/Cas9 would be ideal.

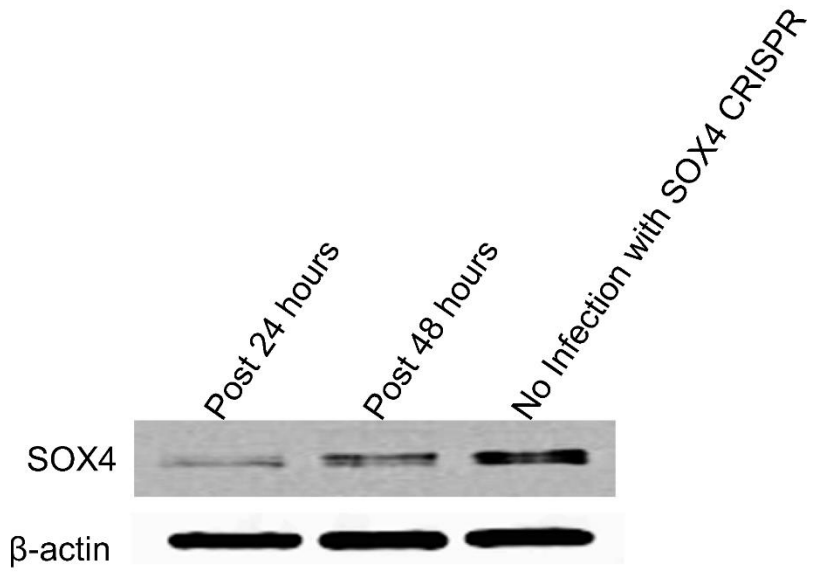


Figure 4.4. SOX4 expression returns after knockdown with CRISPR/Cas9. SOX4 Western blot in RWPE1 cells suggests that cells that expressed SOX4 may have outcompeted those with SOX4 knockdown, possibly due to the lethality of SOX4 deletion.

We believe that the PCAF complex (including TRRAP) are necessary coactivators that promote SOX4 transcriptional activity to drive EMT. We hypothesize that SOX4 recruits the PCAF HAT complex to promoters to activate gene expression during EMT (Figure 4.5).

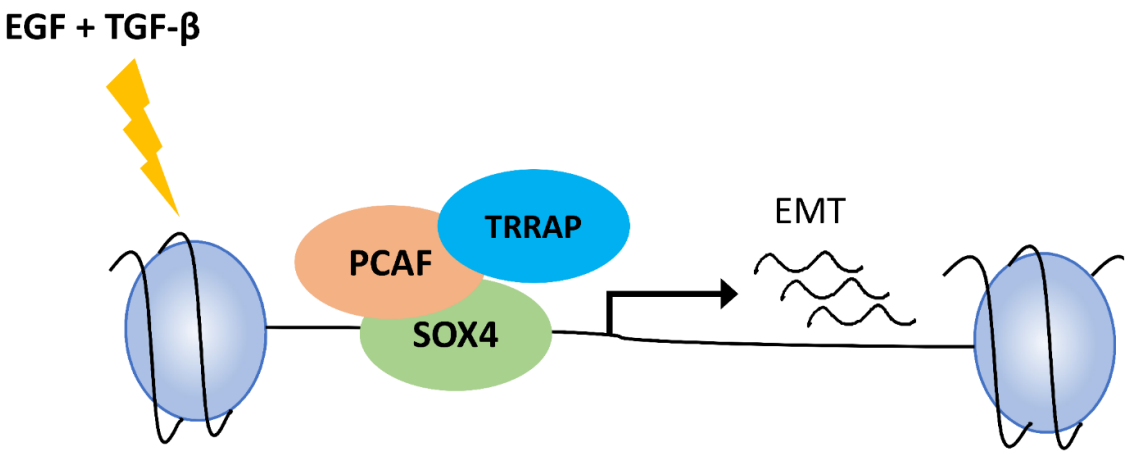


Figure 4.5. Proposed mechanism. Upon TGF-β and EGF induction of EMT, SOX4 recruits PCAF and TRRAP to the promoter regions of EMT-inducing genes.

Chapter 5. TFBSET: A New Tool for the Identification of Enriched Transcription Factor Binding Sites in the Promoter Regions of Human Genes[267]

Kenneth A. Watanabe, Selvi Ramalingam, Joshua Bell, Nitya V. Sharma, David A. Gutman, and Carlos S. Moreno

Under review for publication

My Contribution: As a coauthor on this manuscript, I contributed to the curation of the MATCH output data of transcription factor binding sites from the TRANSFAC database.

Abstract

Transcription Factor Binding Site Enrichment Tool (TFBSET) is a publicly available, web-based tool that identifies enriched transcription factor binding sites (TFBSs) in the promoter regions of a user-specified set of human genes. TFBSET limits its scans for TFBSs to the H3K27ac and DNaseI hypersensitive promoter regions of specified genes to reduce false positive results. TFBSET selects up to 1000 random gene sets equal in number to the user-specified gene set to generate a distribution of TFBSs, compute empirical false discovery rates, and identify overrepresented TFBSs in the query gene set. TFBSET was tested on control gene sets that are induced by specific transcription factors or ligands to confirm its accuracy, and TFBSET compared favorably to other popular TFBS enrichment software. We applied TFBSET to prostate cancer datasets and two TFBSs (E2F and NFY) were identified as overrepresented in patients with the most aggressive prostate cancer subtype. This software may provide insights into signaling pathways and the sets of transcription factors that regulate coordinated gene expression patterns. The website is free and open to all users and there is no login requirement. Address: http://tfbset.bmi.emory.edu/tfbset.html.

5.1 Introduction

Transcription factors are proteins that bind to specific sequences in the promoter and enhancer regions of genes to regulate gene expression to produce mRNA and non-coding RNA transcripts. There are hundreds of transcription factors (TFs) within the human genome [268] and each one binds to specific transcription factor binding sites (TFBSs) composed of specific DNA sequences. A single transcription factor (TF) can regulate many genes, and TFs that regulate many important genes in disease progression can be candidate targets for novel therapies [269]. Over 10% of currently prescribed drugs directly target the class of TFs called nuclear receptors [270]. For example, enzalutamide [84, 271], which blocks the nuclear receptor AR, is used for the treatment of prostate cancer and tamoxifen [272], which blocks the nuclear receptor ESR1, is used for the treatment of breast cancer. TFs other than nuclear receptors are more difficult to target. However, inhibitors of some TFs such as NF-κB [273, 274], CBFβ [275], Pax2 [276], GLI [277], Myc [278, 279] β-catenin [280, 281] and others have been identified.

Given a set of genes that show coordinated regulation, such as genes upregulated in a specific type of cancer, or genes that are upregulated by overexpression of a growth factor or developmental signal transduction pathway ligand, finding the TFBSs that are enriched among the promoters of these genes may be indicative of the responsible TFs. These TFs may be potential targets for therapy or give a better understanding of the signaling pathways involved.

In this study, we present the Transcription Factor Binding Site Enrichment Tool (TFBSET) for identifying TFBSs that are over-represented in the promoter regions of a given gene set. This is performed by querying the open chromatin regions around the TSS of the genes in the given gene set for TFBSs. To determine whether a TFBS is enriched in the gene set, a random set of genes equal in number to the query gene set, is analyzed for TFBSs. This selection process is repeated up to 1000 times to generate a distribution for each TFBS. From these distributions, false discovery rates (FDRs) for each TFBS can be estimated.

In order to limit the scanning to regions of DNA that are available for TF binding and likely functional, we limited the regions to those marked by H3K27 acetylation [282] and DNaseI hypersensitivity [283], known markers for open chromatin and functional promoters and enhancers using data from seven cell lines analyzed by the ENCODE Consortium [226]. We hypothesized that scanning only the regions where TFs can bind would produce fewer false positive TFBSs and reduce computing time. TFBSET also allows the user to select the size of the region around the TSS, up to ±5000 bp, and allows restriction to only those matches that meet a specified score.

Current software tools for identification of enriched TFBSs for a given list of genes include DiRE [284], PASTAA [285], oPOSSUM [286] and PSCAN [287]. None of these tools take advantage of the available experimental data that defines open chromatin regions. Most, however, do limit their queries to evolutionary conserved regions (ECRs) of DNA. DiRE limits the scanning regions to ECRs but also looks at distal regulatory regions. PASTAA has the option to limit the query to ECRs in the promoter but can also scan the entire promoter regions. The oPOSSUM single site analysis also queries ECRs based on the phastCons score [288] but only queries TFBSs defined in the JASPAR database. PSCAN only accepts Refseq gene IDs for input rather than official gene symbols but can query for either JASPAR or TRANSFAC TFBSs. PSCAN does not have the option to limit the queries to ECRs. Though the use of ECRs is popular for reducing the search space for the identification of TFBSs, the chromatin regions around these ECRs may still be closed and inaccessible to TFs resulting in false positive results.

To determine whether restricting scanning regions to open chromatin reduces false positive identifications of enriched TFBSs, TFBSET was tested on gene sets upregulated by known transcription factors (Myc, TP53, NFκB, E2F) and ligands (Hh, Wnt, TGFβ). From these known gene sets, TFBSET properly identified the expected TFs with high sensitivity. Comparison of TFBSET to the other available software programs shows that TFBSET is favorably comparable

in the identification of TFBSs, outperforming each of them to varying degrees. By applying TFBSET to genes upregulated in prostate cancer subtype 1 (PCS1) [218], we were able to identify two TFBSs that were highly overrepresented, NFY and E2F. These TFs may be targets for therapy and treatment for patients with PCS1. TFBSET represents a promising tool in the identification of TFs that may regulate gene sets differentially expressed under a variety of experimental or disease conditions

5.2 Materials and methods

The regions of H3K27Ac and DNaseI hypersensitivity within the human genome (build hg19) were obtained from ENCODE data tracks [226] in the UCSC human genome database [227] (Fig. 5.1C). The regions within 5,000bp of the TSS of a gene were passed through the MATCH software [228] to identify TFBS based on the TRANSFAC database [229, 289]. In total, 216 million vertebrate TFBSs were identified in these regions with a minimum core matrix score of 0.75 and a minimum overall matrix score of 0.70.

To identify enhancer regions, regions outside the promoter that were marked by H3K4me1 and H3K27Ac were identified and passed through MATCH and 57.3 million TFBSs were identified. The TFBSs identified in the enhancer regions were assigned to regulate the most proximal gene. The promoter and enhancer TFBSs were loaded into a MariaDB database and a Web interface was created to query the data (Fig. 5.1A).

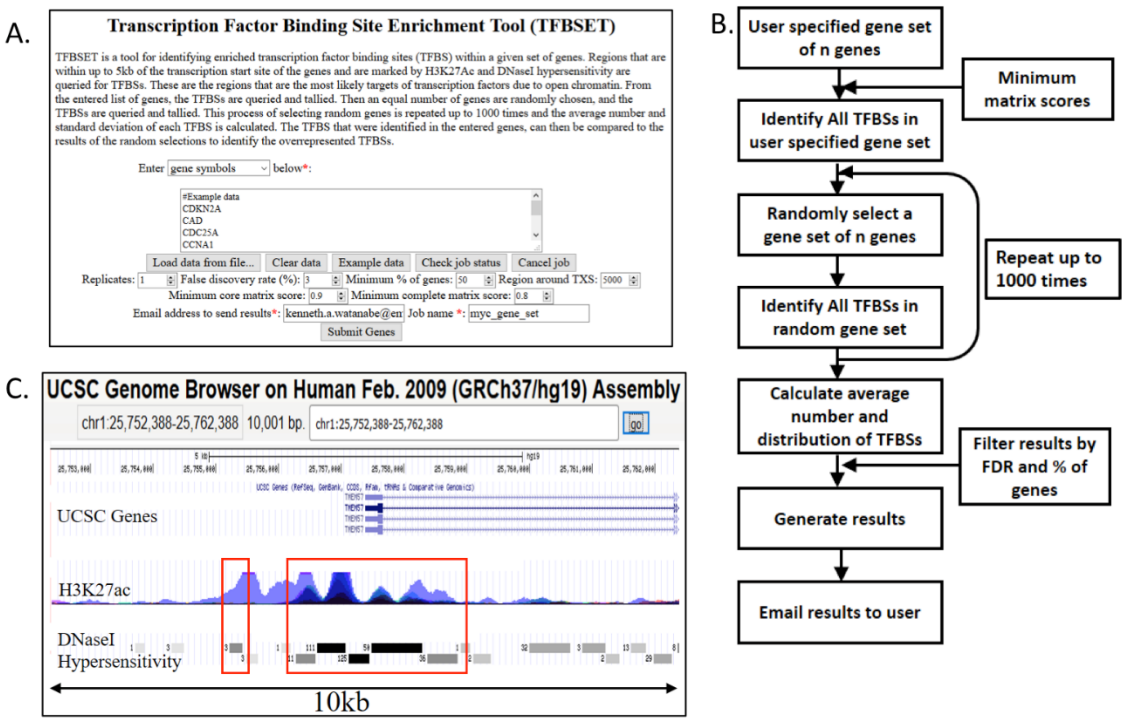


Figure 5.1. TFBSET tool characteristics. A) Screenshot of the TFBSET tool webpage; B) Flow chart of TFBSET logic; C) Screenshot of UCSC Browser indicating regions used for TFBSET analysis. Regions with both H3K27Ac and DNase I hypersensitivity within 5kb of the TSS are included in the TFBSET promoter analysis. Intergenic regions containing both H3K27Ac and H3K4me1 (not shown) were used for analysis of enhancer regions.

