Looking Beyond the Androgen Receptor for Transcriptional Drivers of Prostate Cancer

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctorate of Philosophy in the Graduate School of The Ohio State University

By

Benjamin Douglas Sunkel, B.S.

The Ohio State Biochemistry Program

The Ohio State University

2016

Dissertation Committee:

Qianben Wang, PhD, Advisor

Steven Clinton, MD/PhD

Christin Burd, PhD

Michael Ostrowski, PhD
Abstract

Prostate cancer remains one of the leading causes of cancer-related deaths in American men. Further understanding of prostate carcinogenesis therefore remains a top priority in the cancer research field, and may lead to the development of novel diagnostic and treatment tools. As androgen receptor (AR) is a crucial component in initial, androgen-dependent prostate cancer (ADPC) progression, targeting AR is the current mainstay for combatting the disease in its early stages. However, resistance to AR-targeting therapy too often develops, which results in the lethal, castration-resistant growth phase of the disease. While more potent, second-generation therapies indicated for castration-resistant prostate cancer (CRPC) treatment have reached the clinic, there is still a need to identify therapeutic targets beyond the AR-signaling axis. It is increasingly understood that AR activity occurs in concert with a diverse cohort of pioneer transcription factors and coregulators that define disease stage-specific AR transcriptional output. This effectively produces a moving target and complicates our ability to efficiently observe and disrupt AR activity. Moreover, many of these same factors define the expression of their own clinically relevant gene targets independent of AR. Improving diagnostic and therapeutic success in prostate cancer requires continued focus on defining the fundamental determinants of AR function and the AR-independent transcriptional drivers of this disease.
To this end, we began by defining the molecular mechanism underlying GATA2-mediated AR target gene expression in both ADPC and CRPC cells (Chapter 2). Our work supports a tripartite model of GATA2 pioneer factor activity. Firstly, we show that GATA2 supports AR expression. In AR target gene regulatory regions, GATA2 then establishes an accessible chromatin environment for AR binding. Finally, we find that GATA2 preforms regulatory chromatin loops between AR target gene enhancers and promoters. From a genome-wide perspective, these functions universally support AR genomic binding patterns, suggesting a fundamental role for GATA2 in maintaining AR activity across multiple stages of prostate cancer.

We went on to conduct an integrated genomic analysis of gene expression profiles defined by the master transcription factors, CREB1 and FoxA1, in the absence of androgen (Chapter 3). These efforts establish a novel role for CREB1 in defining disease-relevant transcription profiles throughout prostate cancer progression. We further show that AR-independent FoxA1 activity involves extensive collaboration with CREB1. Genes sets collaboratively regulated by CREB1 and FoxA1 proved useful as biomarkers for predicting treatment failure among primary prostate cancer patients. Finally, we found that CREB1/FoxA1 target gene expression was sensitive to multiple kinase-inhibiting compounds.

In summary, our work provides added evidence for the critical role of factors such as GATA2 in supporting AR activity throughout all stages of prostate cancer. This concept
suggests that targeting AR collaborating factors could efficiently abolish genome-wide AR signaling. Finally, we define prognostic gene expression profiles regulated by the collaborative activity of CREB1 and FoxA1. These gene sets may prove effective in indicating sensitivity to future therapies that do not target AR.
This work is dedicated to John Romine.
Acknowledgments

First and foremost, I would like to extend my sincere gratitude to my advisor, Dr. Qianben Wang, for taking a chance on me, allowing me to independently pursue a well rounded graduate training plan, and guiding me along the way to becoming a confident scientist. I truly appreciate your support and applaud your efforts as an advisor. Present every step of the way, Dr. Steven Clinton deserves equal praise for helping me focus my efforts on rewarding and relevant science, expressing tremendous faith in my abilities, and lending time and attention as a co-advisor during critical phases of my fellowship training. To Drs. Christin Burd and Michael Ostrowski, you went above and beyond your role as committee members, offering career advice when needed and advocating for me every chance you had. Thank you all for your great mentorship.

Drs. Kimberly Specht and Ian Krouse, you taught me how to be a scientist, and you’ve never stopped supporting me.

I would also like to thank all members of the Wang and Clinton labs as well as our collaborators, past and present, for your assistance, training, and camaraderie. In particular, Cassandra Grenade for your unrivaled friendship and support, Zhenqing Ye and Xiangtao Liu for offering your expertise in computational biology and your
teamwork, Dayong Wu and Zhong Chen for your willingness to teach me almost everything I know, and Jennifer Thomas-Ahner for teaching me everything else, being a second set of eyes on my countless fellowship/grant applications, and taking time to talk with me about the finer points of navigating a scientific career.

To the Ohio State Biochemistry Program and the OSU HHMI Med into Grad Scholars Program thank you for facilitating an exceptional graduate training experience. To Pelotonia and Team Buckeye, thank you for giving me the ability to give back to the cancer research effort.

Graduate school would not have been the same without the many friends who made it bearable. Thank you to Horacio Lopez-Nicora and Owen Branson for coffee breaks, lunches, bar hops, statistics help, and valued friendship. Thank you to Aaron Victor for long talks, encouragement, and weightlifting coaching. Gabe Garber, Kyle Hendricks, and Sean Redmond I can’t thank you enough for always being there as my non-science friends to keep things light and remind me that there’s a great big world outside the lab.

Finally, no one deserves more gratitude than my family. Doug and Marsha, you made it all so much easier with your support. You are always encouraging, and you give me the drive to accomplish great things. Maggie and Gwen, I can’t thank you enough for taking care of your little brother as well as you do. Whitney, you’re the entire A-Team, providing round-the-clock support and a firm grounding in the real world. I could not have done this without you all.
Vita

June 2006 ..............................................Columbus North High School

2010......................................................B.S. Biochemistry, Denison University

2011 to present ......................................Graduate Research Fellow, Department of

Molecular Virology, Immunology, and

Medical Genetics, The Ohio State

University

Fellowships Awarded

1F31 CA180535-01A1 08/15/14-08/14/16
NIH/NCI
Inhibiting CREB1/FoxA1 target gene-driven castration-resistant prostate cancer
Role: PI, Mentors: Qianben Wang, Ph.D. and Steven Clinton, M.D./Ph.D.

2014 Pelotonia Graduate Student Fellowship Deferred
Ohio State University Comprehensive Cancer Center
Inhibiting CREB1/FoxA1 target gene-driven castration-resistant prostate cancer
Role: PI, Mentor: Qianben Wang, Ph.D.

Howard Hughes Medical Institute MED into GRAD Scholars Program at The Ohio State
University 07/01/12-06/30/13
Genome-wide chromatin restructuring, facilitated by androgen receptor, alters gene
expression patterns in the progression towards advanced, castration-resistant prostate
cancer
Role: PI, Mentors: Qianben Wang, Ph.D. and Steven Clinton, M.D./Ph.D.
Publications


Fields of Study

Major Field: Biochemistry

Research Area: Cancer Biology
# Table of Contents

Abstract ........................................................................................................................................ ii
Dedication ..................................................................................................................................... v
Acknowledgements ................................................................................................................... vi
Vita ................................................................................................................................................ viii
List of Tables ........................................................................................................................... xiii
List of Figures ............................................................................................................................ xiv

Chapter 1: Introduction

*Androgen Receptor-targeted Therapy in Prostate Cancer and Resistance*

Mechanisms ................................................................................................................................. 1

*Androgen Receptor Function in Prostate Cancer* ................................................................. 6

*GATA2: Multifunctional Pioneer Factor* ................................................................................ 10

*FoxA1: Activity Beyond AR* .................................................................................................. 13

*CREB1: Potential Cancer Driver and Therapeutic Target* .................................................. 16

Chapter 2: Three-tiered role of the pioneer factor GATA2 in promoting androgen-dependent gene expression in prostate cancer

*Abstract* ................................................................................................................................. 21

*Introduction* ........................................................................................................................... 22

*Materials and Methods* ......................................................................................................... 26

*Results* ................................................................................................................................... 30

*Discussion* ............................................................................................................................ 55
Chapter 3: Integrative analysis identifies targetable CREB1/FoxA1 transcriptional co-regulation as a predictor of prostate cancer recurrence

Abstract

Introduction

Materials and Methods

Results

Discussion

Chapter 4: Conclusions

References
List of Tables

Table 1: Correlation of clinical and molecular features of the combined cohort .................................................................95
Table 2: Multivariate proportional hazards model including clinical features and Panel A in the combined cohort.................................96
Table 3: Multivariate proportional hazards model including clinical features and Panel B in the combined cohort .................................96
Table 4: Multivariate proportional hazards model including clinical features and both Panel A and B in the combined cohort .........................96
List of Figures

Figure 1: Genome-wide overlap of AR, GATA2, and FoxA1 ..................31
Figure 2: GATA2 regulates AR expression........................................33
Figure 3: GATA2 facilitates AR-mediated ABCC4 expression ..............38
Figure 4: GATA2 facilitates AR-mediated ADPGK expression ............41
Figure 5: GATA2 mediates androgen-responsive gene expression in CRPC cells.................................................................43
Figure 6: GATA2-bound enhancers exhibit active chromatin signature.......................................................................................45
Figure 7: GATA2 mediates p300 recruitment and chromatin accessibility..................................................................................47
Figure 8: GATA2 facilitates chromatin loop formation .......................49
Figure 9: GATA2 supports genome-wide AR binding patterns ..........54
Figure 10: Genomic distribution of CREB1 binding in ADPC and CRPC cells...............................................................................73
Figure 11: Cell type-specific binding of CREB1 in enhancers.............75
Figure 12: Enhancer-bound CREB1 is associated with cell type-specific, prostate cancer gene expression profiles...............................77
Figure 13: CREB1 and FoxA1 interact preferentially in CRPC cells ......80
Figure 14: CREB1 and FoxA1 define common target genes ................82
Figure 15: Cooperative CREB1 and FoxA1 binding site profiles ............85
Figure 16: CREB1/FoxA1 target genes associated with aggressive prostate cancer........................................................................90
Figure 17: CREB1/FoxA1 target genes predict prostate cancer recurrence..93
Figure 18: Analysis of CREB1/FoxA1 target genes in circulating tumor cells ................................................................. 99
Figure 19: 22rv1 cells exhibit activated CREB1 signaling ......................... 101
Figure 20: Kinase inhibition of CREB1/FoxA1 target gene expression acts through MED1 ................................................................. 104
Figure 21: Kinase inhibitors are functional in a model of low endogenous CREB1 phosphorylation ................................................. 106
Figure 22: AR and AR splice variants have a minor role in CREB1/FoxA1 target gene expression ................................................................. 108
Chapter 1: Introduction

Androgen Receptor-targeted Therapy in Prostate Cancer and Resistance Mechanisms

The androgen receptor (AR), a member of the steroid receptor superfamily (1), is a classic example of a ligand-inducible transcription factor whose activity is tightly linked to numerous physiological processes and disease states (2). Within the various mammalian tissues expressing AR, its essential role in organ development and function has been demonstrated, which ranges from contributions to spermatogenesis, sexual behavior, and skeletal maintenance (3). Full-length AR consists of an N-terminal constitutively active domain (NTD) as well as a zinc-finger motif-containing DNA-binding domain (DBD) and C-terminal ligand-binding domain (LBD) separated by a flexible hinge region (4). Mechanistically, induction of AR activity relies upon binding to male hormones, androgens (e.g. testosterone or the more potent 5α-dihydrotestosterone [DHT]), resulting in release of AR from stabilizing interactions with cytoplasmic heat shock proteins (HSPs) (5). Homodimerization, activation by post-translational modification, and nuclear translocation of ligand-bound AR is essential for its role in determining tissue-, cell type-, and disease stage-specific gene expression patterns (5). Within the nucleus, AR binds genomic regions enriched with its cognate DNA-binding motif, or androgen response element (ARE), consisting classically of the 15 base pair
sequence 5'-TGTTCnAGAACA-3' and providing a degree of specificity to global AR distribution within regulatory elements of androgen responsive genes (6). These regulatory regions further serve as a substrate for the assembly of AR transcription complexes that regulate gene expression.

More than 200 proteins have been identified that assemble within AR transcription complexes, having diverse functionalities ranging from DNA repair to chromatin remodeling to ubiquitin ligases and more (reviewed in (7,8)). In a simplified model of transcription complex assembly, pioneer transcription factors, namely forkhead (FoxA1) and GATA (GATA2) family members (discussed below), occupy compact genomic regions to generate an accessible chromatin environment to which AR may bind upon hormone stimulation (reviewed in (9)). Chromatin remodeling complexes (e.g. SWI/SNF) aid in the formation of open chromatin regions via their ATP-dependent nucleosome remodeling function (10,11). Ligand binding generates structural rearrangements in the AR LBD that produces a binding surface for recruiting transcription coactivators (e.g. p160 family steroid receptor coactivators [SRCs]) or corepressors (e.g. nuclear receptor corepressor [NCoR]), which in turn recruit histone-modifying enzymes such as histone acetyltransferases (e.g. CREB binding protein [CBP]/p300) and histone deacetylases (HDACs) to activate or repress transcription, respectively (8,12-14). AR most often binds to distal enhancer regions, far from the proximal promoter and transcription start site of androgen-responsive genes (15). As such, the multi-subunit Mediator complex is required to mediate the formation of regulatory chromatin loops that bring AR-bound enhancers
within close spatial proximity to AR target gene promoters and to assist in the assembly and function of the pre-initiation complex (PIC) (16-21).

Advances in the understanding of AR activity have had no greater impact than in the human prostate cancer field. Development and normal function of the prostate secretory epithelium is understood to rely on hormone-inducible gene expression mediated by AR (3,22), and knowledge of the subsequent dependence on androgens for the malignant proliferation of prostate adenocarcinoma dates back to the Nobel Prize-winning insights of Dr. Charles Huggins, who demonstrated that surgical castration or application of endogenous estrogens resulted in decreased cellular proliferation of benign prostatic hyperplasia (BPH) and regression of local and metastatic prostate cancer (23-26). Leveraging the sensitivity of prostate cancer to castrate levels of androgens, androgen ablation using therapeutics that target testicular and adrenal androgen synthesis (luteinizing hormone [LH]-releasing hormone agonists/antagonists [e.g. leuprolide and cetrorelix, respectively] and steroidogenic cytochrome P450 [CYP] enzyme inhibitors [e.g. ketoconazole]) would later underlie the medical treatment of advanced disease (27,28). Discovery and characterization of AR (29-33) also allowed for the development of non-steroidal anti-androgens (e.g. flutamide) that could compete with DHT for AR binding, thus inhibiting its transcriptional activity (34).

Today, combination therapy involving androgen ablation and the application of anti-androgens is the mainstay of treatment for metastatic and, increasingly, locally advanced disease and is initially effective against androgen-dependent prostate cancer (ADPC)
cases naïve to hormone therapy (35-37). Despite an early response, the heterogeneous nature of prostate cancer cell populations, characterized by both androgen-dependent and –independent sub-populations, provides an initial rationale for the nearly inevitable progression of the disease to a treatment-insensitive, castration-resistant prostate cancer (CRPC) phenotype. However, recent insights suggest that it is overly simplistic to regard such lethal cases as entirely independent of the activities of AR. As previously reviewed (38-40), a number of mechanisms have been explored that support reestablished AR activity as a prominent driver of CRPC growth even in the presence of the potent second generation anti-androgen (enzalutamide) and CYP inhibitor (abiraterone) (41-47). Briefly, through amplification, overexpression, or perhaps stabilization of AR, cancer cells may exhibit enhanced sensitivity to low levels of circulating androgens (48,49). Mutations within the AR LBD or alternative splicing that excludes the LBD altogether can allow cells to utilize anti-androgens as ligands or deviate completely from the need for ligand-induction for AR transactivation, respectively (50-55). Alternately, intratumoral androgen synthesis, via up-regulation of steroidogenic enzymes (56-58), or aberrant activation of AR by upstream kinases responding to growth factor stimulation, such as epidermal growth factor (EGF) and insulin-like growth factor 1 (IGF1), may stimulate AR activity following androgen ablation (59,60). More recently, it was demonstrated that a related nuclear receptor family member, glucocorticoid receptor (GR), is up-regulated in response to hormone therapy and adopts a regulatory role over oncogenic transcripts that would normally be driven by androgens/AR in the absence of treatment (61). Overexpression or activation of AR coactivators/collaborating transcription factors and lost activity of AR corepressors may contribute to the
redistribution of AR to new binding sites on chromatin, promoting a disease-stage specific transcription profile, or to the activation of AR by anti-androgens (62-65). Finally, recent, high-resolution profiling of AR binding sites using chromatin immunoprecipitation exonuclease sequencing (ChIP-exo) revealed that AR is capable of recognizing a DNA sequence motif entirely distinct from the canonical ARE when bound to antagonistic ligands bicalutamide and enzalutamide. Upon occupying sites exhibiting this novel motif, AR is able to stimulate expression of oncogenic gene targets in response to antiandrogen treatment (66).

It has been demonstrated that the selective pressure applied by the therapeutic strategies described above can result in the outgrowth of treatment-insensitive cancer cell populations utilizing the aforementioned resistance mechanisms (67). Still other CRPC cells lose expression of AR and are thus entirely refractory to compounds targeting the androgen signaling axis. Our current inability to address the molecular determinants of CRPC growth is one reason why prostate cancer is the second leading cause of cancer-related deaths in American men (68). It is clear that a broader view of the factors that define prostate cancer-promoting gene expression profiles, either independently or in collaboration with AR, is required to effectively diagnose, stage, and treat this disease.

The following sections introduce important concepts in global transcription regulation mediated by AR in the course of prostate cancer development and progression, and further highlight additional transcription factors that have been the focus of my dissertation research. The first, GATA2, offers an example of a fundamental determinant
of AR activity that may represent a therapeutic target for more complete disruption of AR-mediated, prostate cancer-relevant gene expression profiles (69). The remaining two, FoxA1 and CREB1, comprise an AR-independent, collaborative transcription factor axis that determines prognostic gene expression patterns, opens the door to potential therapeutic opportunities, and highlights previously unappreciated aspects of prostate cancer biology (70).

**Androgen Receptor Function in Prostate Cancer**

A prominent mechanism by which androgen/AR contributes to ADPC growth and proliferation is through regulation of the G1/S-phase cell cycle transit. Reviewed previously, androgen starvation of prostate cancer cells results in G1 arrest as the expression level of cyclin D1 (CCND1) is diminished, active cyclin/cyclin-dependent kinase (CDK) complex formation is inhibited, and retinoblastoma (RB) tumor suppressor activity is enhanced (71). However, Wang *et al.* demonstrated in a 2009 study that AR supports the androgen-independent progression of CRPC cells through the G2/M phase of the cell cycle as a primary means of enhancing CRPC proliferation in the absence of hormone (15). The group characterized the gene expression profiles of the ADPC cell line model LNCaP and its derivative CRPC model, abl (72), over a time course of androgen stimulation and after silencing of AR to find that the basal AR-mediated gene expression profile of CRPC cells differs remarkably from the hormone-stimulated gene expression profile of ADPC cells. This result suggested that continued AR function in CRPC is directed towards the definition of a unique gene expression profile. In fact, ADPC-specific DHT-up-regulated genes were robustly enriched within the biological
process, “cellular lipid metabolism,” while CRPC-specific basal AR-up-regulated genes were enriched in, “mitotic cell cycle.” A more recent comprehensive characterization of a core set of androgen-stimulated, direct AR target genes revealed metabolic pathways under the control of this signaling axis in two separate cell line models of ADPC: LNCaP (characterized by AR LBD mutation) and VCaP (characterized by AR overexpression) (73). Thus, metabolic processes that may be required to support the rapid proliferation of cancer cell populations appear to be driven by AR activity across distinct molecular subtypes of ADPC. This suggests a therapeutic focus on these downstream products of androgen signaling may be broadly applicable to the initial treatment of prostate cancer.