A flowchart of TFBSET is depicted in Figure 5.1B. To run the program, the user enters a list of genes either by gene symbol, ensembl gene ID, Entrez gene ID or Refseq ID. The software then queries the database to identify the TFBS that are within a specified distance from the TSS (default ±5000 bp) and have the minimum specified matrix scores.

The software randomly selects genes equal in number to the number of genes the user entered and the database is queried to identify the TFBSs. This process is repeated up to 1000 times and the workload is distributed over multiple CPUs to improve performance. Then the standard deviation and average number of occurrences of each TFBS is calculated.

The FDR for a given TFBS is estimated by the number of times the replicated data had more occurrences of the TFBS than the sample data divided by the number of replicates. To limit the number of false positive results, the user can specify the maximum FDR.

A rank is assigned to each TFBS identified. The rank is based on the FDR and the number of standard deviations the TFBS counts in the gene set exceeds the average of the replicate value. The highest ranking TFBS (rank 1) has the lowest FDR and the most number of standard deviations above the average (highest Z score). The lower ranking TFBSs are less overrepresented.

Sometimes, a single gene of the user supplied gene set may have an abundance of a particular TFBS which may bias the results. The user can specify that only TFBSs that occur in at least a specified percentage of the supplied gene set (default 50%) be reported to avoid this bias.

Most genes within the human genome have multiple isoforms due to alternative splicing and these splice variants can have different promoter regions. To prevent double counting of TFBS, the isoform containing the largest region of H3K27Ac and DNaseI hypersensitivity is used for

analysis when querying by gene symbol, ensembl gene ID, or Entrez gene ID. This is not an issue when querying via Refseq ID since each isoform has a unique Refseq ID.

Example gene sets were obtained from the Molecular Signatures Database (MSigDB) [290, 291].

5.3 Usage and implementation

TFBSET can be accessed from our website at: https://tfbset.bmi.emory.edu/tfbset.html. From the main screen (Fig. 1A) the user can enter the genes they wish to identify enriched TFBS. The format must be one gene per line. Lines beginning with a hash (#) or exclamation point (!) will be ignored. The genes can be pasted into the main window or the “Load data from file” button can be used to load the genes from a text file. The user can then specify the number of replicates necessary to calculate the average count and standard deviation of each TFBS. The user can also specify the limits for filtering, either the maximum FDR, minimum number of genes, a minimum core matrix score and total matrix score.

Since it may take some time to generate the results, the user must enter an email address where the results will be sent. When the “Submit Genes” button is pressed, the job is submitted to a background process. The progress of the job can be checked by the “Check job status” button. Jobs can be terminated by selecting the job from the job status window and clicking the “Cancel Selected Jobs” button.

5.4 Results

To validate TFBSET, a set of 24 genes that are upregulated by the c-Myc transcription factor [292] were entered into TFBSET. By adjusting the parameters, it was determined that the parameters shown in Supplemental Table S5.1 were optimal for identifying the Myc target sites. The results of running TFBSET on six independent Myc gene sets from the MSigDB database [293, 294, 295, 296, 297] are shown in Supplemental Table S2.

The Ben Porath dataset resulted in the largest number of Myc TFBSs identified as overrepresented. There were 11 Myc TFBSs identified, 10 of which had a FDR of 0.00% and one (V$NMYC_02) had a FDR of 0.20%. V$NMYC_02 also ranked near the bottom of the list at 612 out of 793. When adjusting the region around the TSS, values less than ±1000 bp resulted in fewer than 11 Myc TFBSs identified as overrepresented. When TFBSET was run using ±1000 bp on other gene sets known to be regulated by a TF, the number of significant TFBSs was less than when run at ±5000 bp (Supplemental Table S5.3). Values from ±2000 bp to ±5000 bp showed no significant difference in results so the default region was set at ±5000 bp around the TSS to ensure the most number of significantly enriched TFBS are identified.

The Ceballos dataset identified only one Myc TFBS (V$NMYC_Q3). This gene set is comprised of genes upregulated in chronic myeloid leukemia expressing TP53 and Myc. The identification of Myc TFBS may have been hindered due to the presence of TP53. The small gene set (21 genes) may also be a contributing factor. V$NMYC_Q3 was the only TFBS that was identified in all of the Myc gene sets.

Using the same parameters optimized with the six Myc gene sets, TFBSET was run on several additional gene sets, and the results are summarized in Table 5.1.

Parameter	Setting
Number of replicates	1000
Maximum FDR rate	3 %
Minimum percent of genes	50 %
Minimum matrix core score	0.9
Minimum overall matrix score	0.8
Include enhancer regions	No
Region around TSS	±5000 bp

Supplemental Table S5.1. Optimal parameters for identification of Myc target sites in the Dang gene set. There is no significant difference in results when region around TSS is varied from ±2000 bp to ±5000 bp.

Matrix Id	Ref (31)	Ref	Ref (33)	Ref (34)	Ref (35)	Ref (36)
V$CMYC_01					0.0(21)	
V$CMYC_O6_01	1.3(499)		1.1(35)	0.0(19)	0.0(15)	0.4(240)
V$MYC_01	0.2(69)		0.8(22)	0.0(2)	0.0(13)	0.0(25)
V$MYC_O2	1.5(562)		1.1(32)	0.0(12)	0.0(4)	
V$MYCMAX_01	1.1(412)			0.0(5)	0.0(7)	
V$MYCMAX_02	0.7(281)		2.8(122)	0.0(64)	0.0(27)	
V$MYCMAX_B	0.6(220)		0.4(6)	0.0(7)	0.0(38)	0.0(22)
V$MYCN_01	0.6(223)		2.3(90)	0.0(3)	0.0(10)	0.3(205)
V$NMYC_01	2.6(750)		1.4(46)	0.0(39)	0.0(6)	
V$NMYC_02				0.7(612)	0.3(934)	
V$NMYC_O3	1.1(418)	0.8(7)	0.4(7)	0.0(43)	0.0(30)	0.2(172)
Highest rank	807	51	142	793	1095	592

Supplemental Table S5.2. **Comparison of results from 6 different gene sets upregulated by Myc. The numbers in the table represent the FDR.** The rank of each TFBSs is in parentheses. The lower the rank, the more significant the TFBS is over-represented.

Gene Set	Reference	# of genes	TFBSET ±1000 bp	TFBSET ±5000 bp
			TRANSFAC	TRANSFAC
Myc	-34	230	Very high	Very high
TP53	-38	96	High	Very high
E2F	-37	200	Very high	Very high
NFKB	-39	91	Med1	Very high
TGFB-EGR	-37	54	Med1	Very high
TGFB-SMAD	-37	54	Low	Med2
Wnt-TCF	-37	42	Low	Med2
Wnt-LEF	-37	42	Med1	Low
PCS1-NFY	-23	85	Very high	Very high
PCS1-E2F	-23	85	Med1/2	Very high
Genome			hg19	hg19
Region			±1000 bp	±5000 bp
Filter			H3K27ac, DNaseI	H3K27ac, DNaseI
Speed			30 min	30 min

Supplemental Table S5.3. TFBSET produces better results at ±5000 bp than ±1000 bp around the TSS. When TFBSET is on gene sets known to be regulated by a transcription factor or ligand, fewer significant TFBS were identified at ±1000 bp than at ±5000 bp around the TSS. Sections marked in green have at least 2 enriched TFBSs in the top 10. Sections marked in pink have no enriched TFBSs in the top 10 and fewer than 10 enriched TFBSs overall.

Gene	Expected TFBS	TFBS(s) identified	Total # TFBSs	# in Top 10	# in Top 50	Total # targets	Reference
Sig24	unknown	FOX	50	4	20	20	[298]
PCS1	unknown	NFY/E2F	777	10	15	31	[218]
PCS2	unknown	IRF	10	3	NA	3	[218]
PCS3	unknown	KLF	11	2	NA	2	[218]
Wnt	LEF/TCF	LEF/TCF	319	0	0	16	[299]
Hh	GLI	GLI	639	0	0	4	[299]
TGFB	SMAD	SMAD	349	0	2	11	[299]
TGFB	EGR	EGR	349	3	6	9	[299]
Myc	MYC	MYC	1095	4	10	11	[296]
TP53	TP53	TP53	649	3	5	5	[300]
NFKB	NFKB	NFKB	187	3	5	5	[301]
E2F	E2F	E2F	1036	10	26	32	[299]
FOX	FOX	FOX	327	2	10	43	[302]
Myb	MYB	MYB	341	1	2	9	[303]
Sox4	SOX4	SOX4	739	0	1	36	[56]

Table 5.1. **TFBSET was run on multiple gene sets to identify enriched TFBSs in proximal promotor regions.** Items highlighted in gray are ligands. The Total # TFBSs" is the total

number of significantly overrepresented TFBSs identified. The “# in top 10” and “# in top 50” are the number of times the expected TFBS (based on previous literature) was identified in the top 10 or top 50 ranked TFBSs respectively. The “Total # Targets” is the total number of expected TFBS identified out of all the TFBSs identified. TFBSs are ranked by Z score.

The analysis was performed with proximal promoter and enhancer regions (Supplemental Table S5.4) but inclusion of the enhancer regions did not improve performance. This may be due to the fact that enhancer regions may not regulate the closest gene as is assumed by the algorithm.

For the unknown gene sets and the genes sets with expected TFBSs, TFBSET identified at least 1 expected TFBS in the top 10 ranked TFBSs except for the Sox4 gene set. For the ligand gene sets, TFBSET was not as successful. For the TGFβ gene set, three EGR TFBSs were identified in the top 10 TFBSs and no SMAD TFBSs were identified in the top 10. Two SMAD TFBSs were identified in the top 50 TFBSs. For the gene sets for the Wnt and Hedgehog ligands, TFBSET did not find any of the expected TFBSs in the top 50 TFBSs, though the expected TFBSs were identified among the enriched TFBSs identified.

		Enhancer regions excluded					Enhancer regions included				
Gene	**Expected TFBS**	**TFBS identified**	**Total # TFBSs**	**# in Top 10**	**# in Top 50**	**Total # targets**	**TFBS Identified**	**Total TFBS**	**# in top 10**	**# in top 50**	**total #**
Sig24	unknown	FOX	50	4	20	20	MEF	133	1	1	2
PCS1	unknown	NFY	777	5	5	5	NFY	730	5	5	5
PCS1	unknown	E2F	777	5	10	26	E2F	730	4	10	25
PCS2	unknown	IRF	10	3	NA	3	MEF	106	1	1	2
PCS3	unknown	KLF	11	2	NA	2	TBX	99	1	1	1
Wnt	LEF	LEF	319	0	0	5	LEF	274	0	0	4
Wnt	TCF	TCF	319	0	0	11	TCF	274	0	0	13
Hh	GLI	GLI	639	0	0	4	GLI	716	0	1	4
TGFB	SMAD	SMAD	349	0	2	11	SMAD	420	0	0	10
TGFB	EGR	EGR	349	3	6	9	EGR	420	0	1	9
Myc	MYC	MYC	1095	4	10	11	MYC	1112	0	5	11
TP53	TP53	TP53	649	3	5	5	TP53	609	1	1	5
NFKB	NFKB	NFKB	187	3	5	5	NFKB	323	0	1	5
E2F	E2F	E2F	1036	10	26	32	E2F	1034	5	23	32
FOX	FOX	FOX	327	2	10	43	Fox	403	0	1	46
Myb	MYB	MYB	341	1	2	9	Myb	316	0	1	8
Sox4	SOX4	SOX4	739	0	1	36	Sox	735	0	0	37

Supplemental Table S5.4. Comparison of TFBSET results of gene sets run excluding and including enhancer regions shows that the results are generally less accurate when the enhancer regions are included. Gene sets highlighted in gray are ligands.

Comparison to Other Software

TFBSET was compared to four other TFBS enrichment programs, DiRE [284], PASTAA [285] , oPOSSUM [286] and PSCAN [287] (Table 5.2). DiRE queries evolutionary conserved regions (ECRs) for TFBSs and then uses an F^t scoring function [304] by assigning weights to the TFBSs. oPOSSUM also uses ECRs based on a technique called phylogenetic footprinting [305]. PASTAA uses a TF affinity approach. The affinities of TFs to the gene promoters is predicted by TRAP (Transcription factor Affinities Prediction) [306]. The higher a TF affinity to a promoter, the more likely the TF plays an important role. PSCAN does not resort to limiting queries to ECRs. DiRE, PASTAA and oPOSSUM were run using ECR. TFBSET and oPOSSUM queried ±5000bp around the TSS. PASTAA and PSCAN did not have the option for ±5000bp so the setting with the largest region around the TSS was used.

All programs were run on gene sets of known TF targets (Myc, TP53, E2F, NFκB), genes upregulated by ligands (TGFβ, Wnt) and the genes upregulated in PCS1 prostate cancers [218]. All the programs successfully identified the appropriate TFBS for the genes upregulated by E2F and NFκB. Of the top 10 ranked TFBSs, there were at least 2 associated with E2F and NFκB for all programs except for oPOSSUM which found only 1 occurrence of E2F in the top 10. For the gene sets upregulated by Myc and TP53, TFBSET did the best with at least 3 of the top 10 ranked TFBSs. Both PASTAA and PSCAN identified at least one Myc and one TP53 among the 10 top ranked TFBS. DiRE identified Myc one time at rank 20 and TP53 at rank 110. oPOSSUM identified one Myc in the top 10 and one TP53 at rank 69.

Gene Set	Reference	# of genes	TFBSET TRANSFAC	DiRE TRANSFAC	PASTAA TRANSFAC	PSCAN-TRANSFAC	PSCAN-JASPAR	oPOSSUM JASPAR
Myc	(34)	230	Very high	Low	High	Very high	High	Med1
TP53	(38)	96	Very high	Low	Med1	Med1	Low	Low
E2F	(37)	200	Very high	Very high	Very high	Very high	Very high	Med1
NFKB	(39)	91	Very high	High	Very high	Very high	High	High
TGFB-EGR	(37)	54	Very high	Med1	High	High	Very high	Low
TGFB-SMAD	(37)	54	Med2	Low	Low	None	Low	None
Wnt-TCF	(37)	42	Med2	Med1	Low	Low	Med2	Low
Wnt-LEF	(37)	42	Low	Low	Low	None	Low	None
PCS1-NFY	(23)	85	Very high	Low	Med1	Very high	Very high	Med1
PCS1-E2F	(23)	85	Very high	Low	Very high	High	Med2	Med1
Genome			Hg19	Hg18,Mouse,Rat	Human, Mouse	Human	Human	Human
Promoter region			±5000 bp	Promoter + Distal	-10,000 to +0 bp	-1000 to +0 bp	-1000 to +0 bp	±5000 bp
Filter			H3K27ac, DNaseI	ECR	ECR	None	None	ECR
Speed			30 min	<1 min	<1 min	<1 min	<1 min	4 min

Category	Description
Very high	3 or more in top 10
High	2 in top 10
Med1	1 in top 10
Med2	10 or more hits
Low	less than 10 hits
None	0 hits

Table 5.2. Comparison of TFBSET to other popular programs. Opossum can only query against the JASPAR database. PSCAN can query both TRANSFAC and JASPAR. TFBSET, DiRE and PASTAA query against the TRANSFAC database. Benchmark speed was based on the Myc gene set but may vary depending on network speed, network connection and number of users. Promoter regions are based on distance from TSS.

The results for the ligands were not as successful as for TF perturbation studies. For genes upregulated by overexpression of the TGFβ, both TFBSET, PASTAA and PSCAN identified at least 2 EGR TFBSs in the top 10 overrepresented TFBSs. DiRE identified only one EGR TFBS (EGR1) with a ranking of 8 while oPOSSUM did not find any in the top 10. Since oPOSSUM queries the JASPAR database, it was possible that there were fewer TFBS matrices in JASPAR than in TRANSFAC. To demonstrate the two databases are comparable, PSCAN was also run on the JASPAR database and 4 EGR TFBSs were among the top 10 overrepresented TFBSs. All programs had a difficult time identifying the SMAD TFBSs as no SMAD matrices were identified in the top 10 for any program. However, TFBSET did identify 11 SMAD matrices as overrepresented.

For genes upregulated from Wnt overexpression, TFBSET found more than 10 TFBSs for TCF and 6 for LEF, but the ranks were quite low (from 61 to 396). DiRE, PASTAA and oPOSSUM found fewer than 10 TFBSs for TCF and LEF. DiRE did identify one TCF TFBS (TCF4) as the top ranked TFBS, but it was the only TCF TFBS identified. PSCAN identified 10 TCF sites when using the JASPAR database, none of which were in the top 10.

None of the programs found more than 10 LEF TFBS among the promoters of the Wnt gene set. PSCAN did not find any LEF sites when using the TRANSFAC database, but PSCAN did find one occurrence at rank 534 when using the JASPAR database. oPOSSUM did not find any LEF sites.

For the PCS1 gene set, TFBSET and PSCAN strongly identified both the NFY and E2F TFBSs. PASTAA also strongly identified E2F and identified NFY as the top enriched TFBSs; however, this was the only occurrence of NFY identified. oPOSSUM identified one NFY and one E2F TFBSs in the top 10. DiRE did identify both E2F and NFY but the frequency of the TFBSs was low.

5.5 Discussion

TFBSET was used to analyze gene sets upregulated by TFs and ligands and the identified overrepresented TFBSs coincided with the expected results with higher precision than the other software tools. PASTAA and PSCAN performed well in comparison but were weaker in identifying the TFBS for the ligands TGFB and Wnt. DiRE and oPOSSUM appeared to be the least precise. Though oPOSSUM only queries the JASPAR database, PSCAN performed well when run on either the TRANSFAC or JASPAR databases.

When TFBSET was used to analyze genes upregulated in prostate cancer subtype 1 (PCS1), the most aggressive subtype, TFBSET was able to identify very clearly that NFY and E2F were key TFs involved in regulation of the PCS1 gene signature. All of the top 10 enriched TFBSs were either NFY or E2F. This result suggests that TFBSET can identify TFs important in disease progression. When TFBSET was run on PCS2 or PCS3, the enriched TFBSs identified were IRF and KLF respectively. However, these results were not as strong as for PCS1. For PCS2, though the highest ranking TFBS identified was for IRF, there were three IRF TFBSs within the top 10 ranked TFBS. For PCS3, there were only two KLF TFBSs in the top 10 ranked TFBS. Though IRF and KLF were overrepresented in PCS2 and PCS3, the lower correlation may indicate that PCS2 and PCS3 are not as dependent on these TFs for growth and proliferation. Not all diseases are caused by aberrant TFs, so it is not expected that TFBSET can identify TF targets in all diseases; however, the fact that NFY and E2F were identified in PCS1 provides new insights into the understanding of the PCS1 subtype and the differences between the three prostate cancer subtypes.

PASTAA and PSCAN also identified NFY and E2F as overrepresented among the promoters of the PCS1 gene set and oPOSSUM to a lesser extent. DiRE had difficulty identifying NFY and E2F in the PCS1 gene set.

When TFBSET was used to analyze gene targets of signal transduction pathways, TFBSET was, for the most part, able to identify the expected TFBSs. However, there were many additional significantly enriched TFBSs identified. For instance, the known downstream TF mediators of the Wnt signaling pathway are TCF and LEF. TFBSET analysis of a gene list of 42 genes upregulated by overexpression of Wnt identified significant enrichment of 20 TCF and LEF TFBSs. However, there were 380 other TFBSs that were also significantly enriched. Also, the TCF/LEF TFBSs were not the most significant (highest ranking) TFBSs. This may be due to the fact that the Wnt signaling pathway is complex, and multiple TFs besides TCF/LEF can play important roles in mediating the effects of Wnt signals. These additional factors may affect other TFs and thus dilute the results. Additional analysis of the other 380 TFBSs may suggest new hypotheses that could be tested to provide a better understanding of the Wnt signaling pathway.

The enhanced performance of TFBSET may be due to differences in the query regions used for TFBS analysis. DiRE, PASTAA and oPOSSUM limit the query region for TFBSs to evolutionary conserved regions (ECRs) between human and mouse, but do not account for chromatin accessibility or marks of active promoter regions. TFBSET queries the proximal promoter regions that are marked by H3K27ac and DNaseI hypersensitivity, thus ensuring that only open chromatin regions are being queried and does not require that these regions be conserved evolutionarily. DiRE queries the promoter regions of genes and potential enhancer regions but does not allow the user to enter a range. PASTAA allows the user to select from one of five different regions around the TSS to query: -10,000bp to 0bp, ±400bp, ±200bp, -400bp to 0, and -200bp to 0. TFBSET is substantially more flexible than both programs, allowing the user to specify the query regions around the TSS from ±100bp to ±5000bp in increments of 100bp.

An additional feature of DiRE is the ability to query distal regulatory elements in addition to proximal promoter regions using a method called Enhancer Identification. The sensitivity and precision are only 28% and 50% respectively, thus limiting its usefulness. Enhancer regions

may activate genes that are at quite a distance (hundreds of thousands of base pairs) from the TSS. TFBSET also has an option to include enhancer regions in the enrichment calculations; however, inclusion of enhancer regions appears to reduce its accuracy using the known gene sets (Supplemental Table S2), possibly because TFBSET assumes that the enhancer regions regulate the nearest gene. Though mammalian studies have not been performed, in *Drosophila*, only 88% of enhancers regulate their closest neighbor [307] whereas promoters virtually always regulate their target gene. Thus, including enhancer regions that do not regulate genes in the gene set may reduce overall accuracy. Future research examining 3D topology of enhancer/TSS interactions will be necessary to identify the gene targets of enhancer regions in mammalian genomes.

One caveat for TFBSET is that there are some genes that do not have regions of H3K27ac or DNaseI hypersensitivity in their promoters, and these genes are excluded from the TFBSET analysis. It is possible that there are open chromatin regions in the promoter regions of these genes, but they were not identified in the current ENCODE datasets. These genes may also be controlled by distal enhancer regions. Further examination of these genes should be performed to identify the regulatory regions of these genes. These regulatory regions can be added to the TFBSET database to improve its accuracy.

Another issue is that TFBSET is much slower than other similar tools, but this limitation can likely be overcome with more powerful hardware, parallelization of the software, or both.

In conclusion, TFBSET was very accurate when used to analyze gene sets containing genes upregulated by a known TF. TFBSET was more accurate than currently available software in the identification of downstream TFs of ligand binding receptors. TFBSET was able to identify two TFBSs overrepresented in PCS1, but not PCS2 and PCS3. This demonstrates that TFBSET can be a valuable tool in determining differences between various subtypes of

diseases. For diseases that do rely on TFs for progression, the TFs identified by TFBSET may be potential targets for therapy.

DATA AVAILABILITY

TFBSET is freely available for non-commercial users at https://tfbset.bmi.emory.edu/tfbset.html

ACKNOWLEDGEMENTS

The authors want to thank James P. Kinney III for technical assistance.

FUNDING

This work has been supported by the Department of Biomedical Informatics, Emory University School of Medicine, the Graduate Division of Biological and Biomedical Sciences, and the Winship Cancer Institute.

Conflict of Interest: none declared.