AR chromatin immunoprecipitation combined with tiled oligonucleotide microarray (ChIP-on-chip) analysis of LNCaP and abl cells in the Wang et al. study provided information regarding the genome-wide redistribution of AR binding sites from ADPC to CRPC. Increased enrichment of AR binding within the enhancers of cell cycle M phase genes (CDK1, CDC20, and UBE2C) in abl cells indicated a direct mechanism of AR-mediated CRPC cell cycle progression distinct from the androgen-stimulated G1/S phase progression of ADPC cells. It was further demonstrated that epigenetic definition of abl-specific enhancers, marked by histone H3 lysine 4 mono- and dimethylation (H3K4me1/2), provided a cell type-specific chromatin environment amenable to androgen-independent, pioneer factor (FoxA1)-mediated AR recruitment to the UBE2C locus. Expression of UBE2C, a ubiquitin-conjugating enzyme involved in mitotic cell cycle progression by promoting the accumulation and activation of the anaphase-promoting complex/cyclosome (APC/C) (74,75), was finally shown to be strongly
correlated with clinical cases of CRPC vs. ADPC and necessary for androgen-independent CRPC but not ADPC cell proliferation. A 2012 study further expanded on this point, demonstrating that a LBD-deficient AR splice variant (AR-V7) that accumulates in CRPC cells following enzalutamide treatment promotes a gene expression profile that includes \textit{UBE2C}, but is otherwise largely unique from the full-length AR-mediated gene profile (76). Importantly, in CRPC cells treated with enzalutamide and in CRPC xenografts treated with abiraterone, expression of AR-V7, but not full-length AR, correlated with UBE2C expression. This relationship was further supported by the analysis of AR, AR-V7, and UBE2C expression in CRPC tissue microarrays.

In light of the high prevalence of prostate cancer progression following therapy targeting the androgen/AR signaling axis and the overwhelming evidence supporting sustained, disease-driving AR activity in the presence of such treatments, it is desirable to identify direct targets of AR that contribute to cancer cell proliferation. The Wang et al. study suggests that these targets may be disease state-specific and encourages further characterization of unique AR-transcriptional targets across a broad spectrum of prostate cancer severity. This process may offer opportunities to address unique drivers of late stage disease. A recent investigation from Myles Brown’s lab describes a role for enhancer of zeste homolog 2 (EZH2) as a coactivator of AR-driven gene expression in CRPC cells, and reveals that enhancer co-occupancy by AR and EZH2 facilitates the expression the DNA-repair proteins, KIAA0101 and RAD51C (77,78). As FoxA1 pioneer factor activity was found to support AR-driven G2/M phase gene expression, these combined results indicate that interactions with specific factors may allow AR to
drive the expression of genes within non-overlapping biological processes, adding complexity to its role in supporting CRPC growth. More recent analyses of AR genomic distribution in primary and CRPC clinical samples also found that stage-specific AR binding sites may occur alongside context-specific collaborating transcription factors (e.g. MYC, E2F, STATs, HOXB13, and IKZF1) (79,80). Altogether, these findings suggest that targeting CRPC-specific AR collaborators/coregulators could prove a viable means of inhibiting significant portions of AR-mediated transcription profiles in treatment-resistant contexts.

There is growing evidence to suggest that AR redistribution, resulting from interactions with specific pioneer and licensing factors and leading to altered genome-wide expression profiles (65), may accompany each phenotypic transition along the path from a normal prostate to advanced prostate cancer. Tomlins et al. noted a subtle elevation in the expression level of a set of androgen-induced genes during the initial progression from benign prostate tissue to early PIN lesions (81), and interestingly, these same genes were significantly down-regulated in later-stage metastatic CRPC tissue samples. More recently, Pomerantz et al. found by ChIP-seq that the distribution of AR binding sites is markedly altered between normal prostate tissues and primary prostate cancer samples (82). Similar redistribution patterns could be induced in immortalized prostate epithelial cells upon forced overexpression of the pioneer factors, HOXB13 and FoxA1. These studies support a changing role for AR throughout the progression of prostate cancer and implicate disease stage-specific collaborating factors in defining the spectrum of AR-mediated, cancer-promoting gene expression patterns. Further understanding of these
collaborating factors might offer new means of disrupting AR activity at all stages of disease progression, which, in combination with existing AR-targeted therapies, could improve therapeutic cure rates.

**GATA2: Multifunctional Pioneer Factor**

The first collaborating factor to be considered in this work is GATA2. Considerable interest in collaborating factors that contribute to AR transcriptional activity initially evolved from analyses of genome-wide AR binding sites in prostate cancer cells (15,83). These analyses revealed a number of co-occurring DNA motifs within AR-occupied regions. Chief among the proteins that recognize these motifs are the pioneer factors FoxA1 and GATA2, whose role in determining nuclear receptor (NR, ex. AR and estrogen receptor [ER]) binding patterns has long been appreciated (reviewed in (9)). In short, pioneer factors are understood to engage regions of compact chromatin, facilitating chromatin decondensation in advance of ligand-stimulated NR binding. While a mechanism by which FoxA1 facilitates chromatin relaxation has been suggested (discussed below), little has been learned mechanistically about GATA2 pioneer activity despite clear correlations between GATA2 and AR/ER binding (15,83,84), abundant evidence of its activity upstream of AR/ER binding (85-88), and the increasing body of literature reporting its role in prostate cancer progression (89,90).

Prior to our work outlined in Chapter 2 (69), it was understood that GATA2 served some role in establishing nucleosome depleted regions amenable to AR binding (85), supporting prostate cancer-relevant gene expression profiles, and that GATA2 expression
levels were positively correlated with aggressive disease phenotypes (89). Beyond these concepts and work demonstrating pioneer activity of GATA2 upstream of AR from the Wang and Knudsen groups (15,83,87), little was understood about the genome-wide function of GATA2, its exact mechanism of pioneer activity, or its other roles in supporting prostate cancer progression beyond facilitating AR. A role for GATA factors in supporting NR-dependent fetal liver development had been established (91,92), and their indispensible role in supporting ER expression (86), luminal cell phenotypes in the mammary gland (93,94), and genome-wide ER binding patterns had also been demonstrated (84). Perhaps most importantly, it had been shown that the anticancer effects of the natural compound, curcumin, in prostate cancer worked at least partially through the disruption of GATA2-mediated AR binding and target gene expression (87), suggesting a means to disrupt AR signaling in prostate cancer via targeting an essential collaborator.

Since our work was published, continued interest in GATA2 activity has revealed additional insights and expanded upon the concept of abrogating prostate cancer growth via disruption of GATA2 both within and beyond the realm of AR signaling. First, Zhao et al. have further clarified our findings to show that FoxA1, but not GATA2, is capable of recruiting AR to binding sites throughout the genome that lack the canonical ARE, while GATA2 may play a supporting role via mechanisms described in Chapter 2 (95). In the end, collaborative binding of AR, GATA2, and FoxA1 was shown to stabilize each factor on chromatin, but the study revealed an important aspect of the differential roles GATA2 and FoxA1 play as AR pioneers. From a therapeutic perspective, a recent study
from He et al. demonstrated an association between GATA2 expression and aggressive prostate cancer phenotypes downstream of GATA2-mediated AR signaling and went on to show that treatment with the reported GATA inhibitor, K7174, reduced GATA2 and AR expression, AR target gene transcription, and AR+/GATA2+ prostate cancer cell growth in vitro and in vivo (96). In truth, the mechanism of K7174 activity remains poorly understood, and while early reports defined it as a specific GATA inhibitor capable of disrupting the binding of an unidentified factor to DNA probes containing a GATA binding motif (97), it has more recently been regarded as a proteasome inhibitor likely owing to its ability to induce an unfolded protein response (97-99). Consistent with this, GATA2 was previously shown to directly drive the expression of components of the proteasome in non small cell lung cancer (NSCLC), and in the same study it was demonstrated that combinatorial inhibition of downstream GATA2 targets (proteasome and Rho signaling) could effectively inhibit KRAS driven NSCLC. Finally, in line with the approach taken in NSCLC, it was recently shown that direct transcriptional regulation of IGF-2 by GATA2 in aggressive prostate cancer, leading to autocrine induction of a multitude of oncogenic kinase signaling cascades, represented a novel opportunity to inhibit a downstream target of GATA2 in order to sensitize prostate cancer cells to chemotherapy (90).

Thus there is ample evidence to support a disease driving role for GATA2 in prostate cancer, both in support of oncogenic AR activity and perhaps as an independent determinant of cancer-promoting signaling events. In either case, focusing on GATA2 function, or that of its downstream effectors, could afford new insights into targetable drivers at all stages of this disease.
FoxA1: Activity Beyond AR

Mentioned previously, pioneer factors possess the unique ability among transcription factors to bind regions of compacted chromatin, leading to chromatin relaxation required for the binding of more typical transcription factors. FoxA1 accomplishes this by displacing the linker histone H1, owing to the structural similarity of its DNA-binding domain to H1 (100-102). Early, genomic studies of AR revealed an important role for FoxA1 in androgen-responsive gene transcription (83), and subsequent work has suggested that FoxA1 is among the most potent AR pioneer factors specifically in the prostate (95,103). FoxA1 has been implicated in the large-scale redistribution of AR binding sites in CRPC compared to ADPC cells and in primary prostate cancer compared to normal prostate, leading to disease stage-specific AR-mediated transcription profiles (15,82). As briefly mentioned above, Wang et al. found that enrichment of the epigenetic marks H3K4me1/2, known to partially determine FoxA1 chromatin occupancy (104), within CRPC cell-specific UBE2C enhancers facilitates androgen-independent pioneer factor binding within these gene regulatory elements. The net result was androgen-independent, FoxA1-mediated AR binding at the UBE2C locus and enhanced expression of this clinically relevant oncogene, suggesting that an important role for FoxA1 in CRPC is the reprogramming of AR activity to drive G2/M transit, rather than G1/S cell cycle progression as in ADPC (15). This finding along with subsequent studies demonstrating a genome-wide effect of FoxA1 silencing on AR binding in ADPC cells (105) and the most recent work revealing the ability of FoxA1 to reprogram AR binding during the transition from normal prostate to primary prostate cancer (82), provides clear evidence that FoxA1
has a wide-spread role in defining AR activity throughout all stages of prostate cancer progression.

Though FoxA1 has a prominent role in driving prostate cancer development and progression via support of AR activity, the abundance of FoxA1 genomic binding sites that do not overlap with AR binding sites (69,95,105) imply that this pioneer factor may have other important roles in this disease. Recent exome sequencing studies have revealed FoxA1 mutations occurring in primary prostate cancer and CRPC cases at a rate between 3-4% (106,107), and further showed that FoxA1 mutation status defines the expression of unique transcriptional targets driving enhanced rates of prostate cancer cell proliferation. It is therefore of tremendous interest and potential value to define the activities of FoxA1 both in concert with AR, and independent of AR, to provide a complete picture of how this pioneer factor drives disease progression, the factors with which it interacts, and its utility as a prognostic biomarker or therapeutic target in advanced disease.

In a 2011 study, Zhang et al. investigated the observation that, in contrast to FoxA1 silencing in G2/M phase-synchronized CRPC cells resulting in G2/M accumulation, unsynchronized cells exhibited significant G1 arrest following FoxA1 knockdown (108). This suggests that FoxA1 may have functional significance throughout the cell cycle. mRNA and protein expression analysis of several G1 phase genes identified CCNE2 and CCNA2 as being robustly overexpressed in CRPC compared to ADPC cells in an androgen-independent, yet FoxA1-dependent manner. Silencing of CCNE2 and CCNA2
significantly decreased CRPC cell proliferation in the absence of androgen, and overexpression of either *CCNE2* or *CCNA2* in LNCaP cells enhanced their sensitivity to and ability to grow in the presence of subphysiological androgen concentrations.

As in the Wang *et al.* study (15), FoxA1 ChIP-on-chip analyses in LNCaP and abl cells were utilized to provide a picture of genome-wide FoxA1 binding patterns and evidence of its transcriptional control over these G1 phase genes. In general, CRPC cell-specific FoxA1 binding was observed within the vicinity of genes differentially regulated (up-regulated and, to a lesser extent, down-regulated) in clinical samples of CRPC vs. ADPC (109-111). Increased FoxA1 binding in abl cells was observed at enhancer elements within the *CCNE2* locus, corresponding to enhanced coactivator/transcription complex assembly (Pol II, CREB binding protein [CBP], and MED1) and chromatin accessibility in CRPC cells, suggesting direct regulation of *CCNE2* by FoxA1. On the other hand, the work determined that *CCNA2* up-regulation was the result of enhanced E2F1 binding to the *CCNA2* promoter subsequent to direct regulation of *E2F1* in CRPC cells by enhanced FoxA1 occupancy of *E2F1* regulatory elements. Importantly, motif analysis of CRPC-specific FoxA1 binding sites revealed a significant co-occurrence with CREB and MYB recognition motifs in this cellular context. Zhang *et al.* then demonstrated that MYBL2/FoxA1 and CREB1/FoxA1 co-occupancy of *CCNE2* and *E2F1* regulatory elements occurs in a codependent fashion and that silencing of either CREB1 or MYBL2 resulted in reduced enhancer chromatin accessibility, androgen-independent expression of *CCNE2*, *CCNA2*, and *E2F1* in CRPC cells, and CRPC cell proliferation. They finally showed that binding of CREB1/MYBL2 was also necessary for the androgen-
independent expression of AR/FoxA1 target genes *UBE2C* and *CDK1*, both of which are involved in G2/M phase of the cell cycle (15).

Together these results identified CREB1/MYBL2 as potential FoxA1 collaborators facilitating diverse, disease-relevant transcriptional outcomes in a CRPC context. Positive correlation between MYBL2/CREB1 expression and disease progression from ADPC to CRPC has been reported and may provide some basis for the observed redistribution of FoxA1 to regions co-occupied by these factors in a model of late stage disease (112,113). Looking back to the Wang *et al.* AR study (15), these results also suggest that FoxA1 engages in distinct collaborative relationships as it determines the expression of genes responsible for progression through multiple phases of the cell cycle. Previous studies have shown that AR expression and transactivation potential are significantly reduced in cells immediately following mitosis, providing some rationale for its selective involvement in FoxA1-mediated cell cycle transit during G2/M phase alone (114,115). It will therefore be important to comprehensively identify collaborators common to FoxA1 transcriptional complexes throughout the cell cycle (*e.g.* CREB1 and MYBL2) that may serve as therapeutic targets in the inhibition of this master cell cycle regulator.

**CREB1: Potential Cancer Driver and Therapeutic Target**

The results of the 2011 study described above (108) encourage a comprehensive characterization of the cAMP response element binding protein (CREB1) as it relates to AR- and FoxA1-driven target gene expression throughout prostate cancer, but also as an independent determinant of cancer-relevant transcription profiles. Elaborated upon in
Chapter 3, several lines of evidence suggest that this second-messenger responsive transcription factor plays an important role in prostate carcinogenesis, and owing to its unique model of transactivation, it may also present a novel therapeutic target in prostate cancer.

Prior to our work described in Chapter 3, Zhang et al. was the only study to suggest a genome-wide role for CREB1 in supporting cancer-relevant gene expression in prostate cancer, and in truth, that work provided mechanistic insights limited to a few gene loci and a potential genomic role based purely on motif analysis (108). Still CREB1 has been the focus of considerable research efforts supporting its disease-driving function. Briefly, in addition to promoting cell cycle progression, CREB1 has been implicated in prostate cancer-relevant functions including steroidogenesis (116), neuroendocrine differentiation (117), and bone metastasis (118). CREB1 has also been defined as a key mediator of metabolic programs in cancer cells, responding to and orchestrating the response to metabolic stress and multiple extracellular hormonal signals (119). Mentioned above, CREB1 expression levels have been correlated with prostate cancer progression (112), and more recently, it has been suggested that CREB1 phosphorylation level could be a useful biomarker of advanced disease (120).

This latter point introduces an important concept in CREB1 function, wherein CREB1 transactivation potential is dramatically enhanced by phosphorylation of a serine residue (Ser133) within its kinase inducible domain (KID) (121). While multiple phosphorylation sites exist throughout CREB1, Ser133 appears to be a site with an unambiguous
activating potential. Responding to myriad extracellular stimuli, a multitude of kinase signaling cascades, classically cyclic nucleotide (cAMP and cGMP) dependent protein kinases (PKA and PKG) as well as PI3K/Akt, terminate with the phosphorylation of Ser133 within an AGC kinase family recognition motif (R-R/K-X-S/T, where X represents any amino acid) (122) within the KID domain (121). Ser133 phosphorylation produces a recognition site for the KID interaction (KIX) domain in the histone acetyltransferase CREB binding protein (CBP), resulting in its recruitment to CREB1 target gene loci and accumulation of the transcription-activating histone modification H3 lysine 27 acetylation (H3K27ac) (123). Thus, interest in CREB1 Ser133 phosphorylation level stems from the potential to utilize this single molecular feature as a proxy for CREB1-driven, cancer relevant gene expression indicative of aggressive disease.

Beyond its utility as a biomarker, current understanding of CREB1 also points to a potential therapeutic strategy in prostate cancer via disruption of the transactivation mechanism described above. One potential treatment strategy involves targeting upstream kinases that mediate Ser133 phosphorylation. Efforts to support this avenue have focused on the ability of kinase inhibitors to disrupt Ser133 phosphorylation subsequent to external stimulation (112,124), though additional efforts should be made to investigate the ability of such compounds to decrease steady-state Ser133 phosphorylation levels. A 2004 study by Best et al. presented an intriguing strategy for inhibiting CREB1 transactivation using a small molecule inhibitor that occupies the CBP KIX domain, thereby preventing its interaction with the phosphorylated CREB1 KID domain (125).
Each of these therapeutic strategies might inherently suffer from a lack of specificity, as kinase targets like PKA and Akt mediate diverse cellular processes, and KIX-KID domain interactions are not limited to CBP-mediated CREB1 transactivation. Additionally, it is well understood that CREB1 exhibits basal transcriptional activity in the absence of Ser133 phosphorylation that would not be sensitive to the above measures. This is due to additional protein-protein interactions occurring within its N-terminal transactivation domain (TAD) as well as its C-terminal bZIP DNA-binding domain (DBD). Beyond the KID, the CREB1 TAD is also home to a glutamine-rich (Q2) constitutive activation domain that mediates the recruitment of general, basal transcription components (TAF4/TAF1130) of the TFIID complex (126,127). Defined by Conkright et al. in 2003, Transducers of Regulated CREB activity (TORCs), subsequently renamed cAMP Regulated Transcription Co-activators (CRTCs), also moderate basal and cAMP-stimulated CREB1 activity through interactions that stabilize bZIP-DNA binding as well as the recruitment of TAF4/TAF1130 to the Q2 domain (119,128). It may therefore require a combinatorial approach to successfully disrupt all CREB1 transcriptional output, and an alternative strategy might be to target downstream targets of CREB1 activity that drive prostate cancer development and progression as suggested above for AR, GATA2, and FoxA1.

To summarize, a myriad of transcription factors and co-activators/co-repressors have been shown to influence AR activity throughout the onset and progression of prostate cancer. These factors influence both AR genomic distribution as well as its transcriptional output, suggesting their utility as therapeutic targets for efficient
disruption of AR signaling. Additionally, each factor outlined here appears to exhibit functionality beyond directly supporting AR, and understanding their transcriptional output could prove beneficial for improving our ability to diagnose clinically relevant prostatic malignancy, predict therapeutic response, and treat this complex disease with an increasing array of rational therapeutic compounds aimed at disrupting activities driven by the factors.
Chapter 2: Three-tiered role of the pioneer factor GATA2 in promoting androgen-dependent gene expression in prostate cancer

Abstract

In prostate cancer, androgen receptor (AR) binding and androgen-responsive gene expression are defined by hormone-independent binding patterns of the pioneer factors, FoxA1 and GATA2. Insufficient evidence of the mechanisms by which GATA2 contributes to this process precludes complete understanding of a key determinant of tissue-specific AR activity. Our observations suggest that GATA2 facilitates androgen-responsive gene expression by three distinct modes of action. By occupying novel binding sites within the AR gene locus, GATA2 positively regulates AR expression before and after androgen stimulation. Additionally, GATA2 engages AR target gene enhancers prior to hormone stimulation, producing an active and accessible chromatin environment via recruitment of the histone acetyltransferase p300. Finally, GATA2 functions in establishing and/or sustaining basal locus looping by recruiting the Mediator subunit, MED1, in the absence of androgen. These mechanisms may contribute to the generally positive role of GATA2 in defining AR genome-wide binding patterns that determine androgen-responsive gene expression profiles. We also find that GATA2 and FoxA1 exhibit both independent and codependent co-occupancy of AR target gene enhancers. Identifying these determinants of AR transcriptional activity may provide a foundation
for the development of future prostate cancer therapeutics that target pioneer factor function.

Introduction

Androgen signaling mediates diverse and complex functions throughout the body ranging from skeletal development and maintenance to spermatogenesis upon ligand activation of the androgen receptor (AR) (3). A member of the nuclear hormone receptor family, AR mediates androgen-dependent gene expression following release from cytoplasmic heat-shock proteins, receptor phosphorylation, and nuclear translocation whereupon AR homodimers bind to recognition sites within regulatory elements of target genes (4). This mechanism of hormone-dependent gene expression facilitates the fetal development of the prostate and regulates maintenance and normal function of the prostate secretory epithelium (3,22). In androgen-dependent prostate cancer (ADPC), androgen-stimulated AR function plays a vital role in the aberrant proliferation of epithelial cells, thus androgen deprivation therapy (ADT) precipitates marked disease regression (22,26). As is too often the case, however, reestablished AR activity leads to ADT resistance, marking the fatal progression to castration resistant prostate cancer (CRPC) (39,40,129). How tissue-, cell type-, and disease stage-specific patterns of androgen-stimulated gene expression are generated is an area of considerable research focus, and a complete picture of the activities determining this specificity is required for our understanding of androgen-driven prostate cancer progression and for the rational design of potent prostate cancer therapeutics.
In nuclear receptor (NR)-regulated gene transcription, pioneer factor function has been studied extensively, providing information on the manner in which tissue-specific, hormone-responsive gene expression is controlled (9,130,131). Pioneer factors are characterized by their ability to engage compact chromatin, initialize local chromatin decondensation, and provide an environment amenable to the recruitment of transcription factors and activation of transcription (9). Their role in determining NR target gene expression was suggested upon analysis of common DNA sequence motifs enriched within genome-wide NR binding sites identified utilizing chromatin immunoprecipitation combined with tiled oligonucleotide microarrays (ChIP-on-chip) or high-throughput sequencing technology (ChIP-seq) (83,132-135). Two key pioneer factors emerged from complementary analyses in distinct hormone-related cancers, breast and prostate, driven by estrogen receptor (ER) and AR transcriptional activity, respectively. Significantly enriched within ER and AR binding sites were motifs recognized by the Forkhead box (FOX) and GATA pioneer transcription factor families (9,15,83,133,136).

FOX proteins, specifically the FOXA subclass, have attracted considerable attention for their capacity to impact local chromatin architecture, their essential role in priming for temporal patterns of gene expression observed in early organogenesis, and their influence over tissue-specific patterns of NR target gene expression, of particular interest in hormone-related cancers (9,101,102,131). In particular, FoxA1, expressed alongside ER and AR in the developing and mature mammary and prostate ductal epithelia, respectively (137-140), has been shown to contribute significantly to maintaining the
oncogenic functions of these nuclear receptors in both treatment-sensitive and -resistant breast and prostate cancers (15,132). While the role of FoxA1 in ER-positive breast cancers appears to be the positive regulation of ER-chromatin binding, resulting in canonical and noncanonical ligand-dependent ER target gene expression and breast cancer cell proliferation (132,141), a more complex relationship between AR and FoxA1 exists in prostate cancer. Initial studies described an essential role of FoxA1 in directing AR-chromatin binding for the activation of androgen-dependent gene expression (15,142). However, a complementary analysis later revealed that FoxA1 silencing resulted in the expected loss of many AR binding events as well the surprising gain of additional AR binding sites not observed in parental cells, suggesting both exclusionary and facilitative roles for FoxA1 in directing AR binding events and determining androgen-dependent gene expression (105,143).

While GATA family proteins have long since been identified as putative pioneer factors acting in concert with FOX proteins in establishing temporal gene expression patterns required for fetal liver development (91,92), progress in understanding GATA pioneer function in NR-driven cancers has been largely limited to ER activity. GATA3 expression patterns positively correlate with both ER and FoxA1 in the developing, mature, and cancerous mammary epithelium, and are in fact necessary for determining hormone-responsive gene expression and breast cancer cell proliferation (86,93,94,144,145). As stated, prostate cancer-specific AR binding sites are also significantly enriched in GATA motifs, and GATA2, overexpressed in high-risk prostate cancer, has been shown to play an essential role in productive AR-chromatin binding
resulting in androgen-responsive gene expression (15,83,89). Given existing evidence for the central role of GATA family members in defining transcription factor activity, the demonstrated function of GATA factors in facilitating hormone-dependent gene expression, and the dearth of information on the mechanism by which GATA proteins contribute functionally to NR-transcriptional regulation, further analyses are needed to elucidate the means by which GATA pioneer factors function in hormone-related cancers.

In the present work, we focus specifically on the role of GATA2 in recruiting AR to distal enhancers elements of androgen-responsive, AR target genes in androgen-responsive prostate cancer cell lines. We provide evidence that GATA2 acts at multiple levels to contribute positively to AR transcriptional activity by enhancing $AR$ expression itself, facilitating AR-enhancer binding by establishing an accessible local chromatin environment, and enhancing AR target gene expression through involvement in the formation and maintenance of regulatory chromatin loops between AR-bound distal enhancers and AR target gene promoters. These mechanisms may account for the generally positive role of GATA2 in defining global AR binding that regulates AR target gene expression. Additionally, we reveal a complex relationship between FoxA1 and GATA2 in mediating $AR$ expression and in site-specific AR-recruitment facilitating AR target gene expression.
Materials and Methods

Cell culture

The LNCaP prostate cancer cell line was purchased from the American Type Culture Collection (ATCC) and C4-2B cells were purchased from ViroMed Laboratories. These cells were maintained in RPMI 1640 media (Invitrogen) supplemented with 10% FBS at 37 C in 5% CO2. For individual experiments, the medium was replaced by phenol red-free RPMI 1640 medium containing 5% charcoal-stripped FBS. The cells were passaged in our laboratory for less than 6 months after receipt.

RNA interference

Control siRNA (siControl) and siRNAs targeting AR, GATA2, FoxA1 (ON TARGET plus™ siRNA) were purchased from Dharmaco (Dharmacon, Lafayette, CO). siRNA transfections were performed using Lipofectamine 2000 (Invitrogen, Carlsbad, CA). 24 hrs before transfection, LNCaP cells were seeded at a density of 5x10^5 cells/well in a 6 well plate, or 8x10^6 cells/150 mm dish. The cells were transfected with 40 nM siRNA and maintained for 72 hrs in hormone free medium. The cells were then harvested or treated with R1881 for an additional 4 hrs before harvest.

Western blots

Western blot analyses were carried out as previously described (108). Briefly, LNCaP cells were collected and lysed in RIPA buffer (1% NP-40, 0.1% SDS, 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% sodium deoxycholate, 1 mM EDTA, proteinase inhibitor
cocktail (Roche)) for 20 min on ice and proteins were resolved on 10% SDS-polyacrylamide gels before being transferred onto nitrocellulose membrane (Bio-Rad). The membrane was blocked with 5% milk powder (Bio-Rad) in 1× TBS containing 0.5% Tween-20 for 1 hr, washed with TBS/Tween, and incubated with specific antibodies for 2 hrs.

*Quantitative real-time RT-PCR*

Real-time reverse transcriptase (qRT)-PCR was carried out as previously described (108). Total RNA was isolated from LNCaP cells with the RNeasy Mini kit (Qiagen, 74104). qRT-PCR was conducted on 2 µg of RNA using the MultiScribe Reverse Transcriptase and Power SYBR Green PCR Master Mix reagents (Applied Biosystems) according to the manufacturer's instructions. Each reaction was performed in triplicate.

*ChIP and ChIP-sequencing*

ChIP was performed as previously described (146) with a few modifications. Briefly, LNCaP cells (1 × 10⁷ cells/150 mm dish) were plated and grown in phenol red-free RPMI 1640 supplemented with 5% charcoal-stripped FBS for 3 days (Cells were treated with 1 nM R1881 for an additional 4 hrs as required). Cells were crosslinked with 1% formaldehyde for 10 min at room temperature. Chromatin was then sonicated, diluted, and immunoprecipitated with specific antibodies at 4°C overnight. Protein A-Sepharose beads were added and incubated for 1 hr with rotation. The beads were washed
sequentially for 10 min each in TSE I (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl), TSE II (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl), and buffer III (0.25 M LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1) and finally 2 times with TE buffer. After being drained with a 27 G x ½ inch needle (Becton Dickinson NJ), the beads were extracted 2 times with 60 ul 1% SDS, 0.1 M NaHCO3 by vortexing in a Thermomixer (Eppendorf) and pooled eluates were heated at 65°C for 16 hrs to reverse the cross-linking. DNA fragments were purified with the QIAquick PCR purification kit (Qiagen 28104) and used as the template in quantitative PCR reactions. Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA) was used and the samples were amplified with the StepOnePlus Real Time PCR System (Applied Biosystems). ChIP-seq library preparation was performed using the Truseq ChIP sample preparation kit (part# 15023092) with 10 ng of purified ChIP-DNA. The libraries were amplified using 15 PCR cycles and sequencing was performed on the Illumina HiSeq 2500 platform at the Ohio State University Comprehensive Cancer Center (OSUCCC) sequencing core (50 bp read length, single-end) with 4 multiplexed samples per lane. ChIP-seq reads were aligned against the standard hg19 build of the human reference genome with Bowtie v1.0.0 (147) allowing at most two mismatches. Reads with more than one valid alignment were removed, leaving only uniquely mapped reads for further analysis. Peaks were called using MACS v1.4.2 (148) with P-value threshold 1.0e-8. Heatmap counts were generated using HOMER v4.3 (149) with bin size 50 bp. The raw ChIP-seq data have been submitted to the Gene Expression Omnibus (GEO) repository under the accession number GSE52725.
Formaldehyde-assisted isolation of regulatory elements

Formaldehyde-assisted isolation of regulatory elements (FAIRE) was carried out as previously described (150). Briefly, formaldehyde was added directly to the cell culture medium at room temperature to a final concentration of 1% and incubated for 10 min. 1 M glycine was added to a final concentration of 125 mM for 5 min at room temperature to quench the formaldehyde. Cells were rinsed with phosphate buffered saline and collected in 1.5 mL tubes. After brief centrifugation, the cell pellets were resuspended in 400 ul of lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-Cl at pH 8.0, 1 mM EDTA) and incubated on ice for 5 min. Samples were then sonicated for 7 min (30 sec on/off cycles) using a Bioruptor (Diagenode) at the highest intensity. The cell lysates were centrifuged at 13000 rpm for 10 min at 4°C to precipitate cellular debris and the soluble chromatin was transferred to a new tube. After 2 rounds of phenol-chloroform extraction, the aqueous phase was combined and incubated at 65°C overnight to reverse crosslinking. DNA was purified with the Qiagen PCR purification Kit (Qiagen 28104) and eluted in 100 ul TE buffer. Relative enrichment in the FAIRE-treated DNA was calculated with DNA from untreated cells serving as the control.

Quantitative Chromosome Conformation Capture Assay (3C)

3C qPCR assays were performed as previously described (15) with some modifications. Briefly, nuclei were first cross-linked with formaldehyde at a final concentration of 1%
for 10 min, then digested with 400 units of *BglII* (NEB) and ligated under extra-diluted conditions. After reversing the crosslinking, DNA was purified by phenol-chloroform extraction followed by ethanol precipitation. Real time PCR was performed using TaqMan® Universal PCR Master Mix (Applied Biosystems). The data was normalized for primer efficiency differences using BAC RP11-197J19 that covers the ABCC4 locus. GADPH loading control was used to normalize DNA concentration. The interaction of two *BglII* sites in the GADPH locus was used for comparison between different 3C assays.

**Results**

*Genome-wide Overlap of AR/GATA2/FoxA1 Binding*

To assess the global overlap of pioneer factor chromatin occupancy with hormone-stimulated AR-DNA binding, we began by performing ChIP-seq in the androgen-responsive prostate cancer cell line LNCaP. GATA2 and FoxA1 ChIP-seq assays were performed in the absence of hormone, while AR ChIP-seq was performed in the presence of hormone to provide a picture of the extent to which basal pioneer factor occupancy determines androgen-dependent AR binding. The four lowermost segments of Figure 1A display the density of aligned ChIP-seq reads for each factor within a 10 kb window centered over AR binding site locations. Notably, only 45% of all AR binding events occur in regions without prior pioneer factor occupancy, indicating that while additional
Figure 1: Genome-wide overlap of AR, GATA2, and FoxA1. (A) Heatmap displaying the density of aligned ChIP-seq reads for AR (left), GATA2 (center), and FoxA1 (right) within a 10 kb window centered over individual and shared factor binding sites. Lettered segments indicate factor occupancy in four classes of AR binding sites; A – AR only, B – AR/GATA2, C – AR/FoxA1, and D – AR/GATA2/FoxA1. (B) Venn diagram representing the proportions of shared factor binding sites relative to total factor binding sites. Lettered regions are consistent with panel (A), and binding site numbers are provided for each.
factors likely facilitate AR genome-wide binding patterns, in this prostate cancer cell context, coordinate binding of GATA2, FoxA1, or both is a key feature of AR chromatin occupancy. That combinatorial pioneer factor activity is a critical determinant of AR distribution is suggested by the observation that 13,457 sites exhibit AR/GATA2/FoxA1 binding, while 15,342 and only 5,036 AR binding sites overlap only with FoxA1 or GATA2, respectively (Figure 1A-B). Based upon these findings, we went on to characterize potential mechanisms by which these pioneer factors, in particular the lesser-characterized GATA2, facilitate AR binding to DNA in advance of androgen-responsive gene expression.

*GATA2 Promotes Androgen Receptor Expression*

A previous study reported putative GATA2 binding sites within the *AR* gene promoter in LNCaP cells, suggesting a potential influence of GATA2 over AR-target gene expression via direct regulation of *AR* itself (89). Surprisingly, our GATA2 ChIP-seq and standard ChIP analyses in LNCaP cells did not reveal a significant enrichment of GATA2 binding within this previously described region, constituting the promoter of the minor AR(A) isoform (151) (Figures 2A-B). In contrast, two significant GATA2 binding peaks were observed within regulatory elements of the full-length AR(B) isoform (referred to henceforth simply as AR): one 4.6 kb upstream of the TSS, overlapping a FoxA1 binding site (Site 1), and the second within the AR promoter region (+600 bp, Site 2) (Figure 2A). While GATA2 binding to the promoter of the minor AR(A) isoform was previously
Figure 2: GATA2 regulates AR expression. (A) GATA2 and FoxA1 binding sites in the AR locus are shown in IGV. Previously reported GATA2 binding site is located immediately downstream of the AR(A) TSS. (B) ChIP assays were performed with the indicated antibody in LNCaP cells after 4 hour treatment with vehicle or 1 nM R1881. (C) AR mRNA was measured by qRT-PCR in LNCaP cells transfected with indicated siRNAs and treated for 4 hours with vehicle or 1 nM R1881. (D) GATA2 ChIP was conducted in LNCaP cells transfected with indicated siRNAs and treated for 4 hours with vehicle or 1 nM R1881. (E) FoxA1 ChIP was conducted in LNCaP cells transfected with indicated siRNAs and treated for 4 hours with vehicle or 1 nM R1881. (F) Western blot was conducted on protein lysates from LNCaP cells transfected with indicated siRNAs and treated for 4 hours with vehicle or 1 nM R1881. (G) ChIP assays using indicated antibodies in LNCaP cells transfected with indicated siRNAs. (H) ChIP assays using indicated antibodies in LNCaP cells transfected with indicated siRNAs and/or treated for 4 hours with vehicle or 1 nM R1881. (I) Cycloheximide (CHX) chase assays performed in LNCaP cells transfected with indicated siRNAs. *p<0.05, **p<0.01 – two-sided t-test. Results of ChIP and qRT-PCR assays reported as the mean of 2-4 replicates with error bars representing the standard deviation.
shown to depend on androgen stimulation (89). ChIP assays for GATA2 and FoxA1 in LNCaP cells revealed no such reliance on hormone stimulation, revealing that these pioneer factors engage AR regulatory elements prior to androgen induction of AR expression (Figure 2B).

To evaluate the relative contribution of each pioneer factor to the regulation of AR expression we individually reduced GATA2 and FoxA1 expression by siRNA transfection and measured AR mRNA and protein levels before and after stimulation with the synthetic androgen R1881. siGATA2 transfection resulted in a significant reduction in AR mRNA levels prior to and following hormone stimulation (Figure 2C), concordant with the observed reduction in GATA2 recruitment to the AR promoter and upstream regulatory elements (Figure 2D). Surprisingly, in hormone-depleted conditions, siFoxA1 transfection resulted in an increase in AR mRNA (Figure 2C) despite loss of FoxA1 binding to the upstream regulatory element (Figure 2E). These results were reflected in the depletion of AR protein levels following siGATA2 transfection in the presence and absence of hormone and in the moderate accumulation of AR protein following siFoxA1 transfection in the absence of hormone (Figure 2F). Notably, while siFoxA1 transfection resulted in an accumulation of GATA2 protein in the absence of hormone, this observation does not initially explain the concomitant increase in AR protein level as no significant increase in GATA2 binding to the upstream AR regulatory element was observed under these conditions (Figure 2G).
A previous report from Cai et al. found that FoxA1-mediated binding of AR to a downstream repressor region establishes a negative feedback loop to control AR expression (152). We performed AR and FoxA1 ChIP using primers for the reported repressor element and found that, consistent with previous results, hormone stimulation enhances occupancy of both factors to this region, enhancing the negative feedback loop. Following siFoxA1 transfection, basal AR occupancy of the repressor was significantly reduced in accordance with enhanced AR expression due to feedback loss. Hormone stimulation following FoxA1 knockdown was sufficient to enhance AR repressor binding and reestablish appropriate feedback signaling (Figure 2H), explaining our observed expression patterns following pioneer factor knockdown. We also performed cycloheximide (CHX) chase assays in LNCaP cells transfected with control-, GATA2-, or FoxA1-targeting siRNAs to rule out the impact of AR protein stabilization and/or destabilization on the observed pattern of AR accumulation, though only a marginal effect of pioneer factor silencing on AR half-life was observed (Figure 2I). While additional factors must be considered to gain a full understanding of pioneer factor function in regulating AR expression, these results provide evidence that GATA2 is capable of binding AR regulatory elements independent of hormone status and is integral to both basal and androgen-stimulated AR expression patterns. In this way, GATA2 contributes positively to androgen-responsive AR target gene expression by direct up-regulation of AR.
**GATA2 Enhances AR Target Gene Expression: Complexity in Combinatorial Pioneer Factor Occupancy**

While previous studies have demonstrated the importance of pioneer factors in androgen-stimulated AR target gene expression, rigorous analyses have been limited to the role of FoxA1 (15,105,143). As GATA motifs, and indeed, GATA2 binding events are significantly enriched alongside FOX motifs/FoxA1 binding events overlapping AR occupied regions, and as GATA2 knockdown or chemical inhibition is understood to result in reduction of AR target gene expression (83,87), we chose to systematically evaluate the interplay between GATA2 and FoxA1 in facilitating AR binding to androgen-responsive gene regulatory elements and the impact these factors have on androgen-dependent gene expression. As the vast majority of AR/GATA2 overlapping binding sites occur alongside FoxA1 (Figure 1), we focused our attention on genes apparently coregulated by these three factors. Referring once again to our ChIP-seq datasets, we identified overlapping AR, GATA2, and FoxA1 binding events within regulatory elements of two androgen-responsive genes that provided informative cases of combinatorial pioneer factor activity: *ABCC4* and *ADPGK*.