Chapter 6. Discussion and future directions

The heterogeneous nature of prostate cancer tumors requires complex and dynamic changes in transcriptional networks to adapt to the various stages of cancer progression. While numerous studies have investigated how tumors acquire ADT-resistance, or become metastatic, via genomic mutations or fluctuations in expression profiles, less is known about the regulatory adaptations that promote aggressiveness. The work presented in this dissertation combines a top-down computational approach that integrates multiple, curated, datatypes to uncover key drivers that mediate differences in tumor biology, and a bottom-up experimental approach that investigates the role of SOX4 in EMT, a critical step of the metastatic cascade.

6.1 Predicting transcriptional regulators of ADT response and metastasis

6.1.1 Early divergent transcriptional response to ADT could reflect an ADT mediated clonal selection

Using a novel cohort in which all but one patient received neoadjuvant ADT alone, we interrogated the changes in tumor biology after ADT. The high impact group displayed a substantially more pronounced transcriptional response than the low impact group, and exhibited an upregulation of a distinct set of genes. All the patients' tumors had decreased expression of the AR-dependent PCS2 subtype, which is an expected response to initial ADT treatment. Similarly, it is also logical that the AR-dependent PCS2 subtype signature would decrease after inhibition of AR signaling. Interestingly, expression of the aggressive PCS1 subtype genes was uniquely retained, and increased in the high impact group after ADT. We interpret these data as suggesting that the ADT is not only selecting against populations of cells sensitive to AR signaling, but also selecting for cells able to activate transcriptional programs and adapt to the inhibition of AR signaling. Moreover, recent evidence from single cell sequencing analysis of circulating tumor cells (CTCs) suggests that ADT can mediate the expansion of clonal populations with multiple genomic amplifications, such as AR, in response

to treatment[308]. Intriguingly, despite relatively short ADT exposure times, we identified drastically divergent transcriptional responses. Concordantly, Dago *et al.* found that ADT induced copy number changes in metastatic CTCs arose relatively quickly after initial ADT administration[308]. These findings may indicate that administration of drugs earlier in prostate cancer development may have greater beneficial effects.

6.1.2 Distinct transcription factor associations may drive transcriptional programs that influence prostate cancer aggressiveness.

The initiation and progression of any cancer depends on the dysregulation of normal transcriptional programs. This underscores the importance of identifying the key regulators of these programs. Many studies utilize expression profiling alone to elucidate critical gene signatures important to prostate cancer[232]. We aimed to find common transcription factor combinatorial relationships that putatively regulate the transcriptional programs associated with ADT response and metastasis. These TFCG associations may serve as critical upstream hubs that promote ADT-resistance and metastasis.

The common TFCGs between the high impact group and Met.PCS1 networks can reflect both spatial and temporal interactions that are maintained after ADT, and during metastasis. Intriguingly, these TFCG relationships remained despite targeting a distinct set of genes between the networks. This is consistent with the findings of Yanez-Cuna *et al.* who observed that combinatorial binding of the pro-EMT, developmental transcription factors Twist and Zelda forced the relocalization of the complex to a different set of loci that were bound when the factors were not interacting with each other[309]. Furthermore, direct PPIs between Jun and Fos, members of an overlapping TFCG in this analysis, are known to promote the relocalization of the complex to sites that are independent of the individual factors[310, 311]. These findings support the biological relevance of our data, and demonstrate that TFCG associations may remain intact despite context-dependent changes in predicted target genes.

Additionally, the promoters of predicted target genes do not need to contain the binding site motifs of all TFCG members. Instead, transcription factors may collaboratively bind at enhancers to promote expression of cognate genes, suggesting that direct interactions are not necessary to regulate the same genes[239]. Therefore, interrogating the binding site motifs in enhancers of target genes may be one way to validate predicted TFCG associations.

Finally, members of a TFCG may also be temporally related, and bind the same target genes sequentially. For example, the FOXA and GATA families are well-known pioneer factors. These factors bind to nucleosomal DNA in closed chromatin regions to facilitate access for additional TFs[312]. Interestingly, pioneer factors, and non-pioneer factors such as SOX2, may require the presence of other specific transcription factors to bind target genes by “mass action”, highlighting one of the ways transcription factors within a group can be interdependent[239].

6.1.3 Future Directions

The data presented in this analysis provide the framework for developing new, and more robust, treatment strategies and therapeutics to prevent resistance and metastasis. This top-down predictive computational approach integrates multiple curated datatypes containing experimental evidence and uncovers important putative hubs. Consequently, there are numerous avenues one can pursue to investigate the translational potential of the overlapping TFCGs.

We observed drastic changes in tumor biology that occurred relatively rapidly after initial ADT administration. Our analysis revealed putative drivers that, we predict, become active at an early stage and remain active throughout progression to metastasis. One limitation is that we could not acquire a dataset of matched metastatic and primary tumors of patients that had failed to respond to ADT. Therefore, future studies could investigate whether TFCG associations are maintained at sequential timepoints in matched samples throughout progression of the hormone-naive primary tumor, to CRPC and metastasis via tumor sampling, or isolation of circulating tumor cells. Performing these analyses in multiple patients could also inform the ubiquitous influence of the TFCGs and support their use as robust drug targets. Furthermore, it would be interesting to see whether TFCGs differed based on the type of ADT GnRH agonists/antagonists, AR-inhibitors, and androgen synthesis inhibitors, administered.

Another noteworthy finding in our analysis was that the Met.PCS1 network formed a subset of the larger high impact network. This suggested an ADT-mediated refinement of TFCG associations that may also drive metastasis. It is unknown whether the overlapping TFCGs characterize existing cell populations in the hormone naïve tumor that were selected for by ADT, or whether these TFCGs associate with ADT-mediated differentiation of *de novo* clonal populations. Gundem *et al.* used mutant allele fractions to characterize subclonal populations in tumors to infer the fraction of tumor cells that harbored the same mutations[313]. It would be

intriguing to investigate the dynamic changes in clonal populations in the post-ADT and CRPC metastatic tumors, and determine whether these cell populations existed in the hormone naïve localized tumor. This might inform the likelihood that overlapping TFCG associations, found in our analysis, were likely present in minor clonal populations before ADT. Therefore, these groups could be utilized as potential prophylactic drug targets, prior to ADT, to inhibit metastatic CRPC.

Metastatic tumors are highly heterogeneous, varying in the degree of intra-tumor AR expression, suggesting that other AR-independent signaling pathways can drive CRPC and metastasis[37, 313, 314, 315]. The overlapping TFCGs provide initial insights into actionable pathways that could be targeted to supplement ADT treatment. Indeed, others have found signaling pathways that can either supplement or bypass AR-signaling upon initial repression of AR activity[16]. Accordingly, there may be critical signaling pathways that are overrepresented in the members of a TFCG. Future studies could utilize combination inhibitors to repress these signaling pathways, and interrogate whether there are synergistic effects that reverse an ADT-resistant, CRPC and/or metastatic phenotype.

It is also possible to interrogate the function and effects of TFCG associations via molecular techniques. Moreover, because the overlapping TFCGs are significantly associated with metastasis, these transcription factors may also be experimentally evaluated for involvement in osteoblast development, or in promotion of a metastatic tumor microenvironment. The findings from our analysis can be used as a resource for experimental biologists who study specific transcription factors that are also found in a TFCG. The results from these experimental studies would be a valuable supplement to our analysis.

Finally, inhibiting TFCG activities *in vitro*, and *in vivo* might also reveal putative high yield therapeutics that may evaluated via clinical trials. For example, it would be intriguing to ascertain the signaling pathways enriched in an overlapping TFCG, and evaluate whether

inhibition of these pathways prevented CRPC and/or metastasis phenotypes. Our IPA analysis found a strong upregulation of genes associated with MAPK signaling, including FOS, FOSB, and JUN. Concordantly, we observed that JUND, FOS, FOSB, and several other Fos/Jun family members were members of a TFCG that associated with both the high impact and metastatic networks. This suggests a possible compensatory activation of MAPK signaling in response to AR signaling repression. Consequently, concomitant inhibition of AR- and MAPK signaling pathways may have prevent castration resistance. The Moreno lab is currently exploring the synergistic effects of combinatorial administration of enzalutamide and inhibitors of the MAPK signaling downstream kinases, MEK, ERK and JNK. Thus, the overlapping TFCGs in this analysis may aid in prioritizing treatment strategies and combination therapies.

6.2 SOX4 promotes EMT in prostate cancer cells

Previous evidence, and our preliminary data support the compelling argument that SOX4 may be an attractive drug target. The role of SOX4 as a transforming oncogene in prostate cancer is well established[56], and our data imply that SOX4 continues to remain active to regulate EMT. Previous evidence suggests that SOX4 is upstream of quintessential EMT transcription factors Snail, Zeb, and Twist[73]. Our data demonstrate that SOX4 functions downstream of the TGF-β signaling pathway, and is sufficient, to initiate the EMT program in prostate cancer cells. We contend that SOX4 is also necessary for EMT, and that it recruits PCAF and/or TRRAP to transcriptionally activate target genes important for the process. To fully understand how SOX4 regulates EMT it will be necessary to identify direct targets in the pathway from indirect and downstream events. Identification of SOX4 target genes critical for EMT induction will yield new insights into the mechanisms by which SOX4 regulates EMT, and may unveil new therapeutic targets or relevant molecular markers of metastatic disease.

While the ARACNE analysis predicted many SOX4 epigenetic related gene targets across multiple cancer types, less is known regarding how SOX4 impacts epigenome in response to EMT. Our preliminary data indicate that SOX4 targets the DNA methyltransferases DNMT1, DNMT3A, and DNMT3B. Assuming SOX4 deletion indeed inhibits EMT, resulting in a subsequent downregulation of DNA methyltransferases, it would be interesting to observe changes in DNA methylation profiles when SOX4 is restored. Moreover, if forced expression of DNA methyltransferases without SOX4 also restored EMT, then genes that regain methylation and are silenced, may potentially be important for maintaining an epithelial state.

The first obvious and important step to supporting our hypothesis that SOX4 recruits PCAF/TRRAP to promote EMT would be to test if EMT is inhibited when these interactions are inhibited. These interactions suggest that SOX4 may play an active role in facilitating open chromatin. It would be interesting to determine how the interactions among PCAF, TRRAP and

SOX4 impact global changes in promoter and enhancer accessibility, and whether these changes affected the expression of pro-EMT SOX4 targets. If the interactions among these three proteins were indeed critical for EMT, then identification of critical domains that mediate the interactions may also serve as intriguing drug targets.

The future experiments outlined here could identify critical mechanisms through which SOX4 regulates transcriptional activity during EMT, and the important target genes that may be essential for metastatic prostate cancer.

1. Greenlee RT, Hill-Harmon MB, Murray T, Thun M. Cancer statistics, 2001. *CA Cancer J Clin* **51**, 15-36 (2001).

2. Hjelmborg JB*, et al.* The heritability of prostate cancer in the Nordic Twin Study of Cancer. *Cancer Epidemiol Biomarkers Prev* **23**, 2303-2310 (2014).

3. Gleason DF. Histologic grading of prostate cancer: a perspective. *Hum Pathol* **23**, 273-279 (1992).

4. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin* **67**, 7-30 (2017).

5. Roy AK*, et al.* Regulation of androgen action. *Vitam Horm* **55**, 309-352 (1999).

6. Anderson KM, Liao S. Selective retention of dihydrotestosterone by prostatic nuclei. *Nature* **219**, 277-279 (1968).

7. Kishimoto M*, et al.* Nuclear receptor mediated gene regulation through chromatin remodeling and histone modifications. *Endocr J* **53**, 157-172 (2006).

8. Mangelsdorf DJ*, et al.* The nuclear receptor superfamily: the second decade. *Cell* **83**, 835-839 (1995).

9. Chen Y, Clegg NJ, Scher HI. Anti-androgens and androgen-depleting therapies in prostate cancer: new agents for an established target. *Lancet Oncol* **10**, 981-991 (2009).

10. Crona DJ, Whang YE. Androgen Receptor-Dependent and -Independent Mechanisms Involved in Prostate Cancer Therapy Resistance. *Cancers (Basel)* **9**, (2017).

11. Jensen EV, Suzuki T, Kawashima T, Stumpf WE, Jungblut PW, DeSombre ER. A two-step mechanism for the interaction of estradiol with rat uterus. *Proc Natl Acad Sci U S A* **59**, 632-638 (1968).

12. Centenera MM, Harris JM, Tilley WD, Butler LM. The contribution of different androgen receptor domains to receptor dimerization and signaling. *Mol Endocrinol* **22**, 2373-2382 (2008).