*ABCC4*, which encodes a member of the multi-drug resistance protein (MRP) subfamily of ATP-binding cassette transporter proteins has been shown to exhibit androgen-dependent expression (153). Importantly, three independent studies have suggested that while ABCC4 mRNA and protein expression is enhanced in primary prostate cancer compared to nonneoplastic prostate tissue, lost expression over time significantly
correlates with metastatic disease progression, biochemical recurrence following androgen-deprivation therapy, and high Gleason score (153-155). As such ABCC4 is a clinically relevant androgen-responsive gene product with potential prognostic value in predicting the development of androgen-independence in advanced prostate cancer. Intronic, androgen-stimulated AR binding in the ABCC4 locus (29.9 kb downstream of the TSS) was found to occur alongside hormone-independent GATA2 and FoxA1 binding sites (Figure 3A-B). Basal and androgen-stimulated AR recruitment to this site is significantly reduced following siFoxA1 transfection, and following siGATA2 transfection, AR recruitment is reduced nearly to the level observed following RNA interference of AR itself, demonstrating that each factor plays a central role in AR recruitment at this site (Figure 3C). Consonantly, androgen-stimulated ABCC4 mRNA expression in siGATA2 and siFoxA1 transfected cells was reduced nearly to siAR levels compared with control knockdown (Figure 3D). To assess the hierarchical relationship between AR, GATA2, and FoxA1 binding at this locus, we first characterized GATA2 and FoxA1 recruitment to the ABCC4 regulatory element following siAR transfection. Neither FoxA1 nor GATA2 enrichment was affected by AR knockdown in the presence or absence of androgen, demonstrating that both factors engage this site prior to androgen stimulation and independently of AR (Figure 3E). In addition, FoxA1 recruitment was unaffected by siGATA2 transfection as was GATA2 recruitment following siFoxA1 transfection, revealing that in this instance, these pioneer factors function independently of one another in binding this regulatory element prior to and following hormone stimulation (Figure 3F).
Figure 3: GATA2 facilitates AR-mediated ABCC4 expression. (A) AR, GATA2 and FoxA1 binding sites are shown in IGV. (B) LNCaP cells were grown for 72 hr in the absence of hormone and were then treated with 1 nM R1881 or vehicle for 4 hr before harvest. ChIP assays were performed using antibodies against AR, GATA2 or FoxA1. (C, D) LNCaP cells were transfected with siControl or factor-specific siRNA in the absence of hormone. After 72 hr, cells were treated with 1 nM R1881 or vehicle for an additional 4 hr. ChIP assays were conducted with AR-specific antibody, or (D) Real-time RT-PCR was performed using gene-specific primers. (E) LNCaP cells were transfected with siControl or siAR in the absence of hormone. 72 hr after siRNA transfection, cells were treated with 1 nM R1881 or vehicle for an additional 4 hr. ChIP assays were performed with indicated antibodies. (F) LNCaP cells were transfected with siControl, siGATA2, or siFoxA1 in the absence of hormone. 72 hr after siRNA transfection, cells were treated with 1 nM R1881 or vehicle for an additional 4 hr. ChIP assays were performed with indicated antibodies. *p<0.05, **p<0.01 – two-sided t-test. Results of ChIP and qRT-PCR assays reported as the mean of 2-4 replicates with error bars representing the standard deviation.
ADP-dependent glucokinase (encoded by \textit{APDGK}) was recently reported as a novel androgen-dependent AR target gene with a likely role in defining cellular energy metabolism in LNCaP cells (73). While mechanistic studies of ADPGK have not been conducted in a model of prostate cancer, a general role for the enzyme in supporting glycolytic energy production under hypoxic stress conditions characteristic of many solid tumors has been proposed, and in fact, \textit{ADPGK} null H460 human lung cancer cells exhibited decreased clonogenic survival in hypoxic conditions compared to parental cells (156). Of potential relevance to prostate cancer is the recent finding that ADPGK activity can induce NF-κB signaling in response to ROS production (157). The interplay between NF-κB and AR in the progression of prostate cancer has been of considerable interest and one report suggests that NF-κB2/p52 up-regulation may be a basis for resistance to the second-generation antiandrogen, enzalutamide (158). Within the \textit{ADPGK} locus, robust AR, GATA2, and FoxA1 binding sites overlap 49.6 kb downstream of the TSS, at an intergenic regulatory element (Figure 4A). Here, androgen stimulation resulted not only in enhanced AR binding but also in increased FoxA1 and, to a lesser extent, GATA2 binding (Figure 4B). This is consistent with previous studies showing significant increases in pioneer factor binding at sites exhibiting adjacent AR and pioneer factor motifs following hormone treatment and likely reflects global changes in the chromatin landscape, allowing for enhanced factor occupancy, or stabilized pioneer factor occupancy owing to interactions with fully assembled DNA-binding transcriptional complexes (83). However, this observation does raise the question of whether androgen-stimulated increases in FoxA1 and GATA2 are a result of pioneer factor-like AR activity,
as has been reported for the glucocorticoid receptor (GR), a related hormone-inducible transcription factor (159,160). AR knockdown revealed no impact on GATA2 or FoxA1 occupancy in the absence of hormone, indicating that these factors bind independently of AR at this location before androgen treatment. Yet in the presence of hormone, enhanced AR binding and transcription complex assembly likely stabilizes pioneer factor occupancy, as GATA2 and FoxA1 binding levels following hormone treatment were diminished by siAR transfection (Figure 4C and 4D). That GATA2 and FoxA1 act to recruit AR to this regulatory element was demonstrated by AR ChIP following siGATA2 or siFoxA1 transfection, as basal and androgen-stimulated AR enrichment was inhibited by knockdown of these pioneer factors (Figure 4E). Basal and androgen-stimulated expression of ADPGK was also significantly inhibited by knockdown of AR, GATA2, or FoxA1, indicating that all three factors are essential for transcriptional activation at this site (Figure 4F). The mechanism behind this strong dependence was readily explained by assessing the interplay between GATA2 and FoxA1 binding to this region. siFoxA1 transfection resulted in a significant loss GATA2 enrichment before and after hormone treatment, and similar results were obtained for FoxA1 enrichment following siGATA2 transfection (Figure 4G and 4H). These results indicate that androgen-stimulated AR binding to this regulatory element is facilitated following the co-dependent binding of GATA2 and FoxA1 to the site, and that loss of either of these pioneer factors results in lost binding of the complementary pioneer factor, binding of AR, and expression of ADPGK in this context.
Figure 4: GATA2 facilitates AR-mediated ADPGK expression. (A) AR, GATA2, and FoxA1 binding sites are shown in IGV. (B) LNCaP cells were grown for 72 hr in the absence of hormone and were then treated with 1 nM R1881 or ethanol for 4 hr before harvest. ChIP assays were performed using antibodies against AR, GATA2 or FoxA1. (C, D) LNCaP cells were transfected with siControl or siAR in the absence of hormone. After 72 hr, cells were treated with 1 nM R1881 or ethanol for an additional 4 hr. ChIP assays were conducted with indicated antibodies. (E, F) LNCaP cells were transfected with siControl or factor-specific siRNA in the absence of hormone. After 72 hr, cells were treated with 1 nM R1881 or ethanol for an additional 4 hr. ChIP assays were conducted with AR-specific antibody, or (F) Real-time RT-PCR was performed using gene-specific primers. (G, H) LNCaP cells were transfected with siControl, siGATA2, or siFoxA1 in the absence of hormone. 72 hr after siRNA transfection, cells were treated with 1 nM R1881 or ethanol for an additional 4 hr. ChIP assays were performed with indicated antibodies *p<0.05, **p<0.01 – two-sided t-test. Results of ChIP and qRT-PCR assays reported as the mean of 2-4 replicates with error bars representing the standard deviation.
To further support these findings, we performed a series of experiments in an additional androgen-responsive prostate cancer cell line C4-2B. Our results suggest that elements of this mechanism of AR target gene coregulation by GATA2 and FoxA1 are conserved across these two systems. A high degree of concordance was observed for patterns of hormone-independent pioneer factor occupancy of AR regulatory elements and regulation of AR expression (Figure 5A-C). Similarly, hormone-independent pioneer factor occupancy was observed in advance of androgen-stimulated AR binding to the ABCC4 and ADPGK regulatory regions, and loss of either pioneer factor significantly abrogated subsequent AR binding (Figure 5D-G). Despite the agreement between ADPGK expression changes following AR, GATA2, and FoxA1 knockdown, siFoxA1 transfection of C4-2B cells actually induced ABCC4 expression while AR and GATA2 knockdown resulted in reduced ABCC4 expression (Figure 5H-I), revealing cell line differences affecting context-specific pioneer factor function. This initial investigation of a small subset of androgen-responsive genes reveals two important characteristics of pioneer factor function in AR transcriptional activation. First, the results support a model that GATA2 serves a generally positive role in mediating androgen-stimulated binding of AR to target gene regulatory elements subsequent to direct regulation of AR. Secondly, GATA2 and FoxA1 exhibit two distinct models of combinatorial binding to AR target gene regulatory elements: independent and codependent.
**Figure 5:** GATA2 mediates androgen-responsive gene expression in CRPC cells. Experiments from Figures 2-4 were replicated in a CRPC cell line C4-2B. (A-I) C4-2B cells were grown in the absence of hormone for 72 hr followed by 4 hr treatment with ethanol or 1 nM R1881. (A,B,D-G) ChIP assays were performed using specified antibodies. (C,H,I) qRT-PCR assays were performed using mRNA-specific primers.
A well-established role for pioneer factors is in facilitating local chromatin reorganization, allowing for improved accessibility for additional DNA-binding transcription factors (9). We asked whether GATA2 might act in this way to enhance AR binding to distal gene regulatory elements, thus allowing for androgen-dependent AR target gene expression. Our histone mark ChIP-seq data (Z.C. and Q.W.) provided a picture of the chromatin environment within the vicinity of shared AR/GATA2/FoxA1 binding sites within AR target gene loci of interest, illustrating that these sites are enriched with active enhancer-specific histone modifications (histone H3 lysine 4 mono- and dimethylation, and lysine 27 acetylation – H3K4me1, H3K4me2, and H3K27ac, respectively) and void of the canonically repressive histone mark, histone H3 lysine 27 trimethylation (H3K27me3), and the promoter-specific histone H3 lysine 4 trimethylation (H3K4me3) (Figure 6A-B) (161).

We confirmed the ChIP-seq results for the ABCC4 and ADPGK regulatory elements, demonstrating that these sites exhibit typical enhancer characteristics. Relative to two control regions (GATA2 binding sites enriched in H3K27me3), androgen-independent levels of H3K4me1 and H3K4me2 were significantly enriched, while H3K4me3 was only modestly enriched (Figure 6C). Additionally, H3K27ac, an indicator of active chromatin, was enriched within ABCC4 and ADPGK regulatory elements prior to androgen-stimulation, compared to control loci (Figure 6D). Finally, we confirmed that these sites are depleted of the repressive H3K27me3, which showed robust enrichment
Figure 6: GATA2-bound enhancers exhibit active chromatin signature. (A, B) AR, GATA2, and FoxA1 binding sites at the ABCC4 and ADPGK loci are shown in IGV along with epigenetic marks; H3K27me3, H3K27ac, H3K4me1, H3K4me2, and H3K4me3. (C-E) LNCaP cells were grown for 72 hr in the absence of hormone. ChIP assays were performed using specified antibodies. (F) LNCaP cells were grown in the absence of hormone for 72 hr, and FAIRE-qPCR was performed using locus-specific primers.
within the control regions (Figure 6E). These results suggest that a relatively active chromatin environment can be found at the regulatory elements of androgen-dependent genes prior to hormone stimulation. In addition to an active chromatin signature, the \textit{ABCC4} and \textit{ADPGK} regulatory loci also exhibit enhanced accessibility as measured by Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) (150), indicating relatively high dissociation of DNA from nucleosomes within these regions (Figure 6F). That androgen-independent pioneer factor occupancy of these loci correlates with this active and accessible chromatin signature supports the established role for these factors in priming AR target genes for transcriptional responsiveness to hormone stimulation.

While the mechanisms by which FoxA1 influences chromatin accessibility is understood to derive from its structural similarity to the linker histone H1, allowing FoxA1 to displace H1 and disrupt heterochromatin (101,102), the impact of GATA2 on chromatin structure and histone modification status in AR signaling has been suggested through correlative studies, though not thoroughly described (87,88). Having observed an enrichment of the active histone mark H3K27ac, we asked what impact GATA2 and FoxA1 have on the recruitment of histone acetyltransferases (HATs) to the \textit{ABCC4} and \textit{ADPGK} loci prior to hormone stimulation. We focused on p300, as this histone-modifying enzyme showed the most robust enrichment within these sites prior to androgen treatment (Figure 7A-B) (13).
Figure 7: GATA2 mediates p300 recruitment and chromatin accessibility. (A,B) LNCaP cells were grown for 72 hr in the absence of hormone and were then treated with 1 nM R1881 or vehicle for 4 hr before harvest. ChIP assays were performed using specified antibodies. (C-E) LNCaP cells were transfected with siControl, siGATA2, or siFoxA1 in the absence of hormone. After 72 hr, (C-D) ChIP assays were performed using specified antibodies or (E) FAIRE-qPCR was performed using locus-specific primers. (F-I) LNCaP cells were transfected with siControl, siGATA2, or siFoxA1 in the absence of hormone. After 72 hr, ChIP assays were performed using specified antibodies. *p<0.05, **p<0.01 – two-sided t-test. Results of ChIP and FAIRE-qPCR assays reported as the mean of 2-4 replicates with error bars representing the standard deviation.
p300 enrichment was significantly influenced by pioneer factor status, as siGATA2 transfection resulted in decreased p300 recruitment to both gene regulatory elements in the absence of hormone, while siFoxA1 transfection only significantly decreased p300 levels at the ADPGK locus (Figure 7C). Additionally, levels of H3K27ac and patterns of chromatin accessibility at each gene locus mimicked p300 recruitment (Figure 7D-E). To our surprise knockdown of either pioneer factor resulted in a significant loss in H3K4me1/2 at both the ABCC4 and ADPGK loci (Figure 7F-I). While a similar observation was made previously upon FoxA1 depletion (105,143), this is a novel observation for GATA2. Though the mechanism is not immediately apparent, pioneer factor occupancy likely affects recruitment of H3K4me1/2-specific histone methyltransferases and/or demethylases. Interestingly, trends in hormone independent GATA2-mediated recruitment of p300 and establishment of active, accessible chromatin closely resemble trends in androgen-stimulated AR binding (Figure 3C and Figure 4E), suggesting a central role for GATA2 in providing an amenable environment for AR binding subsequent to hormone stimulation. That FoxA1-independent GATA2 occupancy of the ABCC4 locus is sufficient for the recruitment of p300 and the establishment of active, accessible chromatin, suggests that this is an important mechanism by which GATA2 functions upstream of AR to mediate androgen-responsive gene expression. Together, these results demonstrate that GATA2, with variable input from FoxA1, plays a general role in enhancing androgen-stimulated AR-mediated expression of these target genes by engaging gene regulatory elements in the absence of hormone and recruiting histone modifying enzymes that activate local chromatin. In this way, androgen-dependent gene loci are primed for expression following hormone stimulation.
Figure 8: GATA2 facilitates chromatin loop formation. (A) LNCaP cells were grown for 72 hr in the absence of hormone and were then treated with 1 nM R1881 or ethanol for an additional 4 hr before cell harvest. 3C assays were performed using locus-specific primers. (B) LNCaP cells were transfected with siControl, siFoxA1 or siGATA2 and grown in the absence of hormone for 72 hr. 3C assays were performed using locus-specific primers. (C-F) LNCaP cells were transfected with siControl, siGATA2, or siFoxA1 in the absence of hormone. After 72 hr, ChIP assays were performed using specified antibodies. *p<0.05, **p<0.01 – two-sided t-test. Results of ChIP and 3C assays reported as the mean of 2-4 replicates with error bars representing the standard deviation.
**GATA2 Facilitates Regulatory Chromatin Loop Formation**

A prominent feature of AR binding is the tendency towards occupancy outside of proximal promoter regions of its target genes. To facilitate transcription, long-distance interactions form between AR-bound distal enhancers and androgen-responsive gene promoters, requiring a host of intermediary protein-protein interactions involving the Mediator coregulatory complex (15,83,162,163). While it is understood that pioneer factor residence on regulatory DNA elements persists beyond the priming period and into periods of active transcription (9), their role in facilitating and maintaining activating chromatin loops is not well established. A previous study showed that FoxA1 is partially responsible for establishing regulatory chromatin loops in castration resistant prostate cancer cell lines (146), suggesting that this pioneer factor is able to recruit requisite loop-forming activities including the Mediator subunit MED1. We asked whether GATA2 and FoxA1 are required in regulatory loop formation for androgen-dependent AR target gene expression by performing quantitative chromosome conformation capture (3C) assays (164) at the *ABCC4* locus in LNCaP cells.

We assessed the interaction between the downstream *ABCC4* regulatory element, bound by AR, GATA2, and FoxA1, with the promoter region. The downstream site exhibited enrichment of MED1 prior to hormone treatment, reflected by the relatively high crosslinking frequency observed in the absence of hormone. Following androgen treatment, MED1 occupancy was enhanced, congruent with increased crosslinking frequency indicating stronger long-distance interaction between this site and the *ABCC4*
promoter (Figure 7A and Figure 8A). Importantly, siGATA2 and siFoxA1 transfection resulted in significant decreases in crosslinking frequency in the absence of hormone (Figure 8B), suggesting that these pioneer factors play important roles in facilitating basal chromatin loop formation, which may combine with basal active chromatin signatures and accessibility to prime ABCC4 for androgen-stimulated expression.

To assess the mechanism by which these pioneer factors contribute to chromatin loop formation and maintenance, we performed MED1 ChIP at the ABCC4 enhancer following siControl, siGATA2 and siFoxA1 transfection. Our results support previous findings that FoxA1 functions in the recruitment of MED1 to distal regulatory elements (146), as FoxA1 knockdown precipitated a significant reduction in MED1 occupancy. Importantly, siGATA2 transfection produced similar results, revealing that GATA2 also functions in the basal recruitment of chromatin loop forming activities (Figure 8C). As previous work has shown that phosphorylation of MED1 is required for locus looping and AR target gene expression (146), we performed p-MED1 ChIP to show that both FoxA1 and GATA2 are central to the recruitment of this activated MED1 form (Figure 8D). It is worth recalling that GATA2 and FoxA1 bind independently of one another at the downstream ABCC4 regulatory element (Figure 3F), thus decreases in MED1, p-MED1, and crosslinking frequencies are the result of independent pioneer factor knockdown and reflect the individual impact of each pioneer factor on loop formation and maintenance. MED1 and p-MED1 recruitment to the ADPGK locus following siFoxA1 and siGATA2 transfection mimicked results at the ABCC4 locus (Figure 8E-F).
though co-dependent binding of FoxA1 and GATA2 at this site prevents our evaluating the independent contribution of each pioneer factor in this instance.

**GATA2 Supports Genome-wide AR Binding**

Having identified three prominent mechanisms by which GATA2 contributes to androgen-responsive AR target gene expression, which suggests a generally positive role for this pioneer factor in recruiting AR to target gene loci subsequent to activation of chromatin within regulatory DNA elements, we sought to evaluate the prevalence of this activity in determining AR binding patterns genome-wide. To address this question, we performed AR ChIP-seq following hormone stimulation in LNCaP cells transfected with Control- or GATA2-targeting siRNA molecules. 49,082 (79.5%) of the AR binding sites identified in siControl conditions were lost following GATA2 knockdown (Figure 9A). Among the 12,529 (20.2%) AR binding sites that were maintained following siGATA2 transfection, a significant loss in ChIP-seq read density within these regions was observed. The loss and decrease of AR binding after GATA2 silencing presumably resulted from the combined loss of AR expression (Figure 2C&F) and GATA2-established active and accessible chromatin structures at these locations. Corroborating our single gene analyses, AR binding within the ABCC4 and ADPGK loci was severely reduced (Figure 9B). Additional, indirect effects on AR binding are also likely, as the nearly 80% loss in binding significantly outnumbers the direct overlap of the AR and GATA2 cistromes (Figure 1A-B), though these effects cannot be distinguished from the upstream effect on AR expression. In stark contrast to the tremendous gain in AR binding
sites previously reported following FoxA1 knockdown in prostate cancer cell lines (105,143), only 131 AR binding sites were unique to siGATA2 conditions. These siGATA2-specific AR binding sites fall into two categories: 1) direct GATA2 exclusion of AR binding and 2) trans-repression of AR binding from a distance. Notably only 19 sites exhibited a gain in AR binding at sites occupied by GATA2 in parental cells, indicating that more than 85% of gained AR binding events are the result of relief of indirect or trans-repressive GATA2 effects. Figure 9C gives examples of each of the three categories of AR binding sites. These results further substantiate our claim that through its important roles in up-regulating AR expression and recruiting AR to activated chromatin, GATA2 serves a generally, if not exclusively, positive role in defining AR genome-wide binding patterns that determine androgen-responsive gene expression profiles.
Figure 9: GATA2 supports genome-wide AR binding patterns. (A) Aligned reads heatmaps from AR ChIP-seq assays performed in LNCaP cells transfected with Control-(left) or GATA2-targeting (right) siRNA within a 10 kb window centered over AR peaks. Binding site totals are provided for each category, as are the average signal plots. (B) IGV images displaying significantly reduced AR binding within the ABCC4 (top) and ADPGK (bottom) loci. (C) IGV images of examples cases of each category of AR binding site. Chromosomal coordinates are provided.
Discussion

In the present study, we sought to close the gap in knowledge between mechanisms by which the prominent pioneer factors, GATA2 and FoxA1, contribute to AR-mediated gene expression in androgen-responsive prostate cancer cells. While FoxA1 has been a major focus in previous studies of AR activity, limited evidence has been provided to illustrate a mode of action for GATA2, though its contribution to AR target gene expression has long been appreciated. Our initial genome-wide analysis provided compelling evidence that GATA2 plays a pivotal role alongside FoxA1 in defining androgen-stimulated AR binding patterns and encouraged an in-depth investigation into the mechanisms employed by GATA2 both in dictating AR binding and in modulating AR target gene expression (Figure 1). We showed that GATA2 operates first by directly regulating expression of AR itself via occupation of novel upstream and promoter regulatory elements identified by ChIP-seq. While previous work from Böhm et al. reported direct regulation of AR by hormone-stimulated GATA2 binding at the promoter of the minor AR(A) isoform (89), we observed hormone-independent occupancy of GATA2 within regulatory elements of the major AR(B) isoform. As GATA2 knockdown resulted in significant loss of AR expression before and after androgen treatment, we suggest that GATA2 mediates both basal and hormone-stimulated expression of full-length AR from these binding sites (Figure 2).

Zaret and Carroll described a role for pioneer factors in “establishing competence for gene expression” by engaging gene regulatory elements in advance of transcription
factors, thereby priming genes for responsiveness to appropriate stimuli (9). FoxA1 facilitates NR-mediated gene expression by engaging compact chromatin, displacing H1 to generate localized accessibility, and also serves to recruit architectural components of transcription regulatory chromatin loops (101,146). These roles have been shown in various systems to contribute to NR target gene expression, but whether GATA2 similarly primes for ligand-inducible AR target gene expression has not been fully demonstrated. By investigating the contribution of GATA2 to androgen-responsive AR target gene expression we observed GATA2 binding in advance of hormone treatment and AR binding, allowing for androgen-stimulated AR recruitment to target gene regulatory loci and gene expression (Figures 3 and 4). Within our subset of AR target genes, we observed a positive impact of GATA2 on androgen-dependent gene expression, though it is reasonable to suspect that a repressive model of target gene coregulation may be adopted in a site-specific manner. Additionally, we have provided evidence that, like the documented ability of FoxA1 to preclude AR binding at certain genomic locations (105,143), a small class of GATA2 binding sites appear exclusionary of AR occupancy and thus also play a negative role in preventing AR transactivation in these regions (Figure 9). However, this subset is negligible in comparison to the FoxA1 study, and only a minor fraction of these sites appear to be the result of direct exclusion of AR by GATA2 occupancy. Future analysis of the repressive mechanisms of GATA2 and/or FoxA1 in determining AR-mediated gene expression profiles is therefore of great interest.
Our analysis has provided some insight into the mechanisms by which GATA2 contributes to AR target gene expression. Previous studies have suggested a notable correlation between GATA2 and p300 recruitment to AR target gene loci: a model corroborated by concurrent loss in binding of the pioneer factor and histone modifier following treatment with a natural compound (87). Additionally, it has been demonstrated in model systems that GATA2 DNA-binding and transactivation potential rely in part upon acetylation by p300, providing evidence of their direct interaction (165). We have extended these findings to show that GATA2 has a central role in recruiting p300 to target gene loci, whereupon this HAT activates chromatin via acetylation of H3K27, generating local chromatin accessibility prior to hormone stimulation. Within the \textit{ABCC4} gene locus, we can observe the individual contribution of pioneer factors to chromatin activation and we see that GATA2, but not FoxA1, knockdown results in lost p300 recruitment, diminished H3K27ac levels, and chromatin compaction relative to siControl transfection (Figure 7). Thus, we demonstrate that GATA2 establishes active and accessible AR target gene enhancers, facilitating AR binding and subsequent gene expression.

We also show that GATA2, via recruitment of the Mediator coregulatory complex subunit, MED1, can establish and/or maintain basal regulatory chromatin loops between AR-bound distal enhancers and androgen-responsive gene promoters. Again, at the \textit{ABCC4} locus, we find that knockdown of either FoxA1 or GATA2 independently results in significant loss of MED1/p-MED1 occupancy, which destabilizes chromatin looping prior to androgen stimulation (Figure 8). This function has been previously demonstrated...
for FoxA1 (146), yet the contribution of GATA2 to chromatin loop formation has been more elusive. Occupancy of the GATA family member, GATA1, along with its coactivator, FOG1, at the β-globin locus has been shown to correlate with loop formation (166), providing the most definitive evidence to suggest a direct involvement of GATA factors in establishing and/or maintaining chromatin loops. While additional factors may be identified that can stabilize MED1-enhancer occupancy and basal chromatin loop formation, we suggest that androgen-independent GATA2 occupancy of AR target gene enhancers is a critical determinant of MED1-mediated locus looping in preparation for hormone stimulated gene expression. Previous studies have implied significant roles for various transcription factors and coactivators in loop formation (20, 163, 166-168) and our current findings suggest that GATA2 plays a central role in these processes via recruitment of p300, allowing for basal chromatin remodeling, and MED1, facilitating basal locus looping.

Our genome-wide analysis of the impact of GATA2 silencing on AR distribution reflects two key functions of GATA2 in these prostate cancer cells. Firstly, GATA2 knockdown significantly reduces AR protein levels and subsequently, AR ChIP-seq signal intensity across the genome. Second, GATA2 silencing results in the loss of active chromatin signatures in regions primed for androgen-stimulated AR binding in parental cells, thus global ChIP-seq enrichment is diminished. While the comprehensive impact of GATA2 loss on transcription-regulatory chromatin looping remains in question, we have provided evidence that the loss of this critical GATA2 function, in combination with the dramatic impact on AR binding patterns, will broadly influence AR target gene expression.
Our findings support, and further contribute to, the growing body of knowledge of GATA family pioneer transcription factors that facilitate nuclear receptor-mediated, hormone-responsive gene transcription. Demonstrating that GATA2 is capable of facilitating open and active chromatin prior to hormone stimulation is concordant with previous observations that nucleosome-depleted regions (NDRs) in AR target gene enhancers overlap GATA2 binding sites that appear to be necessary for NDR maintenance in the absence of androgen (85). Additionally, the ER-collaborating pioneer factor, GATA3, important for breast cancer gene expression profiles, exhibits binding patterns that correlate with genome-wide ER-anchored chromatin loops associated with estrogen-responsive gene transcription (84,167). Further analysis is required to reveal whether or not GATA2 shares common functionality with GATA3, which appears capable of activating and/or silencing local chromatin in a site specific manner leading to enhanced and/or inhibited chromatin loop formation (84).

In summary, we have provided evidence that GATA2 contributes positively to androgen-dependent AR target gene expression by three distinct mechanisms. First, through binding regulatory elements of the AR locus GATA2 directly upregulates basal and hormone-stimulated AR expression. Secondly, by recruiting the histone acetyltransferase, p300, GATA2 facilitates basal AR target gene enhancer activity and promotes chromatin accessibility, allowing for subsequent androgen-stimulated AR binding and target gene expression. Finally, GATA2 recruits MED1/p-MED1 to AR target gene enhancers prior to hormone stimulation, leading to basal chromatin loop formation/maintenance at
androgen-responsive gene loci. These findings encourage future focus on elucidating additional means by which GATA2 contributes to AR-mediated gene expression to further our understanding of how this pioneer transcription factor influences AR binding patterns and downstream target gene expression in prostate cancer. Revealing critical determinants of AR activity offers an important opportunity to direct new therapeutic strategies at upstream components of the androgen signaling axis. As current prostate cancer therapeutic modalities too often fail in preventing biochemical recurrence in advanced cases, focusing future treatments towards activities of factors like GATA2, correlated with advanced prostate cancer (89), may prove a potent strategy.
Chapter 3: Integrative analysis identifies targetable CREB1/FoxA1 transcriptional co-regulation as a predictor of prostate cancer recurrence

Abstract

Identifying prostate cancer-driving transcription factors (TFs) in addition to the androgen receptor promises to improve our ability to effectively diagnose and treat this disease. We employed an integrative genomics analysis of master TFs CREB1 and FoxA1 in androgen-dependent prostate cancer (ADPC) and castration-resistant prostate cancer (CRPC) cell lines, primary prostate cancer tissues, and circulating tumor cells (CTCs) to investigate their role in defining prostate cancer gene expression profiles. Combining genome-wide binding site and gene expression profiles we define CREB1 as a critical driver of pro-survival, cell cycle, and metabolic transcription programs. We show that CREB1 and FoxA1 colocalize and mutually influence each other’s binding to define disease-driving transcription profiles associated with advanced prostate cancer. Gene expression analysis in human prostate cancer samples found that CREB1/FoxA1 target gene panels predict prostate cancer recurrence. Finally, we showed that this signaling pathway is sensitive to compounds that inhibit the transcription co-regulatory factor MED1. These findings not only reveal a novel, global transcriptional co-regulatory function of CREB1 and FoxA1, but also suggest CREB1/FoxA1 signaling is a targetable driver of prostate cancer progression and serves as a biomarker of poor clinical outcomes.
Introduction

The complexity, heterogeneity, and plasticity of prostate cancer have proven major obstacles in our understanding of the etiology and progression of this disease (106,107), and have provided a rich source for discovery of novel cancer concepts and a platform for the development new analytical methods (169). Critical questions remain as to the optimal approaches to characterize aggressive vs. indolent disease in the clinically localized setting, the factors that predict treatment response and failure, and the mechanisms underlying therapeutic failure that reveal novel targets for effective intervention. Specifically, the discovery of targetable prostate cancer drivers outside the androgen/androgen receptor (AR) signaling axis is paramount to achieving cures and improving the duration of therapeutic response. The importance of this concept is made increasingly apparent by the mounting reports of resistance to even the most potent second-generation anti-androgen therapeutics and continually emerging molecular mechanisms underlying such treatment failure (61).

Genomic analyses of primary and advanced metastatic prostate cancers have endeavored to reveal the alterations characterizing aggressive disease in hopes of identifying novel driver genes and pathways (106,107,170,171). While potentially actionable mutations in PI3K (PIK3CB), RAF, DNA repair (ATM and BRCA1/2), and cell cycle (CDK4) pathways were discovered in metastatic CRPC patient samples (172), primary tumor studies have returned few therapeutic candidates beyond AR itself. Additionally, while genome-wide copy number alterations (CNA) alone, or in combination with additional tumor characteristics, provide valuable prognostic information for early risk stratification
(173,174), they do not immediately identify the therapeutic compounds from which a high-risk prostate cancer patient may benefit. These observations encourage an alternate, mechanistic approach to the discovery of targetable disease-driving pathways that can simultaneously predict risk and indicate potential therapeutic strategies early in the course of treatment.

Our recent analysis of the pioneer transcription factor FoxA1 revealed its collaborative relationship with the cAMP response element binding protein (CREB1) in determining the expression of G1/S phase cell cycle progression genes (108). Additionally, FoxA1 and CREB1 supported the androgen-independent AR-mediated expression of G2/M phase cell cycle genes in CRPC cells. These results suggest that comprehensive analysis of genome-wide CREB1/FoxA1 cooperativity may define disease-relevant gene expression profiles during the evolution and progression of human prostate cancer.

Individually, while FoxA1 has been well characterized, CREB1-mediated transcription remains largely unexplored in prostate cancer. This important second messenger-responsive transcription factor (TF), in addition to promoting cell cycle progression, has been implicated in several prostate cancer-relevant functions including steroidogenesis (116) and bone metastasis (118). CREB1 is also emerging as a key mediator of metabolic programs in cancer cells, orchestrating the response to metabolic stress and multiple extracellular hormonal signals (119). Its kinase-inducible transactivation (i.e. via Serine 133 phosphorylation) results in the recruitment of CREB binding protein (CBP)/p300 to target gene loci and can be stimulated by multiple signaling pathways, some of which are
deregulated in prostate cancer (e.g. IGF-I and PI3K/Akt) (121). Consequently, CREB1 signaling represents an attractive therapeutic candidate. Complete profiling of CREB1 target genes should therefore expand on these insights and provide a rationale for the development and use of direct or indirect CREB1-inhibiting compounds in prostate cancer prevention or therapy.

Utilizing the LNCaP and abl cell lines, which model hormone-dependent and -independent growth characteristics of ADPC (androgen-dependent prostate cancer) and CRPC (castration-resistant prostate cancer), respectively (15,77,108,146), the current study is the first to explore global gene regulation by CREB1 individually and in collaboration with FoxA1. Our integrated genomics approach compiles CREB1 and FoxA1 binding site data with gene expression profiling to characterize cooperative CREB1/FoxA1 transcriptional activity. We also utilize these genomic datasets in a knowledge-driven approach to discover prognostic gene panels. We have demonstrated that coordinated CREB1/FoxA1 signaling contributes to disease-relevant transcription programs in both ADPC and CRPC cells. We further show that CRPC-specific, CREB1/FoxA1 target genes provide valuable tools for predicting prostate cancer recurrence. Importantly, we demonstrate that CREB1/FoxA1 transcriptional activity is sensitive to kinase inhibitors that impact the transcription co-regulator MED1. We propose that this work reveals a potentially targetable transcription regulatory signaling axis that may serve as a robust biomarker for disease recurrence.
Material and methods

Cell culture

LNCaP and 22rv1 cells were purchased from American Type Culture Collection (ATCC), and maintained in RPMI 1640 or DMEM, respectively, supplemented with 10% FBS. The abl cell line was kindly provided by Zoran Culig (Innsbruck Medical University, Austria; (72)), and maintained in phenol red-free RPMI 1640 supplemented with 10% charcoal-stripped FBS. Individual experiments were conducted in phenol red-free RPMI 1640 or DMEM supplemented with 5% charcoal-stripped FBS.

Quantitative RT-PCR

Real-time reverse transcriptase (qRT)-PCR was carried out as previously described (69). Cells were treated with vehicle, 5 or 10 µM H89, 10 µM U0126, or 3 µM MK-2206 for 24 hours prior to collection or with Control- (Dharmacon ON-TARGETplus Non-targeting Pool, D-001810-10-20) or AR- (Dharmacon ON-TARGETplus, L-003400-00-0020) targeting siRNA for 72 hours prior to collection. Total RNA was isolated with the RNeasy Mini kit (Qiagen, 74104). qRT-PCR was conducted using the MultiScribe Reverse Transcriptase and Power SYBR Green PCR Master Mix reagents (Applied Biosystems) according to the manufacturer's instructions. Each assay was repeated 3-4 times.
Western blotting

Western blot analyses were carried out as previously described (69). Briefly, cells were collected and lysed in RIPA buffer (1% NP-40, 0.1% SDS, 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5% sodium deoxycholate, 1 mM EDTA, 1X proteinase inhibitor cocktail (Roche), 1X PhosSTOP phosphatase inhibitor cocktail (Roche)) for 20 min on ice and the proteins were resolved on 8% SDS-polyacrylamide gels before being transferred onto nitrocellulose membrane (Bio-Rad). Membranes were blocked with 5% milk powder (Bio-Rad) then incubated with specific antibodies Ser133 phospho-CREB (87G3), CREB1 (48H2), Thr202/Tyr204 phospho-Erk1/2 (9101), Erk1/2 (9102), Ser472 phospho-Akt (D9E), and Akt (C67E7) (Cell Signaling Technology), AR (N-20), GAPDH (6C5), and TRAP220/MED1 (C-19) (Santa Cruz), FoxA1 (ab23738) (Abcam), Calnexin (ADI-SPA-860) (Enzo), or our own Thr1032 phospho-MED1 antibody (YenZyme) (146) for 2 hours at room temperature. Following incubation with secondary antibodies, immunoblots were visualized using the C-DiGit Chemiluminescent Western Blot Scanner (Li-Cor).

ChIP-qPCR

ChIP was performed as previously described (69). For kinase inhibitor assays, cells were treated with vehicle, 10 µM H89, or 10 µM U0126 24 hours prior to collection. For siCREB1 FoxA1 ChIP, cells were transfected with Control- (Dharmacon ON-TARGETplus Non-targeting Pool, D-001810-10-20) or CREB1- (Dharmacon,
SMARTPool: ON-TARGETplus, L-003619-00-0005) targeting siRNA 72 hours before collection using Lipofectamine 2000. For CREB1 overexpression ChIP, cells were transfected with Control (pCMV-LacZ - Clontech) or wild-type CREB1 (pCMV-CREB - Clontech) expression vectors 48 hours prior to collection with Lipofectamine 2000. Cells were then crosslinked with 1% formaldehyde for 10 min at room temperature and chromatin was collected, sonicated, diluted, and immunoprecipitated with 4 µg specific antibodies CREB1 (ab31387) and H3K27ac (ab4729) (Abcam), CBP (C-20), p300 (C-20), and RNA polymerase II (Pol II) (N-20) (Santa Cruz), or pMED1 (YenZyme) (146) at 4°C overnight. Protein A-Sepharose beads were added to the lysates and incubated for 1 hour with rotation. The beads were washed sequentially for 10 min each in TSE I (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl), TSE II (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl), and buffer III (0.25 M LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1) and finally 2 times with TE buffer. Chromatin complexes were eluted in 1% SDS, 0.1 M NaHCO3 and reverse crosslinked at 65°C for 16 hours. DNA fragments were purified with the QIAquick PCR purification kit (Qiagen 28104) and used as the template for quantitative PCR reactions with Power SYBR Green PCR Master Mix reagents (Applied Biosystems). ChIP assays were replicated 2-3 times.