13. Bennett N, Hooper JD, Lee CS, Gobe GC. Androgen receptor and caveolin-1 in prostate cancer. *IUBMB Life* **61**, 961-970 (2009).

14. Lamont KR, Tindall DJ. Minireview: Alternative activation pathways for the androgen receptor in prostate cancer. *Mol Endocrinol* **25**, 897-907 (2011).

15. Huggins CaH, C.V. Studies on Prostate Cancer. *Cancer Research* **1**, 5 (1941).

16. Watson PA, Arora VK, Sawyers CL. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat Rev Cancer* **15**, 701-711 (2015).

17. Harris WP, Mostaghel EA, Nelson PS, Montgomery B. Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. *Nat Clin Pract Urol* **6**, 76-85 (2009).

18. Visakorpi T*, et al.* In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* **9**, 401-406 (1995).

19. Chen CD*, et al.* Molecular determinants of resistance to antiandrogen therapy. *Nat Med* **10**, 33-39 (2004).

20. Denis LJ, Griffiths K. Endocrine treatment in prostate cancer. *Semin Surg Oncol* **18**, 52-74 (2000).

21. Agus DB*, et al.* Prostate cancer cell cycle regulators: response to androgen withdrawal and development of androgen independence. *J Natl Cancer Inst* **91**, 1869-1876 (1999).

22. Taplin ME*, et al.* Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. *N Engl J Med* **332**, 1393-1398 (1995).

23. Wilson CM, McPhaul MJ. A and B forms of the androgen receptor are expressed in a variety of human tissues. *Mol Cell Endocrinol* **120**, 51-57 (1996).

24. Li Y, Alsagabi M, Fan D, Bova GS, Tewfik AH, Dehm SM. Intragenic rearrangement and altered RNA splicing of the androgen receptor in a cell-based model of prostate cancer progression. *Cancer Res* **71**, 2108-2117 (2011).

25. Dehm SM, Schmidt LJ, Heemers HV, Vessella RL, Tindall DJ. Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res* **68**, 5469-5477 (2008).

26. Hu R*, et al.* Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res* **69**, 16-22 (2009).

27. Li Y, Chan SC, Brand LJ, Hwang TH, Silverstein KA, Dehm SM. Androgen receptor splice variants mediate enzalutamide resistance in castration-resistant prostate cancer cell lines. *Cancer Res* **73**, 483-489 (2013).

28. Tepper CG*, et al.* Characterization of a novel androgen receptor mutation in a relapsed CWR22 prostate cancer xenograft and cell line. *Cancer Res* **62**, 6606-6614 (2002).

29. Martin SK, Banuelos CA, Sadar MD, Kyprianou N. N-terminal targeting of androgen receptor variant enhances response of castration resistant prostate cancer to taxane chemotherapy. *Mol Oncol* **9**, 628-639 (2015).

30. Gregory CW*, et al.* A mechanism for androgen receptor-mediated prostate cancer recurrence after androgen deprivation therapy. *Cancer Res* **61**, 4315-4319 (2001).

31. Takahashi S*, et al.* Noncanonical Wnt signaling mediates androgen-dependent tumor growth in a mouse model of prostate cancer. *Proc Natl Acad Sci U S A* **108**, 4938-4943 (2011).

32. Korpal M*, et al.* An F876L mutation in androgen receptor confers genetic and phenotypic resistance to MDV3100 (enzalutamide). *Cancer Discov* **3**, 1030-1043 (2013).

33. Cao B*, et al.* Androgen receptor splice variants activating the full-length receptor in mediating resistance to androgen-directed therapy. *Oncotarget* **5**, 1646-1656 (2014).

34. Ford OH, 3rd, Gregory CW, Kim D, Smitherman AB, Mohler JL. Androgen receptor gene amplification and protein expression in recurrent prostate cancer. *J Urol* **170**, 1817-1821 (2003).

35. Karantanos T, Corn PG, Thompson TC. Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches. *Oncogene* **32**, 5501-5511 (2013).

36. Carreira S*, et al.* Tumor clone dynamics in lethal prostate cancer. *Sci Transl Med* **6**, 254ra125 (2014).

37. Robinson D*, et al.* Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell* **162**, 454 (2015).

38. Grasso CS*, et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239-243 (2012).

39. Vervoort SJ, van Boxtel R, Coffer PJ. The role of SRY-related HMG box transcription factor 4 (SOX4) in tumorigenesis and metastasis: friend or foe? *Oncogene* **32**, 3397-3409 (2013).

40. Schilham MW*, et al.* Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. *Nature* **380**, 711-714 (1996).

41. Cheung M, Abu-Elmagd M, Clevers H, Scotting PJ. Roles of Sox4 in central nervous system development. *Brain Res Mol Brain Res* **79**, 180-191 (2000).

42. Hunt SM, Clarke CL. Expression and hormonal regulation of the Sox4 gene in mouse female reproductive tissues. *Biol Reprod* **61**, 476-481 (1999).

43. Scharer CD, McCabe CD, Ali-Seyed M, Berger MF, Bulyk ML, Moreno CS. Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells. *Cancer Res* **69**, 709-717 (2009).

44. van de Wetering M, Oosterwegel M, van Norren K, Clevers H. Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO J* **12**, 3847-3854 (1993).

45. Ruebel KH*, et al.* Effects of TGFbeta1 on gene expression in the HP75 human pituitary tumor cell line identified by gene expression profiling. *Endocrine* **33**, 62-76 (2008).

46. Lowry WE, Blanpain C, Nowak JA, Guasch G, Lewis L, Fuchs E. Defining the impact of beta-catenin/Tcf transactivation on epithelial stem cells. *Genes Dev* **19**, 1596-1611 (2005).

47. Kobielak K, Stokes N, de la Cruz J, Polak L, Fuchs E. Loss of a quiescent niche but not follicle stem cells in the absence of bone morphogenetic protein signaling. *Proc Natl Acad Sci U S A* **104**, 10063-10068 (2007).

48. Prior HM, Walter MA. SOX genes: architects of development. *Mol Med* **2**, 405-412 (1996).

49. Dy P, Penzo-Mendez A, Wang H, Pedraza CE, Macklin WB, Lefebvre V. The three SoxC proteins--Sox4, Sox11 and Sox12--exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res* **36**, 3101-3117 (2008).

50. Rhodes DR, Chinnaiyan AM. Integrative analysis of the cancer transcriptome. *Nat Genet* **37 Suppl**, S31-37 (2005).

51. Sinner D*, et al.* Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity and proliferation of colon carcinoma cells. *Mol Cell Biol* **27**, 7802-7815 (2007).

52. Beekman JM*, et al.* Syntenin-mediated regulation of Sox4 proteasomal degradation modulates transcriptional output. *Oncogene* **31**, 2668-2679 (2012).

53. Ramezani-Rad P*, et al.* SOX4 enables oncogenic survival signals in acute lymphoblastic leukemia. *Blood* **121**, 148-155 (2013).

54. Pramoonjago P, Baras AS, Moskaluk CA. Knockdown of Sox4 expression by RNAi induces apoptosis in ACC3 cells. *Oncogene* **25**, 5626-5639 (2006).

55. Huang YW*, et al.* Epigenetic repression of microRNA-129-2 leads to overexpression of SOX4 oncogene in endometrial cancer. *Cancer Res* **69**, 9038-9046 (2009).

56. Liu P*, et al.* Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells. *Cancer Res* **66**, 4011-4019 (2006).

57. Bello D, Webber MM, Kleinman HK, Wartinger DD, Rhim JS. Androgen responsive adult human prostatic epithelial cell lines immortalized by human papillomavirus 18. *Carcinogenesis* **18**, 1215-1223 (1997).

58. Tavazoie SF*, et al.* Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* **451**, 147-152 (2008).

59. Bilir B, Kucuk O, Moreno CS. Wnt signaling blockage inhibits cell proliferation and migration, and induces apoptosis in triple-negative breast cancer cells. *J Transl Med* **11**, 280 (2013).

60. Zhang J*, et al.* SOX4 induces epithelial-mesenchymal transition and contributes to breast cancer progression. *Cancer Res* **72**, 4597-4608 (2012).

61. Wang L*, et al.* SOX4 is associated with poor prognosis in prostate cancer and promotes epithelial-mesenchymal transition in vitro. *Prostate Cancer Prostatic Dis* **16**, 301-307 (2013).

62. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer* **9**, 265-273 (2009).

63. Zhu Z*, et al.* Anoikis and metastatic potential of cloudman S91 melanoma cells. *Cancer Res* **61**, 1707-1716 (2001).

64. Xu J, Lamouille S, Derynck R. TGF-beta-induced epithelial to mesenchymal transition. *Cell Res* **19**, 156-172 (2009).

65. Pena C*, et al.* The expression levels of the transcriptional regulators p300 and CtBP modulate the correlations between SNAIL, ZEB1, E-cadherin and vitamin D receptor in human colon carcinomas. *Int J Cancer* **119**, 2098-2104 (2006).

66. Fujita N, Jaye DL, Kajita M, Geigerman C, Moreno CS, Wade PA. MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. *Cell* **113**, 207-219 (2003).

67. Bolos V, Peinado H, Perez-Moreno MA, Fraga MF, Esteller M, Cano A. The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. *J Cell Sci* **116**, 499-511 (2003).

68. Burk U*, et al.* A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep* **9**, 582-589 (2008).

69. Hajra KM, Chen DY, Fearon ER. The SLUG zinc-finger protein represses E-cadherin in breast cancer. *Cancer Res* **62**, 1613-1618 (2002).

70. Yang J*, et al.* Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927-939 (2004).

71. Vandewalle C*, et al.* SIP1/ZEB2 induces EMT by repressing genes of different epithelial cell-cell junctions. *Nucleic Acids Res* **33**, 6566-6578 (2005).

72. Bindels S*, et al.* Regulation of vimentin by SIP1 in human epithelial breast tumor cells. *Oncogene* **25**, 4975-4985 (2006).

73. Tiwari N*, et al.* Sox4 is a master regulator of epithelial-mesenchymal transition by controlling ezh2 expression and epigenetic reprogramming. *Cancer cell* **23**, 768-783 (2013).

74. Chen H, Tu SW, Hsieh JT. Down-regulation of human DAB2IP gene expression mediated by polycomb Ezh2 complex and histone deacetylase in prostate cancer. *J Biol Chem* **280**, 22437-22444 (2005).

75. de Groot AE, Roy S, Brown JS, Pienta KJ, Amend SR. Revisiting Seed and Soil: Examining the Primary Tumor and Cancer Cell Foraging in Metastasis. *Mol Cancer Res* **15**, 361-370 (2017).

76. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *J Clin Invest* **119**, 1420-1428 (2009).

77. Halfon MS*, et al.* Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell* **103**, 63-74 (2000).

78. Wang Q*, et al.* Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138**, 245-256 (2009).

79. Yuan X, Balk SP. Mechanisms mediating androgen receptor reactivation after castration. *Urol Oncol* **27**, 36-41 (2009).

80. Fiore C, Cohen BA. Interactions between pluripotency factors specify cis-regulation in embryonic stem cells. *Genome Res* **26**, 778-786 (2016).

81. Lian F, Sharma NV, Moran JD, Moreno CS. The biology of castration-resistant prostate cancer. *Curr Probl Cancer* **39**, 17-28 (2015).

82. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA: a cancer journal for clinicians* **64**, 9-29 (2014).

83. Zarour L, Alumkal J. Emerging therapies in castrate-resistant prostate cancer. *Current urology reports* **11**, 152-158 (2010).

84. Enzalutamide shows strong activity in prostate cancer. *Cancer Discov* **4**, OF2 (2014).

85. Agarwal N, Di Lorenzo G, Sonpavde G, Bellmunt J. New agents for prostate cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **25**, 1700-1709 (2014).

86. Ezzell EE, Chang KS, George BJ. New agents in the arsenal to fight castrate-resistant prostate cancer. *Current oncology reports* **15**, 239-248 (2013).

87. Saad F, Miller K. Treatment options in castration-resistant prostate cancer: current therapies and emerging docetaxel-based regimens. *Urologic oncology* **32**, 70-79 (2014).

88. Beltran H*, et al.* New therapies for castration-resistant prostate cancer: efficacy and safety. *European urology* **60**, 279-290 (2011).