ChIP-seq

ChIP-seq was performed as previously described (69). Briefly, 10⁷ cells were cross-linked in 1% formaldehyde for 10 minutes at room temperature. Chromatin was sheared,
diluted, incubated with specific antibodies against CREB1 (ab31387) or FoxA1 (ab23738) (Abcam), and precipitated with Protein A-Sepharose beads. ChIP-seq library preparation was performed using the Illumina Truseq ChIP Sample Preparation Kit (part# 15023092) with up to 10 ng of purified ChIP-enriched DNA. The libraries were amplified using 15 PCR cycles, and 50 bp single-end reads were generated on the Illumina HiSeq 2500 platform at the Ohio State University Comprehensive Cancer Center (OSUCCC) sequencing core with four multiplexed samples per lane. ChIP-seq reads were aligned to the human reference genome (hg19) with Bowtie v1.0.0 (147), allowing at most two mismatches. Uniquely mapped reads were utilized for peak calling using MACS v1.4.2 with p-value threshold of $10^{-8}$ (148). Aligned reads heatmaps and average signal plots were generated on the Galaxy platform using the Heatmap and SitePro functions (175). Motif analysis was performed using SeqPos on the Galaxy platform as well as Homer (149). Motif scanning was performed using Bio.motifs (176). For siCREB1 and siFoxA1 ChIP-seq, abl cells were transfected with Control- (Dharmacon ON-TARGETplus Non-targeting Pool, D-001810-10-20), CREB1- (Dharmacon, SMARTPool: ON-TARGETplus, L-003619-00-0005), or FoxA1- (Dharmacon, SMARTPool: ON-TARGETplus, L-010319-00-0005) targeting siRNA using Lipofectamine 2000 per the manufacturer’s protocol. 72 hours following transfection, CREB1 ChIP-seq was performed as described above in siControl and siFoxA1 transfected cells while FoxA1 ChIP-seq was performed in siControl and siCREB1 transfected cells. All ChIP-seq experiments were conducted in duplicate.
LNCaP and abl cells were transfected with Control-, CREB1-, or FoxA1-targeting siRNAs using Lipofectamine 2000 per the manufacturer’s protocol. 72 hours following transfection, RNA was harvested using the RNeasy Mini Kit (Qiagen). cDNA libraries were prepared using the Illumina Truseq RNA Sample Prep Kit (part# RS-122-2001) according to the manufacturer’s protocol. 50 bp single-end reads were generated on the Illumina HiSeq 2500 platform at the OSUCCC sequencing core with three multiplexed samples per lane. Read alignment was conducted using TopHat 2.0.8 (Bowtie2 2.1.0) (177), and relative transcript abundances were calculated using HTSeq 0.5.3p9 (178). edgeR 3.0.6 (179) was used to identify differentially expressed genes between LNCaP and abl cells (siControl vs. siControl) and after TF silencing in both LNCaP and abl cells (siCREB1 vs. siControl and siFoxA1 vs. siControl). All RNA-seq assays were conducted in duplicate.

**Target gene identification**

The Genomic Regions Enrichment of Annotations Tool (GREAT) (180) was used to collect all genes ±50 kb from all CREB1 or FoxA1 binding sites in LNCaP and abl cells. Resulting gene lists were filtered based on responsiveness to siCREB1 or siFoxA1 transfection compared to siControl ([Fold Change] ≥ 1.2, p-value ≤ 0.05, FDR ≤ 5%). Complementary gene identification approach began by classifying three sets of genes by their response to siCREB1 and siFoxA1 transfection (up-regulated, down-regulated, and
unchanged by silencing of both TFs). ±20 kb from the transcription start site of all such genes, we identified solo CREB1 or FoxA1 binding sites, non-overlapping CREB1+FoxA1 binding sites, or overlapping CREB1/FoxA1 binding sites. Responsive genes with either overlapping or non-overlapping binding were defined as co-regulated genes.

*Gene expression analysis in circulating tumor cells*

Patient consent and blood collection were processed according to IRB protocol approved at the University of Texas Health Science Center at San Antonio. Fifteen castration-resistant (CR) and ten castration sensitive (CS) patients were included in our cohort, with approximately twelve circulating tumor cells (CTCs) measured per patient. CTCs were isolated by microfiltration and immunostaining for EpCAM and CD45 following the method described previously with minor modifications (181). CTCs were retrieved from ScreenCell filters with a micromanipulator and stored in 4.7 \( \mu \)L 2X reaction buffer at -80 °C until further processing. Expression profiling of the 25 Panel A and B genes and a control gene, UBB, in CTCs was carried out in duplicate using microfluidic qRT-PCR as described previously (182). Primer sequences were obtained from PrimerBank (183) and synthesized by Sigma. Frozen, single CTCs were thawed on ice and the volume was diluted with 5 \( \mu \)L \( H_2O \). 3 \( \mu \)L of lysate was used for preamplification and microfluidic qRT-PCR. Average gene expression levels were calculated using the \(-\Delta\Delta C_t\) method. A Mann-Whitney U test was applied to analyze differences in expression levels between CR and CS patients for each gene.
Data analysis

Statistical tests, clustering analysis, Cox regression, and logistic regression were performed using JMP version 10.0.2 and R version 3.1.1 with publicly available expression array data downloaded from the Gene Expression Omnibus (GEO) (Taylor – GSE21036) or provided by Glinsky et al. (184). Gene ontology was performed using DAVID (185). To assemble our gene panels, we first prioritized up- and down-regulated genes that were common to multiple ONCOMINE (186) concepts to extract four relatively small gene lists representing CREB1/FoxA1 up- and down-regulated genes from LNCaP and abl cells. Utilizing cBioPortal (187), we further refined these lists to include only those genes exhibiting highly coordinated expression within each panel using data from Taylor et al. (171). Gene panel expression was calculated as the sum of gene ranks (with respect to all genes measured in a given data set) for all genes within a panel. For genes with multiple probes, average gene ranks were utilized. Panel A and Panel B rank sums were then converted to a Z-score for each patient.

Results

CREB1 transcriptional activity in prostate cancer models

Being uncharacterized with respect to its genome-wide transcriptional activity during the evolution of prostate cancer, we began the current study with a comprehensive identification of CREB1 target genes in LNCaP and abl cells. CREB1 ChIP-seq assays
identified a large set of common (n = 18,433) and cell type-specific (abl: n = 9,196, LNCaP: n = 3,391) binding sites occupied by CREB1 in the absence of androgen at a significance threshold level of $10^{-8}$ (Figure 1A-B). We had previously reported that total CREB1 expression is elevated in abl compared to LNCaP cells (108). Noting that the majority of cell type-specific CREB1 binding sites were found in abl, we asked if elevated CREB1 expression might explain this observation. Overexpression of wild-type CREB1 in LNCaP cells induced a marginal increase in CREB1 binding at abl-specific sites, though LNCaP CREB1 occupancy never achieved the level observed in abl at these same regions (Figure 10C-D). On the other hand, at common CREB1 binding sites in the CCNE2 and E2F1 promoter regions, which exhibit elevated CREB1 occupancy in abl, CREB1 overexpression in LNCaP resulted in stronger CREB1 binding that matched the levels observed in abl cells (Figure 10D). These results suggest that CREB1 expression level is just one factor determining the distribution and intensity of CREB1 binding sites.
Figure 10: Genomic distribution of CREB1 binding in ADPC and CRPC cells. (A) Aligned reads heatmap of CREB1 ChIP-seq signal intensity centered over LNCaP-specific (black), common (dark gray), and abl-specific (light gray) CREB1 binding sites in descending order of peak strength within each category. Average signal plots for LNCaP and abl are provided (right) for each category. (B) Overlap of CREB1 peaks identified in LNCaP and abl cells. Common binding site numbers reflect the use of abl or LNCaP CREB1 peaks as bait. (C) Western blot analysis of LNCaP cells transiently transfected for 48 hours with pCMV vectors driving control gene (LNCaP-LacZ) or wild-type CREB1 (LNCaP-CREB) expression. (D) CREB1 ChIP assays (n = 3) were performed in LNCaP-LacZ, LNCaP-CREB (48 hours post-transfection), and parental abl cells. qPCR analysis was performed with primers for 8 abl-specific CREB1 binding sites as well as the CCNE2 and E2F1 promoter regions (2 Common CREB1 binding sites with enhanced CREB1 binding in abl cells). * p-value < 0.05, NS – not significant, two-sided t-test.
We found that common CREB1 binding sites were primarily located in proximal, gene promoter regions (51.6%), while cell line-specific binding sites most often mapped further than 5 kb from the nearest transcription start site (TSS) (cumulatively 83.5% and 85.1% for LNCaP-specific and abl-specific, respectively) (Figure 11A). While the overall proportion of CREB1 distal binding events in LNCaP cells (~13.2%) is consistent with previous genome-wide analyses for this TF in other systems (188), the proportion of distal binding sites increases more than two-fold in abl cells (~28.6%), providing an example of high frequency, distal CREB1 binding not previously reported. Analyzing additional ChIP-seq datasets (77,189) (Z.C. and Q.W.), we observed that common CREB1 binding sites indeed resembled active promoters marked by concurrent histone H3 lysine 4 di- and trimethylation (H3K4me2 and H3K4me3, respectively), with robust RNA Polymerase II (Pol II) enrichment (Figure 11B). These regions were depleted of repressive H3 lysine 27 trimethylation (H3K27me3), which shows robust accumulation over the TSS of silent compared to highly expressed genes in LNCaP and abl cells (Figure 11C). On the other hand, cell line-specific CREB1 binding sites displayed an active enhancer signature characterized by H3K4me2 and Pol II enrichment in the absence of both H3K4me3 and H3K27me3 (Figure 11B) (161).
Figure 11: Cell type-specific binding of CREB1 in enhancers. (A) Absolute distance to the nearest TSS was calculated for CREB1 binding sites in each category using the Genomic Regions Enrichment of Annotations Tool (GREAT) (180). (B) Average signal plots for H3K4me2, H3K4me3, H3K27me3 (Brown lab) (77, 189), and Pol II ChIP-seq assays in LNCaP and abl cells were plotted over CREB1 binding site categories. (C) H3K27me3 ChIP-seq dataset quality was assessed by mapping signal intensities to 4 kb windows centered over the transcription start sites of genes that were silent or highly expressed in both LNCaP and abl cells as assessed by RNA-seq.
This intriguing result lead us to ask whether cell line-specific enhancer binding of CREB1 corresponds with differential gene expression profiles in LNCaP versus abl cells. We therefore conducted RNA-seq following transfection with CREB1- or control-targeting siRNA and identified the genes nearest common and cell type-specific CREB1 binding sites that were responsive to CREB1 silencing (i.e. putative target genes) using the genomic regions enrichment of annotations tool (GREAT) (180). Using the siControl RNA-seq data in LNCaP and abl cells, we then analyzed the basal expression level of these CREB1 target genes to find that cell line-specific enhancer binding of CREB1 corresponds to a higher degree of differential gene expression (proportion of differentially expressed genes: LNCaP-specific = 0.327, abl-specific 0.245) compared to common CREB1 binding in promoter regions (proportion of differentially expressed genes = 0.118) (Exact binomial test p-value < 10^{-16}) (Figure 12, left). Gene ontology analysis revealed that differentially expressed genes up-regulated by abl-specific CREB1 binding sites were enriched for cell cycle processes, and genes down-regulated in either cell line by LNCaP- or abl-specific CREB1 binding were in enriched in apoptotic signaling. On the other hand, genes occupied by common CREB1 binding sites and exhibiting similar expression patterns between LNCaP and abl appear to maintain metabolic gene expression in both cell lines (Figure 12, right) (119). This initial investigation of CREB1 in distinct models of prostate cancer suggests a broad and important role for this TF in determining cancer-relevant cell cycle, pro-survival, and metabolic gene expression patterns via promoter as well as enhancer binding activity.
Figure 12: Enhancer-bound CREB1 is associated with cell type-specific, prostate cancer gene expression profiles. Nearest genes returned by GREAT for LNCaP-specific, common, and abl-specific CREB1 binding sites were filtered based on response of each gene to siCREB1 transfection compared to siControl transfection in LNCaP and abl cells measured by RNA-seq ([Fold Change] ≥ 1.2, pvalue ≤ 0.05, FDR ≤ 5%). The log2(Fold Change[LNCaP/abl]) for the remaining putative CREB1 target genes is plotted for each category (Expression values from siControl data sets). Gene ontology analysis of putative CREB1 target genes differentially or commonly expressed in LNCaP vs. abl cells was performed using DAVID. The top five ontologies for each gene set are provided.
FoxA1 and CREB1 interact preferentially in CRPC cells

Recent investigations have documented the phenomenon of “indirect peaks” identifiable in large-scale binding site data sets (190). This class of peaks are a newly appreciated result of the ChIP-seq protocol, which enriches not only the DNA fragments to which a TF is immediately bound through direct protein-DNA interactions, but also regions of the genome with which a TF may interact indirectly through scaffolding proteins involved in chromatin loop formation. Indirect peaks are generally characterized by lower affinity binding in comparison to direct peaks, and importantly, they lack the recognition motif that defines sequence-specific TF-DNA binding. As we identified relatively weak CREB1 binding in cell type-specific enhancer regions (Figure 10A), we next performed motif analysis to determine whether these regions satisfy both criteria for indirect peaks.

Motif searching and \textit{de novo} motif analysis of common and cell type-specific CREB1 binding sites identified previously reported half- and full-length cAMP response elements (CREs) (Figure 13A) (188,191). Specific enrichment of these relatively short motifs in all binding site categories was confirmed by motif scanning in all CREB1 binding sites. Nearly all common CREB1 binding sites exhibited some version of the CRE (11% - Full CRE, 85% Half CRE), and most cell line-specific CREB1 binding sites also contained a CRE (LNCaP-specific: 8% Full CRE and 75% Half CRE, abl-specific: 7% Full CRE and 73% Half CRE). We also conducted scans for full and half CREs in random genomic intervals matched to our three binding site categories and found that these motifs were significantly enriched in CREB1 binding sites compared to background (Figure 13B).
Together, these results suggest that CREB1 occupies distal enhancer regions in addition to proximal promoter regions via direct, sequence-specific DNA binding.

We also found the FoxA1 recognition motif among the most significantly enriched sequences in each CREB1 binding site category (Figure 13A), thus we performed FoxA1 ChIP-seq in LNCaP (69) and abl cells to characterize their genome-wide interaction. By plotting the FoxA1 ChIP-seq signal over CREB1 binding site categories, we get a picture of the extent to which these factors interact in both LNCaP and abl cells (Figure 13C). We found that abl FoxA1 peaks are significantly enriched compared to LNCaP FoxA1 peaks within abl CREB1 binding sites (both common and abl-specific). On the other hand, there was no significant enrichment of LNCaP FoxA1 binding events within LNCaP-specific CREB1 binding sites (Figure 13D). The preferential CREB1/FoxA1 interaction in CRPC cells was further supported by an inverse relationship between CREB1 and FoxA1 ChIP-seq signal intensity in LNCaP-specific CREB1 bound regions alone (Figure 13E). Large scale efforts aimed at unraveling complex transcription factor networks have previously revealed the context specificity with which such proteins interact to drive unique gene expression programs in various cancer models (192). Our results provide an interesting glimpse at such a phenomenon occurring in a model of disease progression within a single cancer type, and it will be of interest to elucidate which of the many candidate TFs identified by motif searching preferentially interact with CREB1 in the ADPC context.
Figure 13: CREB1 and FoxA1 interact preferentially in CRPC cells. (A) Motif analysis of CREB1 binding site categories. (B) Motif scanning for the Full and Half CRE in CREB1 binding site categories and background locations. P-values (Fisher’s exact test) for the enrichment of motifs within CREB1 sites vs. background are provided below each plot. (C) Heatmap of FoxA1 ChIP-seq signal over CREB1 binding sites ordered as in Figure 10A. (D) FoxA1 peak overlap with CREB1 binding site categories in LNCaP and abl cells (Fisher’s exact test - * indicates p-value < 10^{-5}, N.S. = not significant). (E) LNCaP and abl FoxA1 ChIP-seq average signal is plotted over cell type-specific CREB1 binding sites.
CREB1/FoxA1 co-regulatory activity

We then sought to understand how differences in the spatial overlap between CREB1 and FoxA1 binding in abl compared to LNCaP cells might impact on their functional overlap in defining common gene expression profiles. We performed RNA-seq in LNCaP and abl cells following FoxA1 silencing for inclusion in our analysis. We then employed a global target gene identification approach for CREB1 and FoxA1, which acknowledged promoter and/or enhancer binding of each TF near potential target genes. Integrating binding site and gene expression profiles from LNCaP and abl cells, we identified genes ±50 kb from all CREB1 and FoxA1 binding sites in LNCaP and abl cells. We filtered the resulting gene lists based on responsiveness to siCREB1 or siFoxA1 transfection, respectively (edgeR: ≥ 1.2 fold change compared to siControl, p-value < 0.05, and false discover rate (FDR) < 5%). Figure 14A reports the number of CREB1- and FoxA1-responsive genes identified as well as the degree of overlap between CREB1- and FoxA1-regulated genes in LNCaP and abl cells. Consistent with the heightened physical overlap of CREB1 and FoxA1 binding sites in abl cells, these TFs appear to co-regulate a much larger set of common transcriptional targets in abl cells compared with LNCaP cells (cumulatively 4,117 and 1,867 up-/down-regulated genes in abl and LNCaP, respectively) (Figure 14B). A complementary target gene identification approach (see Materials and Methods) (104) revealed that CREB1/FoxA1 co-regulated genes in LNCaP cells were preferentially driven by non-overlapping CREB1 and FoxA1 binding events, while CREB1/FoxA1 target genes in abl cells were specifically enriched with overlapping TF binding sites (Figure 14C).
Figure 14: CREB1 and FoxA1 define common target genes. (A) GREAT was used to identify genes ±50 kb from all CREB1 or FoxA1 binding sites in LNCaP and abl cells. Resulting lists were filtered based on response of each gene to siCREB1 or siFoxA1 transfection compared to siControl transfection in LNCaP and abl cells measured by RNA-seq ([Fold Change] ≥ 1.2, p-value ≤ 0.05, FDR ≤ 5%). siCREB1 and siFoxA1 responsive genes from each cell line. (B) Overlap of CREB1 and FoxA1-responsive genes from each cell line. (C) Complementary, gene-centric integration of ChIP-seq and RNA-seq datasets. CREB1 and FoxA1 binding sites were identified within 20 kb of CREB1/FoxA1 responsive genes. Genes were plotted according to their pattern of CREB1/FoxA1 co-regulated genes in LNCaP vs. abl cells.
While the complementary approach was useful in revealing this intriguing aspect of cooperative TF activity, and identified 100% of the co-regulated genes found from our initial target gene identification approach, it was less conservative overall. Gene ontology analysis revealed that cooperative CREB1/FoxA1 transcriptional activity likely supports important, disease-driving expression profiles in both ADPC and CRPC cells. For example, in abl cells, CREB1/FoxA1 enhance cell cycle, more specifically mitotic, signaling while repressing differentiation programs (Figure 14D).