89. Dayyani F, Gallick GE, Logothetis CJ, Corn PG. Novel therapies for metastatic castrate-resistant prostate cancer. *Journal of the National Cancer Institute* **103**, 1665-1675 (2011).

90. Antonarakis ES, Carducci MA. Future directions in castrate-resistant prostate cancer therapy. *Clinical genitourinary cancer* **8**, 37-46 (2010).

91. Cai C*, et al.* Androgen receptor gene expression in prostate cancer is directly suppressed by the androgen receptor through recruitment of lysine-specific demethylase 1. *Cancer cell* **20**, 457-471 (2011).

92. Yuan X, Cai C, Chen S, Chen S, Yu Z, Balk SP. Androgen receptor functions in castration-resistant prostate cancer and mechanisms of resistance to new agents targeting the androgen axis. *Oncogene*, (2013).

93. Taplin ME*, et al.* Selection for androgen receptor mutations in prostate cancers treated with androgen antagonist. *Cancer research* **59**, 2511-2515 (1999).

94. Guo Z*, et al.* Regulation of androgen receptor activity by tyrosine phosphorylation. *Cancer cell* **10**, 309-319 (2006).

95. Sharma NL*, et al.* The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. *Cancer cell* **23**, 35-47 (2013).

96. Hsieh AC*, et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55-61 (2012).

97. Brenner JC, Chinnaiyan AM. Disruptive events in the life of prostate cancer. *Cancer cell* **19**, 301-303 (2011).

98. Carver BS*, et al.* Reciprocal feedback regulation of PI3K and androgen receptor signaling in PTEN-deficient prostate cancer. *Cancer cell* **19**, 575-586 (2011).

99. Gonzalez-Billalabeitia E*, et al.* Vulnerabilities of PTEN-p53-deficient prostate cancers to compound PARP/PI3K inhibition. *Cancer discovery* **4**, 896-904 (2014).

100. Chin YM, Yuan X, Balk SP, Toker A. Pten-deficient tumors depend on akt2 for maintenance and survival. *Cancer discovery* **4**, 942-955 (2014).

101. Hong SW*, et al.* NVP-BEZ235, a dual PI3K/mTOR inhibitor, induces cell death through alternate routes in prostate cancer cells depending on the PTEN genotype. *Apoptosis : an international journal on programmed cell death* **19**, 895-904 (2014).

102. Nakabayashi M*, et al.* Phase II trial of RAD001 and bicalutamide for castration-resistant prostate cancer. *BJU international* **110**, 1729-1735 (2012).

103. Templeton AJ*, et al.* Phase 2 trial of single-agent everolimus in chemotherapy-naive patients with castration-resistant prostate cancer (SAKK 08/08). *European urology* **64**, 150-158 (2013).

104. de Muga S*, et al.* Molecular alterations of EGFR and PTEN in prostate cancer: association with high-grade and advanced-stage carcinomas. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **23**, 703-712 (2010).

105. Mimeault M, Johansson SL, Batra SK. Pathobiological implications of the expression of EGFR, pAkt, NF-kappaB and MIC-1 in prostate cancer stem cells and their progenies. *PloS one* **7**, e31919 (2012).

106. Aggarwal RR, Ryan CJ, Chan JM. Insulin-like growth factor pathway: a link between androgen deprivation therapy (ADT), insulin resistance, and disease progression in patients with prostate cancer? *Urologic oncology* **31**, 522-530 (2013).

107. Fahrenholtz CD, Beltran PJ, Burnstein KL. Targeting IGF-IR with ganitumab inhibits tumorigenesis and increases durability of response to androgen-deprivation therapy in VCaP prostate cancer xenografts. *Molecular cancer therapeutics* **12**, 394-404 (2013).

108. Armstrong K*, et al.* Upregulated FGFR1 expression is associated with the transition of hormone-naive to castrate-resistant prostate cancer. *British journal of cancer* **105**, 1362-1369 (2011).

109. Mathew P*, et al.* Platelet-derived growth factor receptor inhibition and chemotherapy for castration-resistant prostate cancer with bone metastases. *Clinical cancer research : an official journal of the American Association for Cancer Research* **13**, 5816-5824 (2007).

110. Najy AJ*, et al.* Cediranib inhibits both the intraosseous growth of PDGF D-positive prostate cancer cells and the associated bone reaction. *The Prostate* **72**, 1328-1338 (2012).

111. Knudsen BS*, et al.* High expression of the Met receptor in prostate cancer metastasis to bone. *Urology* **60**, 1113-1117 (2002).

112. Lee RJ, Smith MR. Targeting MET and vascular endothelial growth factor receptor signaling in castration-resistant prostate cancer. *Cancer journal* **19**, 90-98 (2013).

113. Varkaris A, Corn PG, Gaur S, Dayyani F, Logothetis CJ, Gallick GE. The role of HGF/c-Met signaling in prostate cancer progression and c-Met inhibitors in clinical trials. *Expert opinion on investigational drugs* **20**, 1677-1684 (2011).

114. Mulholland DJ*, et al.* Pten loss and RAS/MAPK activation cooperate to promote EMT and metastasis initiated from prostate cancer stem/progenitor cells. *Cancer Res* **72**, 1878-1889 (2012).

115. Sridhar SS*, et al.* A multicenter phase II clinical trial of lapatinib (GW572016) in hormonally untreated advanced prostate cancer. *American journal of clinical oncology* **33**, 609-613 (2010).

116. Whang YE*, et al.* A phase II study of lapatinib, a dual EGFR and HER-2 tyrosine kinase inhibitor, in patients with castration-resistant prostate cancer. *Urologic oncology* **31**, 82-86 (2013).

117. Cathomas R*, et al.* Efficacy of cetuximab in metastatic castration-resistant prostate cancer might depend on EGFR and PTEN expression: results from a phase II trial (SAKK 08/07). *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 6049-6057 (2012).

118. de Bono JS*, et al.* Phase II randomized study of figitumumab plus docetaxel and docetaxel alone with crossover for metastatic castration-resistant prostate cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **20**, 1925-1934 (2014).

119. Margel D*, et al.* Metformin use and all-cause and prostate cancer-specific mortality among men with diabetes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **31**, 3069-3075 (2013).

120. Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. *Nature reviews Cancer* **11**, 85-95 (2011).

121. Koppenol WH, Bounds PL, Dang CV. Otto Warburg's contributions to current concepts of cancer metabolism. *Nature reviews Cancer* **11**, 325-337 (2011).

122. Giatromanolaki A, Koukourakis MI, Koutsopoulos A, Mendrinos S, Sivridis E. The metabolic interactions between tumor cells and tumor-associated stroma (TAS) in prostatic cancer. *Cancer biology & therapy* **13**, 1284-1289 (2012).

123. Altenberg B, Greulich KO. Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics* **84**, 1014-1020 (2004).

124. Quinn BJ, Kitagawa H, Memmott RM, Gills JJ, Dennis PA. Repositioning metformin for cancer prevention and treatment. *Trends in endocrinology and metabolism: TEM* **24**, 469-480 (2013).

125. Tomlins SA*, et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-648 (2005).

126. Tomlins SA*, et al.* Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**, 595-599 (2007).

127. Helgeson BE*, et al.* Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. *Cancer research* **68**, 73-80 (2008).

128. Tomlins SA*, et al.* TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer research* **66**, 3396-3400 (2006).

129. Wang XS*, et al.* Characterization of KRAS rearrangements in metastatic prostate cancer. *Cancer discovery* **1**, 35-43 (2011).

130. Taylor BS*, et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22 (2010).

131. Lee WH, Isaacs WB, Bova GS, Nelson WG. CG island methylation changes near the GSTP1 gene in prostatic carcinoma cells detected using the polymerase chain reaction: a new prostate cancer biomarker. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **6**, 443-450 (1997).

132. Kong D*, et al.* Epigenetic silencing of miR-34a in human prostate cancer cells and tumor tissue specimens can be reversed by BR-DIM treatment. *American journal of translational research* **4**, 14-23 (2012).

133. Wang X, Gao H, Ren L, Gu J, Zhang Y, Zhang Y. Demethylation of the miR-146a promoter by 5-Aza-2'-deoxycytidine correlates with delayed progression of castration-resistant prostate cancer. *BMC cancer* **14**, 308 (2014).

134. Kirmizis A*, et al.* Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev* **18**, 1592-1605 (2004).

135. Varambally S*, et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**, 624-629 (2002).

136. Varambally S*, et al.* Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer cell* **8**, 393-406 (2005).

137. Bachmann IM*, et al.* EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24**, 268-273 (2006).

138. Berger MF*, et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214-220 (2011).

139. Bradley D*, et al.* Vorinostat in advanced prostate cancer patients progressing on prior chemotherapy (National Cancer Institute Trial 6862): trial results and interleukin-6 analysis: a study by the Department of Defense Prostate Cancer Clinical Trial

Consortium and University of Chicago Phase 2 Consortium. *Cancer* **115**, 5541-5549 (2009).

140. Schneider BJ*, et al.* Phase I study of vorinostat (suberoylanilide hydroxamic acid, NSC 701852) in combination with docetaxel in patients with advanced and relapsed solid malignancies. *Investigational new drugs* **30**, 249-257 (2012).

141. Rathkopf DE*, et al.* A phase 2 study of intravenous panobinostat in patients with castration-resistant prostate cancer. *Cancer chemotherapy and pharmacology* **72**, 537-544 (2013).

142. Whyte WA*, et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319 (2013).

143. Filippakopoulos P*, et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067-1073 (2010).

144. Chapuy B*, et al.* Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer cell* **24**, 777-790 (2013).

145. Loven J*, et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-334 (2013).

146. Cho H*, et al.* RapidCaP, a novel GEM model for metastatic prostate cancer analysis and therapy, reveals myc as a driver of Pten-mutant metastasis. *Cancer discovery* **4**, 318-333 (2014).

147. Asangani IA*, et al.* Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature*, (2014).

148. Wyce A*, et al.* BET inhibition silences expression of MYCN and BCL2 and induces cytotoxicity in neuroblastoma tumor models. *PloS one* **8**, e72967 (2013).

149. Bubendorf L*, et al.* Metastatic patterns of prostate cancer: an autopsy study of 1,589 patients. *Hum Pathol* **31**, 578-583 (2000).

150. Valcourt U, Kowanetz M, Niimi H, Heldin CH, Moustakas A. TGF-beta and the Smad signaling pathway support transcriptomic reprogramming during epithelial-mesenchymal cell transition. *Molecular biology of the cell* **16**, 1987-2002 (2005).

151. Wang L*, et al.* ERG-SOX4 interaction promotes epithelial-mesenchymal transition in prostate cancer cells. *Prostate* **74**, 647-658 (2014).

152. Huang Z*, et al.* Sox9 is required for prostate development and prostate cancer initiation. *Oncotarget* **3**, 651-663 (2012).

153. Cai C*, et al.* ERG induces androgen receptor-mediated regulation of SOX9 in prostate cancer. *The Journal of clinical investigation* **123**, 1109-1122 (2013).

154. Thomsen MK*, et al.* SOX9 elevation in the prostate promotes proliferation and cooperates with PTEN loss to drive tumor formation. *Cancer research* **70**, 979-987 (2010).

155. Wang H*, et al.* SOX9 regulates low density lipoprotein receptor-related protein 6 (LRP6) and T-cell factor 4 (TCF4) expression and Wnt/beta-catenin activation in breast cancer. *The Journal of biological chemistry* **288**, 6478-6487 (2013).

156. Moreno CS. The Sex-determining region Y-box 4 and homeobox C6 transcriptional networks in prostate cancer progression: crosstalk with the Wnt, Notch, and PI3K pathways. *The American journal of pathology* **176**, 518-527 (2010).

157. Kypta RM, Waxman J. Wnt/beta-catenin signalling in prostate cancer. *Nature reviews Urology*, (2012).

158. Zheng D*, et al.* Role of WNT7B-induced noncanonical pathway in advanced prostate cancer. *Molecular cancer research : MCR* **11**, 482-493 (2013).

159. Wu L, Zhao JC, Kim J, Jin HJ, Wang CY, Yu J. ERG is a critical regulator of Wnt/LEF1 signaling in prostate cancer. *Cancer research* **73**, 6068-6079 (2013).

160. Wang G, Wang J, Sadar MD. Crosstalk between the androgen receptor and beta-catenin in castrate-resistant prostate cancer. *Cancer research* **68**, 9918-9927 (2008).