**Complex hierarchy characterizes CREB1-FoxA1 interaction**

As CREB1 and FoxA1 have been widely regarded as promoter- and enhancer-binding factors, respectively, it was intriguing to reveal the breadth of their physical interaction in CRPC cells, and we wondered if these factors might influence each other’s binding patterns. We have previously explored the hierarchical nature of CREB1 and FoxA1 binding near a small number of genes in abl cells (108), but we sought to define this relationship at the genome-wide level. To this end, we conducted CREB1 ChIP-seq in abl cells following siControl or siFoxA1 transfection as well as FoxA1 ChIP-seq following siControl or siCREB1 transfection. We found that a minority of FoxA1 binding sites (3,531/61,423) were lost upon CREB1 silencing (Figure 15A). Though just 10% of these sites overlapped with CREB1 peaks, we were able to find the CREB1 sequence motif within this category of siControl-specific FoxA1 binding sites, suggesting some small level of direct CREB1 influence on FoxA1 binding. While we observed a striking gain of 16,196 siCREB1-specific FoxA1 binding sites primarily outside of promoter regions,
only 8.3% of these regions were occupied by CREB1 under siControl conditions, and motif analysis failed to discover the CREB1 binding motif in this FoxA1 binding site category. We validated a set of siControl- and siCREB1-specific FoxA1 binding sites by standard ChIP-qPCR and found that the influence of CREB1 on FoxA1 binding patterns was robust and reproducible (Figure 15B). Similar patterns were observed in our analysis of CREB1 ChIP-seq data (Figure 15C). Approximately 20% of CREB1 sites were lost upon FoxA1 silencing. These sites existed primarily outside of promoter regions, overlapped considerably with FoxA1 binding sites, and were highly enriched for the FoxA1 binding motif. In contrast, CREB1 binding sites that were gained upon FoxA1 silencing occurred more frequently within promoters, did not overlap with FoxA1 binding sites, and were depleted of the FoxA1 motif.

Altogether, these observations demonstrate that while the vast majority of CREB1 and FoxA1 binding sites are independent of the other factor, subsets of each factor’s cistrome are influenced both positively and negatively by the direct as well as indirect actions of the other factor. Our group and others have similarly explored the effect of cooperative transcription factors upon one another as they colocalize to common genomic regions (69,84). These investigations have generally revealed unidirectional patterns of influence in which pioneer factors like FoxA1 appear unaffected by the loss of their binding partners. Other works have also noted the high prevalence of trans binding repression observed here (105), wherein silencing of one binding partner allows for new binding of the second partner in regions unoccupied by either factor under standard conditions.
Figure 15: Cooperative CREB1 and FoxA1 binding site profiles. (A) Heatmap of FoxA1 ChIP-seq signal intensity over siControl-specific, Common, and siCREB1-specific FoxA1 binding sites. Boxes to the right display binding site numbers, motif occurrences, promoter occupancy, and overlap of CREB1 binding with siControl and siCREB1 specific FoxA1 binding sites. (B) Standard ChIP-qPCR (n = 2) validation of siControl- and siCREB1-specific FoxA1 binding sites identified by ChIP-seq. * indicated two-sided t-test p-value < 0.05. (Figure 15 caption continued on page 86)
(Figure 15 caption continued from page 85) (C) Heatmap of CREB1 ChIP-seq signal intensity over siControl-specific, Common, and siFoxA1-specific CREB1 binding sites. Boxes to the right display binding site numbers, motif occurrences, promoter occupancy, and overlap of FoxA1 binding with siControl and siFoxA1 specific CREB1 binding sites. (D-E) Gene ontology analysis of genes ±50 kb from CREB1-influenced FoxA1 (D) and FoxA1-influenced CREB1 (E) binding sites that were responsive to CREB1 or FoxA1 silencing, respectively.

It is interesting to note that *trans* repression was the primary manner in which CREB1 influenced FoxA1 binding in abl cells (Figure 15A). Consistent with a previous study of the impact of FoxA1 silencing on AR genomic binding patterns (105), FoxA1 exerts both *cis* recruitment and *trans* repressive influence over CREB1 binding (Figure 15C). Our current investigation suggests that complex relationships exist between factors whose functional overlap in defining gene expression programs may be more profound than the physical overlap in their binding patterns, as appears to be the case for CREB1 and FoxA1. To provide further support for the notion of a functional CREB1/FoxA1 overlap resulting from their mutual impact on each other’s binding patterns, we identified genes regulated by FoxA1-influenced CREB1 binding sites (either lost or gained) and CREB1-influenced FoxA1 binding sites using our integrated target gene identification approach. This revealed nearly absolute concordance in the pathways influenced by the complex hierarchical relationship between CREB1 and FoxA1 (Figure 15D-E).

*CREB1/FoxA1 target genes associated with recurrent prostate cancer*

Having revealed that the CREB1/FoxA1 target genes identified in LNCaP and abl cells are enriched within disease-relevant biological processes, we next asked if these genes were associated with human prostate cancer progression. CREB1/FoxA1 up- and down-
regulated genes were uploaded for ONCOMINE Concept Analysis (186) to assess their enrichment in gene sets whose differential expression characterizes diverse patient cohorts of varying disease severity. We found that CREB1/FoxA1 target genes were highly enriched in numerous ONCOMINE Concepts, and that targets identified in LNCaP and abl cells mapped almost entirely to non-overlapping concepts (Figure 16A). For example LNCaP up-regulated targets were among the top 5% overexpressed genes in cancer vs. normal prostate tissue as measured in a study by Taylor et al. (171), while abl up-regulated genes were among the top 1% overexpressed genes in patients from the same study who had recurrent disease within five years after prostatectomy.

Prioritizing genes that were common to multiple ONCOMINE Concepts and therefore broadly characteristic of diverse patient populations, we then refined our large lists of CREB1/FoxA1 target genes identified in LNCaP and abl cells (see Materials and Methods) to arrive at two gene panels we hypothesized would correlate with poor outcomes: the 19-gene Panel A (CREB1/FoxA1 up-regulated genes identified in abl cells: TPX2, CDKN3, BUB1, TOP2A, NCAPG, UBE2C, PLK1, ASPM, CENPF, C15orf42, KIF4A, PBK, NUSAP1, HMMR, CIT, ANLN, SKA3, BIRC5, PRC1) and the 6-gene Panel B (CREB1/FoxA1 down-regulated genes identified in abl cells: AZGP1, CRISPLD2, EPHX2, HSPB8, PTRF, MFGE8). For analysis of the prognostic/predictive power of our final gene panels, we selected two cohorts with published gene expression microarray data in primary prostate cancer tissue from relatively large cohorts of mainly low and intermediate risk patients: Taylor et al. and Glinsky et al. (184). These studies include extensive clinical information for each patient including median follow-up ≥ 5
years, and importantly, these studies were the only two identified that measured a majority of our panel genes (exceptions being the absence of data for C1orf42, ANLN, and SKA3 in the Glinsky cohort). Kaplan-Meier survival analysis demonstrated that high expression of Panel A and low expression of Panel B correlated strongly with shorter time to biochemical recurrence in the Taylor and Glinsky cohorts (Figure 16B) (184). We were unable to assemble any such prognostic gene panel from LNCaP CREB1/FoxA1 target genes. To address the possibility that expression patterns of Panel A and B genes generally correspond with poor outcomes in solid malignancies rather than characterizing prostate cancer outcomes specifically, we used cBioPortal (187) to perform survival analysis in several additional cancers, and found that these gene panels do not appear to relate to shorter overall survival (breast, colorectal, head and neck, and lung) or disease-free survival (ovarian) (193-195). By simulating the selection of 2,000 random 19- and 6-gene sets, we found that the probabilities of randomly arriving at similarly predictive gene panels (i.e. having log-rank p-values in the Taylor cohort of 5.3x10^{-4} and 0.0022, respectively) were just 0.0015 and 0.025 for Panel A and Panel B, respectively.

Ontology analysis of Panel A genes revealed that this cohort of CREB1/FoxA1 targets are primarily responsible for enhancing mitotic cell cycle. This supports our previous findings that both G1/S and G2/M phase progression are mediated in part by cooperative CREB1/FoxA1 transcriptional activation (involving our previously described Panel A member, UBE2C) (15,108,146). Interestingly, our Panel A genes overlap considerably (10 out of 19) with a 31-gene panel used to develop a cell cycle progression score for determining risk of biochemical recurrence (196). Additionally, CENPF, recently
reported as a top-tier driver of advanced prostate cancer, is also present in our panel (197). Together, these data support the concept that enhanced cell cycle gene expression patterns are a hallmark of aggressive prostate cancer. The fact that multiple investigations have demonstrated the ability of cell cycle gene expression profiles to predict the recurrence of primary prostate cancers suggests that the further development of these panels would be of definite clinical utility.

Panel B genes, are largely absent from reported prognostic gene panels for prostate cancer. A notable exception is AZGP1, which can be found among the genes measured in the Oncotype DX® Prostate Cancer Assay used to more accurately stratify clinical disease risk. AZGP1 has also been reported to be a prostate tumor suppressor (198,199). Tumor suppressive functions of additional Panel B genes have also been reported. In prostate cancer models, restored PTRF expression was shown to diminish angiogenesis and metastatic potential (200,201). Additionally, reversal of DNA methylation-dependent HSPB8 silencing in prostate cancer cells induced apoptosis (202). Other Panel B genes appear to have more complex functionality. MFGE8 may serve as a tumor suppressor in ER+/HER2+ breast cancer cells, while its silencing in triple negative breast cancer cells results in enhanced sensitivity to DNA crosslinking chemotherapy (203). EPHX2 has been reported as a putative tumor suppressor in breast cancer, though its role as a potent oncogene, corresponding to AR-mediated signaling, has also been reported in prostate cancer (204-206). Finally, while its related CAP protein family member CRISP3 exhibits enhanced expression in prostate cancer compared to normal tissue, suggesting its role as an oncogene in primary cancer, no such report was found for CRISPLD2 (206).
Figure 16: CREB1/FoxA1 target genes associated with aggressive prostate cancer. (A) ONCOMINE Concept Analysis of CREB1/FoxA1 co-regulated genes in LNCaP vs. abl cells. P-value/odds ratio for ONCOMINE Concept enrichment is provided for each gene category. (B) Kaplan-Meier curves comparing recurrence-free survival of two patient cohorts (Taylor et al. and Glinsky et al.) distinguished by low (bottom quartile) vs. high (top quartile) expression of Panel A or Panel B genes. (C-D) IGV plots of CREB1 and FoxA1 ChIP-seq signal over example Panel A (C) and Panel B (D) gene loci. Vertical scales are the same for each factor between cell lines.
We found that the recently reported TUSON model, which is based on genetic alteration rates in cancer vs. normal tissue, predicts few oncogenic or tumor suppressor functions for any of our Panel A or B genes, respectively, in either a cross-cancer or prostate cancer-specific analysis, (207). This finding suggests that transcription regulation mediated by CREB1/FoxA1 activity, rather than mutation, could provide a means of oncogene activation and tumor suppressor loss in advanced prostate cancer. Additional inspection of mutation data from six prostate cancer cohorts using cBioPortal (187) revealed extremely low mutation rates for Panel A and Panel B genes; ASPM was the most frequently mutated, exhibiting entirely discrete alterations in 1.54% of the 972 patients analyzed across all cohorts. This further supports the notion that our genes panels represent so-called “epi-driver genes,” which are distinguishable from commonly mutated cancer drivers and characterized by abnormal expression patterns in cancerous tissue (208).

To understand the basis for their deregulated expression from a mechanistic perspective, we inspected CREB1 and FoxA1 occupancy at Panel A gene loci. This revealed that while CREB1 binding remains relatively constant at Panel A promoters between ADPC and CRPC cells, FoxA1 binding accumulates in the distal enhancer regions of Panel A genes in abl cells (Figure 16C). Around Panel B genes, we found several examples of abl-specific CREB1 and/or FoxA1 binding sites in the distal regulatory regions of Panel B genes (Figure 16D). This indicates a potential disruption of Panel B gene maintenance in CRPC cells via disease stage-specific, CREB1/FoxA1 enhancer binding. Reviewing the siCREB1/siFoxA1 ChIP-seq data, we found examples of CREB1-trans-repressed
FoxA1 binding sites near NCAPG, C15orf42, PBK, HMMR, CIT, and MFGE8 as well as a FoxA1-cis-recruited CREB1 binding site near AZGP1, further demonstrating these predominant modes of mutual influence between CREB1 and FoxA1 within our gene panels.

**CREB1/FoxA1 co-regulated genes predict prostate cancer recurrence**

At this point, we performed clustering analysis of patients in the Taylor and Glinsky cohorts to determine whether Panel A and Panel B gene expression levels could identify distinct patient subsets (Figure 17A-B). Indeed, contrasting clusters emerged within each cohort. Notably, the clusters marked by orange and red bars in each cohort, representing strong expression of Panel A and weak expression of Panel B genes, capture a larger proportion of patients with a biochemical recurrence event after prostatectomy (Taylor = 24/36, Glinsky = 24/37) than the green cluster representing the opposite gene expression pattern (Taylor = 7/36, Glinsky = 12/37) (Fisher’s Exact: Taylor p-value = 0.0001, Glinsky p-value = 0.01). We also noted that values for serum prostate-specific antigen (PSA) concentration at prostatectomy and Gleason Score differed significantly between the Glinsky and Taylor cohorts (annotated along the top of Figure 17A-B) (PSA – Mann-Whitney U p-value = 0.003, Gleason – Pearson \( \chi^2 \) p-value < 0.0001), thus we chose to pool these patients into a combined cohort to more accurately represent a broader population for subsequent analyses.
Figure 17: CREB1/FoxA1 target genes predict prostate cancer recurrence. (A) Taylor cohort was clustered (Pearson correlation) by Panel A and Panel B expression. Clinical information is annotated along the top of the figure. (B) Glinsky cohort was clustered as in (A). Clinical information is plotted along the top of the figure. (C) Logistic regression models (n = 20 for each combination of model parameters) were trained on a random 75% of the combined cohort and validated on the remaining 25%. AUC, Specificity, and Sensitivity values for each model are displayed as box plots. Significant difference assessed by Mann-Whitney U with α = 0.0083 to correct for multiple comparisons (* - Significant (p-value displayed), NS – Not Significant).
We hypothesized that expression patterns of Panel A and B genes might combine with clinical features to add prognostic value when predicting risk of prostate cancer outcomes, and we first assessed whether Panel A and B levels were independent of PSA, Gleason Score, and stage in our combined cohort. We found that Panel B was statistically correlated with Gleason Score alone, and similarly, Panel A was only correlated with a change in Gleason from 7 to above 7 (Table 1). In general, these results were recapitulated in analyses of the Taylor and Glinsky cohorts individually. Together these results suggested that Panel A and Panel B genes were largely independent of clinical prognosticators, and thus had the potential to provide valuable new information to statistical models of progression. We next performed univariate Cox regression analysis for Gleason, PSA, stage, and both Panel A (Hazard Ratio (HR) = 1.40, 95% Confidence Interval (CI) = 1.21-1.58) and Panel B (HR = 0.56, 95% CI = 0.47-0.67) in the combined cohort, finding that each parameter was highly significant in predicting earlier biochemical recurrence. In a multivariate model of clinical features alone, Gleason, PSA, and stage were all significant model parameters, and when Panel A or Panel B was added to generate a combined clinical/molecular model, all parameters remained highly significant (Tables 2-3). While our panel assembly method selected for within-panel correlation in gene expression patterns, we also found that there was significant between-panel correlation (Spearman’s ρ p-value < 0.0001, Table 1) that suggests our panels may be partially redundant to one another. Concordantly, when we assembled a proportional hazards model including clinical features and both gene panels, individual gene panel effect sizes were reduced or, in the case of Panel A, became only marginally significant (Table 4). As with our earlier correlation analyses, proportional
## Table 1: Correlation of clinical and molecular features of the combined cohort. * denotes significant correlation.
### Combined Cohort Multivariate Model - Clinical + A

<table>
<thead>
<tr>
<th>Model Parameter (p-value)</th>
<th>HR (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gleason (0.0006)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 7</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.81 (1.5-5.58)</td>
<td>0.001*</td>
</tr>
<tr>
<td>&gt; 7</td>
<td>3.82 (1.76-8.51)</td>
<td>0.0007*</td>
</tr>
<tr>
<td>PSA (log[1+baseline]) (0.0023)*</td>
<td>3.36 (1.60-6.29)</td>
<td>0.0023*</td>
</tr>
<tr>
<td>Stage (0.0033)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>1.64 (0.97-2.82)</td>
<td>0.065</td>
</tr>
<tr>
<td>T3</td>
<td>3.56 (1.32-8.1)</td>
<td>0.0147*</td>
</tr>
<tr>
<td>Panel A Score (0.0015)*</td>
<td>1.45 (1.16-1.78)</td>
<td>0.0015*</td>
</tr>
</tbody>
</table>

Table 2: Multivariate proportional hazards model including clinical features and Panel A in the combined cohort. * denotes significant model parameter.

### Combined Multivariate Model - Clinical + B

<table>
<thead>
<tr>
<th>Model Parameter (p-value)</th>
<th>HR (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gleason (0.003)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 7</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.43 (1.29-4.87)</td>
<td>0.0057*</td>
</tr>
<tr>
<td>&gt; 7</td>
<td>3.48 (1.62-7.76)</td>
<td>0.0014*</td>
</tr>
<tr>
<td>PSA (log[1+baseline]) (0.0034)*</td>
<td>3.2 (1.51-6.09)</td>
<td>0.0034*</td>
</tr>
<tr>
<td>Stage (0.0162)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>1.46 (0.88-2.46)</td>
<td>0.15</td>
</tr>
<tr>
<td>T3</td>
<td>2.96 (1.11-6.67)</td>
<td>0.033*</td>
</tr>
<tr>
<td>Panel B Score (0.0002)*</td>
<td>0.63 (0.5-0.8)</td>
<td>0.0002*</td>
</tr>
</tbody>
</table>

Table 3: Multivariate proportional hazards model including clinical features and Panel B in the combined cohort. * denotes significant model parameter.

### Combined Cohort Multivariate Model - Clinical + A/B

<table>
<thead>
<tr>
<th>Model Parameter (p-value)</th>
<th>HR (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gleason (0.0066)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 7</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.47 (1.3-4.94)</td>
<td>0.0051*</td>
</tr>
<tr>
<td>&gt; 7</td>
<td>3.08 (1.39-7.0)</td>
<td>0.0055*</td>
</tr>
<tr>
<td>PSA (log[1+baseline]) (0.0016)*</td>
<td>3.46 (1.66-6.43)</td>
<td>0.0016*</td>
</tr>
<tr>
<td>Stage (0.0076)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>1.6 (0.95-2.74)</td>
<td>0.08</td>
</tr>
<tr>
<td>T3</td>
<td>3.15 (1.17-7.14)</td>
<td>0.026*</td>
</tr>
<tr>
<td>Panel A Score (0.13)</td>
<td>1.22 (0.94-1.56)</td>
<td>0.13</td>
</tr>
<tr>
<td>Panel B Score (0.013)*</td>
<td>0.7 (0.53-0.93)</td>
<td>0.013*</td>
</tr>
</tbody>
</table>

Table 4: Multivariate proportional hazards model including clinical features and both Panel A and B in the combined cohort. * denotes significant model parameter.
hazards modeling in the individual Taylor and Glinsky cohorts largely reflected our analysis of the combined cohort, including the reduction in effect size of each gene panel when considered in combination alongside clinical features. Regarding Panel A, this data supports previous associations of cell cycle gene expression levels with prostate cancer recurrence, showing that this functional gene set adds significant value to prognostic models based on clinical features. On the other hand, our Panel B analysis stands alone in showing that lowered expression of this set of putative tumor suppressors corresponds with more rapid prostate cancer recurrence.