161. Lee E, Madar A, David G, Garabedian MJ, Dasgupta R, Logan SK. Inhibition of androgen receptor and beta-catenin activity in prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 15710-15715 (2013).

162. Carvalho FL, Simons BW, Eberhart CG, Berman DM. Notch signaling in prostate cancer: A moving target. *The Prostate* **74**, 933-945 (2014).

163. Efstathiou E*, et al.* Integrated Hedgehog signaling is induced following castration in human and murine prostate cancers. *The Prostate* **73**, 153-161 (2013).

164. Gowda PS*, et al.* Inhibition of hedgehog and androgen receptor signaling pathways produced synergistic suppression of castration-resistant prostate cancer progression. *Molecular cancer research : MCR* **11**, 1448-1461 (2013).

165. Domingo-Domenech J*, et al.* Suppression of acquired docetaxel resistance in prostate cancer through depletion of notch- and hedgehog-dependent tumor-initiating cells. *Cancer cell* **22**, 373-388 (2012).

166. Martens-Uzunova ES, Bottcher R, Croce CM, Jenster G, Visakorpi T, Calin GA. Long noncoding RNA in prostate, bladder, and kidney cancer. *European urology* **65**, 1140-1151 (2014).

167. Crea F*, et al.* Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. *Oncotarget* **5**, 764-774 (2014).

168. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer discovery* **1**, 391-407 (2011).

169. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77-94 (2014).

170. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nature reviews Molecular cell biology* **14**, 699-712 (2013).

171. Aguilo F, Zhou MM, Walsh MJ. Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. *Cancer research* **71**, 5365-5369 (2011).

172. Pickard MR, Mourtada-Maarabouni M, Williams GT. Long non-coding RNA GAS5 regulates apoptosis in prostate cancer cell lines. *Biochimica et biophysica acta* **1832**, 1613-1623 (2013).

173. Prensner JR*, et al.* The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nature genetics* **45**, 1392-1398 (2013).

174. Ren S*, et al.* Long noncoding RNA MALAT-1 is a new potential therapeutic target for castration resistant prostate cancer. *The Journal of urology* **190**, 2278-2287 (2013).

175. Takayama K*, et al.* Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer. *The EMBO journal* **32**, 1665-1680 (2013).

176. Wang L*, et al.* Linc00963: a novel, long non-coding RNA involved in the transition of prostate cancer from androgen-dependence to androgen-independence. *International journal of oncology* **44**, 2041-2049 (2014).

177. Yang L*, et al.* lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* **500**, 598-602 (2013).

178. Prensner JR*, et al.* The lncRNAs PCGEM1 and PRNCR1 are not implicated in castration resistant prostate cancer. *Oncotarget* **5**, 1434-1438 (2014).

179. Adair TH, Montani JP. In: *Angiogenesis* (ed^(eds). Morgan & Claypool Life Sciences (2010).

180. Karagiannis GS, Saraon P, Jarvi KA, Diamandis EP. Proteomic signatures of angiogenesis in androgen-independent prostate cancer. *The Prostate* **74**, 260-272 (2014).

181. Mak P*, et al.* ERbeta impedes prostate cancer EMT by destabilizing HIF-1alpha and inhibiting VEGF-mediated snail nuclear localization: implications for Gleason grading. *Cancer cell* **17**, 319-332 (2010).

182. Small AC, Oh WK. Bevacizumab treatment of prostate cancer. *Expert opinion on biological therapy* **12**, 1241-1249 (2012).

183. Olsson A, Bjork A, Vallon-Christersson J, Isaacs JT, Leanderson T. Tasquinimod (ABR-215050), a quinoline-3-carboxamide anti-angiogenic agent, modulates the expression of thrombospondin-1 in human prostate tumors. *Molecular cancer* **9**, 107 (2010).

184. Krupitskaya Y, Wakelee HA. Ramucirumab, a fully human mAb to the transmembrane signaling tyrosine kinase VEGFR-2 for the potential treatment of cancer. *Current opinion in investigational drugs* **10**, 597-605 (2009).

185. Kelly WK*, et al.* Randomized, double-blind, placebo-controlled phase III trial comparing docetaxel and prednisone with or without bevacizumab in men with metastatic castration-resistant prostate cancer: CALGB 90401. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **30**, 1534-1540 (2012).

186. Tannock IF*, et al.* Aflibercept versus placebo in combination with docetaxel and prednisone for treatment of men with metastatic castration-resistant prostate cancer (VENICE): a phase 3, double-blind randomised trial. *Lancet Oncol* **14**, 760-768 (2013).

187. Yakes FM*, et al.* Cabozantinib (XL184), a novel MET and VEGFR2 inhibitor, simultaneously suppresses metastasis, angiogenesis, and tumor growth. *Molecular cancer therapeutics* **10**, 2298-2308 (2011).

188. Smith DC*, et al.* Cabozantinib in patients with advanced prostate cancer: results of a phase II randomized discontinuation trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **31**, 412-419 (2013).

189. Basch E*, et al.* Effects of Cabozantinib on Pain and Narcotic Use in Patients with Castration-resistant Prostate Cancer: Results from a Phase 2 Nonrandomized Expansion Cohort. *European urology* **pii: S0302-2838(14)00132-8.**, [Epub ahead of print] (2014).

190. Nguyen HM*, et al.* Cabozantinib inhibits growth of androgen-sensitive and castration-resistant prostate cancer and affects bone remodeling. *PloS one* **8**, e78881 (2013).

191. Crawford ED*, et al.* Challenges and recommendations for early identification of metastatic disease in prostate cancer. *Urology* **83**, 664-669 (2014).

192. Andriole GL*, et al.* Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *Journal of the National Cancer Institute* **104**, 125-132 (2012).

193. Schroder FH*, et al.* Prostate-cancer mortality at 11 years of follow-up. *The New England journal of medicine* **366**, 981-990 (2012).

194. Miller AB. New data on prostate-cancer mortality after PSA screening. *The New England journal of medicine* **366**, 1047-1048 (2012).

195. Chou R*, et al.* Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. *Annals of internal medicine* **155**, 762-771 (2011).

196. Chou R, LeFevre ML. Prostate cancer screening--the evidence, the recommendations, and the clinical implications. *JAMA : the journal of the American Medical Association* **306**, 2721-2722 (2011).

197. Leyten GH*, et al.* Prospective multicentre evaluation of PCA3 and TMPRSS2-ERG gene fusions as diagnostic and prognostic urinary biomarkers for prostate cancer. *European urology* **65**, 534-542 (2014).

198. Danila DC*, et al.* TMPRSS2-ERG status in circulating tumor cells as a predictive biomarker of sensitivity in castration-resistant prostate cancer patients treated with abiraterone acetate. *European urology* **60**, 897-904 (2011).

199. Martinez-Pineiro L, Schalken JA, Cabri P, Maisonobe P, de la Taille A, Triptocare Study G. Evaluation of urinary prostate cancer antigen-3 (PCA3) and TMPRSS2-ERG score changes when starting androgen-deprivation therapy with triptorelin 6-month formulation

in patients with locally advanced and metastatic prostate cancer. *BJU international* **114**, 608-616 (2014).

200. Park K*, et al.* TMPRSS2:ERG gene fusion predicts subsequent detection of prostate cancer in patients with high-grade prostatic intraepithelial neoplasia. *J Clin Oncol* **32**, 206-211 (2014).

201. Qu X*, et al.* A three-marker FISH panel detects more genetic aberrations of AR, PTEN and TMPRSS2/ERG in castration-resistant or metastatic prostate cancers than in primary prostate tumors. *PLoS One* **8**, e74671 (2013).

202. Vaananen RM*, et al.* Cancer-associated changes in the expression of TMPRSS2-ERG, PCA3, and SPINK1 in histologically benign tissue from cancerous vs noncancerous prostatectomy specimens. *Urology* **83**, 511 e511-517 (2014).

203. Yao Y, Wang H, Li B, Tang Y. Evaluation of the TMPRSS2:ERG fusion for the detection of prostate cancer: a systematic review and meta-analysis. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **35**, 2157-2166 (2014).

204. Dijkstra S*, et al.* Prostate cancer biomarker profiles in urinary sediments and exosomes. *The Journal of urology* **191**, 1132-1138 (2014).

205. Irshad S*, et al.* A molecular signature predictive of indolent prostate cancer. *Sci Transl Med* **5**, 202ra122 (2013).

206. Aytes A*, et al.* Cross-Species Regulatory Network Analysis Identifies a Synergistic Interaction between FOXM1 and CENPF that Drives Prostate Cancer Malignancy. *Cancer cell* **25**, 638-651 (2014).

207. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. *The Journal of clinical investigation* **113**, 913-923 (2004).

208. Sboner A*, et al.* Molecular sampling of prostate cancer: a dilemma for predicting disease progression. *BMC medical genomics* **3**, 8 (2010).

209. Long Q*, et al.* Global Transcriptome Analysis of Formalin-Fixed Prostate Cancer Specimens Identifies Biomarkers of Disease Recurrence. *Cancer research* **74**, 3228-3237 (2014).

210. Cuzick J*, et al.* Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncol* **12**, 245-255 (2011).

211. Bolla M*, et al.* Long-term results with immediate androgen suppression and external irradiation in patients with locally advanced prostate cancer (an EORTC study): a phase III randomised trial. *Lancet* **360**, 103-106 (2002).

212. Marques RB, Dits NF, Erkens-Schulze S, van Weerden WM, Jenster G. Bypass mechanisms of the androgen receptor pathway in therapy-resistant prostate cancer cell models. *PLoS One* **5**, e13500 (2010).

213. Heinlein CA, Chang C. Androgen receptor in prostate cancer. *Endocr Rev* **25**, 276-308 (2004).

214. Holzbeierlein J*, et al.* Gene expression analysis of human prostate carcinoma during hormonal therapy identifies androgen-responsive genes and mechanisms of therapy resistance. *Am J Pathol* **164**, 217-227 (2004).

215. Longo DL. New therapies for castration-resistant prostate cancer. *N Engl J Med* **363**, 479-481 (2010).

216. Hu R*, et al.* Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. *Cancer Res* **72**, 3457-3462 (2012).

217. Sun S*, et al.* Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. *J Clin Invest* **120**, 2715-2730 (2010).

218. You S*, et al.* Integrated Classification of Prostate Cancer Reveals a Novel Luminal Subtype with Poor Outcome. *Cancer Res* **76**, 4948-4958 (2016).

219. Glass K, Huttenhower C, Quackenbush J, Yuan GC. Passing messages between biological networks to refine predicted interactions. *PLoS One* **8**, e64832 (2013).

220. Glass K, Quackenbush J, Spentzos D, Haibe-Kains B, Yuan GC. A network model for angiogenesis in ovarian cancer. *BMC Bioinformatics* **16**, 115 (2015).

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

222. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

223. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297 (2012).

224. Li Z*, et al.* The OncoPPi network of cancer-focused protein-protein interactions to inform biological insights and therapeutic strategies. *Nat Commun* **8**, 14356 (2017).

225. Peri S*, et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* **13**, 2363-2371 (2003).

226. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

227. Kent WJ*, et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).

228. Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* **31**, 3576-3579 (2003).

229. Matys V*, et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* **34**, D108-110 (2006).

230. Favata MF*, et al.* Identification of a novel inhibitor of mitogen-activated protein kinase kinase. *J Biol Chem* **273**, 18623-18632 (1998).

231. Network NCC. National Comprehensive Cancer Network - Prostate Cancer (Version 2.2017). (ed^(eds).

232. Tomlins SA*, et al.* Integrative molecular concept modeling of prostate cancer progression. *Nat Genet* **39**, 41-51 (2007).

233. Tran C*, et al.* Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* **324**, 787-790 (2009).

234. Sharma A*, et al.* The retinoblastoma tumor suppressor controls androgen signaling and human prostate cancer progression. *J Clin Invest* **120**, 4478-4492 (2010).

235. Antonarakis ES*, et al.* AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med* **371**, 1028-1038 (2014).

236. Lawson DA, Zong Y, Memarzadeh S, Xin L, Huang J, Witte ON. Basal epithelial stem cells are efficient targets for prostate cancer initiation. *Proc Natl Acad Sci U S A* **107**, 2610-2615 (2010).

237. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555-568 (2015).

238. Almendro V, Marusyk A, Polyak K. Cellular heterogeneity and molecular evolution in cancer. *Annu Rev Pathol* **8**, 277-302 (2013).

239. Deplancke B, Alpern D, Gardeux V. The Genetics of Transcription Factor DNA Binding Variation. *Cell* **166**, 538-554 (2016).

240. Ruiz C*, et al.* Advancing a clinically relevant perspective of the clonal nature of cancer. *Proc Natl Acad Sci U S A* **108**, 12054-12059 (2011).

241. Culig Z*, et al.* Mutant androgen receptor detected in an advanced-stage prostatic carcinoma is activated by adrenal androgens and progesterone. *Mol Endocrinol* **7**, 1541-1550 (1993).

242. Carro MS*, et al.* The transcriptional network for mesenchymal transformation of brain tumours. *Nature* **463**, 318-325 (2010).

243. Chen JC*, et al.* Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. *Cell* **159**, 402-414 (2014).

244. Cooper LA*, et al.* The tumor microenvironment strongly impacts master transcriptional regulators and gene expression class of glioblastoma. *Am J Pathol* **180**, 2108-2119 (2012).

245. Li J*, et al.* EGF-induced C/EBPbeta participates in EMT by decreasing the expression of miR-203 in esophageal squamous cell carcinoma cells. *J Cell Sci* **127**, 3735-3744 (2014).

246. Wang YH*, et al.* CEBPD amplification and overexpression in urothelial carcinoma: a driver of tumor metastasis indicating adverse prognosis. *Oncotarget* **6**, 31069-31084 (2015).

247. Huan H*, et al.* C/EBPalpha Short-Activating RNA Suppresses Metastasis of Hepatocellular Carcinoma through Inhibiting EGFR/beta-Catenin Signaling Mediated EMT. *PLoS One* **11**, e0153117 (2016).

248. Barakat DJ, Zhang J, Barberi T, Denmeade SR, Friedman AD, Paz-Priel I. CCAAT/Enhancer binding protein beta controls androgen-deprivation-induced senescence in prostate cancer cells. *Oncogene* **34**, 5912-5922 (2015).

249. Bilir B*, et al.* SOX4 Is Essential for Prostate Tumorigenesis Initiated by PTEN Ablation. *Cancer Res* **76**, 1112-1121 (2016).

250. Roe JS*, et al.* Enhancer Reprogramming Promotes Pancreatic Cancer Metastasis. *Cell* **170**, 875-888 e820 (2017).

251. Bluemn EG*, et al.* Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. *Cancer Cell* **32**, 474-489 e476 (2017).

252. ACFa. F. *American Cancer Society*, (2012).

253. Kuwahara M*, et al.* The transcription factor Sox4 is a downstream target of signaling by the cytokine TGF-beta and suppresses T(H)2 differentiation. *Nat Immunol* **13**, 778-786 (2012).

254. Liu Y, Zeng S, Jiang X, Lai D, Su Z. SOX4 induces tumor invasion by targeting EMT-related pathway in prostate cancer. *Tumour Biol* **39**, 1010428317694539 (2017).

255. Fu W*, et al.* MicroRNA-132/212 Upregulation Inhibits TGF-beta-Mediated Epithelial-Mesenchymal Transition of Prostate Cancer Cells by Targeting SOX4. *Prostate* **76**, 1560-1570 (2016).

256. Zhang J*, et al.* Metformin inhibits epithelial-mesenchymal transition in prostate cancer cells: involvement of the tumor suppressor miR30a and its target gene SOX4. *Biochem Biophys Res Commun* **452**, 746-752 (2014).

257. Wang Z*, et al.* Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**, 1019-1031 (2009).

258. Jin Q*, et al.* Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO J* **30**, 249-262 (2011).

259. Mizuguchi Y*, et al.* Cooperation of p300 and PCAF in the control of microRNA 200c/141 transcription and epithelial characteristics. *PLoS One* **7**, e32449 (2012).

260. Zhang B*, et al.* KLF5 activates microRNA 200 transcription to maintain epithelial characteristics and prevent induced epithelial-mesenchymal transition in epithelial cells. *Mol Cell Biol* **33**, 4919-4935 (2013).

261. Zhau HE*, et al.* Epithelial to mesenchymal transition (EMT) in human prostate cancer: lessons learned from ARCaP model. *Clin Exp Metastasis* **25**, 601-610 (2008).

262. Margolin AA*, et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7 Suppl 1**, S7 (2006).

263. Margolin AA, Wang K, Lim WK, Kustagi M, Nemenman I, Califano A. Reverse engineering cellular networks. *Nat Protoc* **1**, 662-671 (2006).

264. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet* **37**, 382-390 (2005).

265. Sawan C*, et al.* Histone acetyltransferase cofactor Trrap maintains self-renewal and restricts differentiation of embryonic stem cells. *Stem Cells* **31**, 979-991 (2013).

266. Lai YH*, et al.* SOX4 interacts with plakoglobin in a Wnt3a-dependent manner in prostate cancer cells. *BMC Cell Biol* **12**, 50 (2011).

267. Watanabe KA, Ramalingam S, Bell J, Sharma NV, Gutman DA, Moreno CS. TFBSET: A New Tool for the Identification of Enriched Transcription Factor Binding Sites in the Promoter Regions of Human Genes. *PloS One*, submitted. (2018).

268. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252-263 (2009).

269. Darnell JE, Jr. Transcription factors as targets for cancer therapy. *Nat Rev Cancer* **2**, 740-749 (2002).

270. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov* **5**, 993-996 (2006).

271. Scher HI*, et al.* Increased survival with enzalutamide in prostate cancer after chemotherapy. *N Engl J Med* **367**, 1187-1197 (2012).

272. Fisher B, Costantino JP, Redmond CK, Fisher ER, Wickerham DL, Cronin WM. Endometrial cancer in tamoxifen-treated breast cancer patients: findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. *J Natl Cancer Inst* **86**, 527-537 (1994).

273. Lin YZ, Yao SY, Veach RA, Torgerson TR, Hawiger J. Inhibition of nuclear translocation of transcription factor NF-kappa B by a synthetic peptide containing a cell membrane-

permeable motif and nuclear localization sequence. *J Biol Chem* **270**, 14255-14258 (1995).

274. Natarajan K, Singh S, Burke TR, Jr., Grunberger D, Aggarwal BB. Caffeic acid phenethyl ester is a potent and specific inhibitor of activation of nuclear transcription factor NF-kappa B. *Proc Natl Acad Sci U S A* **93**, 9090-9095 (1996).

275. Illendula A*, et al.* Chemical biology. A small-molecule inhibitor of the aberrant transcription factor CBFbeta-SMMHC delays leukemia in mice. *Science* **347**, 779-784 (2015).

276. Grimley E, Liao C, Ranghini EJ, Nikolovska-Coleska Z, Dressler GR. Inhibition of Pax2 Transcription Activation with a Small Molecule that Targets the DNA Binding Domain. *ACS Chem Biol* **12**, 724-734 (2017).

277. Agyeman A, Jha BK, Mazumdar T, Houghton JA. Mode and specificity of binding of the small molecule GANT61 to GLI determines inhibition of GLI-DNA binding. *Oncotarget* **5**, 4492-4503 (2014).

278. Jeong KC*, et al.* Intravesical instillation of c-MYC inhibitor KSI-3716 suppresses orthotopic bladder tumor growth. *J Urol* **191**, 510-518 (2014).

279. Soodgupta D*, et al.* Small Molecule MYC Inhibitor Conjugated to Integrin-Targeted Nanoparticles Extends Survival in a Mouse Model of Disseminated Multiple Myeloma. *Mol Cancer Ther* **14**, 1286-1294 (2015).

280. Lenz HJ, Kahn M. Safely targeting cancer stem cells via selective catenin coactivator antagonism. *Cancer Sci* **105**, 1087-1092 (2014).

281. Eguchi M, Nguyen C, Lee SC, Kahn M. ICG-001, a novel small molecule regulator of TCF/beta-catenin transcription. *Med Chem* **1**, 467-472 (2005).

282. Creyghton MP*, et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936 (2010).

283. Cockerill PN. Structure and function of active chromatin and DNase I hypersensitive sites. *FEBS J* **278**, 2182-2210 (2011).

284. Gotea V, Ovcharenko I. DiRE: identifying distant regulatory elements of co-expressed genes. *Nucleic Acids Res* **36**, W133-139 (2008).

285. Roider HG, Manke T, O'Keeffe S, Vingron M, Haas SA. PASTAA: identifying transcription factors associated with sets of co-regulated genes. *Bioinformatics* **25**, 435-442 (2009).

286. Kwon AT, Arenillas DJ, Worsley Hunt R, Wasserman WW. oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. *G3 (Bethesda)* **2**, 987-1002 (2012).

287. Zambelli F, Pesole G, Pavesi G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res* **37**, W247-252 (2009).

288. Siepel A*, et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-1050 (2005).

289. Wingender E, Dietze P, Karas H, Knuppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* **24**, 238-241 (1996).

290. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011).

291. Subramanian A*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

292. Dang CV. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol Cell Biol* **19**, 1-11 (1999).

293. Ceballos E*, et al.* Inhibitory effect of c-Myc on p53-induced apoptosis in leukemia cells. Microarray analysis reveals defective induction of p53 target genes and upregulation of chaperone genes. *Oncogene* **24**, 4559-4571 (2005).

294. Acosta JC*, et al.* Myc inhibits p27-induced erythroid differentiation of leukemia cells by repressing erythroid master genes without reversing p27-mediated cell cycle arrest. *Mol Cell Biol* **28**, 7286-7295 (2008).

295. Bild AH*, et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353-357 (2006).

296. Ben-Porath I*, et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* **40**, 499-507 (2008).

297. Alfano D*, et al.* Modulation of cellular migration and survival by c-Myc through the downregulation of urokinase (uPA) and uPA receptor. *Mol Cell Biol* **30**, 1838-1851 (2010).

298. Long Q*, et al.* Global transcriptome analysis of formalin-fixed prostate cancer specimens identifies biomarkers of disease recurrence. *Cancer Res* **74**, 3228-3237 (2014).

299. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).

300. Kannan K*, et al.* DNA microarrays identification of primary and secondary target genes regulated by p53. *Oncogene* **20**, 2225-2234 (2001).

301. Hinata K, Gervin AM, Jennifer Zhang Y, Khavari PA. Divergent gene regulation and growth effects by NF-kappa B in epithelial and mesenchymal cells of human skin. *Oncogene* **22**, 1955-1964 (2003).

302. Bakker WJ*, et al.* Differential regulation of Foxo3a target genes in erythropoiesis. *Mol Cell Biol* **27**, 3839-3854 (2007).

303. Liu F, Lei W, O'Rourke JP, Ness SA. Oncogenic mutations cause dramatic, qualitative changes in the transcriptional activity of c-Myb. *Oncogene* **25**, 795-805 (2006).

304. Pennacchio LA, Loots GG, Nobrega MA, Ovcharenko I. Predicting tissue-specific enhancers in the human genome. *Genome Res* **17**, 201-211 (2007).

305. Lenhard B, Sandelin A, Mendoza L, Engstrom P, Jareborg N, Wasserman WW. Identification of conserved regulatory elements by comparative genome analysis. *J Biol* **2**, 13 (2003).

306. Roider HG, Kanhere A, Manke T, Vingron M. Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics* **23**, 134-141 (2007).

307. Kvon EZ*, et al.* Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91-95 (2014).

308. Dago AE*, et al.* Rapid phenotypic and genomic change in response to therapeutic pressure in prostate cancer inferred by high content analysis of single circulating tumor cells. *PLoS One* **9**, e101777 (2014).

309. Yanez-Cuna JO, Dinh HQ, Kvon EZ, Shlyueva D, Stark A. Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. *Genome Res* **22**, 2018-2030 (2012).

310. Jolma A*, et al.* DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527**, 384-388 (2015).

311. Kerppola TK, Curran T. Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations: implications for transcription factor cooperativity. *Cell* **66**, 317-326 (1991).

312. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**, 2227-2241 (2011).

313. Gundem G*, et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353-357 (2015).

314. Efstathiou E*, et al.* Effects of abiraterone acetate on androgen signaling in castrate-resistant prostate cancer in bone. *J Clin Oncol* **30**, 637-643 (2012).

315. Roudier MP*, et al.* Phenotypic heterogeneity of end-stage prostate carcinoma metastatic to bone. *Hum Pathol* **34**, 646-653 (2003).