We also asked if our gene panels might be utilized in a logistic regression analysis for the prospective, binary classification of patients who will or will not experience a biochemical recurrence event. Utilizing the combined cohort, univariate analysis showed that PSA, stage, Gleason score, Panel A, and Panel B could all be used to classify patients by predicted recurrence, with Gleason score being the strongest clinical predictor as assessed by the receiver operating characteristic (ROC) curve (area under the curve (AUC) = 0.738) and Panel A and B performing similarly (AUC = 0.683 and 0.710, respectively). As in our proportional hazards model, when Panel A and Panel B were considered in combination, Panel A became only marginally significant, though the model itself outperformed individual gene panel models (AUC = 0.722). In a multivariate clinical model, PSA, stage and Gleason score were all significant model parameters, outperforming the best univariate models (AUC = 0.784). In combined clinical/molecular multivariate models, we found that the addition of Panel A, Panel B, or both Panel A and B improved upon the clinical-only model (AUC = 0.821, 0.810, and 0.822, respectively).
To better understand how our gene panels were able to improve the accuracy of these recurrence prediction models, we needed to validate each model’s performance in an independent cohort. To this end, we randomly divided the combined cohort into a training subset representing 75% of the total cohort, and a validation subset representing the remaining 25%. The training subset was utilized to generate logistic regression models for clinical features alone or in combination with Panel A and/or Panel B. The coefficients of each model were then applied to the validation subset, and the process was repeated 20 times. As in our analysis of the entire combined cohort, we found that the AUC metric improved significantly upon inclusion of molecular information to existing clinical prognosticators (Figure 17C). Assessing the performance of these models in the validation cohorts, we determined that this improvement could be attributed to an increase in model sensitivity, while our gene panels did not significantly impact model specificity. This important outcome demonstrates that our newly defined, independent molecular features are capable of identifying at-risk patients not captured by current clinical prognosticators without an increase in false positives.

To find if our gene panels also correlated with a second clinical outcome, the development of resistance to hormone therapy, we conducted an analysis of Panel A and Panel B gene expression in circulating tumor cells (CTCs) isolated from a cohort of prostate cancer patients who had received hormone therapy for metastatic or recurrent disease. In this nested study, we measured an average of 12 CTCs from 10 patients whose disease remained sensitive to therapy (Castration-Sensitive (CS)) and 15 patients who had developed resistance (Castration-Resistant (CR)) as indicated by rising serum PSA
Figure 18: Analysis of CREB1/FoxA1 target genes in circulating tumor cells. (Figure 18 caption continued on page 100)
(Figure 18 caption continued from page 99) A) Expression data is displayed for Panel A and Panel B genes in CTCs grouped by patients defined as CR (blue) vs. CS (yellow). Expression levels were normalized to UBB and plotted as the average of duplicate microfluidic qPCR runs for each cell analyzed. Undetectable transcripts are displayed in white and were excluded from subsequent statistical analysis. (B-C) Box plots of Panel B (C) and Panel A (B) gene expression in CR vs. CS patients (Mann-Whitney U - * - p-value < 0.05, ** - p-value < 0.01, *** - p-value < 0.001, NS – Not Significant).

countention (Figure 18A). Panel B genes were often significantly under-expressed in CR compared to CS patient-derived CTCs (Figure 18B), though this trend was not maintained at the patient level. We observed more similar expression patterns of Panel A genes in CTCs isolated from CR vs. CS patients (Figure 18C), and again found no between-patient differences in Panel A expression levels. Though our sample size was small, these results suggest that the primary value of our gene panels is in predicting recurrence among primary prostate cancer patients and that they may not further distinguish patients who have or will develop treatment resistance among a population of recurrent prostate cancer patients in which gene panel expression changes have likely already occurred.

Disrupting CREB1/FoxA1 target gene expression via phospho-MED1 inhibition

We finally explored a strategy to target CREB1/FoxA1 signaling in advanced prostate cancer cell lines. Though its basal transcriptional activity can occur in the absence of an activating phosphorylation event at serine 133 (Ser133), CREB1 signaling presents an attractive therapeutic target as this TF’s transactivation potential is dramatically enhanced by druggable kinases including PKA (119,121). Ser133-phosphorylated-CREB1
Figure 19: 22rv1 cells exhibit activated CREB1 signaling. (A) Western blot analysis of CREB1 and pCREB1 in abl, 22rv1, C4-2B, and PC-3 cells. (B) qRT-PCR analysis of CREB1/FoxA1 target genes expression in LNCaP, abl, and 22rv1 cells. (C) ChIP-qPCR assays were conducted at CREB1/FoxA1 target gene promoters in abl and 22rv1 cells using the indicated antibodies (n = 3). (* indicates two-sided t-test p-value < 0.05)
(pCREB1) has been proposed as a potential prostate cancer biomarker (112), with increased pCREB1 levels correlating with more aggressive disease, and several studies have explored the use of various kinase inhibitors to abrogate, for example, parathyroid hormone- or epidermal growth factor-stimulated CREB1 (124). We set out to inhibit endogenous CREB1 or pCREB1 activity in the absence of external stimuli.

We conducted our initial analysis in 22rv1 cells, as these exhibited similar protein expression, but much stronger phosphorylation of CREB1 compared to abl, C4-2B, and PC-3 cell lines (Figure 19A). 14 of our Panel A genes, which were significantly over-expressed in abl compared to LNCaP cells, were further up-regulated in 22rv1 cells, and we selected eight Panel A genes (ASPM, CENPF, CIT, HMMR, PBK, PRC1, SKA3, and UBE2C) and four of our previously identified CREB1/FoxA1 target genes (CCNA2, CCNE2, E2F1, and CDK1) (108) for further analysis (Figure 19B). Standard ChIP-qPCR assays revealed nearly identical binding of CREB1 to the promoter regions of all 12 genes in abl and 22rv1 cells (Figure 19C). Ser133 phosphorylation in the CREB1 KID domain generates a substrate for histone acetyltransferase CBP/p300 recruitment (123), thus we observed significantly higher enrichment of both CBP and p300 as well as enhanced histone H3 lysine 27 acetylation (H3K27ac) levels at all gene promoters in 22rv1 compared to abl cells (Figure 19C). Enhanced transcriptional activity was also indicated by higher RNA polymerase II (Pol II) recruitment to 10/12 gene promoters in 22rv1 cells (Figure 19C). Consistent with the role of the Mediator co-regulatory complex subunit, MED1, in forming transcription regulatory chromatin loops between enhancer- and promoter-bound transcription factor complexes (i.e. FoxA1 and CREB1,
respectively) (146), we also observed binding of threonine 1032–phosphorylated MED1 (pMED1) at all CREB1-regulated gene promoters in abl cells, and found significantly enhanced pMED1 occupancy in 22rv1 cells (Figure 19C).

Though the kinase inhibitor H89 has been used extensively to study and target CREB1 activity enhanced by stimuli such as forskolin, we found that H89 treatment alone could not reliably reduce endogenous pCREB1 levels in 22rv1 cells (Figure 20A). Paradoxically, 22rv1 cells exhibited a dose-dependent reduction in CREB1/FoxA1 target gene expression in response to H89 treatment (Figure 20B). These results suggest that H89 has some impact on CREB1/FoxA1 signaling without directly affecting CREB1 per se. This notion was supported by the maintenance of CREB1 as well as CBP and p300 recruitment to target gene promoters following H89 treatment (Figure 20C). We hypothesized that H89 may have some effect on the modification/activity of other CREB1/FoxA1 co-regulators and were intrigued to find that pMED1 but not total MED1 expression levels were dramatically reduced by H89 treatment (Figure 20A). pMED1 binding to target gene promoters was also attenuated by H89 treatment, which corresponded to Pol II destabilization across most genes (Figure 20C). This is consistent with another known role of the Mediator complex in establishing and maintaining the stability of the pre-initiation complex (PIC) at Pol II-transcribed genes promoters (21).
Figure 20: Kinase inhibition of CREB1/FoxA1 target gene expression acts through MED1. (A) Representative western blot of 22rv1 cells treated for 24 hours with vehicle or H89. (B) CREB1/FoxA1 target gene expression was measured by qRT-PCR in 22rv1 cells treated for 24 hours with vehicle or H89 (n = 3-4). (C) ChIP-qPCR was conducted at the promoters of CREB1/FoxA1 target genes using the indicated antibodies in 22rv1 cells treated for 24 hours with vehicle or H89 (n = 2). (D) Representative western blot of 22rv1 cells treated for 24 hours with vehicle or U0126. (E) CREB1/FoxA1 target gene expression was measured by qRT-PCR in 22rv1 cells treated for 24 hours with vehicle or U0126 (n = 3-4). (F) ChIP-qPCR was conducted at the promoters of CREB1/FoxA1 target genes using the indicated antibodies in 22rv1 cells treated for 24 hours with vehicle or H89 (n = 2). * Indicates significant p-value < 0.05, two-sided Student’s t-test.

104
To help verify our results, we sought to reproduce our findings by inhibiting known MED1 kinases, Akt and Erk (146,209). As 22rv1 cells are PTEN wild-type and exhibit very low levels of phosphorylated Akt (Figure 21A), the Akt inhibitor MK-2206 was not effective in diminishing CREB1/FoxA1 target gene expression in 22rv1 cells (Figure 21B). Treatment with the MEK inhibitor, U0126, reduced Erk phosphorylation and recapitulated the results of H89 treatment, including reduced pMED1 levels, CREB1/FoxA1 target gene expression, and pMED1 and Pol II recruitment to promoters (Figure 20D-F). As this treatment strategy appears not to act directly via inhibition of CREB1 phosphorylation, we hypothesized that it would remain effective in cells with low endogenous pCREB1 levels. Concordantly, abl cells exhibited dramatically reduced CREB1/FoxA1 target gene expression levels following H89 and U0126 treatment (Figure 21C). Additionally, as abl cells have activated Akt (Figure 21A), we found that this signaling axis was also sensitive to MK-2206 treatment in abl cells (Figure 21C). We then asked if these kinase inhibitors, which impact broad-ranging cellular processes, exhibit similar transcription inhibitory activity over non-CREB1/FoxA1 targets. We selected 10 genes that failed to meet either criteria of a target gene (TF binding and TF-responsive expression) but were still highly expressed in our RNA-seq data. These genes were largely unaffected, and in some cases induced, by all compounds tested in both 22rv1 and abl cells (Figure 21D), suggesting that the CREB1/FoxA1 axis is a prominent target of their chemical activity.
Figure 21: Kinase inhibitors are functional in a model of low endogenous CREB1 phosphorylation. (A) Western blot analysis of Akt and pAkt in abl, 22rv1, and PC-3 cells. (B) qRT-PCR analysis of CREB1/FoxA1 target gene expression in 22rv1 cells treated for 24 hours with vehicle or 3 µM MK-2206 (n = 2). (C) qRT-PCR analysis of CREB1/FoxA1 target gene expression in abl cells treated for 24 hours with vehicle or 5 or 10 µM H89 (left), 10 µM U0126 (middle), or 3 µM MK-2206 (right) (n = 2). (*) indicates two-sided t-test p-value < 0.05. (D) qRT-PCR expression analysis of 10 control, non-CREB1/FoxA1 target genes in 22rv1 and abl cells treated for 24 hours with the indicated compounds.
AR has been shown to play at least a partial role in supporting the expression of several of our CREB1/FoxA1 target and Panel A genes (108). Additionally, Hu et al. reported that forced expression of the constitutively active AR V7 splice variant in LNCaP cells caused the up-regulation of several Panel A genes in the absence of hormone (76). This presents a potential confounding factor in 22rv1 cells, which express both full-length AR (AR FL) as well as AR splice variants (AR Vs) (Figure 22A). To assess whether CREB1/FoxA1 target gene expression is driven by AR activity in this model, we transfected 22rv1 cells with AR-targeting siRNA and measured the expression of PSA as well as our CREB1/FoxA1 target genes (Figure 22B-C). While PSA expression after AR knockdown dropped to less than 50% of control, expression of reported AR-regulated genes that we identified here as CREB1/FoxA1 targets (CCNA2, UBE2C, CDK1) (15,108,210) was only marginally, though significantly inhibited by AR knockdown, and other CREB1/FoxA1 targets were entirely unaffected (Figure 22C). We therefore suggest that AR plays only a minor role in supporting the expression of these genes.

Together, our results reveal that prostate cancer-driving CREB1/FoxA1 transcriptional activity is sensitive to broad inhibition of the indispensable co-activator MED1. Additionally, these results provide an interesting example of how CREB1 transcriptional activity can be altered in the absence of any apparent effect on its phosphorylation (121).
Figure 22: AR and AR splice variants have a minor role in CREB1/FoxA1 target gene expression. (A) Western blot assay depicting the expression of full-length AR (AR FL) and AR splice variants (AR Vs) in LNCaP, abl, and 22rv1 cells. (B) Western blot demonstrating the knockdown of both AR FL and AR Vs by siAR transfection for 72 hours in 22rv1 cells. (C) qRT-PCR analysis of CREB1/FoxA1 target genes as well as PSA (AR target gene) expression in 22rv1 cells 72 hours after siControl or siAR transfection (n = 2). * indicates two-sided t-test p-value < 0.05.
Discussion

Our analysis has provided several intriguing outcomes and a functional view of how transcription factor networks acting outside the androgen/AR signaling axis contribute to prostate cancer growth and progression. This study enhances the growing body of research defining CREB1 as an important driver of various cancers including leukemia, non-small cell lung cancer, breast cancer, and prostate cancer (113,211,212), and encourages continued development of novel CREB1-inhibiting compounds that either affect kinase-dependent CREB1 transactivation or impinge upon its relationship with collaborators/co-regulators such as FoxA1 and MED1 (213). Our integrative genomics approach identified CREB1/FoxA1 activity as a targetable signaling axis defining the expression of prognostic/predictive biomarkers for primary prostate cancer, in which the discovery of clinically actionable genetic alterations has been challenging.

While prognostic gene signatures that aid in the difficult task of stratifying risk in prostate cancer patients continue to emerge, the mechanism of their deregulated expression is often not addressed, leaving the extent of their clinical utility partially untapped (196,214). We found that CREB1/FoxA1 signaling assumes direct regulatory control over a large proportion of genes that were assembled in developing a prognostic, cell cycle progression score reported previously (10/19 Panel A genes) (196). While some overlap between these cell cycle genes and CREB1/FoxA1 co-regulated genes could be expected based on our knowledge that these TFs regulate both G1/S (CCNE2, CCNA2, and E2F1) and G2/M (UBE2C and CDK1) phase gene expression (108), the more than 50% overlap provides evidence for the central role of CREB1 and FoxA1 in
driving a disease-relevant cell cycle gene expression program associated with poor clinical outcomes. The fact that Panel A is enriched within commonly over-expressed genes among aggressive prostate cancers, further suggests that enhanced cell cycle signaling is a hallmark of advanced disease. Thus, our findings provide additional justification for developing predictive and prognostic tools based on cell cycle signaling downstream of CREB1/FoxA1, and we further suggest that prostate cancer patients characterized by heightened cell cycle signaling might be inherently sensitive to CREB1/FoxA1 inhibition.

Our analysis also revealed a small set of known as well as potential tumor suppressors, inactivated by CREB1/FoxA1 transcriptional repression rather than mutation, that were predictive of recurrence (207). As exact mechanisms of CREB1-mediated transcriptional repression remain elusive, the therapeutic implications of this prognostic gene panel are undetermined (121). Interestingly, the lowest collective expression of Panel B genes generally occurred in patients with the strongest collective expression of Panel A genes. This seems to indicate that heightened CREB1/FoxA1 activity results in coordinated up- and down-regulation events and that focusing on just a few genes that universally indicate CREB1/FoxA1 signaling could be sufficient for characterizing a given patient’s risk. Based on its strong performance in prediction models, we suggest that Panel B gene expression could also be further developed into a tool to define an aggressive subset of primary prostate cancers in patients destined to relapse and who may benefit from trials of adjuvant therapy.
Tremendous enthusiasm has developed for the use of circulating tumor cells for real-time monitoring of dynamic events in both primary and metastatic cancer settings without the need for invasive tissue biopsy. A great deal has been learned through the analysis of circulating tumor content regarding the prevalence and clonality of common prostate cancer deletions and mutations in cases with multiple metastatic foci (215), the enhanced expression of EMT-related genes in CRPC patients (182), AR mutation and amplification status (216,217), and sustained AR signaling in the presence of hormone therapy (218). Our analysis showed that Panel A and B expression values were similar across recurrent prostate cancer patients regardless of whether or not they have developed resistance to antiandrogen therapy, thus they may have limited value beyond their strong performance in the primary cancer setting for predicting recurrence. Some controversy exists relating to the number of CTCs that can be isolated from primary vs. metastatic prostate cancer patients (219), but future efforts should be made to determine whether CTC gene expression analysis could be applied to primary prostate cancer patients for the prediction of subsequent recurrence.

Finally, our findings collectively suggest that pCREB1 alone may not be an all-inclusive indicator of aggressive prostate cancer. While the phosphorylation of CREB1 is a partial marker of CREB1 transcriptional activity with demonstrated utility in identifying aggressive disease (112,120), it does not account for the phosphorylation-independent mechanisms of transactivation, which we now see can be potentiated by the cooperative function of FoxA1 and by the co-regulatory role of MED1. Thus, measuring total CREB1
transcriptional output, or that of a representative subset of disease-relevant genes, will likely afford increased sensitivity in predicting outcomes over pCREB1 levels alone.
Chapter 4: Conclusion

The work described in this thesis demonstrates several key concepts that are of importance in understanding the mechanisms underlying the expression of cancer-supporting transcription profiles during the progression of prostate cancer as well as their implications for future cancer diagnostic and therapeutic tools. We find that GATA2 is unique among AR pioneer factors for its apparent universal support of AR expression and binding, whereas factors like FoxA1 are known to exhibit a more complicated relationship with AR. We show that mechanistically, GATA2 is capable of pre-binding androgen-stimulated AR binding sites to establish an accessible chromatin environment in these regions via direct recruitment of p300 and to prime these loci for hormone-stimulated expression via preformation of regulatory chromatin loops. These findings encourage the development of GATA2-targeting compounds capable of disrupting this indispensable AR cofactor for efficient, combinatorial inhibition of AR activity.

We also define a gene regulatory role for the prominent AR pioneer factor, FoxA1, occurring beyond its interaction with AR, and in so doing establish the genome-wide activity of CREB1 as a driver of prostate cancer-relevant gene expression profiles across disease stages. Genes cooperatively regulated by CREB1 and FoxA1 in advanced disease models are strongly associated with aggressive disease phenotypes and poor clinical
outcomes among prostate cancer patients. We believe our findings strongly support the
development of novel diagnostic tools for identifying patients in need of more aggressive
treatment for primary prostate cancer. Our work further demonstrates a mechanism-based
strategy for disrupting the expression of CREB1/FoxA1 target genes that might
ultimately prove effective in patient cohorts characterized by enhanced CREB1/FoxA1
signaling. Perhaps most importantly, our efforts shed additional light on the concept of
utilizing the macro-scale transcriptional output of a transcription factor, or multiple
collaborative transcription factors, when defining biomarkers of a particular disease
outcome. This appears to be a particularly cogent argument when studying transcription
factors like CREB1, whose transactivation potential is stimulated by, but does not
require, a post-translational modification. Comprehensive characterization of a
transcription factor’s output, regardless of the mechanism, may have greater potential to
identify downstream targets for use as biomarkers or therapeutic targets.
References


118


is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. *Cancer research*, 69, 2305-2313.


cancer recurrence after androgen deprivation therapy. *Cancer research, 61*, 4315-4319.


receptor (AR) transcription complex by stabilizing the association between AR and GATA2. *Nucleic acids research.*


7174, reduces the endothelial VCAM-1 induction by inflammatory cytokines, acting through the regulation of GATA. *Biochem Biophys Res Commun, 272*, 370-374.


androgen receptor binding to chromatin, androgen signalling and prostate cancer. *The EMBO journal*, **30**, 3962-3976.


by Nexrutine, a Phellodendron amurense bark extract. *Clinical cancer research: an official journal of the American Association for Cancer Research, 13*, 2784-2794.


phosphorylation in men undergoing prostate biopsies. *Int Urol Nephrol*, 46, 505-510.


